Finance and Economics Discussion Series

Federal Reserve Board, Washington, D.C.
ISSN 1936-2854 (Print)
ISSN 2767-3898 (Online)

How Much Does Racial Bias Affect Mortgage Lending? Evidence from Human and Algorithmic Credit Decisions

Neil Bhutta, Aurel Hizmo, Daniel Ringo

2022-067

Please cite this paper as:
Bhutta, Neil, Aurel Hizmo, and Daniel Ringo (2022). "How Much Does Racial Bias Affect Mortgage Lending? Evidence from Human and Algorithmic Credit Decisions," Finance and Economics Discussion Series 2022-067. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2022.067.

How Much Does Racial Bias Affect Mortgage Lending? Evidence from Human and Algorithmic Credit Decisions*

Neil Bhutta   Aurel Hizmo   Daniel Ringo†

First draft: July 15, 2021
Current draft: August 2, 2022

Abstract

We assess racial discrimination in mortgage approvals using new data on mortgage applications. Minority applicants tend to have significantly lower credit scores, higher leverage, and are less likely than white applicants to receive algorithmic approval from race-blind government automated underwriting systems (AUS). Observable applicant-risk factors explain most of the racial disparities in lender denials. Further, we exploit the AUS data to show there are risk factors we do not directly observe, and our analysis indicates that these factors explain at least some of the residual 1-2 percentage point denial gaps. Overall, we find that differential treatment has played a limited role in generating denial disparities in recent years.

*The views and conclusions expressed are those of the authors, and do not necessarily reflect the views of the Board or other members of the research staff. We thank Daryl Larson and Seamus Lawton for research assistance. We thank Lauren Lambie-Hanson and Scott Nelson for their discussions of the paper, and Bob Avery, Andreas Fuster, Hajime Hadeishi, and Doug McManus, as well as participants at seminars and conferences hosted by Boston College, the American Finance Association, the National Bureau of Economic Research Summer Institute, New York University, the American Real Estate and Urban Economics Association, the Federal Housing Finance Agency, the American Bankers Association, Fannie Mae, the Federal Reserve Bank of Boston, and the Federal Reserve Bank of Philadelphia for helpful comments.

†All three authors are at the Federal Reserve Board. Emails: neil.bhutta@frb.gov; aurel.hizmo@frb.gov; daniel.ringo@frb.gov.

1 Introduction

American families use mortgages to purchase their homes, to lower their housing costs when interest rates decline, and to tap into home equity for a variety of reasons including investments in human capital and small businesses. But not all families can easily get a mortgage; in particular, access to mortgage credit differs sharply by race and ethnicity, which may contribute to the wide racial and ethnic gaps in homeownership and wealth (e.g. Bhutta et al. 2020). For example, in 2018 and 2019, Black mortgage applicants were twice as likely as white applicants to have their application denied by lenders.

In order to craft policies that can address these disparities in credit access, it is crucial to identify what drives them. The landmark study of Munnell et al. (1996) found compelling evidence that discrimination played a major role in mortgage lending decisions in the early 1990's.[1] Since then, the mortgage industry has evolved in many ways, including widespread adoption of technologies such as automated underwriting that can help reduce racially biased credit decisions. Nonetheless, the wide gaps in mortgage denials present in recent data have led many to conclude that discrimination persists. Media reports and survey evidence indicate widespread beliefs that financial institutions do not treat minorities fairly.[2] But it has been challenging to firmly assess the role of discrimination without detailed underwriting data on mortgage applicants similar to what Munnell et al. had collected.

In this paper, we use new confidential supervisory data collected under the Home Mortgage Disclosure Act (HMDA) to estimate the extent to which racial and ethnic discrimination by mortgage lenders continues to generate disparities in denial rates. "Discrimination" here refers to lenders treating applicants with identical observed risk factors differently on the basis of race or ethnicity—including both taste-based and statistical discrimination—which

[1] Other evidence based on detailed lender data from the 1990s also indicates discrimination in application approvals (Courchane et al. 2000).

[2] For example, see the reports by Martinez and Glantz (2018) and Donnan et al. (2022). Results from a survey by the National Association of Realtors (2022) indicates that many minority homebuyers believe there is discrimination in real estate transactions even if they did not experience discrimination themselves. Also, results from the 2019 Survey of Consumer Finances indicate that Black and Hispanic respondents are 2-3 times more likely than white respondents to fear being denied credit.

has been illegal since 1968 under the Fair Housing Act. Overall, we find that differential treatment has played a limited role in generating denial disparities in recent years, consistent with significant progress in fair lending over the last 30 years.

Rather than differential treatment, we find that group differences in risk characteristics drive most of the disparities in credit access. To start, we show that Black and Hispanic applicants tend to be more leveraged and have much lower credit scores. For example, the average credit score for Black applicants is over 40 points lower than white applicants. We also document that Black and Hispanic applicants are less likely to receive algorithmic approval recommendations from government automated underwriting systems (AUS) than white applicants. These AUS recommendations reflect the underwriting and eligibility guidelines of Fannie Mae, Freddie Mac, the Federal Housing Administration (FHA), and the Veterans Administration (VA), and are "color blind" in that race and ethnicity (or proxies like neighborhood location) cannot be used in the algorithm.[3]

Next, we try to explain lenders' rejection decisions via regression analysis. AUS recommendations alone account for about half of the Black-white gap in denial rates. Additionally, lenders frequently impose overlays—deviations from AUS recommendations such as different credit score cutoffs—to further manage risk.[4] To account for overlays, we control carefully for credit score, DTI, and LTV, on top of AUS recommendations, and find a residual Black-white denial gap of two percentage points, and residual Hispanic and Asian gaps of about one percentage point. We refer to these residual gaps as "excess denials". In contrast, Munnell et al. (1996) found excess denials of Black and Hispanic applicants of about 8 percentage points. As Munnell et al. (1996) also controlled for highly detailed applicant underwrit-

[3]We provide more details on AUS and the inputs they do use in Section 2.

[4]Lenders may impose overlays even on loans that are nominally guaranteed by the government out of fear of e.g. putback or litigation risk. For example, Fuster et al. (2021) show that regulation can affect the willingness of lenders to make FHA loans. See also Bhutta et al. (2017). If they have a disproportionate tendency to lead to the denial of minority applicants, overlays could themselves be considered a form of discrimination distinct from the differential treatment definition we consider in this paper. Legally, at least, such overlays would be defensible against a disparate impact claim as a business necessity if they helped reduce default risk. In Section 5.2.1 we provide evidence that lenders with stricter overlays generally experience better loan performance.

ing information—sourced from credit reports and lenders' worksheets—we believe our much lower estimates of excess denials are due to improvements in lender compliance with fair lending laws since their study period, rather than differences in our studies' methodologies.

Moreover, we show evidence that these one to two percentage point excess denials are overestimates of the extent of discrimination. Instead, they likely reflect (at least in part) additional risk factors that are not recorded in HMDA. First, we show that there are racial gaps in AUS recommendations conditional on observables (credit score, DTI, LTV, loan amount, etc.). Given that AUS is color-blind, we interpret these residual AUS gaps as reflecting differences in unobserved (to us) risk or eligibility variables—for example, liquid reserves. Second, we show evidence that lenders' credit standards on these unobserved variables are behind at least some of the unexplained racial denial gaps. We find that the lenders that impose the strictest standards on their applicants regardless of race also have the largest excess denial rates of minorities. Furthermore, these stricter lender also tend to have better unexplained loan performance, consistent with these lenders imposing stricter standards on unobservable risk factors.

Finally, we examine heterogeneity in excess denials across lenders and markets to indirectly test whether excess denials reflect discrimination. Discrimination may be less prevalent at fintech lenders where borrowers have no in-person contact with lenders, or worse in geographies where lenders have more market power (and are thus freer to discriminate, unconstrained by market discipline) and in geographies where the general population displays a greater degree of racial animus. However, we fail to find clear evidence for such correlations.

Several recent papers also study denial gaps and discrimination. Bartlett et al. (2022) estimate excess denials of 7-10 percentage points, but they use the public HMDA data and therefore cannot control for applicant credit scores. Kopkin (2018), also using public HMDA data, finds that denial gaps for conventional mortgages are correlated with geographic measures of racial animus. Park (2021) uses confidential HMDA data to test whether model-based loan-level loss probabilities under severe economic stress can explain racial denial disparities. Finally, Giacoletti et al. (2021) document that Black borrowers are more likely to close at the end of the month, which they argue is due to reduced discrimination as loan officers face binding monthly origination quotas.[5] Relative to this literature, we find smaller racial differences in conditional denial rates by comparing applicants with nearly identical observable risk characteristics and AUS recommendations. Furthermore, we show that the residual gaps can themselves be explained (at least in part) by additional underwriting factors not included in even the confidential HMDA data.

In contrast to testing for equality in approval rates as we do in this paper, economists often propose testing for higher marginal profitability of loans to minorities relative to whites as an indicator of discrimination. Because of the absence of loan-level profitability data, researchers have used ex-post default data as a proxy (e.g., Berkovec et al. 1998; Peter and Pinto 2021), but the connection between default and profitability is unclear given widespread government guarantees of credit risk and racial differences in prepayment speeds (Gerardi et al. 2020). Moreover, group differences in *average ex-post* default rates may not be informative about the *expected* default rates of *marginal* applicants (Ladd 1998; Arnold et al. 2018). Beyond these identification challenges, such "outcome tests" narrowly focus on taste-based discrimination as modeled by Becker (2010), rather than testing generally for race-based credit decisions as we aim to do in this paper (Domínguez et al. 2022).

Our results highlight several important policy implications. First, disparities in denials largely reflect differences in underlying measures of applicant credit risk. Increasing the vigor of fair lending enforcement, therefore, can do little to further reduce the disparities. Instead, policies that aim to improve the observed risk characteristics of minority applicants may be more fruiful. For example, gaps in credit scores could potentially be attenuated through education and financial literacy (e.g., Homonoff et al. 2019), or through improvements in the quality of credit history data (e.g., Blattner and Nelson 2021). Second, our results highlight potential disparate impact issues in lender decisions. We show that lenders often

[5]This conclusion is complicated by the common preference of borrowers to close at the end of the month to avoid prepaid interest. It is also not clear that loan officers are routinely subject to monthly quotas.

impose stricter standards than AUS recommendations; while these stricter standards may be applied in a race-neutral fashion, they can disproportionately affect minority applicants and may not be entirely justified when government takes most of the credit risk. Third, our findings may help to ease fears of discrimination that have contributed to the Black-white homeownership gap by discouraging minority families from applying for loans (Charles and Hurst 2002).

An important caveat to our study is that we only study discrimination in approval decisions conditional on formally applying. But lenders may unfairly discourage or informally reject minorities before a formal application is even submitted (e.g., Hanson et al. 2016). We also recognize that there are other margins along which discrimination may occur which we do not study here. Lenders may provide minority applicants with poor service, which could contribute to some of the racial and ethnic differences we find in verification and completeness of applications, shown in Section 5.4 (Ross et al. 2008; Frame et al. 2022; Jiang et al. 2021). Also, discrimination in pricing has been the subject of several recent papers (Bhutta and Hizmo 2020; Bartlett et al. 2022; Willen et al. 2020); these papers find small or no differences in pricing.

Our findings should also be considered in the context of a large literature that has found evidence of racial bias in other markets and settings.[6] The mortgage market may be unique in how heavily regulated and closely scrutinized it is, due to its perceived social importance. Indeed, the HMDA data we use in this study owe their existence to a concerted policy effort to fight discrimination in mortgage lending in particular. The finding that this effort has been largely successful does not mean that individual racial prejudice has been eliminated, or that overt discrimination would not rear its head if fair lending enforcement were to be relaxed. Furthermore, the racial differences in observable risk characteristics to which we attribute most of the denial disparities could have their roots in discriminatory treatment elsewhere

[6]For example, see Lang and Kahn-Lang Spitzer (2020) for a recent review of studies considering racial discrimination in labor markets and criminal justice, or Butler et al. (2022), Blanchflower et al. (2003), and Howell et al. (2021) for evidence in other credit markets.

(e.g. the labor market) or in the inter-generational persistence of unequal treatment that occurred in years past.

2 Background

Most mortgages are originated through one of three government-related programs: (1) "conventional conforming" loans sold to Fannie Mae and Freddie Mac (government-sponsored enterprises, or GSEs); (2) loans insured by the FHA, which is the main program for borrowers with small down payments and lower credit scores; and (3) loans guaranteed by the VA for military families. Each program has specific eligibility criteria (e.g. maximum loan size) and underwriting standards. Outside of these programs, roughly 20 percent of mortgages are held in the portfolios of banks, credit unions and other financial institutions, including most "jumbo" loans — that is, conventional loans beyond the loan size limits of the GSEs.

In the typical first stage of the mortgage application process, prospective mortgage borrowers contact a lender or a mortgage broker (someone who works with multiple lenders) to inquire about getting a mortgage. Inquiries over the internet have become more common in recent years, along with application processes that are fully online (Buchak et al. 2018; Fuster et al. 2019). Loan officers (or online algorithms) will gauge the needs and resources of the borrower and recommend a particular loan program, and can then quickly conduct a prequalification screen based on a check of their credit score and the stated income and assets of the borrower. At this stage, potential applicants who have a low credit score or who appear to lack income or down payment funds may be dissuaded from moving forward. [7]

The second stage is to submit a formal application, along with documentation of income and assets (e.g. pay stubs, tax returns, account statements, etc.). Loan officers help ensure borrowers provide the right documentation and fill out the application correctly. However,

[7] See Hanson et al. (2016) and Ross et al. (2008) for evidence of differential treatment in pre-application stages. Racial or ethnic bias that impedes minorities from moving beyond this first stage will not be captured in our study.

loan officers do not make final credit decisions. Application information is entered into an AUS, and the associated documents are sent to a separate underwriting department, which will make the ultimate determination of whether the loan will be approved.

In this paper we focus on three AUS used for government backed loans, through which the vast majority of applications are run: Fannie Mae's Desktop Underwriter (DU), Freddie Mac's Loan Product Advisor (LPA), and FHA's credit risk scorecard (TOTAL).[8] These AUS use the application information to provide a recommendation for whether the loan may be approved. The AUS scores loans for credit risk based on statistical default models, and also ensures that loans meet certain eligibility requirements depending on the specific loan program. These models must follow fair lending regulations, and therefore cannot take into account race or ethnicity, or proxies such as neighborhood location or ZIP code. The most commonly used AUS is Fannie Mae's DU. Fannie Mae does not publish the algorithm DU uses, but does report the risk factors it considers. Along with credit history, LTV and DTI ratio, these factors include liquid reserves, housing expense ratio, and self-employment (see Fannie Mae 2021). LPA and TOTAL consider similar factors.[9] Ultimately, the AUS recommendation consists of two indicators: whether the loan is low risk enough to be recommended for approval, and whether the loan is eligible for the program being considered. In this paper we define an "AUS denial" to be an application that did not receive an "accept/eligible" (or equivalent) recommendation.

AUS only provides a recommendation to underwriters. The final determination on a loan applicant is made by an underwriter. This determination reflects several inputs, including AUS results, any additional lender requirements not in the AUS (referred to as "overlays"), and successful verification of all applicant information (income, assets, employment history, etc.). Loans that pass AUS could still be rejected because, for example, income could not be fully verified or because the property appraisal ended up being lower than expected.

[8]In Appendix Table A.1, we report the fraction of applications that were processed by each AUS, broken out by loan program.

[9]Both DU and LPA are subject to quarterly review by the Federal Housing Finance Agency, to ensure compliance with fair lending laws (see FHFA 2021).

Alternatively, loans that do not pass AUS could still be approved; for example the underwriter might be willing to overlook a blemish in one's credit history if the applicant can provide an adequate explanation.[10]

3 Data

We use mortgage application data collected under HMDA (FFIEC 2018-2019), which cover most mortgage lending in the U.S. (Bhutta et al. 2017). These data have long included important applicant socioeconomic characteristics including race and ethnicity, gender, and borrower income, along with basic loan information such as loan amount, census tract of the securing property, loan purpose (i.e. home purchase, refinance or home improvement), and whether the loan carried government insurance.

Until only recently, the HMDA data lacked key underwriting variables that lenders use in determining whether to approve a loan. Beginning in 2018, the HMDA data fields were expanded to include borrower credit score, DTI ratio, and combined LTV ratio. In addition, if the lender used an AUS to assist in the credit decision (as described in Section 2) they must report the output of the AUS as well as the specific model used.

We use the full, confidential version of the expanded HMDA data from 2018 and 2019. There were nearly 15 million first-lien home purchase and refinance mortgage applications for owner-occupied single-family properties, excluding observations where no credit decision was made because the application was either withdrawn or not completed by the borrower. We restrict our attention to applications for typical fixed rate, 30-year loans, and drop any applications with missing or invalid credit scores.[11] We also drop jumbo loans, as they are

[10] Lenders use AUS to assist in underwriting even for loans they do not intend to sell through a government program. Among loans held on the portfolio of the originating lender at least until the end of the year of origination, 77% had their application run through an AUS (sample omits loans originated in October through December of each year). Moreover, lenders also portfolio loans that are eligible for GSE or Ginnie Mae securitization: approximately 90% of these portfolio loans had a positive AUS recommendation. See Section 3 for a fuller description of these data.

[11] Small lenders—those who originated fewer than 500 closed-end mortgage loans in each of the prior two years—are exempt from reporting the new fields, such as credit score.

not eligible for securitization through a government program and so AUS recommendations for these loans are not informative of their risk characteristics. Finally, for our main analysis we focus on the nearly 90 percent of these applications that went through one of the three main AUS (DU, LPA and TOTAL), leaving us with a dataset of nearly 9 million applications. The details of our sample selection and number of observations are described in Appendix table A.2.

4 Estimating "Excess Denials" of Minority Applicants

In this section we assess how much more likely minority mortgage applicants are to be rejected than otherwise similar white applicants. Our empirical analysis first compares lender denial decisions with the algorithmic recommendations from AUS. Then we evaluate how much of the denial gaps can be explained by observable risk factors.

4.1 Comparing Lender Decisions to Algorithmic Decisions

We start by fixing some ideas as to how AUS recommendations, lender decisions and borrower characteristics relate to each other. The binary outcome of an AUS denial recommendation, D_{AUS}, can be written as:

$$D_{AUS} = g(X, u) \tag{1}$$

Where $g(\cdot)$ is a deterministic function of risk characteristics, X, which are observable in the HMDA data, and other risk characteristics, u, which are not observed in the HMDA data.

Lender i's binary denial decisions can similarly be written as:

$$D^{i}_{Lender} = h_i(X^*, u, w, r) + e \tag{2}$$

Lenders also base their decisions on X and u, but lender i's decision function $h_i(\cdot)$ may differ from the AUS function $g(\cdot)$ in its treatment of these inputs. Furthermore, the values of X may be revised during verification after initial underwriting, so equation 2 takes a potentially modified version of X, X^*. Lenders may also take into account risk factors not considered by AUS, w, or they may (illegally) base decisions on race or ethnicity, r. Finally, the error term, e, reflects an idiosyncratic element arising from human error in lender decisions.[12]

The top row of Table 1 shows that lenders denied 17 percent of Black mortgage applicants in 2018-2019, substantially higher than the 8 percent denial rate for white applicants.[13] Lenders also denied Hispanic and Asian applicants at higher rates than white applicants. At the same time, Table 1 indicates that observable applicant characteristics (X or X^* in the equations above) differ on average across groups. Most strikingly, the average credit score of Black applicants is about 40 points lower than that of white applicants; the average for Hispanic applicants is over 20 points lower. Black and Hispanic applicants also have significantly lower incomes on average, and higher LTV's and DTI's on average. These differences could be driving the differences in denial rates, as opposed to racial bias.

As discussed in Section 2, recommendations from government AUS are not a function of race or ethnicity, r. Still, the second row of Table 1 shows that Black and Hispanic applicants are more likely than white applicants to receive an AUS denial recommendation. Even if application decisions were based purely on government credit-risk algorithms, the data suggest that the Black-white denial gap in 2018-19 would still have been about 9 percent.

We can also see in Table 1 that lenders deny each applicant group at a higher rate than AUS. This holds even for white applicants, consistent with lenders denying applicants more often than AUS for reasons other than minority status. As highlighted by equations 1 and

[12] Lenders' decisions may depend directly on AUS recommendations. However, because D_{AUS} is a deterministic function $g(\cdot)$ of X and u, for simplicity we do not include D_{AUS} as a separate argument in $h_i(\cdot)$.

[13] We follow the method described in Bhutta et al. (2017) to designate a race and ethnicity for each application.

2, lenders could have different decision functions than AUS (h versus g), account for other risk factors, w, or use updated measures of risk, X^*.

How much of the gaps in denials by lenders can be traced to the gaps in AUS denial recommendations? In column 2 of Table 2 we regress an indicator of lender denial on applicant race and ethnicity dummies, while conditioning on AUS denial recommendations (interacted with indicators for loan purpose and program). Compared to the unconditional gaps shown in column 1, controlling for AUS recommendations shrinks the Black-white denial gap from 9 to 4.3 percentage points, and the Hispanic-white gap from 3.1 to 2 percentage points, while the Asian-white gap increases slightly. Overall, differential rates of AUS recommendations can explain some of the minority denial gap, but far from all of it.

4.2 Do "Overlays" Explain Racial Disparities?

AUS recommendations are not binding, and lenders may choose to impose tighter underwriting standards, known as overlays, than the government programs require.[14] In terms of equations 1 and 2, $g(\cdot)$ may change from 0 to 1 at different values of X and u than $h_i(\cdot)$ does. Here we investigate how much of the remaining minority denial gaps can be explained by lender overlays on observable characteristics, X.

In column 3 of Table 2 we add in underwriting controls derived from the new HMDA fields.[15] We include a fully interacted set of discretized bins of credit score, LTV, and DTI ratio.[16] We also include the AUS denial recommendation indicator, to help capture some of the potential unobservable risk factors, u. Lastly, we include county-by-month fixed effects, indicators for requested loan amount (discretized into bins of $50,000), an indicator for the

[14] Lenders may also approve and potentially portfolio a loan that the AUS does not recommend accepting. See Table 1 for the frequency of disagreement between lender decisions and AUS recommendations by race and ethnicity.

[15] We treat reported underwriting factors as exogenous. This assumption could mask some forms of discrimination: for example, lenders could be providing less help to minority applicants in fully documenting all sources of income and thereby inflating their DTI ratios. However, using the National Survey of Mortgage Originations linked to HMDA records, we find that minority home buyers are no more likely than white home buyers to self-report an income higher than the income recorded in HMDA for that buyer.

[16] See notes to Table 2 for details on how we bin credit score, LTV and DTI.

presence of a co-applicant, the log of reported income, and a lender fixed effect. All these covariates are fully interacted with the indicators for loan purpose and program.

The estimated Black and Hispanic denial rate gaps are cut in half relative to column 2. Black applicants are 2 percentage points more likely, Asian applicants 1.4 percentage points more likely, and Hispanic applicants 1 percentage point more likely to be denied than comparable white applicants.[17] Comparing columns 1 and 3 indicates that we can explain over three-quarters of the Black-white denial gap.[18] We refer to the remaining gaps as "excess denials".

In the online Appendix, we provide estimates of excess denials across various subsets of the data. Tables A.4 and A.5 indicate that excess denials are of roughly similar magnitude across loan purpose, loan program, lender type and geographic region. Appendix Table A.6 shows that excess denials appear both in the subsample of applicants approved by the AUS as well as the subsample of those with an AUS denial. In other words, lenders are both more likely to override a positive AUS recommendation to deny a minority applicant, and to override a negative AUS recommendation to approve a white applicant. Overall, these results demonstrate that excess denials are not driven by any particular segment of the data.

5 Explanations for Excess Denials

Should these remaining minority denial gaps, or "excess denials," be attributed to racial discrimination by lenders, or can they be explained by other non-discriminatory factors? In this section we test for both discriminatory and non-discriminatory explanations for the

[17]The explanatory power of the new HMDA fields is far greater than with pre-2018 data, as we demonstrate in Appendix Table A.3. In column 2 of Table A.3 we control only for fields available in pre-2018 versions of HMDA. These older fields can explain little of the denial disparities.

[18]Appendix Table A.3 shows that the raw Black-white denial gap *before* we select on applications that were run through AUS is nearly 12 percentage points. Our estimate of Black excess denials of 2 percentage points is over 83 percent lower than this bigger baseline gap. Although racial bias by lenders could drive selection into our AUS-only sample, additional evidence shown in Table A.3 is inconsistent with this story. In particular, columns 5 and 7 of Table A.3 display denial gaps with and without applications that were run through AUS, and the results are quite similar, suggesting a limited role for discriminatory selection into AUS evaluation.

excess denials. First, we investigate the possibility of lender overlays on *un*observable risk factors. Second, we present several indirect tests of whether discrimination drives excess denials, for example exploiting regional variation in racial attitudes. Finally, we end this section by considering the explanations that lenders provide in HMDA for denials.

5.1 Do Unobserved Risk Characteristics Vary by Race and Ethnicity?

As a starting point, we test for racial differences in risk factors that remain unobserved in the HMDA data. Despite the expanded HMDA data, we still do not directly observe various risk factors that might influence credit decisions, such as the applicant's cash reserves, the length of time employed at their current job, or how well they are able to document their income and assets.

To test for disparities in unobservables, we run a regression of AUS recommendations on race and ethnicity, controlling for observable underwriting variables. As explained earlier, AUS results cannot directly depend on the applicant's race and thus unexplained racial or ethnic gaps in AUS recommendations must reflect additional quantifiable factors that we do not observe in the HMDA data. In terms of equations 1 and 2, this would take the form of variation across borrowers in u that is correlated with race and ethnicity, r.

Column 5 of Table 2 shows that even with the full set of controls for observable underwriting variables, Black applicants are 1.5 percentage points less likely to be recommended for acceptance by an AUS than observably identical white applicants. This result implies that Black applicants tend to be considered by AUS to be somewhat riskier along dimensions we do not observe in the HMDA data. However, for Hispanic and Asian applicants the respective unexplained AUS denial gaps are close to zero. This result implies that these two groups, on average, do not differ significantly enough on the unobservable factors considered by the AUS, u, to trigger differential AUS recommendations.

Figure 1 plots unexplained AUS denial gaps as well as lender excess denials for each race and ethnicity by 10-point credit score bins.[19] For Black applicants (top panel), the AUS denial gap relative to white applicants rises substantially as credit score declines, suggesting wider differences in unobserved risk factors at lower credit scores. Strikingly, the same panel also shows that Black-white lender excess denials are highest among the same subset of applicants that AUS consider riskiest along unobservable dimensions. Even though excess denials are conditional on AUS output, the differences in unobservables that drive the Black-white AUS gap could still contribute to Black excess denials due to overlays, as described further in the next section.

In the middle and bottom panels of Figure 1, the Hispanic-white and Asian-white AUS denial gaps are very close to zero throughout the credit score range. Unlike the top panel, we cannot detect meaningful differences in unobservable risk factors anywhere in the credit score distribution.

5.2 Do Overlays on Unobserved Characteristics Help Explain Excess Denials?

Excess denials could be explained by lenders having tougher standards than the AUS on applicant characteristics that are not observed in the HMDA data, if minority applicants more frequently fall short of these overlays. In terms of equations 1 and 2, this would mean that u varies by race and that $g(\cdot)$ and $h_i(\cdot)$ differ in their treatment of u, or that w differs by race. We test this hypothesis by exploiting cross-sectional differences in lender policies. We construct a lender-specific measure of the "strictness" of their underwriting policies, and then correlate lenders' strictness with their excess denials. Different lender overlays on unobservables should create a positive correlation between strictness and excess denials across lenders if unobservable characteristics of minority applicants appear riskier than those of white applicants.

[19] Raw racial denial gaps for each credit score bin, by lender and by AUS, are plotted in Appendix Figure A.1.

To start, we estimate lender-specific excess denials of minority applicants. For this analysis we focus on the 100 largest lenders in our data, as measured by the total count of originations in 2018 and 2019. We run a regression with our full set of controls (as in column 3 of Table 2) but allow the coefficients on race and ethnicity to vary by lender. Importantly, to ensure that our lender-specific excess denial estimates do not simply pick up differences across lenders in their standards on *observable* underwriting factors (credit score, LTV, and DTI ratio), we allow these coefficients to also vary by lender. The lender-specific coefficients on the race and ethnicity dummies are plotted, along with 95 percent confidence intervals, in the left-hand panels of Figure 2. For each of the three minority groups shown, at least 85 of the 100 largest lenders had an excess denial rate greater than zero and there are at least ten lenders that have excess denial estimates of 4 percentage points or more.

Next, we consider whether lenders may differ in their estimated excess denials due to differences in their loan approval "strictness" (i.e. stricter policies may have a disproportionate effect on minority applicants despite being applied equally to all groups). Strictness is estimated as the lender-specific probability of denying a *white* applicant, conditional on the full set of control variables (similar to the column 3 specification in Table 2, but only including white applicants in the estimation). We construct this measure based solely on white applicants to isolate differences in lender policy without contamination by any differential treatment of minority applicants. The lender fixed effects from this white-only regression yield our estimates of lender-specific strictness. Different measures of strictness between two lenders, i and j, indicate that $h_i(\cdot, \cdot, \cdot, white) \neq h_j(\cdot, \cdot, \cdot, white)$. Lender strictness equal to zero means that lender's denial rate of white applicants was exactly average, conditional on observables. Higher strictness means lenders are imposing tougher standards on their borrowers.

If excess denials of minorities are at least partially due to racial and ethnic differences in the ability to meet overlays on unobservables, then we would expect a positive correlation between strictness and excess denials.[20] We show a scatterplot of lender excess denials against the measure of lender strictness in the right-hand column of Figure 2. A tight, positive slope is visually apparent for all three of the minority groups presented, with correlations being 0.63 for Black applicants, 0.5 for Hispanic applicants, and 0.65 for Asian applicants. Lenders that impose the strictest standards on their *white* applicants tend to also have the largest excess denials of minority applicants. This finding suggests that excess denials are at least partly a result of tight lender standards on unobservable factors.

5.2.1 Lender Strictness and Loan Performance

While we cannot identify all the overlays stricter lenders are imposing on the observable and unobservable factors of their applicants, we can test whether these overlays actually lead to a reduction in risk or if they only result in an unjustified disparate impact on minority applicants. This section provides evidence that stricter lenders end up making better-performing loans. Furthermore, the reduction in risk is not entirely explained by observable underwriting characteristics of their borrowers. This suggests both that strictness is partially measuring lenders' overlays on unobservables, and that those overlays have some effect on reducing risk.

To measure the ex-post riskiness of individual lenders' originations, we use performance data from loans securitized in Ginnie Mae pools with origination dates in 2018 and 2019. These data (publicly available on Ginnie Mae's website) track monthly delinquency status for most FHA and VA loans. We match the largest Ginnie issuers by name to their records as lenders in HMDA, and limit the sample to institutions that both directly securitized the large majority of their FHA and VA loans, and that mostly securitized only their own originations. These restrictions ensure a sample for which the Ginnie performance data reflects the issuer's own underwriting criteria as a lender.

[20]To be clear, our measure of lender strictness reflects stringency on both unobservable (u and w) and observable risk factors X (e.g. credit score, LTV, and DTI). However, because our lender-specific excess denial estimates allow for lender-specific coefficients on X, overlays on observables cannot themselves generate a correlation between our measures of lender-specific strictness and excess denials. For more evidence that strictness is picking up overlays on unobservables, see Section 5.2.1.

We end up with 48 matched institutions that directly securitized over 75 percent of their FHA and VA originations through Ginnie, and that also originated at least 75 percent of the loans they securitized. Of these, 23 institutions achieved 90 percent on both the marks. We measure riskiness as the percentage of the issuer's loans that were ever 60 days or more delinquent within one year of origination (normalized to zero for the average issuer), and re-estimate lender strictnesses specific to applications for FHA and VA loans.

The left-hand charts of Panels 1 and 2 in Figure 3 show scatterplots and linear fits of Ginnie riskiness against lender strictness, for both the 75 percent and 90 percent samples. A strong negative relationship is apparent in both. Strictness appears to meaningfully reduce subsequent delinquencies.

To test whether strictness is measuring overlays on unobservables, we estimate issuer-specific residual riskiness conditional on observables. Residual riskiness is estimated as the issuer fixed effect in a regression of delinquency on a flexible function of DTI, LTV, and credit score, as well as month-of-origination dummies.[21] The right-hand charts of Panels 1 and 2 in Figure 3 plot residual riskiness against strictness. While not as strong as the unconditional riskiness correlations shown in the left-hand charts, a negative relationship is still apparent. Lender strictness predicts delinquencies even among loans with similar observable risk characteristics, suggesting that stricter lenders are imposing tighter standards on unobservable as well as observable applicant characteristics. Higher excess denials among stricter lenders are likely (at least in part) a consequence of this tendency.

5.3 Indirect Tests of Whether Discrimination Drives Excess Denials

To further understand whether excess minority denials might reflect differential treatment to any extent, we try several indirect tests of discrimination by testing whether excess denials are larger in circumstances we would have *ex ante* expectations for discrimination to be more prevalent.

[21]The regressions include interactions between dummy variables for single integer buckets of LTV, DTI, and 20 point buckets of credit scores.

First, we compare fintech lenders to traditional mortgage lenders. By automating more of the application process, fintechs cut out some human judgement and consequently have the potential to reduce racial discrimination (Howell et al. 2021). We re-estimate equation 1 on different subpopulations of lenders, including lenders identified as fintechs by Fuster et al. (2019), and present results in Table 3. We find excess denials are, if anything, higher at fintech lenders, the opposite result we would expect if excess denials reflect racially biased human judgement.[22]

Next, we compare outcomes in more- and less-competitive lending markets. In less competitive markets, a few large lenders could potentially leverage their market power to make inefficient decisions, such as indulging in taste-based discrimination. We rerun our denial regressions, including an interaction term between applicant race and the market share of the top 4 lenders in that county. Results are presented in column 2 of Table 3. The estimated interaction effects are all negative. This suggests competitive pressure does not reduce excess denials, in contrast to what we would expect if excess denials were driven by taste-based discrimination.

Finally, we compare outcomes in markets differentiated by a population-level measure of racial animus. If excess denials reflect taste-based discrimination, we might see relatively high excess denials in areas with more racial hostility. We interact applicant race with the frequency of racially-charged Google search terms in a given media market, a measure provided by Stephens-Davidowitz (2014), and re-estimate the denial regressions.

Results are shown in column 3 of Table 3. It does appear that excess denials are somewhat higher in media markets exhibiting greater racial animus. However, when we repeat the exercise for AUS rather than lender excess denials, we observe the same pattern—i.e., higher AUS excess denials in markets of greater racial animus (compare columns 3 and 6 of Table 3). This suggests that white-minority differences in unobservable risk factors are larger in markets with higher racially charged search frequencies, potentially explaining the similar

[22] Appendix Table A.5 provides estimates of excess denials for traditional banks and (non-fintech) nonbanks.

correlation with excess denials. Overall, we do not find compelling evidence that excess denials can be explained by differential treatment of minority applicants.

5.4 How Do Lenders Explain Excess Denials?

Lenders are now required under HMDA to report a denial reason for every denied application. Lenders may choose from a list of nine potential reasons, or use a free text field. Not surprisingly, none of the nine reasons refer to race or ethnicity, and a lender engaged in illegal discrimination would be unlikely to explicitly admit this, so the self-reported reasons may not always reflect reality. Nonetheless, we can use these stated reasons to better understand how lenders justify their excess denials. Details of the data and analysis, and a discussion of what can be inferred from the results, are presented in the online appendix. One main finding from this analysis is that lenders often attribute excess denials to either "incomplete application" or "verification" of applicant information, especially for Asian and Hispanic excess denials. In terms of our earlier conceptual framework, there may be racial and ethnic variation in the probability that $X \neq X^*$. This suggests that minority applicants may experience more difficulties in the latter stages of the mortgage approval process (i.e. after initial underwriting and AUS recommendations have completed), contributing to excess denials.

6 Conclusion

Using newly available HMDA data for 2018-2019, we find that standard underwriting factors can explain most of racial and ethnic disparities in denial rates. Further evidence suggests that the remaining 1-2 percentage point differences in denial rates (what we refer to as "excess denials") are at least partially due to differences in racial and ethnic distributions of unobservable underwriting factors. Overall, racially-biased credit decisions appear less common than has been suggested by previous research. Our results imply significant progress in fair lending for mortgages over the last 30 years.

References

Arnold, David, Will Dobbie, and Crystal S Yang, "Racial Bias in Bail Decisions," *The Quarterly Journal of Economics*, November 2018, *133* (4), 1885–1932.

Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace, "Consumer-lending discrimination in the FinTech Era," *Journal of Financial Economics*, 2022, *143* (1), 30–56.

Becker, Gary S, *The economics of discrimination*, University of Chicago press, 2010.

Berkovec, James A., Glenn B. Canner, Stuart A. Gabriel, and Timothy H. Hannan, "Discrimination, Competition, and Loan Performance in FHA Mortgage Lending," *The Review of Economics and Statistics*, 05 1998, *80* (2), 241–250.

Bhutta, Neil and Aurel Hizmo, "Do Minorities Pay More for Mortgages?," *The Review of Financial Studies*, 04 2020, *34* (2), 763–789.

__ , Andrew C Chang, Lisa J Dettling, and Joanne W Hsu, "Disparities in Wealth by Race and Ethnicity in the 2019 Survey of Consumer Finances," *FEDS Notes*, 2020, (2020-09).

__ , Steven Laufer, and Daniel Ringo, "The decline in lending to lower-income borrowers by the biggest banks," *FEDS Notes*, 2017, (2017-09), 28–1.

Blanchflower, David G., Phillip B. Levine, and David J. Zimmerman, "Discrimination in the Small-Business Credit Market," *The Review of Economics and Statistics*, 11 2003, *85* (4), 930–943.

Blattner, Laura and Scott Nelson, "How costly is noise: Data and disparities in the US mortgage Market," *Working Paper*, 2021.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru, "Fintech, Regulatory Arbitrage, and the Rise of Shadow Banks," *Journal of Financial Economics*, 09 2018, *130.*

Butler, Alexander W, Erik J Mayer, and James P Weston, "Racial Disparities in the Auto Loan Market," *The Review of Financial Studies*, 05 2022. hhac029.

Charles, Kerwin Kofi and Erik Hurst, "The transition to home ownership and the black-white wealth gap," *Review of Economics and Statistics*, 2002, *84* (2), 281–297.

Courchane, Marsha, David Nebhut, and David Nickerson, "Lessons learned: Statistical techniques and fair lending," *Journal of Housing Research*, 2000, pp. 277–295.

Domínguez, Patricio, Nicolás Grau, and Damián Vergara, "Combining discrimination diagnostics to identify sources of statistical discrimination," *Economics Letters*, 2022, *212*, 110294.

Donnan, Shawn, Ann Choi, Hannah Levitt, and Christopher Cannon, "Wells Fargo Left Black Homeowners Behind in the Pandemic Mortgage Refinancing Boom," *Bloomberg News*, March 2022.

Fannie Mae, "Risk Factors Evaluated by DU," https://selling-guide.fanniemae.com/Selling-Guide/Origination-thru-Closing/Subpart-B3-Underwriting-Borrowers/Chapter-B3-2-Desktop-Underwriter-DU-/1032994121/B3-2-03-Risk-Factors-Evaluated-by-DU-09-01-2021.htm 2021. Accessed: 2021-12-31.

FFIEC, "Home Mortgage Disclosure Act Data," 2018-2019.

FHFA, "Order on Fair Lending Compliance and Report Submission," https://www.fhfa.gov/SupervisionRegulation/LegalDocuments/Documents/Orders/FNM-Final-Order-re-Fair-Lending-Reporting.pdf 2021. Accessed: 2021-12-31.

Frame, W Scott, Ruidi Huang, Erik J Mayer, and Adi Sunderam, "The Impact of Minority Representation at Mortgage Lenders," Technical Report, National Bureau of Economic Research 2022.

Fuster, Andreas, Matthew C Plosser, and James I Vickery, "Does CFPB oversight crimp credit?," *CEPR Discussion Paper No. DP15681*, 2021.

__ , Matthew Plosser, Philipp Schnabl, and James Vickery, "The Role of Technology in Mortgage Lending," *The Review of Financial Studies*, 04 2019, *32* (5), 1854–1899.

Gerardi, Kristopher, Paul Willen, and David Hao Zhang, "Mortgage prepayment, race, and monetary policy," *Available at SSRN 3697625*, 2020.

Giacoletti, Marco, Rawley Heimer, and Edison G Yu, "Using High-Frequency Evaluations to Estimate Discrimination: Evidence from Mortgage Loan Officers," *Available at SSRN 3795547*, 2021.

Hanson, Andrew, Zackary Hawley, Hal Martin, and Bo Liu, "Discrimination in mortgage lending: Evidence from a correspondence experiment," *Journal of Urban Economics*, 2016, *92*, 48–65.

Homonoff, Tatiana, Rourke O'Brien, and Abigail B Sussman, "Does Knowing Your FICO Score Change Financial Behavior? Evidence from a Field Experiment with Student Loan Borrowers," Working Paper 26048, National Bureau of Economic Research July 2019.

Howell, Sabrina T, Theresa Kuchler, David Snitkof, Johannes Stroebel, and Jun Wong, "Automation and Racial Disparities in Small Business Lending: Evidence from the Paycheck Protection Program," Working Paper 29364, National Bureau of Economic Research October 2021.

Jiang, Erica Xuewei, Yeonjoon Lee, and Will Shuo Liu, "Disparities in Consumer Credit: The Role of Loan Officers in the FinTech Era," *Available at SSRN*, 2021.

Kopkin, Nolan, "The conditional spatial correlations between racial prejudice and racial disparities in the market for home loans," *Urban Studies*, 2018, *55* (16), 3596–3614.

Ladd, Helen F., "Evidence on Discrimination in Mortgage Lending," *Journal of Economic Perspectives*, June 1998, *12* (2), 41–62.

Lang, Kevin and Ariella Kahn-Lang Spitzer, "Race discrimination: An economic perspective," *Journal of Economic Perspectives*, 2020, *34* (2), 68–89.

Martinez, E and A Glantz, "How reveal identified lending disparities in federal mortgage data. REVEAL," *The Center for Investigative Reporting*, 2018, *15.*

Munnell, Alicia H, Geoffrey MB Tootell, Lynn E Browne, and James McEneaney, "Mortgage lending in Boston: Interpreting HMDA data," *The American Economic Review*, 1996, pp. 25–53.

National Association of Realtors, "Snapshot of Race and Home Buying in America," https://cdn.nar.realtor/sites/default/files/documents/2022-snapshot-of-race-and-home-buying-in-the-us-report-02-23-2022_0.pdf 2022. Accessed: 2022-16-12.

Park, Kevin A., "Measuring Risk and Access to Mortgage Credit with New Disclosure Data," *The Journal of Structured Finance*, 2021, *26* (4), 53–72.

Peter, Tobias and Edward J. Pinto, "The Rest of the Story: The AEI Housing Center's Critique of: 'How We Investigated Racial Disparities in Federal Mortgage Data'," https://www.aei.org/economics/the-rest-of-the-story-the-aei-housing-centers-critique-of-how-we-investigated-racial-disparities-in-federal-mortgage-data/ 2021. Accessed: 2021-12-31.

Ross, Stephen L., Margery Austin Turner, Erin Godfrey, and Robin R. Smith, "Mortgage lending in Chicago and Los Angeles: A paired testing study of the pre-application process," *Journal of Urban Economics*, 2008, *63* (3), 902–919.

Stephens-Davidowitz, Seth, "The cost of racial animus on a black candidate: Evidence using Google search data," *Journal of Public Economics*, 2014, *118*, 26–40.

Willen, Paul, David Hao Zhang et al., "Do Lenders Still Discriminate? A Robust Approach for Assessing Differences in Menus," Technical Report 2020.

Table 1: Summary statistics

	All	White	Black	Hispanic	Asian
Lender Denial Rate	0.10	0.08	0.18	0.12	0.10
AUS Denial Rate	0.06	0.05	0.14	0.08	0.05
Lender-AUS Disagreement Rate	0.09	0.08	0.15	0.10	0.08
Loan Amount (000)	253	246	230	243	333
Income (000)	92	93	78	76	104
Credit Score	720	726	685	703	738
LTV (%)	84.14	83.33	90.70	87.78	79.90
DTI (%)	39.10	37.98	42.76	42.54	39.86
N. Obs.	8,975,213	5,501,800	678,228	810,974	413,046

Note - Table shows average characteristics for purchase and refinance applications in 2018 and 2019 for first lien, 30 year FRM, on owner occupied single unit homes for which an AUS recommendation was reported. Sample excludes withdrawn or incomplete applications. Data source: HMDA.

Table 2: Denial Regressions using the AUS sample

	Lender Denial			AUS Denial	
	(1)	(2)	(3)	(4)	(5)
Black	0.090**	0.043**	0.019**	0.084**	0.015**
	(0.005)	(0.003)	(0.001)	(0.008)	(0.001)
Asian	0.012**	0.021**	0.014**	-0.004*	0.002**
	(0.004)	(0.003)	(0.001)	(0.002)	(0.001)
Hispanic	0.031**	0.020**	0.010**	0.022**	0.000
	(0.004)	(0.003)	(0.001)	(0.002)	(0.001)
Other	0.075**	0.035**	0.018**	0.062**	0.008**
	(0.007)	(0.006)	(0.002)	(0.009)	(0.001)
Joint Race	-0.005*	-0.000	0.003**	-0.000	-0.000
	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
Missing Race	0.064**	0.043**	0.017**	0.015**	0.005**
	(0.010)	(0.008)	(0.002)	(0.005)	(0.001)
AUS Outcome		Yes	Yes		
County by Month FE			Yes		Yes
Loan Amount Bins			Yes		Yes
Co-applicant			Yes		Yes
Log Income			Yes		Yes
FICO-LTV-DTI grid			Yes		Yes
Lender FE			Yes		Yes
R-Squared	0.010	0.233	0.398	0.009	0.355
N. Obs.	8,944,156	8,944,156	8,710,063	8,944,156	8,710,063

Note - All the controls except for the race/ethnicity indicators are fully interacted with program by loan purpose indicators. Credit scores are discretized into buckets of 300-399, 400-499, 500-579, and buckets of 10 points for scores above 580. DTI ratios are discretized into buckets of 5 percentage points for ratios from 0 to 30 percent, single percentage point bins for DTI between 30 and 60 percent, and bins of 20 for DTI between 60 and 100. LTV ratios are discretized into buckets of 10 percentage points from 0 to 80 percent, then in 5 percentage point buckets up to 95 percent, single percentage points up to 100 percent, and then bins of LTV of 101-110, 111-120, and 121-200. The standard errors are clustered at the lender and county levels. Significance: * p<0.1, ** p<0.05. Data source: HMDA.

Table 3: Indirect tests for discrimination

	Lender Denial			AUS Denial		
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.020**	0.023**	0.021**	0.015**	0.014**	0.014**
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
Hispanic	0.011**	0.014**	0.011**	0.001	0.004**	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Asian	0.014**	0.018**	0.014**	0.002**	0.003*	0.002**
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
Fintech						
× Black	0.019**			-0.007		
	(0.007)			(0.005)		
× Hispanic	0.006			-0.002		
	(0.005)			(0.002)		
× Asian	0.003			-0.000		
	(0.005)			(0.002)		
Top 4 Lenders' Share						
× Black		-0.008			0.005	
		(0.009)			(0.009)	
× Hispanic		-0.018**			-0.020**	
		(0.007)			(0.007)	
× Asian		-0.025**			-0.005	
		(0.009)			(0.007)	
Racially Charged Search Rate						
× Black			0.002**			0.003**
			(0.001)			(0.001)
× Hispanic			0.002**			0.002**
			(0.001)			(0.000)
× Asian			0.002**			0.001**
			(0.001)			(0.000)
AUS Outcome	Yes	Yes	Yes			
County by Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Loan Amount Bins	Yes	Yes	Yes	Yes	Yes	Yes
Co-applicant	Yes	Yes	Yes	Yes	Yes	Yes
Log Income	Yes	Yes	Yes	Yes	Yes	Yes
FICO-LTV-DTI grid	Yes	Yes	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.396	0.396	0.396	0.356	0.356	0.356
N. Obs.	7,168,470	7,168,470	7,048,946	7,168,470	7,168,470	7,048,946

Note - The racially charged search rate is constructed by Stephens-Davidowitz (2014) by using Google searches for racially charged terms in 195 designated market areas. The variable is standardized. The list of Fintechs comes from Fuster et al. (2019). The county market share of the top 4 lenders, derived from HMDA data, has a mean of 0.31 and standard deviation of 0.14. This table only includes White, Black, Hispanic and Asian applicants so the number of observations is lower than in Tables 1 and 2. All the controls except for the race/ethnicity indicators are fully interacted with program by loan purpose indicators . See notes to Table 2 for details on FICO, LTV, and DTI bins. The standard errors are clustered at the lender and county levels. Significance: * p<0.1, ** p<0.05. Data source: HMDA.

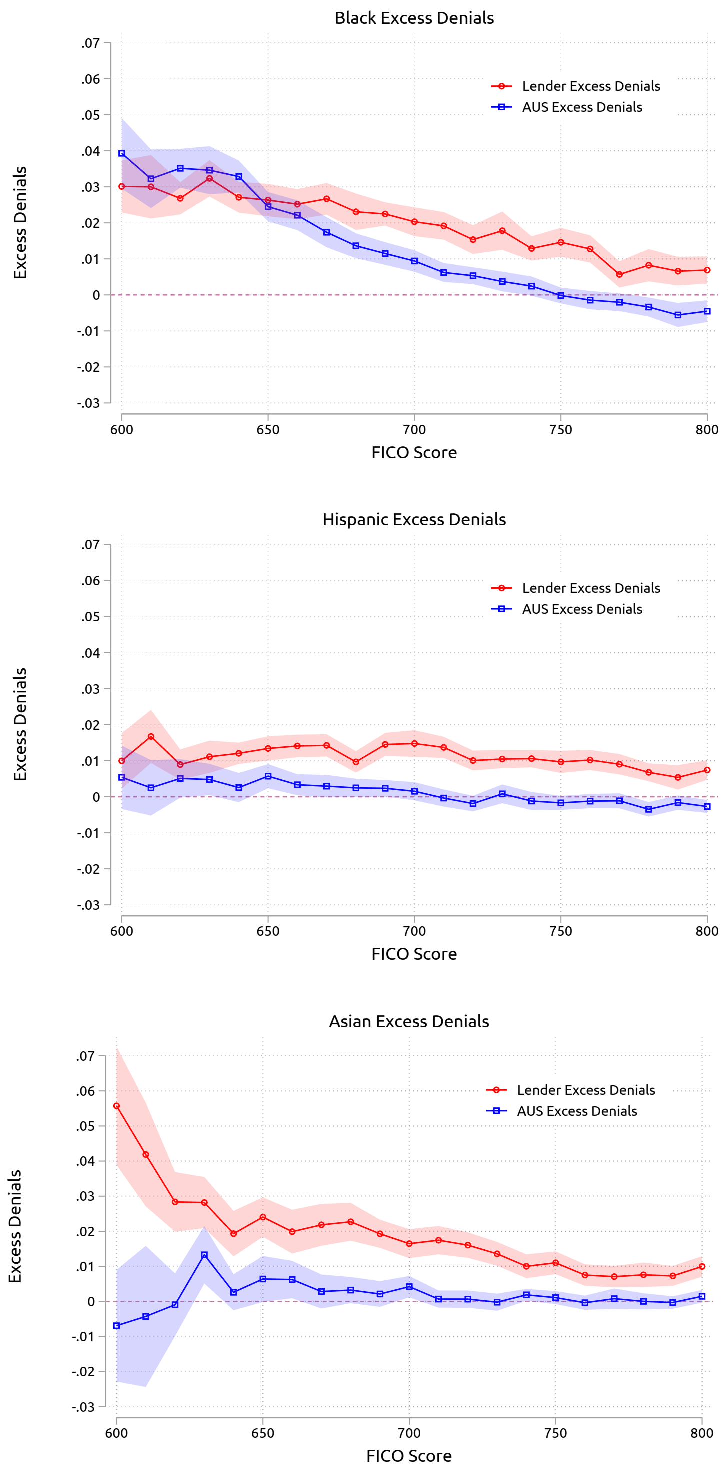


Figure 1: Lender and AUS minority exess denials relative to white by FICO score.

Note: Figure plots race and ethnicity regression coefficients by FICO after controlling for all borrower and loan characteristics as in specifications 3 and 5 of Table 2. Data source: HMDA.

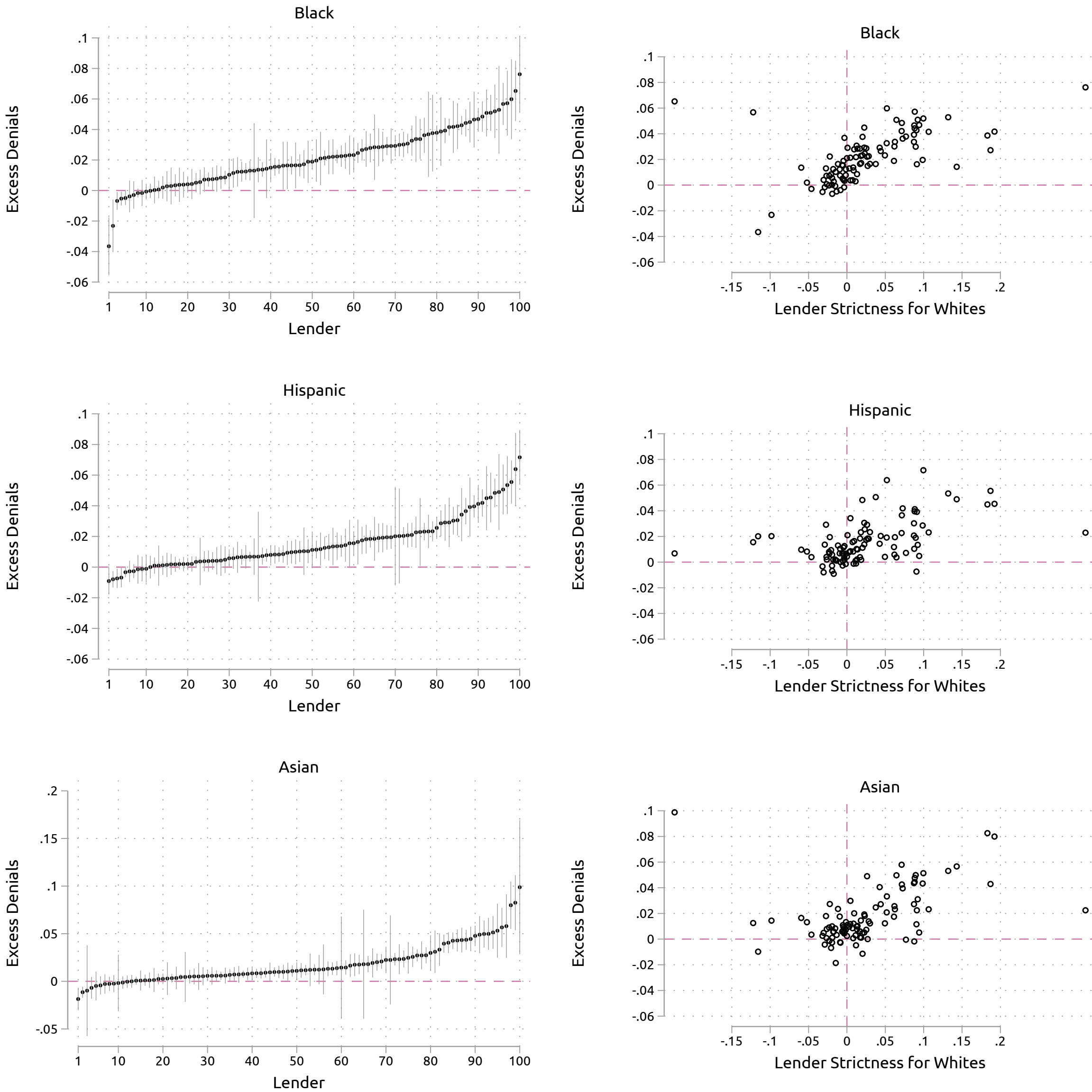
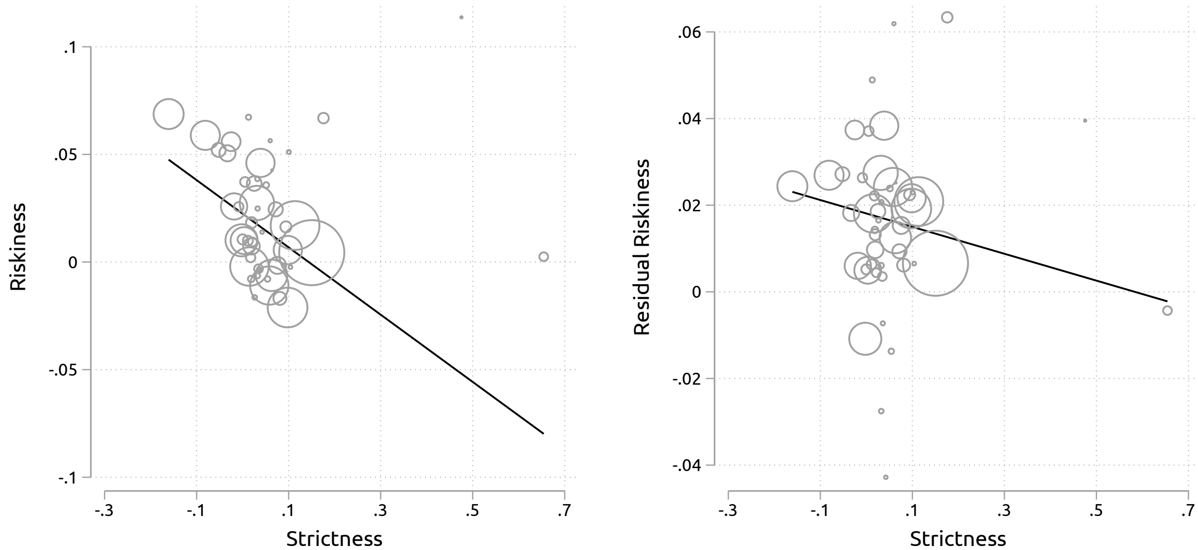


Figure 2: Excess denials for the top 100 lenders

Note: The vertical axes are regression coefficients of lender denials on race after controlling for borrower and loan characteristics as well as AUS outcomes separately for each of the top 100 lenders in our data. Observations in the left-hand charts are sorted by the magnitude of the estimated lender-specific excess denial. Lender strictness coefficients are lender fixed effects from a regression of lender denials on all controls including AUS outcomes for only white borrowers. Data source: HMDA.

Panel 1: Lenders that originate and securitize 75% of their FHA/VA loans through Ginnie



Panel 2: Lenders that originate and securitize 90% of their FHA/VA loans through Ginnie

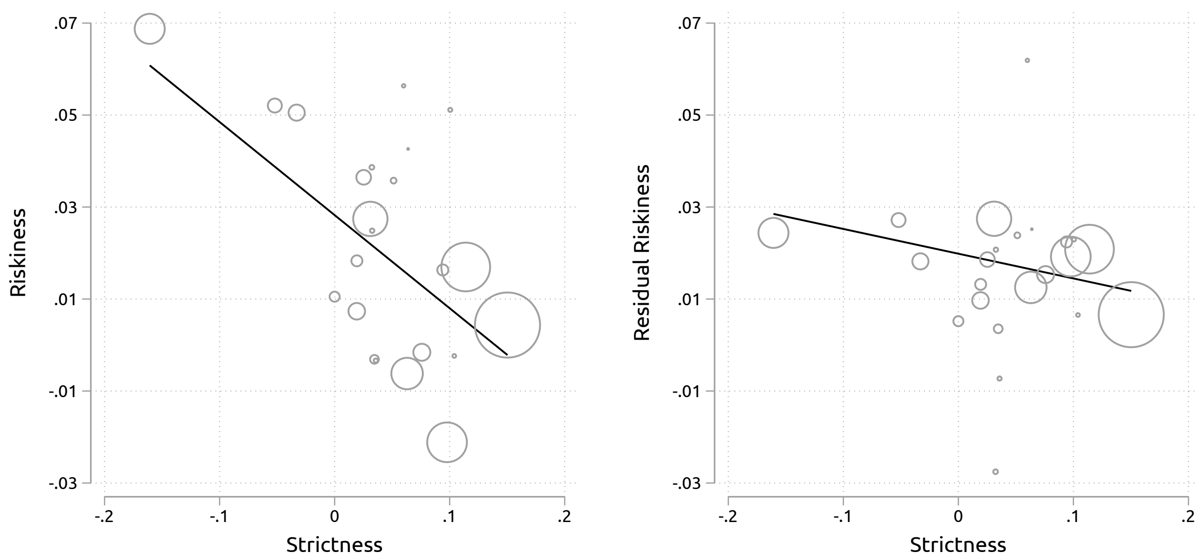


Figure 3: Lender strictness and loan performance

Note: We measure riskiness as the demeaned percentage of the issuer's loans that were ever 60 days or more delinquent within one year of origination. Residual riskiness is estimated as the issuer fixed effect in a regression of delinquency on a flexible function of DTI, LTV, and credit score, as well as month dummies. Lender strictness coefficients are lender fixed effects from a regression of lender denials on all controls including AUS outcomes for only white borrowers who applied for FHA/VA loans. The size of the circle represents the size of the lender/issuer. Data sources: HMDA, Ginnie Mae.

Internet Appendix for

"How Much Does Racial Bias Affect Mortgage Lending? Evidence from Data on Human and Algorithmic Credit Decisions"

Neil Bhutta, Aurel Hizmo, Daniel Ringo

September 21, 2022

A Denial Gaps in the Full Sample and from Alternate Specifications

In our main analysis, we use the sample of mortgage applications that have an AUS recommendation. In Table A.3 we display lender denial regression results using a “full” sample that includes applications without an AUS recommendation. Column 1 shows raw gaps without any controls, and indicates wider disparities in mortgage denials than in our main analysis. For example, the Black-white denial gap is nearly 12 percentage points in Table A.3, compared to 9 percentage points in Table 2. This difference reflects that applications without an AUS recommendation are highly likely to be denied and that minority applicants are more likely to end up without an AUS recommendation.

If the racial disparity in being evaluated by AUS is a result of discrimination, then the results in our main analysis may be biased against finding discrimination. To assess this possibility, columns 5 and 7 display results where we condition on the full set of controls (except for AUS recommendation), with and without applications where the AUS recommendation is missing, respectively. After conditioning on FICO, LTV, DTI, income, etc., the estimated lender denial gaps are quite similar, suggesting a limited role for discriminatory selection into AUS evaluation.

Table A.3 also shows how various controls affect the denial gap estimates. In column 2 we use only those variables that would have been available in pre-2018 HMDA. Next we add in newly available underwriting variables among the controls, starting with credit score in column 3 and fully interacted bins of credit score, LTV, and DTI in column 4. In column 6 we add in census tract fixed effects to test for potential redlining, but the racial gaps are little changed relative to column 5.

B Denial Regressions by Loan Purpose, Geographic Region, Lender Type, and AUS Result

Table A.4 presents lender denial regressions by loan purpose and census region. Each regression includes our full set of controls, including AUS outcome. The first three columns indicate somewhat higher denial gaps for cashout refinance mortgages than for home purchase mortgages, but this is on a higher overall average denial rate as can be seen in the bottom row of the table. In the last four columns, the smallest denial gaps occur in the West Census Region. For Black and Hispanic applicants, the largest denial gaps occur in the Midwest Census Region.

Table A.5 provides estimates of lender denial gaps by type of lender: depositories (i.e. banks and credit unions), nonbank mortgage lenders, and "fintech" lenders. To make these comparisons consistent across lender type we focus only on home purchase loans, and run separate regressions for conforming and FHA loans. We do so because denial rates tend to differ across loan type and purpose, and different lender types may tend to have higher shares of certain types of loans (e.g. depositories have higher concentrations of conventional refinance loans, whereas nonbanks and fintechs focus more heavily on FHA home purchase loans). We find that depositories and nonbanks tend to exhibit similar denial gaps, even though nonbanks arguably face less regulatory scrutiny. As we presented earlier in Table 3, fintechs have somewhat higher denial gaps than other types of lenders. As shown in the last row, fintechs also tend to have higher overall denial rates as well.

In Table A.6 we split our sample by AUS result. As shown at the bottom of the table, the lender denial rate for loans with an AUS "accept" recommendation is about 7 percent, while the lender denial rate for loans with an AUS "reject" recommendation is about 60 percent. In the industry, denials despite a positive AUS result are referred to as "high-side overrides"; and "low-side overrides" refer to when lenders accept a loan despite a negative AUS outcome We find that in both samples, minority applicants are more likely to be denied than white applicants, conditional on all observable risk characteristics. Thus, the overall denial gaps we present in our main tables reflect disparities in both high-side and low-side overrides.

C How Do Lenders Explain Excess Denials? An Analysis of Lenders' Stated Reasons for Denial

In this section, we describe the analysis of lenders' stated reasons for denial as an investigation into the causes of excess denials of minority mortgage applicants.[1] Lenders must report at least one reason for why an application was denied from a default list or select "other" and describe the reason in a free text field.[2] The default reasons are issues with borrower credit history, DTI ratio, the collateral, insufficient cash, employment history, mortgage insurance, verification of applicant data, or an incomplete application.[3]

To infer lenders' explanations for excess denials, we estimate differences by race and ethnicity in the conditional probability of being denied for each of the listed reasons. To do so, we create a new set of outcome variables, Y^D for each reason D, set equal to one if an application was denied and the first stated reason for denial was D, and zero otherwise. For each D we then estimate:

$$Y^D = \Sigma_{r \in R} \beta_r^D \mathbf{1}\{x = r\} + \boldsymbol{W} \boldsymbol{\beta}_{\boldsymbol{W}}^{\boldsymbol{D}} + \varepsilon \tag{A.1}$$

where x is the applicant's race and ethnicity, R is our set of minority race and ethnicities and W is the vector of underwriting controls, including AUS recommendation, identical to specification 3 of Table 2. We estimate equation 2 separately for each of the nine denial reasons.

We plot the estimated contribution of each denial reason to the racial denial gaps in

[1] Along with the new data fields reported, beginning with the 2018 data lenders are required to list at least one reason for denial for all denied applications. Previously, this field was reported at the lender's option.

[2] The most common explanations given under the "other" category are quite general, such as indicating that the lender does not extend credit under the terms requested without further detail.

[3] In Appendix Figure A.2 we show the unconditional breakdown of denial reasons by race.

the stacked bar charts in Figure A.3, shown separately by race and ethnicity. For all three minority groups, “verification” and “incomplete” account for a substantial share of excess denials. In other words, according to these lender-reported denial reasons, minority applicants are conditionally more likely to experience difficulties in the later stages of the application process and when underwriters attempt to verify the applicant’s information. These steps mostly occur after initial underwriting, when an AUS recommendation is obtained, which may help explain why Hispanic and Asian borrowers experience positive excess denials from lenders, but not from AUS recommendations. Unfortunately, we do not have any method to ensure lenders are truthfully reporting these reasons for denial. Furthermore, difficulties with verification and application completion could reflect lenders providing poorer service to minority applicants. We therefore cannot be sure that denials citing “incomplete” or “verification” are not actually due to some form of discrimination.

Indeed, some stated reasons raise suspicions. For example, credit history and DTI ratios are offered as explanations for a sizable fraction of excess denials, particularly for Black applicants, despite the fact that we are controlling very flexibly for credit score and DTI ratio in the denial regressions. Innocent explanations are possible — for example, more Black applicants may be denied due to a consideration of their full set of underwriting characteristics, observed and unobserved in HMDA. Lenders are only required to report one reason, and so many may simply select “DTI” or “credit history” if these were important, but not the only, factors in the decision to deny credit. Moreover, lenders (and AUS) may consider aspects of credit histories for which the credit score is not a sufficient statistic, or consider the front-end as well as the (HMDA reportable) back-end DTI ratio.[4] Recall that excess AUS denials were particularly elevated for low-credit score black applicants. Nevertheless, greater regulatory scrutiny may be warranted for lenders that are more likely to report denying a minority applicant due to credit history than a white applicant with an

[4] Front end DTI refers to the ratio between the applicant’s proposed mortgage debt service payments and income. Back-end DTI refers to the ratio between all debt payments (both mortgage and non-mortgage) and income. While HMDA requires reporting of only the back-end ratio, some lending programs (such as FHA loans) impose restrictions on both front- and back-end DTI ratios.

identical credit score.

Table A.1: Market shares for automated underwriting systems

	All	Conforming	FHA	VA	Jumbo
Desktop Underwriter	0.63	0.71	0.40	0.84	0.26
Loan Product Advisor	0.14	0.21	0.01	0.06	0.03
TOTAL	0.11	0.00	0.53	0.00	0.00
Other	0.04	0.02	0.00	0.01	0.29
N/A	0.07	0.06	0.06	0.08	0.42

Note - The sample is restricted to purchase and refinance applications in 2018 and 2019 for first lien, 30 year FRM, on owner occupied single unit homes. Sample excludes withdrawn or incomplete applications. Data source: HMDA.

Table A.2: Sample selection and observation counts

	N. of applications
(0) First-lien owner-occupied home purchase and refinance applications	14,934,868
(1) from lenders subject to full reporting requirements;	14,543,282
(2) that are for 30-year fixed rate mortgages;	11,194,171
(3) which are conventional conforming, FHA or VA;	10,587,943
(4) and are not missing credit score, LTV, or DTI;	9,760,460
(5) and have an AUS decision (main analysis sample).	8,975,213

Note - Observation counts for the 2018-2019 Confidential HMDA data. For (0) we only keep applications for site-built single-unit properties, and only keep applications that were originated, denied by lenders or approved by lenders but not accepted by the applicant. For (2) FICO is limited to values between 300 and 850, LTV between 0 and 200, and DTI between zero and 100. For (5) we only keep applications run through one of the three government produced AUSs.

Table A.3: Denial Regressions using the full sample and the AUS sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black	0.116**	0.083**	0.050**	0.031**	0.026**	0.022**	0.024**	0.020**
	(0.008)	(0.004)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Asian	0.012**	0.034**	0.033**	0.024**	0.018**	0.017**	0.014**	0.014**
	(0.004)	(0.003)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Hispanic	0.036**	0.026**	0.014**	0.008**	0.011**	0.010**	0.010**	0.009**
	(0.005)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Other	0.104**	0.073**	0.049**	0.035**	0.024**	0.022**	0.020**	0.019**
	(0.011)	(0.005)	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Joint Race	-0.003	0.019**	0.011**	0.008**	0.004**	0.004**	0.003**	0.003**
	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Missing Race	0.070**	0.047**	0.042**	0.034**	0.023**	0.021**	0.019**	0.017**
	(0.008)	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
County by Month FE		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan Amount Bins		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Co-applicant		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log Income		Yes	Yes	Yes	Yes	Yes	Yes	Yes
FICO bins			Yes					
FICO-LTV-DTI grid				Yes	Yes	Yes	Yes	Yes
Lender FE					Yes	Yes	Yes	Yes
Tract FE						Yes		Yes
AUS sample							Yes	Yes
R-Squared	0.012	0.128	0.206	0.365	0.411	0.420	0.349	0.362
N. Obs.	9,718,280	9,467,775	9,467,774	9,441,994	9,439,678	9,354,445	8,710,063	8,627,141

Note - All the controls except for the race/ethnicity indicators are fully interacted with program by loan purpose indicators. Credit scores are discretized into buckets of 300-399, 400-499, 500-579, and buckets of 10 points for scores above 580. DTI ratios are discretized into buckets of 5 percentage points for ratios from 0 to 30 percent, single percentage point bins for DTI between 30 and 60 percent, and bins of 20 for DTI between 60 and 100. LTV ratios are discretized into buckets of 10 percentage points from 0 to 80 percent, then in 5 percentage point buckets up to 95 percent, single percentage points up to 100 percent, and then bins of LTV of 101-110, 111-120, and 121-200. The AUS sample includes all applications that were run through one of the three government produced AUS. The standard errors are clustered at the lender and county levels. Significance: * $p<0.1$, ** $p<0.05$. Data source: HMDA.

Table A.4: Denial regressions for different subsamples of the data

	Loan Purpose			Census Region			
	Purchase	Refi	Cashout	Northeast	Midwest	South	West
Black	0.017**	0.022**	0.028**	0.023**	0.027**	0.018**	0.014**
	(0.001)	(0.003)	(0.003)	(0.002)	(0.003)	(0.001)	(0.001)
Asian	0.012**	0.011**	0.024**	0.014**	0.012**	0.016**	0.011**
	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Hispanic	0.009**	0.009**	0.014**	0.012**	0.014**	0.010**	0.007**
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Other	0.011**	0.028**	0.032**	0.037**	0.019**	0.018**	0.013**
	(0.002)	(0.005)	(0.004)	(0.005)	(0.004)	(0.003)	(0.003)
Joint Race	0.002**	0.002	0.009**	0.003	0.001	0.004**	0.002**
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Missing Race	0.012**	0.023**	0.025**	0.015**	0.017**	0.018**	0.014**
	(0.001)	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)
AUS Outcome	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County by Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan Amount Bins	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Co-applicant	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log Income	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FICO-LTV-DTI grid	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.308	0.464	0.419	0.393	0.413	0.397	0.382
N. Obs.	5,922,743	1,093,879	1,693,441	1,029,362	1,715,664	3,358,565	2,519,893
Average Lender Denial Rate	0.065	0.138	0.197	0.096	0.089	0.106	0.092

Note - All the controls except for the race/ethnicity indicators are fully interacted with program by loan purpose indicators. See notes to Table A.3 for details on FICO, LTV, and DTI bins. The standard errors are clustered at the lender and county levels. Significance: * p<0.1, ** p<0.05. Data source: HMDA.

Table A.5: Denial regressions for different subsamples of the data

	Conforming, Purchase				FHA, Purchase			
	All Lenders	Depository	Nonbank	Fintech	All Lenders	Depository	Nonbank	Fintech
Black	0.015**	0.015**	0.016**	0.023**	0.023**	0.023**	0.022**	0.033**
	(0.001)	(0.001)	(0.002)	(0.008)	(0.001)	(0.002)	(0.002)	(0.005)
Asian	0.011**	0.012**	0.010**	0.014**	0.019**	0.020**	0.018**	0.031**
	(0.001)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.002)	(0.008)
Hispanic	0.007**	0.009**	0.004**	0.004	0.013**	0.015**	0.012**	0.014**
	(0.001)	(0.001)	(0.002)	(0.004)	(0.001)	(0.001)	(0.001)	(0.003)
Other	0.011**	0.013**	0.009**	0.020*	0.016**	0.014**	0.018**	0.041**
	(0.002)	(0.003)	(0.003)	(0.011)	(0.003)	(0.004)	(0.004)	(0.015)
Joint Race	0.001*	0.002*	0.001	0.005	0.004**	0.003**	0.004**	0.006
	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.002)	(0.002)	(0.006)
Missing Race	0.010**	0.008**	0.012**	0.016**	0.021**	0.020**	0.022**	0.023**
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
AUS Outcome	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County by Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan Amount Bins	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Co-applicant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log Income	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FICO-LTV-DTI grid	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.294	0.308	0.285	0.260	0.329	0.359	0.306	0.280
N. Obs.	3,675,072	2,189,590	1,441,460	208,642	1,492,301	716,631	744,114	88,364
Average Lender Denial Rate	0.050	0.048	0.053	0.089	0.100	0.098	0.101	0.153

Note - The depository institutions in columns 2 and 6 are banks and credit unions. The list of fintechs comes from Fuster et al. (2019). Nonbanks are nondespositories and exclude fintechs. All the controls except for the race/ethnicity indicators are fully interacted with program by loan purpose indicators. See notes to Table A.3 for details on FICO, LTV, and DTI bins. The standard errors are clustered at the lender and county levels. Significance: * p<0.1, ** p<0.05. Data source: HMDA.

Table A.6: Lender denial regressions separately by AUS recommendation

	AUS Accept Sample		AUS Reject Sample	
	(1)	(2)	(3)	(4)
Black	0.050**	0.018**	0.025**	0.030**
	(0.004)	(0.001)	(0.011)	(0.002)
Asian	0.008**	0.013**	0.117**	0.028**
	(0.003)	(0.001)	(0.027)	(0.003)
Hispanic	0.017**	0.008**	0.054**	0.021**
	(0.003)	(0.001)	(0.015)	(0.002)
Other	0.047**	0.017**	0.013	0.027**
	(0.006)	(0.002)	(0.038)	(0.005)
Joint Race	-0.003	0.003**	-0.028**	0.002
	(0.003)	(0.000)	(0.010)	(0.004)
Missing Race	0.051**	0.015**	0.123**	0.030**
	(0.012)	(0.002)	(0.015)	(0.003)
County by Month FE		Yes		Yes
Loan Amount Bins		Yes		Yes
Co-applicant		Yes		Yes
Log Income		Yes		Yes
FICO-LTV-DTI grid		Yes		Yes
Lender FE		Yes		Yes
R-Squared	0.006	0.237	0.009	0.564
N. Obs.	8372025	8150776	572131	445871
Average Lender Denial Rate	0.069	0.066	0.604	0.591

Note - The first two specifications limit the sample to applications that were recommended to be accepted by the AUS, while the last two include only applications that were rejected by the AUS. All the controls except for the race/ethnicity indicators are fully interacted with program by loan purpose indicators. See notes to Table A.3 for details on FICO, LTV, and DTI bins. The standard errors are clustered at the lender and county levels. Significance: * p<0.1, ** p<0.05. Data source: HMDA.

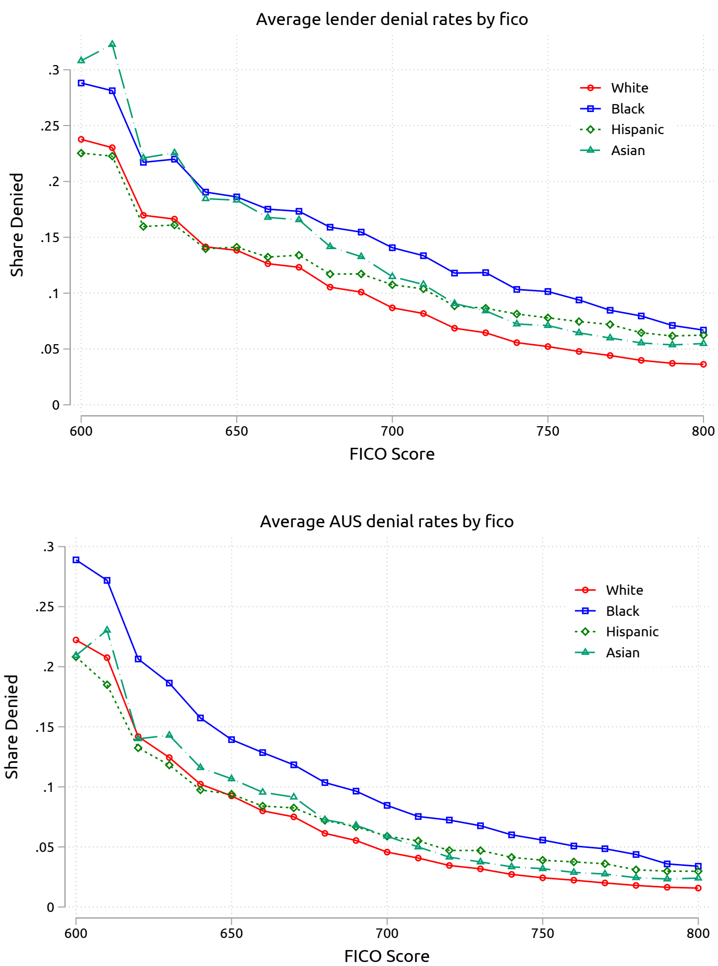


Figure A.1: Lender and AUS average denials by credit score

Note: Line show simple average denials by race. Sample includes all applications that were processed through an AUS. Data source: HMDA.

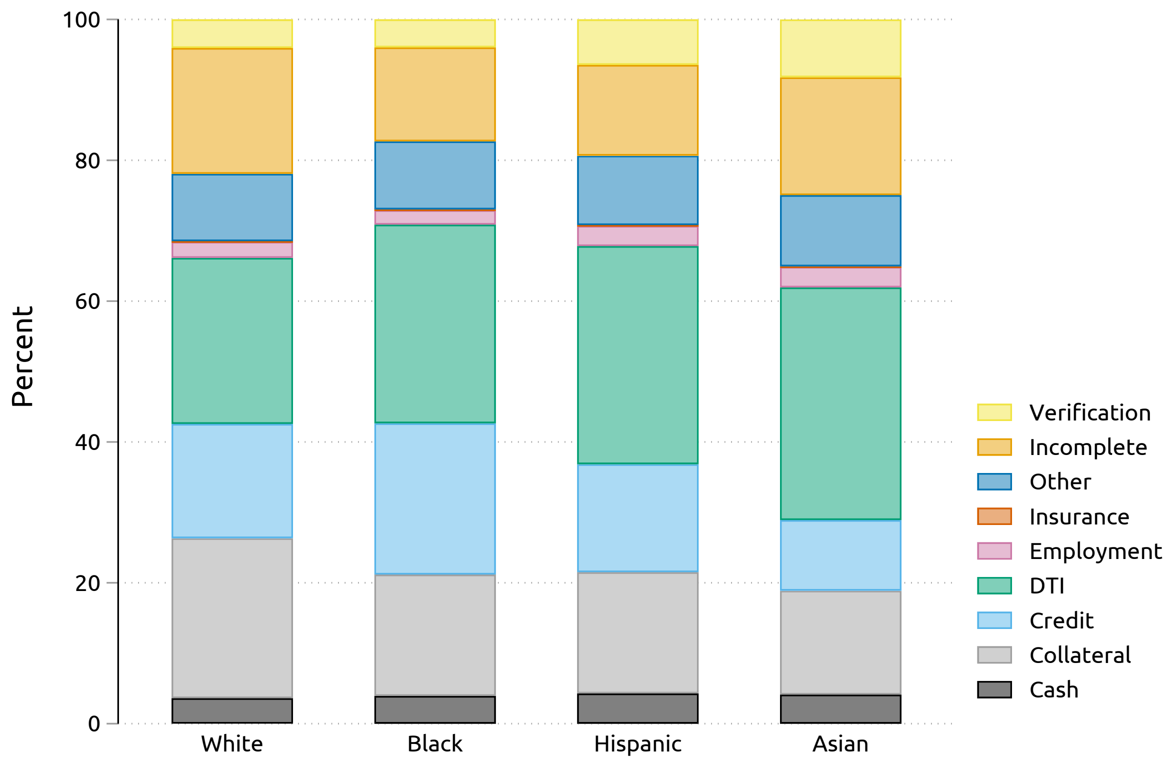


Figure A.2: Raw shares of denial reasons provided by lenders

Note: The figure plots the shares of denial reasons lenders give for denying borrowers of different races/ethnicities. Data source: HMDA.

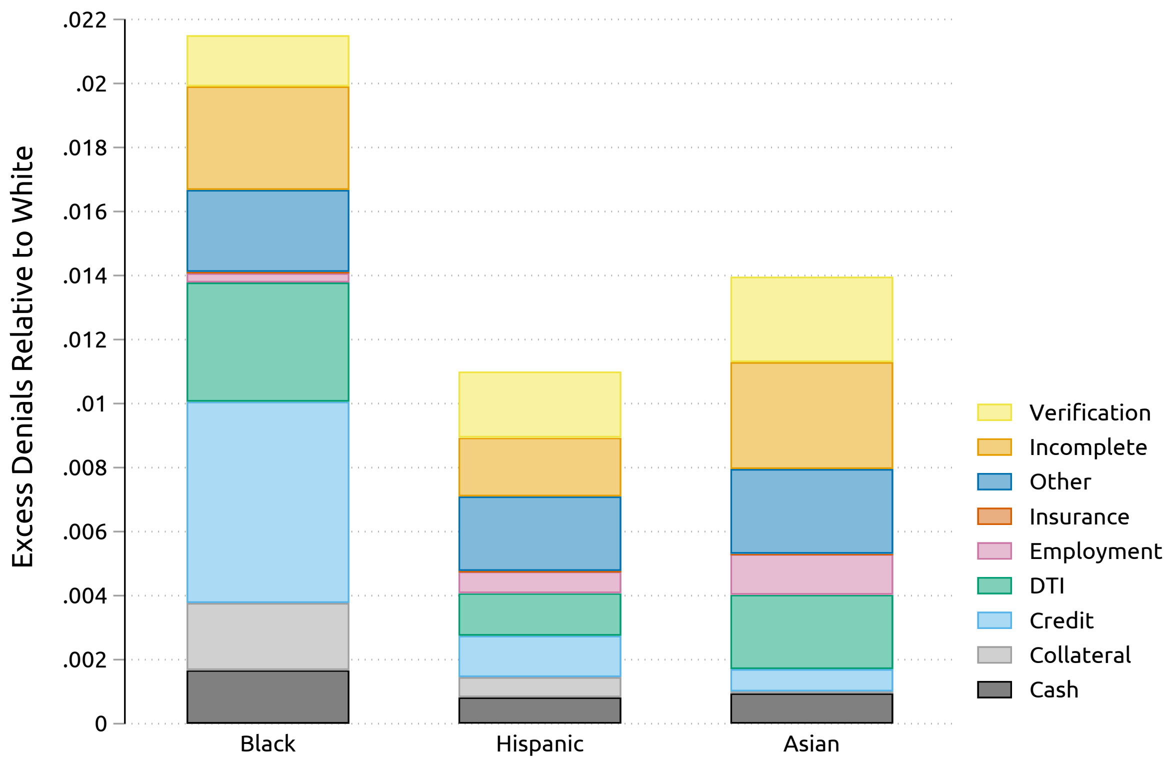


Figure A.3: Excess denial gaps broken down by lender provided denial reason

Note: The figure plots coefficients from separate regressions by denial reason controlling for all borrower and loan characteristics as in columns (3) of Table 2. Data source: HMDA.