The Case of Missing Mortgages: Are Minorities Still Being Disproportionately Denied a Loan and for What Reasons?

Dilyana Dimova

US Department of Housing and Urban Development

Does discrimination in mortgage lending persist even in the age of automated underwriting? Using recently expanded Home Mortgage Disclosure Act (HMDA) data, this paper looks at loan denial rates and finds that even accounting for lender-specific (size and underwriting platform) and loan-specific factors (credit scores, loan-to-value and debt-to-income ratio), demographic indicators still play a considerable role. African Americans and Native Americans are the minority groups that have the greatest probability of their loan application being turned down followed by Hispanics. The analysis shows that if minorities were denied mortgages at rates similar to Caucasians, there would be about 290,000 more mortgages originated in the period 2018-23. Minorities are most often denied a loan for limited credit history, high debt-to-income ratio or insufficient collateral. If approved for a mortgage, they are more likely to be offered a high-cost loan characterized by an interest rate with a spread thus making homeownership costlier for them.

Keywords: residential mortgages, loan denial, discrimination, economic inequality

JEL: G21, R32, R21

I. Introduction

The rise of automated underwriting made lending less onerous and was supposed to reduce discrimination by minimizing human bias. The extent to which this happens has been difficult to quantify so far. Researchers have had to settle for limited data and extrapolation when trying to tease out borrower-specific from loan-specific factors. Since 2018, the Home Mortgage Disclosure Act (HMDA) dataset has expanded to allow the consideration of the role of demographic effects in loan decisions after controlling for risk factors.

Lenders usually care most about minimizing risk, so a mortgage loan denial largely reflects loan-specific characteristics that capture potential loan riskiness. A person with a subpar credit score, a high debt-to-income (DTI) ratio, high combined loan-to value (CLTV) ratio, spotty employment history or insufficient collateral poses a red flag. However, sometimes factors beyond these such as borrower demographics play a role. The motivation for that may be less outright discrimination and more a desire to proxy for missing financial information. Earlier works such as Mian and Sufi (2008), Bayer, Ferreira and Ross (2014), Ards, Ha and Mazas (2015) and Choi, et al. (2019) have documented discrimination among various racial and ethnic groups. Coincidentally African Americans and Hispanics tend to experience shorter credit histories, insufficient employment history in higher rates than Caucasian borrowers or buy properties in areas with lower appreciation prospects. However, the extent to which discrimination in lending to minorities tends to persist, even accounting for these factors, is a continued focus in the economic literature.

A common hurdle in analyzing disparities in mortgage loan denial rates has been the lack of detailed loan data pertaining to each specific loan. Thus, when lenders deny a loan for risk reasons, lack of riskiness information has often led researchers to attribute loan denial to other more easily available data points such as demographics. The National Home Mortgage Disclosure Act (HMDA) dataset that collects mortgage data since 1975 has historically collected few loan-specific factors that quantify loan riskiness until 2018. Many papers have supplemented HMDA with proprietary data aggregated from individual lenders or data obtained directly from them. However, challenges persist.

This study takes advantage of expanded HMDA data starting in 2018 that covers loan-specific details that help quantify riskiness such as DTI, CLTV and credit scores as well as lists the reasons for loan denial as provided by lenders. This allows to separate demographics from riskiness and test their separate effect as well as assess whether the reason given for denial matches the loan characteristics. The paper achieves that using a three-part approach. The first part evaluates the contribution of demographic factors to the probability of loan denial once loan factors associated with riskiness are accounted for. The second part considers individual reasons for denial and how race, ethnicity or gender contribute to their likelihood. The last part looks at approved loan applications and explores whether minorities are more likely to be offered a high-cost loan, i.e., a loan with a spread above the benchmark interest rate.

The findings confirm that loan-specific measurements of riskiness such as a poor credit score, DTI above 43% and CLTV above 90% are the biggest predictor of probability of denial. Applying with a large lender and automatic underwriting reduce denial rates but even accounting for these, demographics still have a role. African Americans, Native Americans and Hispanics tend to be denied a mortgage at a greater rate than Caucasians. The effect is more pronounced for conventional loans than for Federal Housing Administration (FHA) loans which tend to have higher leverage and at higher FICO scores since lower scores tend to experience high denial rates regardless of demographics. Using counterfactual study, the paper estimates that if minorities were denied a loan at rates similar to that of Caucasians accounting for available risk metrics, there would have been 290,000 more mortgages originated in the time period.

The reasons for denial most often given to minorities are limited credit history, high DTI or insufficient collateral. When approved for a mortgage, these groups are also more likely to be offered a high-cost loan. The study confirms that loan denials continue to be associated with demographic factors even accounting for available risk metrics. Rather than attributing these findings to overt discrimination, to some extent they may reflect still missing financial factors such as employment history and financial history that disproportionally affect minorities but are not yet extensively accounted for in loan data.

II. The Literature on Mortgage Denial and Demographics

The literature on loan denials usually focuses on disparities in denial rates among different racial or ethnic groups, regional differences or disparities between different periods owing to looser or tighter underwriting criteria. The literature on lending disparities among different groups of borrowers is by far the largest. Most of those studies document higher denial rates for minorities and people of color. At the frontier of the literature on racial discrimination in mortgage markets are papers like Zhang and Willen (2021), Bhutta and Hizmo (2020), and Courchane and Ross (2019), and Gerardi, Willen, and Zhang (2023). The last one suggests that Black and Hispanic homeowners pay significantly higher mortgage interest rates than white and Asian homeowners because the latter are much more likely to exploit periods of falling interest rates by refinancing their mortgages or moving. Beyond financial literacy, Black and Hispanic borrowers face challenges refinancing owing to lower credit scores, equity and income.

Bhutta and Hizmo (2020) temper the suggestion of entrenched racial discrimination by considering differences in discount points. They show that minorities and whites face identical schedules, but sort to different locations on the schedule. Such sorting may reflect systematic differences in liquidity or preferences. Zhang and Willen (2021) take sorting by showing that Black and Hispanic borrowers getting conforming mortgages are willing to increase their interest rate in order to switch to the options offered to non-Hispanic white borrowers.

Bayer, Ferreira and Ross (2014) examine how high-cost mortgage lending varies by race and ethnicity. The authors use unique panel data that matches a representative sample of mortgages in seven large metropolitan markets between 2004 and 2008 to public records of housing transactions and proprietary credit reporting data. The results reveal a significantly higher incidence of high costs loans for African American and Hispanic borrowers even after controlling for key mortgage risk factors. Significant racial and ethnic differences are widespread throughout the market, even across high and low risk borrowers. These differences are reduced considerably with the inclusion of lender fixed effects, implying that a significant portion of the estimated market-wide racial differences can be attributed to differential access to (or sorting across) mortgage lenders.

Ards, Ha and Mazas (2015) point to two competing explanations for the wide disparities in loan denial rates among different racial/ethnic groups. One argument is that these disparities result from underlying racial disparities in credit worthiness. A competing view is that the disparities arise from a pattern of racial discrimination among mortgage lenders. The authors use the Freddie Mac Consumer Credit Survey dataset to test the hypothesis that measures of discrimination disappear when one accounts for racial differences in credit scores. They demonstrate that one cause of the appearance of poor credit risk among African American applicants is that African Americans with good credit risk underestimate their credit worthiness and apply for loans in lower numbers. The findings suggest that even nondiscriminatory lending behavior has the unintended effect of screening out low-risk African Americans and thereby yields higher denial rates among African Americans. This in turn confirms prior beliefs about the poor credit of average African American applicants. Much, but not all, of the racial disparity in loan outcomes can be explained by racial differences in credit scores and the resulting racial disparity in loan outcomes explains much of the racial difference in false perceptions about bad credit. Thus, a possible self-fulfilling mechanism remains within the credit market that perpetuates views about African American bad credit.

Historically, the HMDA database has been used to document large racial and ethnic differences in the likelihood of having a mortgage approved (Avery, Beeson and Sniderman 1996). While a substantial fraction of the racial differences is certainly attributable to differences in borrower and loan risk factors, Munnell, Tootell, Browne, McEneany (1996) in a seminal, study show that substantial racial differences remain in a sample of HMDA loans in Boston even after controlling for detailed mortgage risk factors. Avery, Canner and Cooke (2005) document large racial and ethnic differences in the incidence of rate spread loans in HMDA but are unable to control for common mortgage risk factors (e.g., borrower credit score), which are not included in the HMDA database at this time.

Considering this limitation of the HMDA database, most of the studies that have documented differences in the prevalence of high-cost loans have used one of two sources: proprietary data aggregated from individual lenders or data obtained directly from individual lenders. Ghent, Hernandez-Murillo and Owyang (2013), Haughwout, Mayer and Tracy (2009), Bocian, Ernst and Lee (2006) use proprietary data aggregated from the reports of many lenders and merged with the HMDA database. While these studies have documented substantial unexplained differences by race, they are often restricted to samples that represent a subset of the market, usually emphasizing loans that are securitized privately or lenders that operate primarily in the subprime sector. Several other studies have examined the mortgage pricing behavior of individual lenders (African American, Boehm, DeGennaro 2003; Nelson 2005; Courshane 2007; Courshane and Nickerson 1997). These studies have found very small, if any, within-lender differences between Caucasian and minority borrowers in the incidence of high-cost mortgage credit.

In this study, we examine racial and ethnic differences in the incidence of high-cost mortgage loans in a market-wide sample covering several large U.S. metropolitan areas or regions. The shift to market-wide data changes the question being asked from whether similar borrowers receive different prices from the same lender (e.g., disparate treatment discrimination) to whether unexplained racial differences exist in market outcomes, a phenomenon that Heckman (1998) describes as market discrimination. Significant market level differences in the price of credit have important consequences for the dynamics of racial and ethnic inequality in homeownership, wealth, and credit worthiness, even if only small differences exist at the lender level.

Bayer, Ferreira and Ross (2014) attempt to avoid the HMDA limitations by linking the HMDA data on home purchase and refinance mortgages between 2004 and 2008 to public records data on housing transactions and liens in seven distinct metropolitan housing markets. The empirical analysis reveals significant unexplained racial and ethnic differences in the incidence of high cost or subprime mortgage credit. These differences persist after controlling for detailed measures of borrower and loan characteristics including credit score, ratio of the loan amount to housing price, presence of subordinate liens, and housing and debt expenses relative to individual income. Relative to a model based on control variables available in HMDA, the inclusion of these additional controls erodes about half of the racial and ethnic differences in loan pricing. Still, the remaining differentials are sizable with African American and Hispanic borrowers. These loan-pricing differences exist across a variety of large metropolitan housing markets including not only faster growing markets in California and Florida that experienced especially severe housing market downturns, but also slower-growing Eastern and Midwestern housing markets.

The lack of detailed information on borrower and loan characteristics typical of most datasets so far can overstate the prevalence of discrimination by attributing denial incidence to discrimination when in fact it is due to loan features. Park (2022) compares denial and default rates of applications for over 7.6 million FHA-insured home purchase mortgages by specifically modeling expected risk by including credit score, assets after closing, receipt of downpayment assistance, self-employment, and more. The analysis suggests that even with advanced loan characteristics, some disparities in endorsement rates cannot be fully explained by differences in default risk.

Ross and Wang (2021) use HMDA rate spread loans to identify lenders involved in riskier lending prior to the foreclosure crisis. They categorize lenders as having a high rate spread share, or as a shorthand “high-cost lenders,” if 20 percent or more of their mortgages were rate spread loans. Then they develop a shift-share measure of changes in high rate spread share lender representation in housing submarkets across origination years. While half the cross-sectional correlation between foreclosure and high rate spread lender share is explained by borrower observables, the authors find robust and stable estimates of the within housing submarket relationship between foreclosure and predicted changes in market share. Estimates are not explained by local housing price variation, rather evidence suggests servicer behavior in response to rising local foreclosure rates as a mechanism.

Beyond racial factors, the literature on loan denials usually focuses on regional differences or disparities between different periods owing to looser or tighter underwriting criteria. Mian and Sufi (2008) demonstrate that between 2001-05, high latent demand zip codes experienced large relative decreases in denial rates, coupled with increases in house price appreciation, even though these zip codes experienced significantly negative relative income and employment growth over this time. The authors attribute these patterns for high latent demand zip codes to a sharp relative increase in the fraction of loans sold by originators shortly after origination, a process referred to as “disintermediation.”

Some studies raise the issue that using denial rates as a measure depends on the composition of borrowers along with how loose or tight credit standards are. Higher denial rates can be the result of either a tighter credit environment or an increase in applications by weaker credit borrowers. To mitigate this problem, Li and Goodman (2014) construct a different measure of the denial rate by dividing applicants into two categories: high credit profile (HCP), and low credit profile (LCP). HCP applicants are those whose credit profiles are strong enough that their mortgage applications are unlikely to be denied by lenders. If the number of HCP applicants receiving loans equals the number of HCP applicants applying for loans, then the difference between the total number of mortgage applicants and the number of HCP applicants equals the number of applicants whose credit profiles are not strong enough to have a zero-denial rate. These are the LCP applicants. The total number of denied applicants divided by the number of LCP applicants equals the denial rate of LCP applicants. The authors call this the real denial rate, or RDR, versus the traditional denial rate of all applicants (HCP and LCP), which is called the observed denial rate, or ODR. The RDR also shows much smaller racial and ethnic gaps in credit accessibility.

Very few studies look at denial rates for high-leverage mortgages that are more frequent in high-cost areas. Bhutta and Keys (2018) consider private mortgage insurance (PMI) which is mandated for high-leverage mortgages purchased by Fannie Mae and Freddie Mac to serve as a private market check on GSE risk-taking. The authors document that PMI firms dramatically expanded insurance on high-risk mortgages at the tail-end of the housing boom, contradicting the industry's own research regarding house price risk. They determine that PMI behavior during the housing boom in part reflects a “moral hazard” incentive inherent to insurance companies in general to underprice risk and be undercapitalized, and that rather than providing discipline, private mortgage insurers facilitated GSE risk-taking.

The present research is closest to Park (2022) in modeling expected risk, but it goes further by considering conventional loans not just FHA loans. Unlike Avery, Canner and Cooke (2005), Munnell, Tootell, Browne, McEneany (1996) and Park (2022) which all use older HMDA data and unlike Bayer, Ferreira and Ross (2014) which attempt to avoid the limitations of HMDA by using other sources, this paper takes advantage of expanded HMA data post-2018. It also explores the likelihood of specific denial reasons and the incidence of high-cost loans, both issues which are not the focus of the aforementioned papers.

This paper makes three contributions to the field of mortgage loan denials. First, it exploits recently available expanded loan-level data in the HMDA dataset to consider factors contributing to a greater likelihood other than applicant demographics. This research includes loan characteristics that increase riskiness as well as spatial factors that could make a loan application more likely to be turned down. The absence of such data in the past may have contributed to denials being attributed to discrimination when in fact they may have reflected loan-specific factors.

Second, the paper looks beyond factors that make denials more likely, and links specific reasons cited for denials to loan factors associated with that reason. It explores the extent to which these reasons are backed by loan-specific factors as opposed to masking demographic factors. Finally, the research considers whether minorities whose loans applications have been approved are more likely to be offered high-cost loans (i.e., loans with a spread[1]) than other borrower groups.

III. Methodology

The methodology is based on a modified version of Ferguson and Peter (1995) theoretical model. The cumulative distribution function of the pool of mortgage loan applicants is *F(θ).* Lenders choose the optimal $\theta*$ to maximize profits. The bank's expected profit on a loan is:

$$E(\pi) = \theta r - (1 - \theta)$$

Where r is the interest charged on the loan. The bank receives profit r for a repaid loan valued at \$1 and zero otherwise. Lenders choose $\theta*$ to satisfy the zero-profit condition implying:

$$\theta^* = \frac{1}{1 - r}$$

The bank earns positive profits from borrowers with credit quality $\theta \in (\theta^*, 1]$. The average creditworthiness of the borrower of an accepted mortgage loan is:

[1] The Consumer Financial Protection Bureau (CFPB) which collects the HMDA data defines higher-price loans as those where the rate spread is greater than 1.5 percentage points. In other words, the difference between annual percentage rate (APR) and the average prime offer rate (APOR) is greater than 1.5. See https://ffiec.cfpb.gov/tools/rate-spread.

$$\hat{\theta} = \int_{\theta^*}^{1} \frac{\theta f(\theta) d\theta}{1 - F(\theta^*)}$$

For randomly received applications, the denial rate R_F is the proportion of the pool of applicants that falls below the cutoff θ^* so $R_F = F(\theta^*)$. The probability of default is $(1-\theta)$ and the expected default rate is D_F is:

$$D_F = \int_{\theta^*}^{1} \frac{(1-\theta) f(\theta) d\theta}{1 - F(\theta^*)} = 1 - \hat{\theta}$$

Lenders then separate the pool of applicants into two groups that differ in credit riskiness. The favorable credit borrowers follow a distribution of $G(\theta)$ while unfavorable credit borrowers follow a distribution of $H(\theta)$ so that $G(\theta) \leq H(\theta)$. $G(\theta)$ stochastically dominates $H(\theta)$. The average credit worthiness of borrowers in the two groups are $\hat{\theta}_G$ and $\hat{\theta}_H$ respectively where $\hat{\theta}_G > \hat{\theta}_H$.

Suppose θ^* decreases, which implies more of the riskier loans are now denied as the maximum acceptable risk decreases. That means that the denial rate of *H* increases relative to *G* because the pool with the greater number of applicants who fall short of the new lower cutoff point of θ^*. The increase in the relative denial rate may incorrectly be interpreted as worsening racial disparities when it is merely the result of risk tolerance on the part of creditors.

No discrimination would imply there is no difference in the probability of approval or denial between subgroups *conditional on risk* so that $E(G|\theta) = E(H|\theta)$. Empirical research often finds denial rate disparities even after accounting for observable risk characteristics. However, factors attributable to discrimination may be due to an incomplete list of risk metrics in HMDA or underwriting requirements. That is, the actual risk θ may not be properly accounted for using only available HMDA data and instead the perceived risk θ' is recorded in the data. If *H* is disproportionately associated with the unobserved risk, $\rho(\theta', H) > 0$, then the approval rate of *H* will be lower than *G* even when conditional on the observed risk, $E(G|\theta) < E(H|\theta)$. The consequence of not observing the true risk may be overstating disparities in denial rates. Both higher denial rates and higher default rates are expected among the higher risk group *H*. Discrimination would occur if *H* had a higher denial rate but a *lower* default rate.

Most studies treat risk and non-financial borrower characteristics as exogenous to one another but that may not necessarily be the case. In the past, “redlining,” has led to the exclusion of minority homeowners from desirable areas with higher appreciation home values. In this case race and ethnicity implied lower homeownership rates in desirable areas where home values tend to appreciate. The reduced homeownership incidence, especially in such areas, implied less intergenerational wealth and collateral which can make a borrower be perceived as riskier. While the Fair Housing Act prohibits such practices, their past incidence means that wealth accumulation by minorities and in turn collateral provision lags behind other groups.

The loan application can either be accepted/funded or denied (withdrawn and accepted but not originated loans are not considered. This binary outcome necessitates a probit model where the dependent variable is the loan outcome, and the independent variables are borrower/loan characteristics and other relevant factors.

$$Denial = \beta_0 + \beta_1 Risk^* + \beta_2 Y_i + \varepsilon$$

Where *Risk* is the true risk associated with the loan application free from lender perceptions. It is not completely accounted for in the that and thus unobservable to the researcher and instead we observe *Risk** that is the perceived risk by the lender. Vector Y_i captures other factors such as whether the loan application was subject to automatic underwriting, the number of mortgage applications a particular lender receives in a year and if the property is located in a high-cost area defined as area where median real estate prices are at least 10% above the national median. In an idealized setting, the denial rate should not depend on borrower status in a protected class.

However, lenders make decisions based on perceived risk based on data at hand. Theoretically, that also may include demographic characteristics on the part of the applicants such as race, ethnicity, age and gender. The true *Risk* is a latent variable that depends on many factors such as debt-to-income ratio, loan-to-value ratio property values, income and so on. Unlike perceived *Risk** it does not depend on any borrower characteristics:

$$Risk^* = \gamma_0 + \gamma_1 X_i + \gamma_2 Risk + \omega$$

Here X_i is a vector of borrower characteristics. Hence *Denial* depends on the true risk and a set of demographic factors that may distort perception of risk:

$$Denial = \beta_0 + \beta_1\gamma_0 + \gamma_1\beta_1 X_i + \beta_1\gamma_2 Risk + \beta_2 Y_i + \varepsilon + \beta_1\omega$$

$$Denial = \delta_0 + \delta_1 X_i + \delta_2 Risk + \beta_3 Y_i + \pi$$

Since $Risk$ is unobservable, it is proxied by loan characteristics associated with riskiness such as high debt-to-income ratio (DTI) and subpar credit scores.

The next part of the analysis focuses on the main reason for denial stated by lenders in the HMDA data. This offers up to four reasons simultaneously and many lenders choose more than one reason. The relationship of interest is whether specific reasons stated are backed by data if available and whether minority groups are more likely to be offered certain reasons for denials relative to others. The hypothesis tested is whether borrowers face the probability of denial for a specific reason versus not having their application denied. In this sense, this part tests probability of denial for a specific reason as opposed to denial for any reason in the first part.

Not all denial reasons can be validated using available data but those relating to high debt-to-value ratios and/or credit ratings can, so the focus is on denials that stated those reasons as primary or secondary motivation. The logit specification below explores whether certain groups are more likely to be cited these reasons for denial than others:

$$Denial\ Reason_i = \gamma_0 + \gamma_1 X_i + \gamma_2 Y_i + \pi$$

Where *Denial Reason* is a vector of testable denial reasons such as high DTI, employment history, credit history, insufficient collateral and cash issues. X_i is a vector of borrower characteristics and Y_i captures loan-specific metrics.

Finally, the last part concerns approved loans, not denied ones. It considers whether minority groups are more likely to be offered high-cost loans relative to other groups for similar loan characteristics. High-cost loans are defined as loans that have a spread i.e., interest rate that is above the 30-year Freddie Mac Fixed Rate Mortgage:

$$High\ Cost == \ \gamma_0 + \gamma_1 X_i + \gamma_2 Y_i + \pi$$

All three parts take advantage of expanded HMDA data on loan characteristics associated with riskiness. Despite this, it is still not sufficient to capture the complete process of loan application decisions. It lacks data on the process lenders employ to make decisions, does not record employment history beyond current income, and does not track the lifetime of past approved loans which is a crucial piece of data that lenders base decisions on.

IV. The Missing Mortgages and Trends in Loan Denial

Using recently expanded HMDA, this paper looks at loan denial rates in the period 2018-23 and the factors that confer greater likelihood of denial. The data is restricted to owner-occupied dwellings to remove investors from the sample. The pool of mortgages also excludes reverse mortgages, interest only mortgages, HELOC loans, mortgages that have balloon payments or mortgages with negative amortization. Fixed rate mortgages (FRM) and adjustable-rate mortgages (ARM) are considered separately. By far most loans are for a 30-year term.

Loans could have a variety of outcomes which are listed in Table 1. We restrict our study to loans that are accepted and originated, approved but not originated or denied by the lender. The first two types constitute the approved loans pool. Removing loans that are approved but not originated does not change the analysis significantly since their share is fairly small.

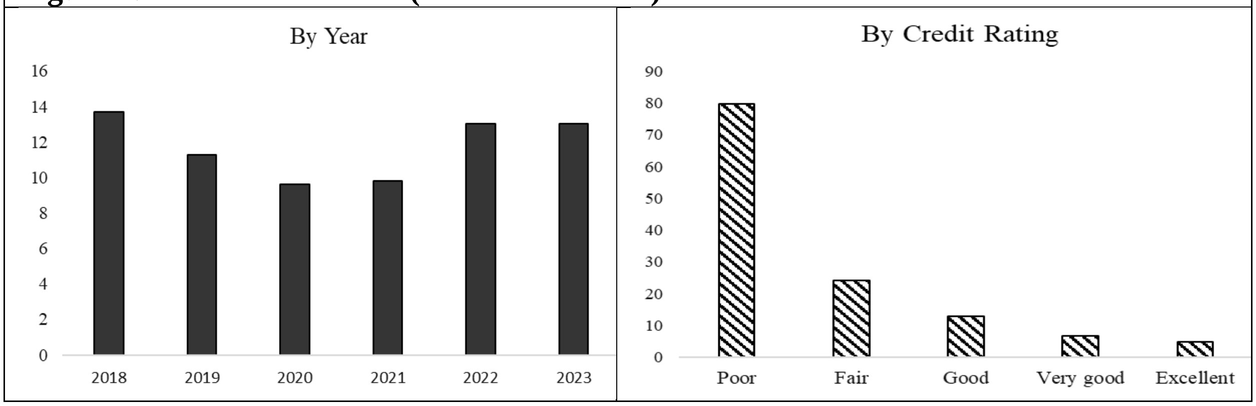
Table 1. Share of Loans by Status/Decision in 2018-23

Loan status	Number of loans	Percent of total loans
Loan originated	30,900,875	66.59
Application approved but not originated	1,132,921	2.44
Application denied by financial institution	4,931,999	10.63
Application withdrawn by applicant	7,065,149	15.22
File closed for incompleteness	2,308,843	4.98
Preapproval request denied by financial institution	17,721	0.04
Preapproval request approved but not accepted	49,054	0.11

For the purposes of the study borrowers are categorized as a particular race/ethnicity if the primary borrower states this as their first race/ethnicity and the coborrower, where present, is not

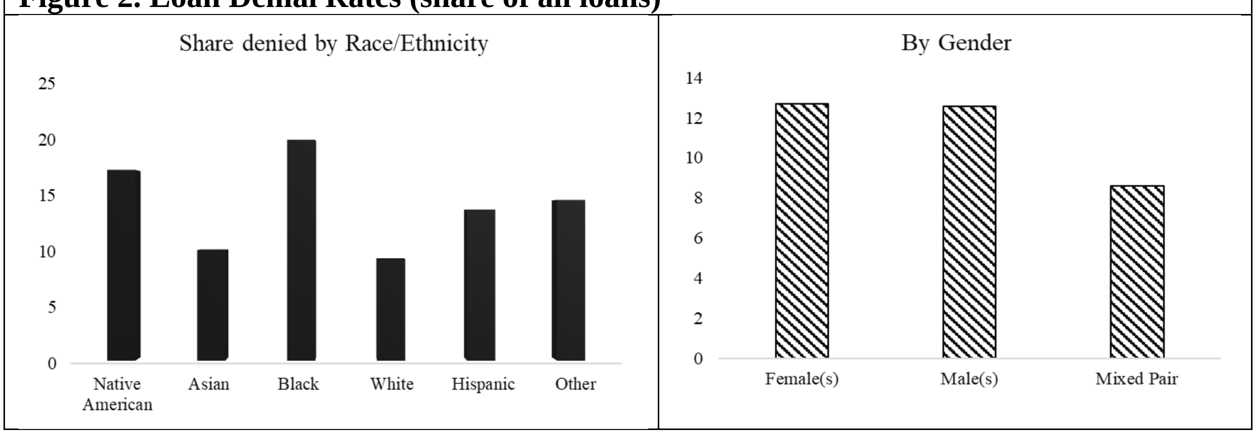
Caucasian. In cases where one borrower is a minority, and the other one is Caucasian the pair is labeled as mixed race. A subset of borrowers has not stated their race/ethnicity. We suspect that minorities may be fearful of disclosing their race expecting discrimination and that category may represent largely borrowers of color. We include those applicants as a different category under other race since we lack data to properly assign them to another group.

Figure 1. Loan Denial Rates (share of all loans)



Every year in the sample period between 9% and 13% of loans are being turned down with the rate declining until 2022 when they jump back to 13%. Denied loan applications overwhelmingly belong to applicants with poor FICO credit scoring, 80% of which are not funded (Figure 1). On the other hand, prospective home buyers with very good or excellent FICO scores are nearly always granted a mortgage with only 5% denial rates. Credit scoring follows FICO convention designating scores in the 300-580 range as Poor, 580-670 range as Fair, 670-740 range as Good, 740-800 range as Very good and above 800 as Excellent.

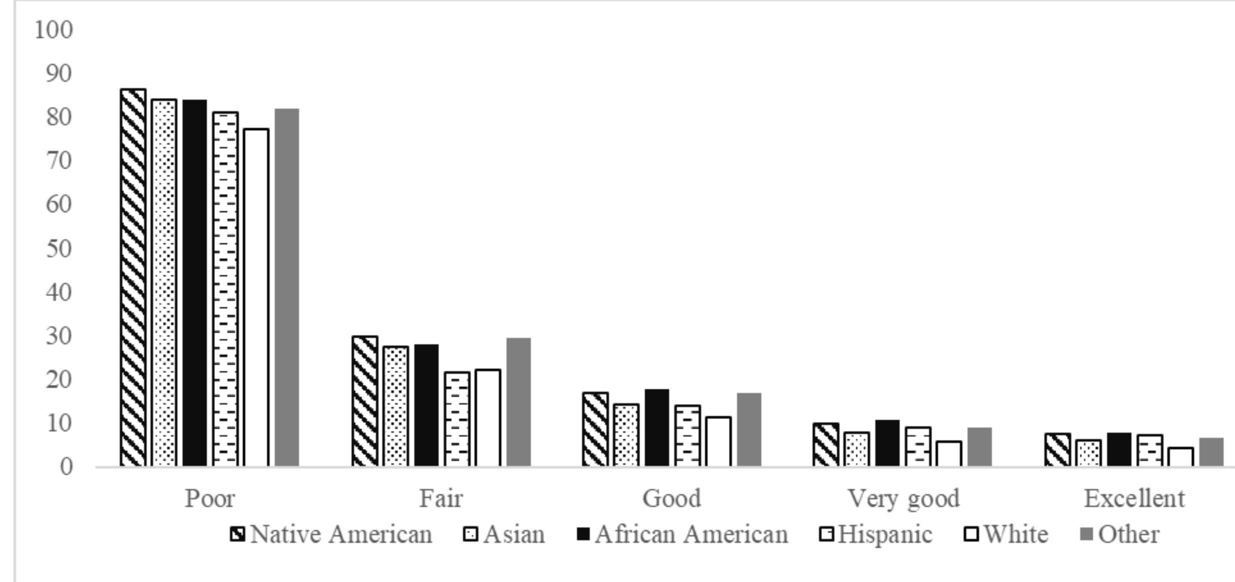
Figure 2. Loan Denial Rates (share of all loans)



Minority loan applicants tend to be refused a loan at higher rates than Caucasians except for Asians (Figure 2). African Americans face the steepest disparity: their denial rate is 20% versus

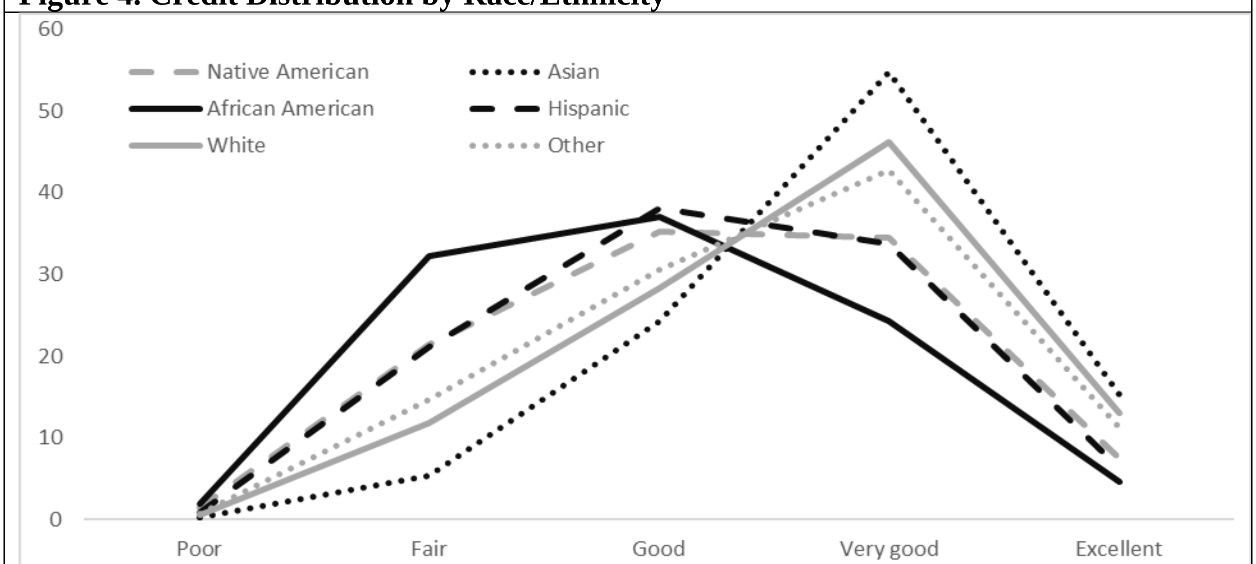
7% for Whites. Native Americans, and Hispanics follow with 17% and 14% respectively. Denial rates are similar for male and female applicants.

Figure 3. Share of Loan Denials by Race/Ethnicity per Credit Score



Even among owners of lower credit scores, minorities of every race/ethnicity tend to be denied a loan at a higher rate than Caucasians and the same is true for holders of credit that's deemed fair or good (Figure 3). However, a look at Figure 4 shows that the credit scores of African Americans is overwhelmingly bunched at fair and good with very few applicants rated very good or excellent. In contrast, Asians predominantly tend to have very good credit rating and followed

Figure 4. Credit Distribution by Race/Ethnicity



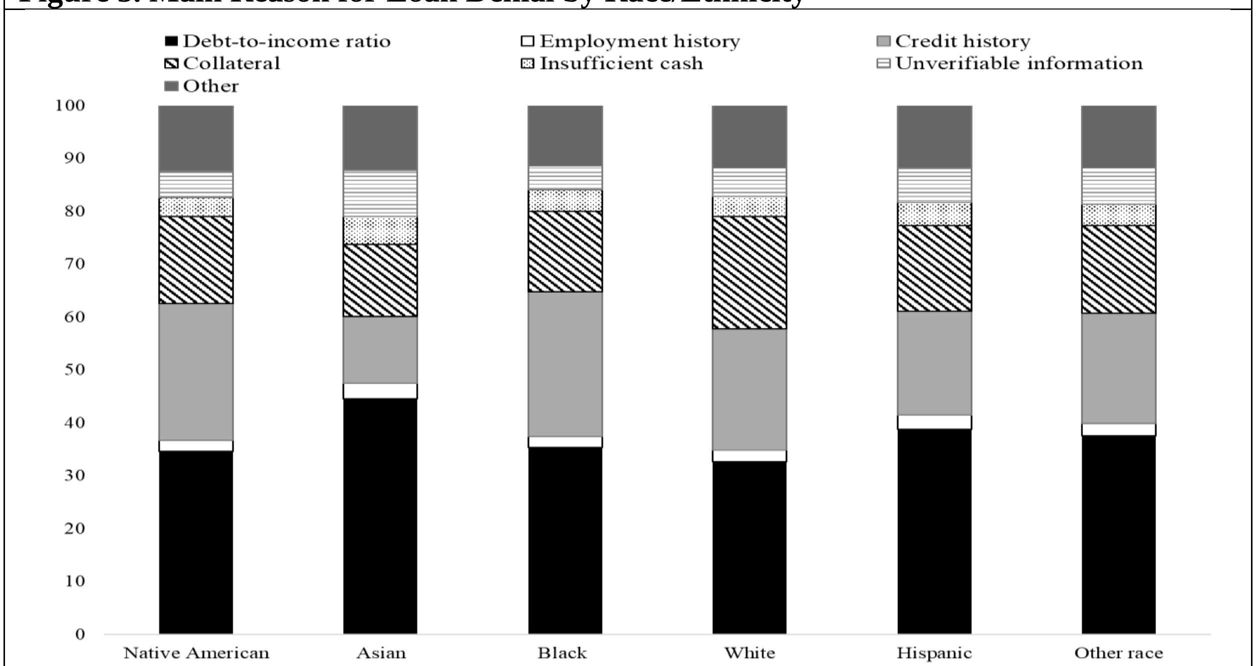
by Caucasians.

Unfortunately, HMDA does not record the full financial picture of each application. Notably there is missing information about collateral and work and financial history. Therefore, it is hard to disentangle completely the role of discrimination from that of missing financial information. Lenders are prompted to provide a reason for the denial, but the reason is not always verifiable given the data provided. The main reason for denial given by lenders is usually high DTI, insufficient credit history or insufficient collateral. The rates have been stable over time with a moderate increase in the debt-to-income ratio given as the primary reason.

Most lenders cite high DTI and insufficient credit history as reason for denial of African American applicants (Figure 5). Credit history is also the leading reason for denial for Native Americans and Caucasians. In contrast, DTI leads among Asians and Hispanics. Asians are the least likely to be denied for an incomplete credit history. Women are most frequently turned down for high DTI. Incomplete credit history is the reason most often cited for denying mortgage to mixed gender pairs.

When approved, minorities tend to be offered a high-cost mortgage which is a mortgage where the interest rate has spread (Figure 6). The rate of high-cost mortgages dropped from 25% in 2018 to 10% in 2023. 33% of African American applicants were offered high-cost loans and 26% of Hispanic borrowers. The corresponding rates for Caucasians and Asians are 12% and 5% respectively. Women are also disproportionally offered high-cost loans, while heterosexual couples get more favorable loan terms.

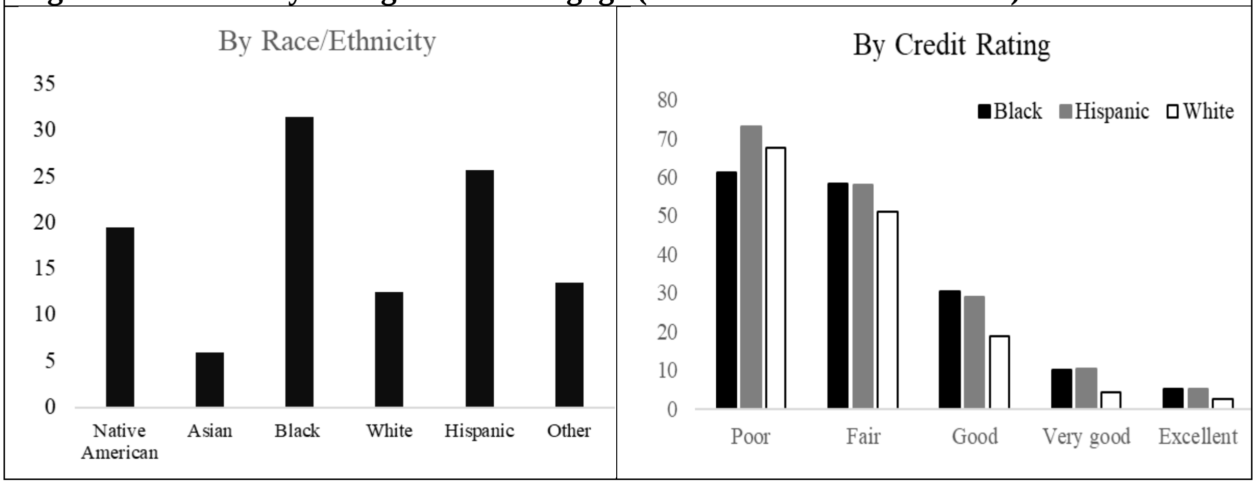
Figure 5. Main Reason for Loan Denial by Race/Ethnicity



High-cost mortgages were overwhelmingly attached to relatively low credit rating applicants. However, there are racial and ethnic disparities. Minority applicants with credit rating Fair (630-

690) or Good (690-720) are offered loans with an interest rate spread at higher rates than Caucasians with similar credit scores. Only for Excellent credit scores does the disparity almost disappear.

Figure 6. Probability of High-Cost Mortgage (share of all endorsed loans)



V. Results

a. Factors Behind Denial and The Missing Applications

This chapter presents the results of the loan denial analysis both for fixed rate mortgages (FRM) and form adjustable-rate mortgages (ARM). It looks at the probability of denial for the entire population of purchase and refinance loans as well as for a subset of high-appreciation Metropolitan Statistical Areas (MSA) areas. Then it assesses the likelihood of being provided with a specific reason for denial versus other reasons. Finally, it considers whether certain groups are more likely to be originated a high-cost loan after loan approval. All specifications have been tested for robustness by including various sets of controls. The specifications are robust to the inclusion of borrower income, time trends and locality specifications. The term and type of loans have also been considered where available. The vast majority of mortgages are FRM, and they produce different results than ARM ones, but term controls do not produce any noticeable difference since the majority of mortgages are for a 360-month term. The results are also robust to substituting the flow versions of credit scores, DTI and CLTV for their categorical counterparts. Regressing by loan amount categories results in nearly identical outcomes.

The main loan-specific metrics included are credit scores, debt-to-income (DTI) and for conventional refinance loans, combined loan-to-value (CLTV). Credit scoring follows FICO convention designating scores in the 300-580 range as Poor, 580-670 range as Fair, 670-740 range as Good, 740-800 range as Very good and above 800 as Excellent. Smaller credit score bins do not produce markedly different outcomes. DTI is separated into categories with the default being DTI below 37%, DTI between 37%-43%, DTI in the 43%-50% range and DTI above 50%. DTI below 43% is an underwriting requirement for a loan to receive a Qualified Mortgage (QM) after 2014 (CFPB, 2019). Mortgages with DTI exceeding 43% would still qualify as QMs if they meet the eligibility requirements to be insured or guaranteed by FHA. CLTV is split into the following bins: under 80%, 80%-90%, 90%-95% and above 95%. Omitting loan risk measurements makes the effect of demographic factors more pronounced confirming the theory of Bayer, Ferreira and Ross (2014) that demographics may exhibit omitted variable bias in the absence of loan risk data.

Other metrics included are whether the loan application was subject to an automatic underwriting system such as Desktop Underwriter (DU), Loan Prospector (LP), Technology Open to Approved Lenders (TOTAL) or Guaranteed Underwriting System (GUS). The size of the lender (proxied by the number of loan applications received in a year) could also point to the extent to which lender cherry-pick borrowers. Smaller lenders may be more inclined to lend to applicants with good financial history which often does not favor minorities in order to minimize losses of their modest portfolios. Larger lenders may be better positioned to absorb losses.

Table 2. Probability of Loan Denial: Fixed Rate Mortgages

	(1)	**(2)**	**(3)**	**(4)**
	Conventional		**FHA**	
	Purchase	**Refinance**	**Purchase**	**Refinance**
African	0.184***	0.340***	0.319***	0.288***
American	(0.006)	(0.005)	(0.012)	(0.014)
Hispanic	0.077***	0.259***	0.087***	-0.100***
	(0.005)	(0.005)	(0.012)	(0.015)
Asian American	0.131***	0.163***	0.250***	0.216***
	(0.005)	(0.005)	(0.027)	(0.030)
Native	0.337***	0.472***	-0.510***	-0.560***
American	(0.017)	(0.012)	(0.022)	(0.024)
Mixed race	0.024***	0.147***	-0.234***	-0.290***
couple	(0.009)	(0.008)	(0.022)	(0.025)
Other race/Not	0.194***	0.402***	0.311***	0.162***
stated	(0.006)	(0.004)	(0.017)	(0.015)
Female(s)	-0.045***	-0.076***	-0.030***	-0.026***
	(0.002)	(0.001)	(0.002)	(0.003)
Mixed gender	-0.147***	-0.205***	-0.077***	-0.166***
couple	(0.002)	(0.001)	(0.002)	(0.003)
Fair credit	-1.489***	-1.285***	-1.209***	-0.610***
	(0.012)	(0.007)	(0.009)	(0.006)
Good credit	-1.991***	-1.865***	-1.442***	-0.842***
	(0.012)	(0.007)	(0.009)	(0.006)
Very good credit	-2.237***	-2.192***	-1.498***	-0.893***
	(0.012)	(0.007)	(0.009)	(0.008)
Excellent credit	-2.281***	-2.325***	-1.461***	-0.842***

	(0.012)	(0.007)	(0.014)	(0.020)
DTI 37%-43%	-0.024***	-0.010***	-0.053***	-0.121***
	(0.002)	(0.001)	(0.003)	(0.004)
DTI 43%-50%	0.122***	0.179***	-0.019***	-0.113***
	(0.002)	(0.001)	(0.003)	(0.003)
DTI 50%+	2.186***	2.201***	0.466***	0.312***
	(0.003)	(0.002)	(0.003)	(0.003)
CLTV 80%-90%	0.122***	0.258***	-0.150***	0.021***
	(0.002)	(0.002)	(0.006)	(0.003)
CLTV 90%-95%	0.134***	0.528***	-0.104***	0.160***
	(0.002)	(0.003)	(0.006)	(0.006)
CLTV 95%+	0.203***	1.479***	-0.229***	0.185***
	(0.002)	(0.006)	(0.005)	(0.006)
Lender with 50-	-0.055***	0.068***	-0.036***	0.072***
100 applications	(0.002)	(0.002)	(0.002)	(0.004)
Lender with 100-	-0.075***	0.159***	-0.021***	0.184***
200 applications	(0.002)	(0.001)	(0.003)	(0.003)
Lender with 200-	-0.092***	0.290***	-0.111***	0.315***
400 applications	(0.003)	(0.002)	(0.004)	(0.004)
Lender with 400-	0.021***	0.235***	-0.152***	0.349***
1,000 applications	(0.004)	(0.002)	(0.008)	(0.006)
Lender with >	0.289***	0.381***	-0.276***	0.230***
1,000 applications	(0.007)	(0.004)	(0.018)	(0.018)
Automatic	-0.697***	-0.759***	-1.141***	-1.294***
Underwriting	(0.002)	(0.002)	(0.007)	(0.006)
Aut. Under.	0.084***	-0.120***	-0.070***	-0.095***
x Black	(0.007)	(0.005)	(0.012)	(0.014)
Aut. Under.	0.103***	-0.159***	0.023*	0.114***
x Hispanic	(0.006)	(0.005)	(0.013)	(0.015)
Aut. Under.	0.099***	-0.057***	-0.038	0.007
x Asian	(0.006)	(0.005)	(0.027)	(0.031)
Aut. Under.	-0.146***	-0.236***	0.703***	0.823***
x Native	(0.019)	(0.013)	(0.023)	(0.026)
Aut. Under.	0.026**	-0.096***	0.280***	0.327***
x Mixed Race	(0.010)	(0.008)	(0.023)	(0.027)
Aut. Under.	0.003	-0.194***	-0.102***	0.052***
X Other Race	(0.006)	(0.005)	(0.017)	(0.016)
High-cost area	-0.026***	-0.038***	-0.072***	-0.083***
	(0.002)	(0.001)	(0.002)	(0.003)
Year FE	Yes	Yes	Yes	Yes
Constant	0.915***	1.292***	1.224***	1.412***
	(0.013)	(0.007)	(0.012)	(0.009)
Observations	10,992,342	15,121,308	3,363,826	1,273,087

Note: Coefficients shown. P-values in brackets. *** p<0.01, ** p<0.05, * p<0.1.

The largest predictor of loan denials in the conventional category is having a Poor credit score which is the default (Table 2). All other credit categories have negative signs implying that they are less associated with denial relative to the lowest rating. The largest magnitude change is from Fair to Good and Good to Very Good. Having Excellent credit does not confer significant advantage over Very Good. For FHA loans, the benefits of a higher credit score peter out after Good rating.

Having DTI below 43% is associated with lower probabilities of denial than the baseline category under 37%. Denial rates are significantly higher for DTI above 50%. The effects for FHA mortgages are considerably more muted than for conventional loans. CLTV above 80% is also modestly associated with higher probability of denial except for a conventional refinance loan where the contribution is quite strong. FHA loans which tend to have higher CLTV values do not exhibit the same pattern.

Lenders receiving 50 to 400 applications for a conventional mortgage a year have a lower likelihood of denying a request than very small financial intuitions receiving less than 50 applications. However, very large lenders, those receiving upwards of 400 applications a year are associated with a greater probability of a loan being turned down. The same is not true for FHA mortgages where the larger the lender, the smaller the probably of loan denial.

Being a minority is associated with greater likelihood of denial ceteris paribus. In terms of a conventional mortgage, the highest probability of denial is faced by Native Americans followed by African Americans. Asian Americans also see their application turned down more often than Caucasians but to a lesser extent than that of the first two groups. Hispanics have a harder time refinancing than purchasing a new home. Mixed-race couples (couples where one borrower is Caucasian, and the other is non-Caucasian) also have slightly elevated rates of denial relative to Caucasians but are quite close. African Americans, Asian Americans and Other races face higher likelihood of denial for FHA loans than for conventional mortgages. Hispanics tend to be treated the same while mixed race couples are treated more favorably when applying for an FHA mortgage. Female applicants and mixed gender couples face lower probabilities of their mortgage loan application being turned down than males only. Automatic underwriting is associated with a lower likelihood of denial, but the odds are more modest for African Americans and Hispanics applying for a conventional new purchase loan. Refinance loans show the opposite pattern for all minorities.

Table 3. Probability of Loan Denial: Adjustable-Rate Mortgages

	(1)	(2)	(3)	(4)
	Conventional		FHA	
	Purchase	Refinance	Purchase	Refinance
African	0.227***	0.202***	0.181	-0.221
American	(0.012)	(0.015)	(0.316)	(0.250)
Hispanic	0.192***	0.242***	0.568	-0.510
	(0.012)	(0.017)	(0.383)	(0.364)
Asian American	0.109***	0.063***	-0.227	3.898
	(0.008)	(0.009)	(0.447)	(114.850)
Native	0.338***	0.315***	4.640	4.013
American	(0.034)	(0.040)	(115.984)	(197.516)
Mixed race	0.110***	0.131***	-4.553	0.306
couple	(0.017)	(0.021)	(168.877)	(0.661)

Other race/Not stated	0.262***	0.249***	0.148	-0.620***
	(0.011)	(0.012)	(0.460)	(0.223)
Female(s)	-0.037***	-0.068***	-0.008	-0.074**
	(0.006)	(0.007)	(0.048)	(0.036)
Mixed gender couple	-0.107***	-0.161***	-0.105**	-0.144***
	(0.005)	(0.005)	(0.049)	(0.041)
Fair credit	-0.770***	-0.438***	-1.930***	-0.506***
	(0.020)	(0.016)	(0.193)	(0.136)
Good credit	-1.322***	-0.886***	-2.264***	-0.619***
	(0.019)	(0.015)	(0.193)	(0.137)
Very good credit	-1.591***	-1.189***	-2.331***	-0.703***
	(0.019)	(0.015)	(0.199)	(0.146)
Excellent credit	-1.684***	-1.356***	-2.700***	-0.897***
	(0.020)	(0.016)	(0.294)	(0.289)
DTI 37%-43%	-0.066***	0.014**	-0.128	0.007
	(0.006)	(0.006)	(0.088)	(0.053)
DTI 43%-50%	0.297***	0.404***	0.029	-0.060
	(0.006)	(0.007)	(0.074)	(0.046)
DTI 50%+	1.709***	1.844***	0.585***	0.366***
	(0.007)	(0.007)	(0.068)	(0.044)
CLTV 80%-90%	0.154***	0.273***	-0.184*	-0.036
	(0.006)	(0.007)	(0.106)	(0.037)
CLTV 90%-95%	0.370***	0.620***	-0.180*	0.183**
	(0.007)	(0.014)	(0.101)	(0.073)
CLTV 95%+	0.124***	0.824***	-0.421***	0.200***
	(0.007)	(0.019)	(0.091)	(0.071)
Lender with 50-100 applications	-0.015**	0.087***	-0.000	0.213***
	(0.006)	(0.006)	(0.053)	(0.052)
Lender with 100-200 applications	0.003	0.152***	0.118**	0.171***
	(0.007)	(0.007)	(0.053)	(0.045)
Lender with 200-400 applications	0.154***	0.170***	0.317***	0.203***
	(0.008)	(0.008)	(0.079)	(0.054)
Lender with 400-1,000 applications	0.482***	0.038***	0.745***	0.117
	(0.011)	(0.013)	(0.139)	(0.096)
Lender with > 1,000 applications	1.465***	0.487***	0.811	0.641
	(0.012)	(0.038)	(0.502)	(0.502)
Automatic Underwriting	-0.280***	-0.154***	-1.144***	-0.950***
	(0.007)	(0.006)	(0.188)	(0.132)
Aut. Under. x Black	0.052**	0.063***	0.035	0.467*
	(0.021)	(0.023)	(0.320)	(0.254)
Aut. Under. x Hispanic	-0.066***	-0.167***	-0.453	0.707*
	(0.021)	(0.024)	(0.388)	(0.370)
Aut. Under. x Asian	0.037***	-0.006	0.425	-3.888
	(0.013)	(0.014)	(0.454)	(114.850)
Aut. Under. x Native	-0.215***	-0.096*	-4.353	-3.675
	(0.058)	(0.055)	(115.984)	(197.516)
Aut. Under. x Mixed Race	-0.069**	-0.111***	4.573	-0.246
	(0.027)	(0.030)	(168.877)	(0.671)
Aut. Under.	-0.119***	-0.073***	0.068	0.852***

X Other Race	(0.017)	(0.017)	(0.465)	(0.227)
High-cost area	0.198***	0.123***	-0.057	-0.177***
	(0.005)	(0.005)	(0.041)	(0.033)
Year FE	Yes	Yes	Yes	Yes
Constant	-0.305***	-0.376***	1.986***	1.224***
	(0.020)	(0.016)	(0.281)	(0.188)
Observations	801,453	563,580	7,114	6,744

Note: Coefficients shown. P-values in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Adjustable-rate mortgages are a far smaller share of the mortgage pool and that is especially true for FHA ARM loans (Figure 3). While the loan products in this category are more varied, they offer results largely similar to that of fixed rate mortgages for conventional loans. DTI above 50% and to a lesser degree CLTV above 90% as well as a poor credit score are still the strongest predictors of denial albeit with lower coefficients across the board compared to FRM mortgages. Demographic effects also vary slightly African Americans and Asians are denied an ARM mortgage at somewhat reduced rates than a FRM mortgage. However, this is likely due to selection bias rather than greater leniency on the part of lenders. ARM borrowers may plan to live in their home for a short period of time or may expect their income or credit standing to increase and to thereby refinance at better terms compared to FRM applicants who may have already achieved their peak financial standing. FHA ARM loans are a very small pool and do not exhibit any demographic patterns.

While these results are indicative of discrimination, the effect of demographics may be somewhat overstated. The main reason is that although HMDA provides expanded loan-specific metrics that capture risk, it still does not include all relevant metrics such as credit history, employment history, the opportunity to shop around for better loan offers and general financial literacy on the part of borrower. In addition, outcomes may reflect disparate impact, being consistent with overt discrimination when in fact it is owing to financial factors that disproportionally concern minorities such as insufficient credit history.

Table 4. Probability of loan denial in Top 20 MSA with Highest Price Appreciation

	(1)	(2)	(3)	(4)
	Conventional		FHA	
	Purchase	Refinance	Purchase	Refinance
African	0.339***	0.480***	0.330***	0.389***
American	(0.012)	(0.009)	(0.028)	(0.025)
Hispanic	0.223***	0.297***	-0.118***	-0.028
	(0.009)	(0.008)	(0.026)	(0.021)
Asian American	0.020***	-0.086***	0.138***	0.339***
	(0.006)	(0.005)	(0.044)	(0.039)
Native	0.336***	0.507***	-0.493***	-0.268***
American	(0.029)	(0.020)	(0.051)	(0.039)
Mixed race	-0.021	0.057***	-0.336***	-0.124***
couple	(0.013)	(0.011)	(0.050)	(0.039)
Other race/Not	0.155***	0.267***	0.144***	0.384***

stated	(0.008)	(0.006)	(0.033)	(0.026)
Female(s)	-0.039***	-0.070***	-0.049***	-0.026***
	(0.003)	(0.002)	(0.005)	(0.006)
Mixed gender	-0.139***	-0.186***	-0.116***	-0.173***
couple	(0.003)	(0.002)	(0.005)	(0.006)
Fair credit	-1.601***	-1.283***	-1.147***	-0.626***
	(0.037)	(0.014)	(0.022)	(0.011)
Good credit	-2.091***	-1.825***	-1.331***	-0.830***
	(0.037)	(0.014)	(0.022)	(0.011)
Very good credit	-2.334***	-2.144***	-1.367***	-0.846***
	(0.037)	(0.014)	(0.022)	(0.015)
Excellent credit	-2.386***	-2.246***	-1.359***	-0.725***
	(0.037)	(0.014)	(0.030)	(0.035)
DTI 37%-43%	-0.022***	-0.012***	-0.055***	-0.127***
	(0.003)	(0.002)	(0.008)	(0.008)
DTI 43%-50%	0.137***	0.174***	-0.029***	-0.111***
	(0.003)	(0.002)	(0.007)	(0.007)
DTI 50%+	2.046***	2.236***	0.435***	0.317***
	(0.005)	(0.003)	(0.007)	(0.006)
CLTV 80%-90%	0.149***	0.330***	-0.130***	-0.069***
	(0.003)	(0.003)	(0.012)	(0.006)
CLTV 90%-95%	0.145***	0.557***	-0.137***	0.100***
	(0.003)	(0.005)	(0.012)	(0.011)
CLTV 95%+	0.261***	1.544***	-0.261***	0.096***
	(0.005)	(0.012)	(0.010)	(0.011)
Lender with 50-	-0.042***	0.045***	-0.032***	0.091***
100 applications	(0.003)	(0.002)	(0.005)	(0.007)
Lender with 100-	-0.063***	0.146***	-0.036***	0.237***
200 applications	(0.003)	(0.002)	(0.006)	(0.006)
Lender with 200-	-0.053***	0.280***	-0.123***	0.345***
400 applications	(0.005)	(0.003)	(0.009)	(0.007)
Lender with 400-	0.040***	0.226***	-0.185***	0.328***
1,000 applications	(0.007)	(0.004)	(0.018)	(0.012)
Lender with >	0.146***	0.347***	-0.166***	0.221***
1,000 applications	(0.013)	(0.007)	(0.035)	(0.030)
Automatic	-0.593***	-0.740***	-1.090***	-1.100***
Underwriting	(0.004)	(0.003)	(0.018)	(0.011)
Aut. Under.	-0.082***	-0.267***	-0.091***	-0.220***
x Black	(0.013)	(0.010)	(0.029)	(0.026)
Aut. Under.	-0.096***	-0.222***	0.169***	-0.003
x Hispanic	(0.010)	(0.008)	(0.027)	(0.022)
Aut. Under.	0.217***	0.197***	0.068	-0.108***
x Asian	(0.007)	(0.006)	(0.045)	(0.041)
Aut. Under.	-0.156***	-0.270***	0.660***	0.502***
x Native	(0.032)	(0.021)	(0.054)	(0.043)
Aut. Under.	0.057***	-0.033***	0.355***	0.127***
x Mixed Race	(0.015)	(0.012)	(0.051)	(0.042)
Aut. Under.	0.013	-0.088***	0.000	-0.193***
X Other Race	(0.009)	(0.007)	(0.034)	(0.027)

Year FE	Yes	Yes	Yes	Yes
Constant	0.907***	1.174***	1.089***	1.133***
	(0.037)	(0.014)	(0.029)	(0.016)
Observations	3,356,654	5,780,838	700,228	346,540

Note: Coefficients shown. P-values in brackets. *** p<0.01, ** p<0.05, * p<0.1.

We also consider a subset consisting of the top 20 MSAs with the highest house price appreciation from 2000 to 2023 (Table 4). Since in high valuation areas borrower leverage may be higher for all borrowers, the effect of demographics may be lower. The metro areas are listed in Appendix I. The effect of demographics is overall the same as in the nationwide sample, but the destruction is different. The effect of race and ethnicity is much higher for loans that are reviewed by a human compared to automatic underwriting which reduces the effect of bias. This is likely due to the fact that high price appreciation areas are such because of limited housing stock so homebuyers face more competition. Hence sellers and lenders could afford to cherry pick their clients when given the opportunity.

Table 5. Counterfactual Analysis in the Contribution of Minorities to Missing Loans

	Percent	**Number of Loans**
African American	0.32	102,332
Hispanic	0.13	41,540
Asian American	0.08	26,340
Native American	0.04	12,530
Mixed race	0.01	4,578
Other/Not stated	0.32	102,368
All non-Caucasian	**0.90**	**289,657**

Lastly, counterfactual analysis considers whether the share of denied loans would change if the applicants were all Caucasian with the credit score, CLTV and DTI unchanged (Table 5). The counterfactual shows that there would have been one percentage point more originated loans given the total number of loans considered is 32,129,467 for the period of 2018-23 that implies about 289,657 or almost 1% of all loans are missing which would have been originated were the applicants Caucasian rather than minority. Over one third of those missing originated loans belong to African Americans. However, the number may be somewhat overestimated considering that HMDA does not capture all financial factors relating to a mortgage application but likely not by a lot since the important risk metrics such as credit scores, CLTV and DTI are already accounted for.

b. Reasons for Denial

Next, we consider the primary reason lenders cite for turning down a loan application. The specific reason is tested against the probability of not denying an application so while the first part tested the likelihood of denial for any reason, this section tests the likelihood of denial for a specific reason and how this matches the loan risk factors that HMDA accounts for. The two options that can be linked to loan data contained in HMDA are high DTI and credit history. Three others are not directly testable, but we consider them anyway: employment history, insufficient collateral, and cash issues. Lenders can specify up to four reasons and we focus on the primary reason given. They are not required to provide a reason so we cannot include cases where none was specified.

As expected for conventional loans, having a DTI ratio above 50% is strongly associated with denial (Table 6). Similarly, having a Poor credit score is highly predictive of denial for insufficient credit history. The other three denial reasons, namely employment history, collateral and insufficient cash are not observable given the data at hand. High DTI is somewhat predictive of denial for employment history and insufficient cash. High CLTV is grounds for denial for insufficient collateral. A Poor (and in some cases Fair) credit score is positively associated with a refused mortgage for all specific reasons but likely these capture the omitted variable effects since the main reason is not observable for the last three categories. Automatic underwriting is associated with a lower probability of denial, especially in the case of credit history.

Table 6. Probability of denial for a specific reason: Conventional Loans

	(1)	(2)	(3)	(4)	(5)
	DTI	**Employment**	**Credit history**	**Collateral**	**Cash**
African	0.217***	0.128***	0.257***	0.014**	0.205***
American	(0.006)	(0.014)	(0.005)	(0.007)	(0.010)
Hispanic	0.090***	0.163***	0.082***	0.074***	0.184***
	(0.006)	(0.012)	(0.005)	(0.006)	(0.009)
Asian American	0.208***	0.113***	-0.071***	-0.048***	0.168***
	(0.006)	(0.012)	(0.006)	(0.006)	(0.008)
Native	0.333***	0.272***	0.409***	0.216***	0.320***
American	(0.017)	(0.035)	(0.013)	(0.018)	(0.025)
Mixed race	0.096***	0.085***	0.131***	0.019*	0.069***
couple	(0.011)	(0.022)	(0.008)	(0.010)	(0.016)
Other race/Not	0.236***	0.179***	0.253***	0.106***	0.261***
stated	(0.006)	(0.013)	(0.005)	(0.006)	(0.009)
Female(s)	-0.062***	-0.039***	-0.027***	-0.038***	-0.056***
	(0.002)	(0.004)	(0.002)	(0.002)	(0.003)
Mixed gender	-0.164***	-0.125***	-0.050***	-0.148***	-0.152***
couple	(0.002)	(0.004)	(0.002)	(0.002)	(0.003)
Fair credit	-0.254***	-0.189***	-1.485***	0.214***	0.244***
	(0.011)	(0.025)	(0.005)	(0.018)	(0.027)
Good credit	-0.632***	-0.380***	-2.398***	-0.031*	0.003
	(0.011)	(0.025)	(0.005)	(0.018)	(0.027)
Very good credit	-0.923***	-0.549***	-3.076***	-0.234***	-0.199***
	(0.011)	(0.025)	(0.005)	(0.018)	(0.027)
Excellent credit	-1.062***	-0.633***	-3.327***	-0.316***	-0.329***
	(0.011)	(0.025)	(0.006)	(0.018)	(0.027)
DTI 37%-43%	0.193***	0.019***	-0.110***	-0.012***	0.072***

	(0.003)	(0.004)	(0.002)	(0.002)	(0.003)
DTI 43%-50%	0.803***	0.115***	-0.060***	0.015***	0.200***
	(0.002)	(0.004)	(0.002)	(0.002)	(0.003)
DTI 50%+	3.347***	1.269***	1.056***	0.728***	0.751***
	(0.002)	(0.005)	(0.003)	(0.004)	(0.006)
CLTV 80%-90%	0.030***	0.034***	-0.044***	0.353***	0.107***
	(0.002)	(0.005)	(0.003)	(0.002)	(0.003)
CLTV 90%-95%	-0.017***	0.078***	0.000	0.235***	0.112***
	(0.003)	(0.005)	(0.003)	(0.002)	(0.004)
CLTV 95%+	-0.016***	0.130***	-0.113***	0.413***	0.111***
	(0.004)	(0.006)	(0.004)	(0.002)	(0.005)
Lender with 50-	-0.008***	0.009**	0.029***	-0.003*	-0.005
100 applications	(0.002)	(0.004)	(0.002)	(0.002)	(0.003)
Lender with 100-	0.029***	0.027***	0.090***	0.046***	-0.007**
200 applications	(0.002)	(0.004)	(0.002)	(0.002)	(0.003)
Lender with 200-	0.115***	0.081***	0.188***	0.128***	0.015***
400 applications	(0.003)	(0.005)	(0.003)	(0.002)	(0.004)
Lender with 400-	0.108***	0.072***	0.325***	0.070***	-0.001
1,000 applications	(0.004)	(0.008)	(0.004)	(0.004)	(0.007)
Lender with >	0.423***	0.269***	0.876***	0.038***	0.198***
1,000 applications	(0.006)	(0.012)	(0.005)	(0.008)	(0.011)
Automatic	-0.512***	-0.250***	-0.861***	-0.283***	-0.421***
Underwriting	(0.003)	(0.005)	(0.002)	(0.002)	(0.004)
High-cost area	-0.017***	-0.035***	-0.032***	-0.013***	0.054***
	(0.002)	(0.003)	(0.002)	(0.001)	(0.003)
Year FE	Yes	Yes	Yes	Yes	Yes
Constant	-1.718***	-2.412***	0.927***	-1.790***	-2.357***
	(0.012)	(0.025)	(0.006)	(0.018)	(0.027)
Observations	25,610,524	24,844,807	25,264,492	25,158,119	24,882,280

Note: Coefficients shown. P-values in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Beyond loan-specific factors, demographic factors are also associated with specific denial reasons. Since the reasons employment history, insufficient collateral and cash issues cannot be directly linked to loan data within HMDA, it is possible that in those cases some demographic factors reflect omitted variable bias. African Americans and Native Americans are most likely to be turned down for high DTI and insufficient credit history, but they are more likely to be turned down for other reasons than their Caucasian counterparts. For Hispanics, cash reasons followed by employment history dominate. Asian Americans are most likely to experience denial for high DTI. Borrowers with unspecified race has the same experience as African Americans likely because there is some race overlap.

FHA loans have the same patterns for the effect of DTI levels and credit scores but more muted (Table 7). African Americans most often are given the high DTI or credit history reasons for loan denial, and the effect is largely in line with conventional loans. Hispanics, Native Americans and mixed-race couples see less not more probability of denial for all reasons. Since FHA loans have higher DTI and lower downpayment to begin with and more often conform to standard underwriting procedures, demographic effects tend to have less contribution to denial levels except for African Americans.

Table 7. Probability of denial for a specific reason: FHA Loans

	(1)	(2)	(3)	(4)	(5)
	DTI	Employment	Credit history	Collateral	Cash
African American	0.264***	0.222***	0.216***	0.119***	0.188***
	(0.013)	(0.027)	(0.012)	(0.019)	(0.027)
Hispanic	-0.066***	0.169***	-0.134***	-0.031	0.045
	(0.014)	(0.027)	(0.013)	(0.019)	(0.028)
Asian American	0.284***	0.397***	0.044	-0.003	0.045
	(0.026)	(0.052)	(0.028)	(0.042)	(0.066)
Native American	-0.143***	-0.370***	-0.499***	-0.492***	-0.417***
	(0.026)	(0.058)	(0.022)	(0.033)	(0.059)
Mixed race couple	-0.201***	-0.100**	-0.238***	-0.314***	-0.193***
	(0.028)	(0.051)	(0.022)	(0.036)	(0.058)
Other race/Not stated	0.243***	0.150***	0.149***	0.102***	0.140***
	(0.016)	(0.036)	(0.015)	(0.023)	(0.036)
Female(s)	-0.009***	-0.024***	-0.016***	-0.022***	-0.036***
	(0.003)	(0.006)	(0.003)	(0.003)	(0.005)
Mixed gender couple	-0.075***	-0.005	-0.034***	-0.153***	-0.118***
	(0.003)	(0.006)	(0.003)	(0.003)	(0.005)
Fair credit	-0.538***	-0.176***	-1.239***	0.139***	-0.158***
	(0.009)	(0.021)	(0.005)	(0.012)	(0.019)
Good credit	-0.729***	-0.283***	-1.694***	0.102***	-0.314***
	(0.009)	(0.022)	(0.006)	(0.012)	(0.020)
Very good credit	-0.758***	-0.346***	-2.044***	0.087***	-0.387***
	(0.010)	(0.023)	(0.009)	(0.013)	(0.021)
Excellent credit	-0.682***	-0.423***	-2.155***	0.114***	-0.447***
	(0.018)	(0.041)	(0.028)	(0.019)	(0.035)
DTI 37%-43%	0.098***	-0.006	-0.108***	-0.101***	0.018***
	(0.007)	(0.008)	(0.004)	(0.004)	(0.007)
DTI 43%-50%	0.502***	0.021***	-0.173***	-0.138***	0.038***
	(0.006)	(0.007)	(0.003)	(0.004)	(0.006)
DTI 50%+	1.552***	0.218***	-0.087***	-0.086***	0.095***
	(0.005)	(0.007)	(0.003)	(0.004)	(0.006)
CLTV 80%-90%	-0.126***	-0.038***	-0.277***	0.087***	0.131***
	(0.004)	(0.010)	(0.004)	(0.004)	(0.009)
CLTV 90%-95%	-0.232***	-0.013	-0.469***	-0.171***	0.159***
	(0.005)	(0.011)	(0.005)	(0.005)	(0.009)
CLTV 95%+	-0.417***	-0.042***	-0.666***	-0.521***	0.112***
	(0.003)	(0.007)	(0.003)	(0.003)	(0.006)
Lender with 50-100 applications	-0.020***	-0.018***	0.016***	0.059***	-0.019***
	(0.003)	(0.006)	(0.003)	(0.004)	(0.005)
Lender with 100-200 applications	0.051***	-0.017***	0.085***	0.154***	-0.008
	(0.003)	(0.006)	(0.003)	(0.003)	(0.006)
Lender with 200-400 applications	0.077***	-0.038***	0.150***	0.247***	-0.030***
	(0.004)	(0.009)	(0.004)	(0.004)	(0.008)

Lender with 400-	0.015*	-0.088***	0.241***	0.297***	-0.115***
1,000 applications	(0.008)	(0.018)	(0.007)	(0.008)	(0.017)
Lender with >	-0.032*	-0.263***	0.142***	-0.016	-0.096**
1,000 applications	(0.019)	(0.052)	(0.018)	(0.024)	(0.037)
Automatic	-1.299***	-0.673***	-1.378***	-0.562***	-0.532***
Underwriting	(0.007)	(0.016)	(0.006)	(0.009)	(0.016)
High-cost area	-0.059***	-0.077***	-0.057***	-0.052***	-0.052***
	(0.003)	(0.006)	(0.003)	(0.003)	(0.005)
Year FE	Yes	Yes	Yes	Yes	Yes
Constant	-0.498***	-1.783***	1.473***	-0.990***	-1.789***
	(0.012)	(0.027)	(0.009)	(0.015)	(0.025)
Observations	4,052,911	3,846,781	4,023,405	3,949,287	3,857,194

Note: Coefficients shown. P-values in brackets. *** p<0.01, ** p<0.05, * p<0.1.

c. The Demographic of High-Cost Loans

The last part concerns the likelihood of being offered a high-cost loan (a loan with a spread) for approved mortgages. As expected, the biggest predictor of a high-cost mortgage is having subpar credit (Table 8). Followed by DTI above 50% and CLTV above 90%. Very large lenders (those with over 1,000 applications a year) also tend to offer high-cost mortgages in more cases but automatic underwriting platforms contribute to fewer loans with a spread. High-cost areas, in contrast, are less likely to imply a positive rate spread on mortgages since other risk factors such as DTI and credit history are already factored in.

Among those who have their loans approved, Hispanics are most likely to be offered a high-cost conventional purchase or refinance loan than Caucasians. African Americans tend to end up with an interest rate markup when applying for a FHA new purchase loan. Native Americans also face higher prospects of a conventional loan with a spread but the same is not true for FHA loans. Women also see a higher probability of high-cost refinance loans, both conventional and FHA.

Table 8. Probability of Being Offered a High-Cost Loan

	(1)	(2)	(3)	(4)
	Conventional		FHA	
	Purchase	Refinance	Purchase	Refinance
African	-0.314***	0.020***	0.210***	0.170***
American	(0.006)	(0.007)	(0.018)	(0.031)
Hispanic	0.832***	0.347***	0.094***	0.160***
	(0.005)	(0.006)	(0.017)	(0.028)
Asian American	0.291***	-0.148***	-0.240***	-0.068
	(0.004)	(0.007)	(0.040)	(0.065)
Native	0.167***	0.149***	-1.475***	-1.031***
American	(0.018)	(0.019)	(0.029)	(0.057)
Mixed race	-0.208***	-0.064***	-0.821***	-0.493***
couple	(0.009)	(0.010)	(0.027)	(0.049)

Other race/Not stated	0.100***	-0.069***	-0.012	-0.081**
	(0.006)	(0.007)	(0.025)	(0.034)
Female(s)	0.000	0.040***	0.014***	0.024***
	(0.002)	(0.002)	(0.002)	(0.004)
Mixed gender couple	-0.017***	-0.046***	0.036***	0.052***
	(0.001)	(0.001)	(0.002)	(0.004)
Fair credit	0.171***	-0.195***	-0.978***	-0.960***
	(0.019)	(0.012)	(0.016)	(0.010)
Good credit	-0.695***	-0.808***	-1.449***	-1.360***
	(0.019)	(0.012)	(0.016)	(0.010)
Very good credit	-1.430***	-1.413***	-1.657***	-1.355***
	(0.019)	(0.012)	(0.016)	(0.011)
Excellent credit	-1.496***	-1.535***	-1.769***	-1.238***
	(0.019)	(0.012)	(0.018)	(0.023)
DTI 37%-43%	0.033***	0.072***	0.025***	-0.041***
	(0.001)	(0.002)	(0.002)	(0.004)
DTI 43%-50%	0.077***	0.074***	0.000	-0.077***
	(0.002)	(0.002)	(0.002)	(0.004)
DTI 50%+	0.243***	0.185***	-0.049***	-0.159***
	(0.005)	(0.005)	(0.002)	(0.004)
CLTV 80%-90%	0.331***	-0.068***	-0.160***	-0.224***
	(0.002)	(0.002)	(0.005)	(0.004)
CLTV 90%-95%	0.650***	0.239***	0.238***	-0.001
	(0.002)	(0.004)	(0.005)	(0.007)
CLTV 95%+	0.840***	0.389***	0.587***	0.277***
	(0.002)	(0.011)	(0.005)	(0.007)
Lender with 50-100 applications	-0.155***	-0.089***	-0.057***	-0.018***
	(0.002)	(0.002)	(0.002)	(0.004)
Lender with 100-200 applications	-0.246***	-0.051***	-0.149***	0.002
	(0.002)	(0.002)	(0.002)	(0.004)
Lender with 200-400 applications	-0.259***	-0.048***	-0.186***	0.009*
	(0.002)	(0.002)	(0.003)	(0.005)
Lender with 400-1,000 applications	-0.052***	-0.164***	-0.133***	0.003
	(0.004)	(0.003)	(0.006)	(0.008)
Lender with > 1,000 applications	0.396***	-0.393***	-0.289***	-0.412***
	(0.006)	(0.009)	(0.013)	(0.026)
Automatic Underwriting	-0.651***	-0.526***	0.079***	0.267***
	(0.002)	(0.002)	(0.010)	(0.013)
Aut. Under. x Black	0.465***	0.117***	-0.201***	-0.182***
	(0.007)	(0.007)	(0.019)	(0.031)
Aut. Under. x Hispanic	-0.606***	-0.305***	-0.054***	-0.220***
	(0.005)	(0.007)	(0.017)	(0.029)
Aut. Under. x Asian	-0.488***	-0.211***	0.035	-0.072
	(0.005)	(0.007)	(0.040)	(0.066)
Aut. Under. x Native	-0.098***	-0.061***	1.398***	0.941***
	(0.019)	(0.020)	(0.030)	(0.059)
Aut. Under. x Mixed Race	0.271***	0.067***	0.803***	0.423***
	(0.010)	(0.011)	(0.028)	(0.050)
Aut. Under.	-0.081***	0.030***	-0.030	0.013

X Other Race	(0.006)	(0.007)	(0.025)	(0.034)
High-cost area	-0.111***	-0.346***	-0.173***	-0.289***
	(0.001)	(0.001)	(0.002)	(0.003)
Year FE	Yes	Yes	Yes	Yes
Constant	0.074***	0.473***	1.494***	1.250***
	(0.019)	(0.012)	(0.019)	(0.016)
Observations	11,045,509	13,746,358	2,982,522	840,050

Note: Coefficients shown. P-values in brackets. *** p<0.01, ** p<0.05, * p<0.1.

The results show that a low credit score and high DTI rates are by far the biggest predictor of loan denials and of ending up with a high-cost loan for approved mortgage applications. However, even accounting for such loan risk factors, borrower demographic characteristics still matter. African Americans, Native Americans and Hispanics tend to be turned down more often than their Caucasian counterparts and when they are approved, they end up more frequently with a high-cost loan. They also have poorer credit history and shorter employment records which weigh them down when applying for a mortgage. Automatic underwriting tends to reduce this bias to some extent, but it still persists. While in high appreciation areas the degree of discrimination is attenuated, it is a visible factor. Gender is not a big predictor of denial rates, and neither is living in a high-cost area.

VI. Conclusion

Discrimination in mortgage application decisions has been a persistent factor that has hampered equal access to mortgages. Until recently, estimating the degree of discrimination has been limited by the lack of detailed loan-level data. In such cases, researchers have frequently overestimated the degree of discrimination or attributed it to discrimination denials which in fact have been due to loan-specific riskiness.

Better data such as the expanded HMDA data allows the separation of demographic factors from loan-specific contributors to the probability of denial. This study takes advantage of the data by contributing to the literature of mortgage denial in three important steps. First, it evaluates the likelihood of loan denial using loan-specific metrics of riskiness as well as applicant demographic data. Second, it estimates the probability of various reasons for denial given risk metrics that relate to the reason provided as well as demographics. Lastly, it looks whether minorities are more likely to be offered a high-cost loan when approved.

It finds that a poor credit score and high DTI and CLTV rates are by far the biggest predictor of loan denials and of ending up with a high-cost loan for approved mortgage applications. Large lenders and automatic underwriting reduce the probability of denial. Despite accounting for such loan-risk and lender-specific factors, being a minority still matters. African Americans, Native Americans and Hispanics tend to be turned down more often than their Caucasian counterparts and when they are approved, they end up more frequently with a high-cost loan. African

Americans and Native Americans are more likely to be turned down for high DTI and insufficient credit history than their Caucasian counterparts. For Hispanics, employment history and cash issues dominate. Gender as well as being in a high-cost area is not equitable with higher denial rates.

The study implies that even accounting for available loan-specific risk factors, the relevance of demographic persists. The paper estimates that 290,000 more mortgages would be expected for minorities based on the observable characteristics in our data set. It is difficult to assess much of this difference is due to discrimination and how much is due to characteristics that are unobserved in our data set. Employment history, downpayment as well as detailed loan structure data are not available in the expanded HMDA set. The ability to consider these factors can further separate the financial from the non-financial contributors to mortgage loan denial.

As such, the literature of loan discrimination is a work in progress. More data on the financial background of the borrowers and the lifetime of originated loans is needed to provide a definite answer to the question of the role of race/ethnicity and mortgage lending.

References

Ards, Sheila, Inhyuck Steve Ha, Jose-Luis Mazas (2015). "Bad Credit and Intergroup Differences in Loan Denial Rates," *The Review of African American Political Economy,* vol 42, iss. 1-2, pp.19-34.

Avery, Robert, Patricia Beeson, and Mark Sniderman (1996). "Accounting for Racial Differences in Housing Credit Markets," In Mortgage Lending, Racial Discrimination, and Federal Policy (J. Goering and R. Wienk, Eds.). Washington, DC: Urban Institute Press.

Avery, Robert, Glen Canner, and Robert Cooke (2005). "New Information Reported under HMDA and Its Implications in Fair Lending Enforcement," Federal Reserve Bulletin (Summer 2005).

Bayer, Patrick, Fernando Ferreira and Stephen Ross (2014). "Race, Ethnicity and High-Cost Mortgage Lending," NBER Working Paper No. 20762

Bhutta, Neil and Benjamin Keys (2018). "Eyes Wide Shut? The Moral Hazard of Mortgage Insurers during the Housing Boom," NBER Working Paper No. 24844

Black, Harold, Thomas Boehm, and Ramon DeGennaro (2003). "Is There Discrimination in Mortgage Pricing? The Case of Overages." *Journal of Banking and Finance*, 27(6), pp. 1139-65.

Choi, Jung Hyun, Alanna McCargo, Michael Neal, Laurie Goodman, Caitlin Young (2019). "Explaining the Black-White Homeownership Gap: A Closer Look at Disparities across Local Markets," Urban Institute Research Report.

CFPB (2019). "Ability-to-Repay and Qualified Mortgage Rule Assessment Report." www. https://www.consumerfinance.gov/.

Courchane, Marsha (2007). "The Pricing of Home Mortgage Loans to Minority Borrowers: How Much of the APR Differential Can We Explain?" *Journal of Real Estate Research*, 29, 399-439.

Freddie Mac (2021). "Racial and Ethnic Valuation Gaps in Home Purchase Appraisals," Research Note http://www.freddiemac.com/research/insight/20210920_home_appraisals.page

Ghent, Andra, Rubén Hernández-Murillo, and Michael Owyang (2013). "Race, Redlining and Subprime Mortgage Pricing," St Louis Federal Reserve Working Paper.

Li, Wei and Laurie Goodman (2014). "A Better Measure of Mortgage Application Denial Rates," Urban Institute Research Report.

Mian, Atif and Amir Sufi (2008). "The Consequences of Mortgage Credit Expansion: Evidence from the 2007 Mortgage Default Crisis," NBER Working Paper No. 13936.

Munnell, Alicia, Geoffrey Tootell, Lynn Browne, and James McEneaney (1996). "Mortgage Lending in Boston: Interpreting HMDA Data," *American Economic Review*, 86, 25-53.

Nelson, Ashlyn Aiko (2005). "Better Credit, Better Schools: Credit Quality, Race, and Discrete Choice Residential Sorting," Doctoral Dissertation. Stanford University.

Park, Kevin (2022). "A Comparison of Mortgage Denial and Default rates by Race, Ethnicity and Gender," Social Sciences Research Network.

Ross, Stephen and Yuan Wang (2021). "Mortgage Lenders and the Geographic Concentration of Foreclosures," NBER Working Paper No. 28781

Appendix I. Top 20 Metro Areas with Largest House Price Index Growth (2000-22)

CBSA	Metro Area	HPI growth
14260	Boise City, ID	324.05
31080	Los Angeles-Long Beach-Glendale, CA	286.68
46520	Urban Honolulu, HI	281.29
12420	Austin-Round Rock-Georgetown, TX	280.21
33100	Miami-Miami Beach-Kendall, FL	278.31
34940	West Palm Beach-Boca Raton-Boynton Beach, FL	259.39
42660	Tacoma-Lakewood, WA	258.83
41620	Salt Lake City, UT	257.49
42660	Seattle-Bellevue-Kent, WA	254.75
45300	Tampa-St. Petersburg-Clearwater, FL	252.14
38060	Phoenix-Mesa-Chandler, AZ	247.39
40140	Riverside-San Bernardino-Ontario, CA	245.83
33100	Fort Lauderdale-Pompano Beach-Sunrise, FL	241.94
41740	San Diego-Chula Vista-Carlsbad, CA	236.69
40140	Anaheim-Santa Ana-Irvine, CA	236.48
35840	North Port-Sarasota-Bradenton, FL	232.03
38900	Portland-Vancouver-Hillsboro, OR-WA	229.82
34980	Nashville-Davidson--Murfreesboro--Franklin, TN	225.38
15980	Cape Coral-Fort Myers, FL	222.09
41860	San Francisco-San Mateo-Redwood City, CA	219.63