FEDERAL RESERVE BANK
PHILADELPHIA

Working Papers
RESEARCH DEPARTMENT

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

Neil Bhutta
Federal Reserve Bank of Philadelphia
Consumer Finance Institute

Aurel Hizmo
Federal Reserve Board

Daniel Ringo
Federal Reserve Board

WP 24-09

PUBLISHED
March 2024

FRB CFI
PHILADELPHIA

ISSN: 1962-5361

DOI: https://doi.org/10.21799/frbp.wp.2024.09

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

Neil Bhutta† Aurel Hizmo‡ Daniel Ringo§

March 7, 2024

Abstract

We assess racial discrimination in mortgage approvals using confidential 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 more limited role in generating denial disparities in recent years than suggested by previous research.

Keywords: mortgage, mortgage approval, discrimination, mortgage lender, automated underwriting, credit score, fair lending

JEL Classification Codes: G21, G28, R30, R51

---

*This Philadelphia Fed working paper represents preliminary research that is being circulated for discussion purposes. The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia, the Board of Governors, or the Federal Reserve System. Any errors or omissions are the responsibility of the authors. No statements here should be treated as legal advice. Philadelphia Fed working papers are free to download at https://philadelphiafed.org/researchand-data/publications/working-papers.. 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, Mallick Hossain, Bob Hunt, Doug McManus, and Raven Molloy, 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, the Federal Reserve Bank of Philadelphia, and the University of Illinois-Chicago for helpful comments.

†Federal Reserve Bank of Philadelphia; neil.bhutta@phil.frb.org

‡Federal Reserve Board; aurel.hizmo@frb.gov.

§Federal Reserve Board; 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, about 18 percent of Black mortgage applicants were denied by lenders, more than twice the rate of White applicants.

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 1990s.[1] Since then, mortgage industry practices have evolved, 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. (1996) had collected.

In this paper, we use 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 has been illegal since 1968 under the Fair Housing Act. This notion of discrimination, often referred to as "disparate treatment," is the same as that in Munnell et al. (1996), and in the resume audit literature (e.g., Bertrand

[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 home buyers 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.

and Mullainathan, 2004; Hanson and Hawley 2011; Kline et al. 2022). Other inequities that cause disparities in mortgage outcomes are also important and have received attention in the literature, but they are beyond the scope of this paper. They include noise in credit report data as explored in Blattner and Nelson (2021), and disparate impact discrimination, i.e., lenders making credit decisions on the basis of characteristics that are correlated with race but unrelated to creditworthiness, as discussed in Bartlett et al. (2022).[3] Racial disparities in applicant risk characteristics could also be due to biased treatment that minority individuals were subjected to throughout their lives (e.g., in education or labor markets) prior to applying for a loan. Our paper confines its focus to investigating the extent to which disparate treatment by mortgage lenders further disadvantages minority households, over and above any previous harms imposed on them by broader society.

Overall, we find a much smaller role for differential treatment in generating denial disparities compared to the benchmark estimates of Munnell et al. (1996), consistent with significant progress in fair lending over the last 30 years. Our main analysis proceeds in three parts. First, we quantify group differences in measures of applicant risk, documenting that Black and Hispanic applicants tend to be more leveraged and have much lower credit scores. For example, the average credit score of Black applicants is over 40 points lower than that of 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 "colorblind" in that race and ethnicity (or proxies like neighborhood location) cannot be used in the algorithm, as we discuss in Section 2.[4]

Second, we try to explain lenders' rejection decisions via regression analysis. AUS recommendations only partially explain the Black-White and Hispanic-White gaps in lender denial rates, as lenders do not always adhere to these recommendations. This lack of adherence raises concerns that lenders could be overriding automated decisions in biased ways. On the other hand, lenders may impose "overlays," or deviations in credit standards from AUS, without regard to race, and these

[3] Related to Blattner and Nelson (2021), Fuster et al. (2022) find that more sophisticated risk prediction models could help slightly reduce denial disparities.

[4] Throughout this paper, we group Fannie Mae and Freddie Mac within the set of U.S. government mortgage institutions. Fannie Mae and Freddie Mac continue to be in conservatorship and thus under the management of the Federal Housing Finance Authority.

overlays can generate different decisions from AUS.[5] To account for overlays, we flexibly control for credit score, debt payment-to-income (DTI) ratio, loan-to-value (LTV) ratio, and other risk factors, on top of AUS recommendations, and find a residual Black-White denial gap of 2 percentage points, and residual Hispanic and Asian gaps of about 1 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 underwriting 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.

Third, we assess whether excess denials might be driven by omitted variables rather than discrimination. We start by showing that there are racial gaps in AUS recommendations after controlling for the risk factors we observe in the HMDA data. Given that AUS is colorblind, we interpret these residual AUS gaps as reflecting differences in unobserved (to us) risk or eligibility variables—for example, applicants' liquid reserves. Then we test whether overlays on these unobservables drive excess denials, by exploiting cross-sectional differences in lender policies. We construct a lender-specific measure of the "strictness" of underwriting policies (a lender's propensity to deny White applicants conditional on observables), and correlate lender strictness with lender-specific excess denials.[6] If minority applicants are relatively riskier in these unobserved dimensions, this should create a positive correlation between strictness and excess denials. Indeed, we find a strong correlation between strictness and excess denials and that the least-strict lenders have excess minority denials close to zero. Furthermore, stricter lenders also tend to have better unexplained loan performance, consistent with stricter lenders imposing tougher standards on unobserved risk factors. These findings suggest that racial disparities in unobserved risk help drive excess denials and therefore the 1-2 percentage points of excess denials overstate the role of disparate treatment in denial disparities.

[5] 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 overlays 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 5.2.2, we provide evidence that lenders with stricter overlays generally experience better loan performance.

[6] Our concept of lender-specific strictness is similar to that of a judge-specific "propensity to release" that Arnold et al. (2018) and Arnold et al. (2022) use in their studies of racial bias in bail decisions.

We also indirectly test whether excess denials reflect discrimination by further examining heterogeneity in excess denials across lenders and markets. 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.[7] However, we fail to find clear evidence for such correlations.

Finally, we test for several other margins of disparate treatment in the application process that would not be picked up in our analysis of HMDA application decisions. First, we draw on the National Mortgage Database (NMDB®) to test if lenders differentially discourage minorities from even submitting a formal application that could be observed in the HMDA data (Hanson et al. 2016; Ross et al. 2008). Second, we use the survey component of the NMDB — the National Survey of Mortgage Originations (NSMO) — to test whether lenders exhibit bias in their measurement of applicant income, which could make minority applicants appear riskier "on paper" and help justify rejections. Third, we again use the NSMO data to test for racial and ethnic differences in lenders' service quality (e.g., on-time closing). To briefly preview the findings, we fail to find clear evidence of differential discouragement or that lenders more frequently understate minority applicants' income. However, we do find evidence that lenders provide lower quality service to minority borrowers, suggestive of potential differential treatment along a dimension that has been largely unexplored in the literature.[8] Further discussion of these findings is available in Section 6, and full details are in the Appendix.

A few recent papers have studied denial gaps and discrimination. Bartlett et al. (2022) estimate denial gaps of 7-10 percentage points, but this estimate does not condition on applicant credit scores and other key underwriting factors. Giacoletti et al. (2023) similarly find a 7 percentage point Black-White denial gap without conditioning on credit score, LTV, or DTI, and estimate that at least half of this 7 pp denial gap reflects disparate treatment, which they argue can be

[7]See Kopkin (2018) for earlier evidence using public HMDA data that denial gaps are correlated with geographic measures of racial animus.

[8]To be sure, these differences in service quality are among those who were accepted for and obtained a new mortgage. In Section 5.4, we use the data on lenders' stated reasons to infer their explanations for excess denials and find that lenders often attribute excess denials to "verification" and "incomplete application." It is possible that difficulties among minorities in completing applications reflects some degree of differential assistance or effort by lenders helping applicants with the application process. Along these lines, Frame et al. (2022) find that minorities are more likely to complete mortgage applications and originate mortgages when paired with a minority loan officer.

identified from within-month variation across race groups in the timing of loan closings. Finally, Park (2021) uses confidential HMDA data to test whether model-based loan-level loss probabilities under severe economic stress can explain racial denial disparities. Relative to this literature, we take advantage of expanded confidential HMDA data to provide the first estimates of contemporary disparate treatment that are comparable to the seminal Munnell et al. (1996) work from the 1990s. We find smaller racial differences in conditional denial rates relative to Munnell and other work. Moreover, we use the data on AUS decisions and a novel lender-level analysis to show that these residual gaps can be further explained (at least in part) by additional underwriting factors still not observed in the HMDA data.

Relatedly, several recent papers focus on discrimination in mortgage interest rates and fees (Bartlett et al. 2022; Bhutta and Hizmo 2020; Mota et al. 2022; Willen and Zhang 2023). In our appendix, we revisit this topic using the 2018-2019 confidential HMDA data. We find unexplained interest rate gaps of 1 and 2 basis points for Black and Hispanic borrowers, respectively (equivalent to a \$1-\$2 increase in the monthly payment for a \$200,000 loan), and unexplained gaps in upfront origination fees of 0.02 and 0.04 percent of the loan amount (i.e., \$40 and \$80 for a \$200,000 loan). These magnitudes are within close range of the recent literature.

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 White borrowers 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. 2023). 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" focus on taste-based discrimination (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 fruitful.

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 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 disparate treatment in mortgage approval decisions, which may have contributed to the Black-White homeownership gap by discouraging minority families from applying for loans (Charles and Hurst 2002).

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.[9] 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 had some success does not imply that individual racial prejudice has been eliminated, or that overt discrimination would not become more common if fair lending enforcement were to be relaxed. Furthermore, it is important to keep in mind that the racial and ethnic differences in observable risk characteristics to which we attribute most of the denial disparities can be traced at least in part to discriminatory treatment elsewhere (e.g., the labor market) or in the intergenerational persistence of historical unequal treatment.

2 The Mortgage Application and Underwriting Process

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,

[9]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.

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 downpayment funds may be dissuaded from moving forward.[10]

The second stage is to submit a formal application, along with documentation of income and assets (e.g., pay stubs, tax returns, account statements). Loan officers help ensure borrowers provide the right documentation and fill out the application correctly. However, 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).[11] 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 it does report the risk factors it considers (see Fannie Mae A 2021). These are:

[10]In Section 6.1, we describe an analysis testing for differential discouragement of minorities. For previous evidence of differential treatment in preapplication stages, see Hanson et al. (2016) and Ross et al. (2008).

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

- **Credit-history risk factors**
  - Length of credit history
  - Delinquent accounts
  - Presence of active installment loan accounts
  - Revolving credit utilization
  - Public records, foreclosures, and collection accounts
  - Credit inquiries
  - Rent payment history (this factor was not included in DU during our study period. See Fannie Mae B 2021)

- **Non-credit-history risk factors**
  - Borrower's equity and LTV ratio
  - Amount of liquid reserves
  - Loan purpose
  - Loan term
  - Loan amortization type (fixed versus adjustable rate)
  - Occupancy type
  - DTI ratio
  - Housing expense ratio
  - Property type
  - Presence of co-borrowers
  - Cash flow assessment (used only when no borrower has a credit score)
  - Variability of income (during our study period, this factor was self-employment income. See Fannie Mae C 2021)

Fannie Mae A (2021) describes how each of these factors influence the predicted risk of a loan. LPA and TOTAL consider similar factors. 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). From Fannie Mae's selling guide: "DU conducts its analysis uniformly, and without regard to race, gender, or other prohibited factors. DU uses validated, statistically significant variables that have been shown to be predictive of mortgage delinquency across all groups." For our purposes, the significant points are that 1) AUS do not directly consider applicant race or ethnicity (or neighborhood location), and 2) they do consider factors beyond those observable to us in HMDA.

Ultimately, the AUS provides a recommendation consisting 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. The AUS recommendation is not binding on lenders. 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. 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.[12]

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.[13] 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

[12]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 percent 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 percent of these portfolio loans had a positive AUS recommendation. See Section 3 for a fuller description of these data.

[13]Lenders often use a FICO score in underwriting, but they may also use a VantageScore, or other scores. Throughout the paper, we use the term "FICO" to refer to credit scores in general, rather than as a reference to any specific type of credit score.

to applications for typical fixed rate, 30-year loans, and drop any applications with missing or invalid credit scores.[14] We also drop "jumbo loans," which have loan sizes above eligibility limits for securitization through government programs. 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 (see Section 2 for the full list of factors considered in DU, the most popular AUS).

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

[14] 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.

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 independent idiosyncratic element arising from human error in lender decisions.[15]

The notion of discrimination we are pursuing in this paper is disparate treatment. We consider a lender to be engaging in disparate treatment if, in expectation, the credit decisions reached for a group of applicants differs from that of an otherwise identical group that presented a different race or ethnicity. In other words, lender i engages in discrimination against Black applicants relative to White applicants if:

$$\int h_i(X^*, u, w, black)dF_B(X^*, u, w) > \int h_i(X^*, u, w, White)dF_B(X^*, u, w), \tag{3}$$

where $F_B(\cdot)$ represents the joint cumulative distribution function of the underwriting factors (X^*, u, and w) that characterize the population of Black mortgage applicants.

Disparate treatment encompasses both taste-based and statistical discrimination. It is distinct, however, from disparate *impact* discrimination. Disparate impact occurs when a lender makes credit decisions on the basis of nonracial characteristics of the applicant that serve as a proxy for race (i.e., they are correlated with race and the lender cannot justify their consideration as a business necessity).[16] This could take the form of the lender considering applicant characteristics (in X, w, or u) that are correlated with race but do not predict loan performance or profitability. An investigation of whether lenders are justified in underwriting loans on the basis of all the non-racial factors they consider is beyond the scope of this paper (although see Section 5.2.2 for some suggestive evidence).

We begin our investigation of disparate treatment by considering racial and ethnic differences in the outcomes of, and inputs to, D^i_{Lender} and D_{AUS}. The top row of Table 1 shows that lenders denied 18 percent of Black mortgage applicants in 2018-2019, substantially higher than the 8 percent denial rate for White applicants.[17] 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

[15]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 $h_i(\cdot)$.

[16]See Federal Reserve (2016).

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

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 LTVs and DTIs on average. These differences could be driving the differences in denial rates, as opposed to racial bias on the part of mortgage lenders.[18]

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. Thus, even if application decisions were based purely on government AUS recommendations, the data suggest that the Black-White denial gap in 2018-19 would still have been about 9 percentage points.

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 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.[19] In terms of equations

[18] These racial disparities in underwriting factors could themselves be due to racial bias applicants suffered previously in their life history. For example, employment discrimination could cause minority applicants to become delinquent on other debts, hurting their credit scores.

[19] 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.

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.[20] We include a fully interacted set of discretized bins of credit score, LTV, and DTI ratio.[21] 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 presence of a co-applicant, indicators for reported income (discretized into deciles), and a lender fixed effect. All these covariates are fully interacted with the indicators for loan purpose and program.[22]

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.[23] Comparing columns 1 and 3 indicates that we can explain over three-quarters of the Black-White denial gap.[24] We refer to the remaining gaps as "excess denials."

In the appendix B, 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

[20]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. See online appendix E for details.

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

[22]In the online appendix, we provide results from an even more flexible specification, in which the controls consist of the full set of interactions between indicators for bins of income, credit score, LTV, and DTI ratio as well as indicators for a co-applicant, AUS denial, loan program, and purpose (the regressions further control for lender fixed effects by loan program and purpose). Results are very similar to those of our main specification—see table A.7.

[23]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.

[24]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.

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.

Table 2 also shows that estimated excess denials for other racial minorities (American Indian, Alaska Native, Hawaiian or other Pacific Islander, or applicants who report multiple minority races or ethnicities) fall between those of Black and Asian applicants. Applicants whose race was not recorded in HMDA ("Missing") have similar excess denials. Applications from households consisting of one non-Hispanic White and one minority applicant ("Joint") have the lowest estimated excess denials across the minority groups.

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 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.[25] 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

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

strictness and excess denials across lenders if unobservable characteristics of minority applicants appear riskier than those of White applicants.

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 deviation from the market average 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 $h_i(\cdot,\cdot,\cdot,White) \neq h_j(\cdot,\cdot,\cdot,White)$. Higher strictness means lenders are imposing tougher standards on their borrowers.[26]

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

[26]This measure of strictness could potentially suffer from selection bias if riskier applicants sort themselves to different lenders. In Section 5.2.1, we provide evidence that our measure of strictness is truly capturing lender overlays, as opposed to selection by applicants to lenders according to their unobservable risk characteristics. We also show that this measure of strictness is highly predictive of overlays imposed on regions of the risk characteristic distribution where most denials of Black applicants occur.

strictness and excess denials.[27] 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. At the same time, the least strict lenders tend to have excess denials close to zero. This finding suggests that excess denials are at least partly a result of tight lender standards on unobserved risk factors, and therefore our 1-2 percentage point excess denials estimate overstates the role of disparate treatment in denial disparities.

5.2.1 Evaluation of Lender "Strictness"

In Section 5.2, we find that lender strictness is highly correlated with excess denials, which suggests that excess denials are driven at least in part by unobserved applicant risk factors. This interpretation rests on two assumptions.

First, we are assuming that the differences in conditional denial rates (i.e., strictness) are due to differences in lender policies, rather than differences in the unobservable risk characteristics of the pools of applicants that apply to each lender. Second, we are assuming that a single index of strictness, measured over each lender's population of White applicants, is informative of the overlays they impose everywhere on the distribution of applicant characteristics. White and minority applicants have different distributions of these characteristics, so it is possible that a lender that appears strict among the White population (e.g., requiring a higher amount of liquid reserves than average for borrowers with credit scores in the typical range for White applicants) might have less strict overlays in the range relevant for minority populations (e.g., requiring lower liquid reserves than average for borrowers with credit scores in the typical range for Black applicants). If so, we could be mismeasuring the effective strictness minority applicants face at each lender.

To test the first assumption, we first consider whether applicants to stricter lenders are observably riskier. The first three rows of Table 3 show the comparisons of credit score, LTV ratio,

[27] 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 5.2.2.

and DTI ratio for White applicants between the top tercile and bottom tercile of lenders in terms of strictness. Applicants to stricter lenders appear, if anything, less risky on observable dimensions than those to the least-strict lenders. In the fifth through seventh rows, these comparisons are repeated for Black applicants. Again, stricter lenders tend to attract slightly less risky applicants than less-strict lenders. The ninth through eleventh rows show that Black-White gaps in risk characteristics appear very similar in high- and low-strictness lenders. Overall, these findings fail to suggest that stricter lenders have riskier applicant pools, or that White and Black applicants differentially sort into lender strictness according to their observable risk characteristics.

Furthermore, we can test whether stricter lenders are attracting applicants with riskier *unobservable* characteristics. To measure unobservable riskiness of the applicant pools, we calculate "AUS residuals" for each lender. AUS residuals are defined by race (analogously to lender strictness) as the idiosyncratic AUS denial rate of applicants of a given race to that lender, conditional on observables. This AUS residual is estimated as the lender fixed effect from running the column 5 specification of Table 2 on the applicant population, by race. Because the AUS algorithms themselves do not differ across lenders, AUS residuals should pick up only differences in applicant risk characteristics—specifically, those not explained by the observables.

From the fourth row of Table 3, we can see that AUS residuals for the White population are no larger at high-strictness lenders than at low-strictness lenders. This suggests little or no sorting by unobservable risk characteristics to lenders of different strictness. Row 8 similarly shows no pattern of such sorting by Black applicants.

Finally, in row 12, we can see that there is no substantive difference in AUS excess denials by lender strictness—that is, stricter lenders do not have larger Black-White gaps in unobservable risk characteristics than less-strict lenders do. This result suggests no differential sorting along unobservable risk characteristics by race to high- or low-strictness lenders. For comparison, *lender* excess denials of Black applicants relative to White are shown in row 13. Row 13 essentially repeats the analysis of Section 5.2 and shows lender excess denials are considerably higher among stricter lenders. All together, these results suggest that 1) measured strictness is picking up differences in lender overlays, not differences in unobserved applicant risk characteristics; and 2) there is no differential sorting by race on unobservable risk characteristics to stricter or less-strict lenders. Stricter overlays are associated with higher excess denials of minorities, even though minority applicants at stricter lenders appear no riskier than those at less-strict lenders.

To test the second assumption, we compare our baseline measure of strictness to one estimated from a subsample of lower-income, lower-credit score applicants. We define this subsample to be applicants with both an income and credit score less than the 90^{th} percentile of those fields among Black applicants who were denied credit by their lender.[28] Overlays imposed on this region of the distribution are relevant for explaining excess denials of minority applicants. Among this subsample, we estimate a new set of lender strictness values (following the same procedure as in Section 5.2) and compare them to our baseline. We plot lender strictness measured on this low-income, low-credit score subsample against our baseline strictness in Panel A of Figure 3. As can be seen, the correlation is extremely tight. Our measure of strictness is a very good predictor of that lender's idiosyncratic probability of denying an applicant whose risk characteristics put them in the range of an average denied Black applicant.

Furthermore, we show that lender strictness estimated on the White population predicts an analogous measure of strictness calculated purely on the Black population. That is, we estimate Black strictness as the lender fixed effect from a denial regression (including our full set of controls) run exclusively on the sample of Black applications. Estimated Black strictness is plotted against our baseline (White) strictness in Panel B of Figure 3, and the correlation can again be seen to be very tight. In other words, lenders who impose tight overlays seem to impose them across the board.

5.2.2 Lender Strictness and Loan Performance

While we cannot identify all the overlays that 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.

[28]This procedure keeps only applicants with credit score < 739 and income less than \$121,000, which drops about 52 percent of our sample.

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 reflect the issuer's own underwriting criteria as a lender.

We end up with 48 matched institutions that directly securitized over 75 percent of their FHA and VA originations through Ginnie Mae, 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 strictness specific to applications for FHA and VA loans.

The left-hand charts of Panels 1 and 2 in Figure 4 show scatterplots and linear fits of Ginnie Mae 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.[29] The right-hand charts of Panels 1 and 2 in Figure 4 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.[30]

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

[30]Our measure of lender-specific strictness is similar to the judge-specific measure of "propensity to release" used by Arnold et al. (2018) and Arnold et al. (2022), in that we both estimate the idiosyncratic tendencies of particular decision makers (lenders and judges, respectively) to make a binary choice (credit and bail denial, respectively)

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.

First, we compare fintech lenders to traditional mortgage lenders. By automating more of the application process, fintechs cut out some human judgment 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 4. We find excess denials are, if anything, higher at fintech lenders, the opposite result we would expect if excess denials reflect racially biased human judgment.[31]

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 4. 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 4. 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 4). This suggests that White-

conditional on controls in the context of racial discrimination. However, the loan performance data we describe in this sub-section are not matchable at the loan level. We therefore do not have loan performance by lender-and-race and so cannot make use of the estimators from those papers, which use judge-by-race *ex-post* misconduct data. Instead, we use the loan performance data to validate the idea that stricter lenders are mitigating risk by imposing overlays on factors both observable and unobservable in HMDA.

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

minority differences in unobservable risk factors are larger in markets with higher racially charged search frequencies, potentially explaining the similar 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 appendix C. 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 been completed), contributing to excess denials. Difficulties with verification and application completion could reflect lenders providing poorer service to minority applicants (e.g., providing less assistance with filling out the application). We show direct evidence of this channel in Section 6.3 by using survey data on borrower experiences and satisfaction with the lending process.

6 Assessing Other Dimensions of Disparate Treatment

In this section, we briefly describe three additional analyses in which we test for other forms of disparate treatment in the application process. A more complete discussion of data sources, methods, and results can be found in the online appendix. Additionally in Appendix Section G, we replicate the findings in recent literature of economically small differences in mortgage pricing by race and ethnicity in the 2018-2019 confidential HMDA data.

6.1 Do Lenders Discourage Minorities from Applying?

Throughout Sections 4 and 5, we have investigated racial and ethnic disparities in the probability of denial conditional on a loan application appearing in the HMDA data. However, as we describe in Section 2, there is typically a pre-screening process involving a credit check, and lenders may discourage some borrowers from applying in these early stages of contact without an explicit credit denial ever being reported under HMDA. Lenders could differentially discourage minorities from applying by being less likely to process applications from minorities conditional on the initial credit check, or by being less likely to even conduct a credit check for minorities (e.g., by not responding at all to initial contact by applicants, as found in Hanson et al. (2016)). If there is substantial discouragement of minority applicants relative to otherwise similar White applicants, our analysis based purely on the HMDA data could be misleading.

To gain insight into this issue, we examine data on mortgage *inquiries*, which are indicators in a consumer's credit file signaling that a mortgage lender has run a credit check because the consumer is shopping around for a new mortgage. We obtain inquiry data by race and ethnicity for existing mortgage borrowers from the National Mortgage Database (NMDB). The NMDB — created by the Federal Housing Finance Agency and Consumer Financial Protection Bureau — matches loan-level panel data on mortgages from several sources (e.g., HMDA, administrative records from the GSEs and FHA) to panel data on credit bureau records of the individual mortgagors, and is designed to be representative of the population of all outstanding mortgages in the U.S. When we observe a mortgage inquiry for a given mortgage borrower, this implies that the borrower is seeking to refinance or buy a new home. To measure lender discouragement, we estimate ratios of HMDA applications to credit inquiries, by race and ethnicity and conditional on credit score; ratios less than one indicate discouragement, in that the number of borrowers seeking a new mortgage (as measured by inquiries) exceeds the number of formal mortgage applications observed in HMDA.[32]

To summarize the findings (see appendix D for additional details), we do not find consistent evidence of bias against minorities in lender discouragement. We find that Black borrowers are somewhat *more* likely than White borrowers to have a HMDA application conditional on having

[32]To be be clear, we do not link HMDA applications to credit inquiries at the borrower level; rather, we compare counts of HMDA applications to counts of inquiries from the NMDB data over a given time frame. To make this comparison meaningful requires a number of data adjustments and data timing considerations, which we describe in detail in the appendix.

had an inquiry, whereas Hispanic borrowers are somewhat less likely than White borrowers to have had a HMDA application conditional on having had an inquiry. We find that application-to-inquiry ratios are much lower at lower credit scores, consistent with lenders discouraging the riskiest candidates from applying; but this discouragement of low-score individuals occurs to a similar extent for both White and minority borrowers. Finally, we find that the conditional likelihood of having an inquiry is similar across racial and ethnic groups. Overall, while there are some data limitations as we discuss in the appendix, our findings fail to indicate that lenders engage in large-scale differential discouragement of minority prospective borrowers at the pre-application stage.

6.2 Do Lenders Understate Minority Incomes?

Lenders must measure borrower income by collecting information about their sources of income; if lenders understate minority incomes, it could cause minority applicants to appear riskier "on paper" and lead to higher minority denial rates. To test whether lenders may be systematically understating the income of minorities, we use data from the National Survey of Mortgage Originations (NSMO), which is a survey component of the NMDB. The NSMO is a quarterly survey of mortgage borrowers who recently got a new mortgage, and gathers data along several dimensions, such as their demographics, their housing and economic expectations, how they found and selected their mortgage, and their experiences in borrowing process. A unique feature of the NSMO is that each survey respondent is linked to their detailed administrative mortgage and credit history data. This feature allows us to compare self-reported income of borrowers to the income measured by lenders for underwriting purposes, conditional on a detailed set of controls, such as credit score, loan type, loan amount, age, co-borrower status, self-employment status, and more. We find that while self-reported income often exceeds lender-reported income, higher self-reported income is not more likely for minority applicants relative to White applicants. See appendix E for further details and the complete results.

6.3 Does Lender Service Quality Vary by Race or Ethnicity?

Many aspects of the mortgage origination process are difficult for regulators to monitor and for lenders to automate, which could lead to greater variation in these processes by race and ethnicity. We use the NSMO data (described above in Section 6.2) to test for differences by race and ethnicity in borrower experiences and satisfaction with their lender and the lending process. First, we find that Black borrowers are more likely than non-Hispanic White borrowers to report that their closing date was delayed or postponed, and far more likely to have experienced delays in processing that caused them to have to redo paperwork. Importantly, these results are conditional on a detailed set of controls, most of which are based on administrative data and therefore are precisely measured. Thus, we can rule out that these differences by race are driven by loan or borrower characteristics that might be associated with delays in the loan process, such as having a high LTV or getting an FHA loan. Second, we also find that Black and Asian borrowers express less satisfaction with their borrower experience (e.g., timeliness of disclosures) and with their lender. Overall, these results are suggestive of significant differences in service quality by race and ethnicity. See appendix F for further details and the complete results.

7 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 underwriting factors that are unobservable in the HMDA data. Thus, we conclude that our 1-2 percentage point estimate of excess denials overstates the role of disparate treatment in generating mortgage denial disparities.

We find a much smaller role for disparate treatment in generating denial disparities compared to the benchmark estimates of Munnell et al. (1996), implying significant progress in fair lending for mortgages over the last 30 years. As such, strengthening fair lending enforcement in the mortgage market may only yield small improvements in mortgage credit access for minority families. Instead, because disparities in credit access largely reflect differences in underlying measures of applicant credit risk, policies that aim to reduce gaps in observed credit risk, potentially through better measurement of credit risk, may be more fruitful.

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.

__ , __ , and Peter Hull, "Measuring racial discrimination in bail decisions," *American Economic Review*, 2022, *112* (9), 2992–3038.

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.

Bertrand, Marianne and Sendhil Mullainathan, "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination," *American Economic Review*, 2004, *94* (4), 991–1013.

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 A, "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.

Fannie Mae B, "Selling Guide Announcement (SEL-2021-08)," "https://singlefamily.fanniemae.com/media/28876/display" 2021. Accessed: 2023-2-17.

Fannie Mae C, "Selling Guide Announcement (SEL-2021-02)," "https://singlefamily.fanniemae.com/media/25256/display" 2021. Accessed: 2023-2-17.

Federal Reserve, "Federal Fair Lending Regulations and Statutes," in "Consumer Compliance Handbook" 2016, chapter V.

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.

__ , Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther, "Predictably unequal? The effects of machine learning on credit markets," *The Journal of Finance*, 2022, *77* (1), 5–47.

Gerardi, Kristopher, Paul S Willen, and David Hao Zhang, "Mortgage prepayment, race, and monetary policy," *Journal of Financial Economics*, 2023, *147* (3), 498–524.

Giacoletti, Marco, Rawley Heimer, and Edison G Yu, "Using High-Frequency Evaluations to Estimate Discrimination: Evidence from Mortgage Loan Officers," *Available at SSRN 3795547*, 2023.

Hanson, Andrew and Zackary Hawley, "Do landlords discriminate in the rental housing market? Evidence from an internet field experiment in US cities," *Journal of Urban Economics*, 2011, *70* (2-3), 99–114.

__ , __ , 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.

Kline, Patrick, Evan K Rose, and Christopher R Walters, "Systemic discrimination among large US employers," *The Quarterly Journal of Economics*, 2022, *137* (4), 1963–2036.

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.*

Mota, Nuno, Mark Palim, and Susan Woodward, "Mortgages are still confusing, and it matters: How borrower attributes and mortgage shopping behavior impact costs," Technical Report, Fannie Mae 2022.

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 and David Hao Zhang, "Testing for Discrimination in Menus," Technical Report 2023.

Table 1: Summary statistics

	All	White	Black	Hispanic	Asian	Other	Joint	Missing
Lender Denial Rate	0.10	0.08	0.18	0.11	0.10	0.16	0.08	0.15
AUS Denial Rate	0.06	0.05	0.14	0.07	0.05	0.11	0.06	0.07
Lender-AUS Disagree	0.09	0.08	0.15	0.10	0.08	0.14	0.08	0.12
Loan Amount (000)	253	246	230	246	333	256	307	268
Income (000)	92	93	78	79	103	82	117	95
Credit Score	720	726	685	703	738	704	714	718
LTV (%)	84.14	83.37	90.70	87.66	79.91	86.04	85.93	82.51
DTI (%)	39.10	37.98	42.76	42.28	39.84	41.23	38.50	39.53
N. Obs.	8,975,213	5,546,031	676,171	876,683	415,911	84,073	167,457	1,208,887

Note - Table shows average characteristics for purchase and refinance applications in 2018 and 2019 for first-lien, 30-year fixed-rate mortgages, on owner-occupied single-unit homes for which an AUS recommendation was reported. Sample excludes withdrawn or incomplete applications. "Other" refers to applicants of other minority groups, including American Indian, Alaska Native, Hawaiian, and other Pacific Islander applicants, as well as individuals reporting more than one minority race or ethnicity. "Joint" refers to applications from multiple applicants, one of whom is non-Hispanic White applicant and one of whom is a minority. Data source: HMDA.

Table 2: Estimates of excess denials

	Lender Denial			AUS Denial	
	(1)	(2)	(3)	(4)	(5)
Black	0.091**	0.043**	0.020**	0.085**	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	0.029**	0.019**	0.009**	0.020**	0.000
	(0.004)	(0.003)	(0.001)	(0.002)	(0.001)
Other Minority	0.070**	0.034**	0.018**	0.058**	0.008**
	(0.007)	(0.006)	(0.002)	(0.009)	(0.001)
Joint Race	-0.001	0.002	0.005**	0.004	-0.000
	(0.003)	(0.002)	(0.001)	(0.003)	(0.001)
Missing Race	0.063**	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
Income Bins			Yes		Yes
FICO-LTV-DTI grid			Yes		Yes
Lender FE			Yes		Yes
R-Squared	0.010	0.233	0.400	0.009	0.357
N. Obs.	8,944,156	8,944,156	8,737,868	8,944,156	8,737,868

Note - All the controls except for the race/ethnicity indicators are fully interacted with program by loan purpose indicators. The FICO-LTV-DTI grid is a set of fixed effects created by interacting credit score bins x LTV bins x DTI bins. 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. Income bins are fixed effects that denote income deciles. The standard errors are clustered at the lender and county levels. Significance: * p<0.1, ** p<0.05. Data source: HMDA.

Table 3: Do riskier applicants sort into “strict” lenders?

	Non-Strict Lender	Strict Lender	Difference	P-value
White				
FICO	711.8	718.3	6.500	0.254
LTV	86.45	81.97	-4.480	0.000
DTI	39.00	39.05	0.050	0.924
Lender Strictness	-0.035	0.096	0.131	0.000
AUS Strictness	0.033	0.019	-0.014	0.263
Black				
FICO	675.1	685.0	9.900	0.064
LTV	92.55	88.16	-4.390	0.000
DTI	43.01	42.78	-0.230	0.655
Lender Strictness	-0.060	0.116	0.176	0.000
AUS Strictness	0.058	0.052	-0.006	0.741
Black-White				
FICO	-36.72	-33.33	3.39	0.663
LTV	6.101	6.188	0.087	0.953
DTI	4.005	3.733	-0.272	0.699
Lender Strictness	-0.025	0.020	0.045	0.012
AUS Strictness	0.024	0.033	0.008	0.714
Black Excess Denials				
Lender Excess Denials	0.007	0.038	0.031	0.000
AUS Excess Denials	0.015	0.019	0.004	0.322

Note - Strict lenders are defined as lenders at the top tercile of the strictness distribution and non-strict as those in the bottom tercile of strictness. AUS strictness is the lender fixed effects from a regression of AUS decisions on lender fixed effects and all borrower and loan controls. Data source: HMDA.

Table 4: Indirect tests for discrimination

	Lender Denial			AUS Denial		
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.020**	0.023**	0.021**	0.016**	0.015**	0.015**
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
Hispanic	0.010**	0.013**	0.011**	0.001	0.004**	0.001*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Asian	0.014**	0.018**	0.014**	0.002**	0.004**	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	0.005			-0.002		
	(0.005)			(0.002)		
× Asian	0.002			0.000		
	(0.005)			(0.002)		
Top 4 Lenders' Share						
× Black		-0.009			0.004	
		(0.009)			(0.009)	
× Hispanic		-0.018**			-0.021**	
		(0.007)			(0.007)	
× Asian		-0.023**			-0.009	
		(0.009)			(0.007)	
Racially Charged Search Rate						
× Black			0.002**			0.003**
			(0.001)			(0.001)
× Hispanic			0.002**			0.002**
			(0.000)			(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
Income Bins	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	0.398	0.398	0.397	0.356	0.356	0.357
N. Obs.	7,294,332	7,294,332	7,172,937	7,294,332	7,294,332	7,172,937

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, DTI, and income bins. The standard errors are clustered at the lender and county levels. Significance: * p<0.1, ** p<0.05. Data source: HMDA.

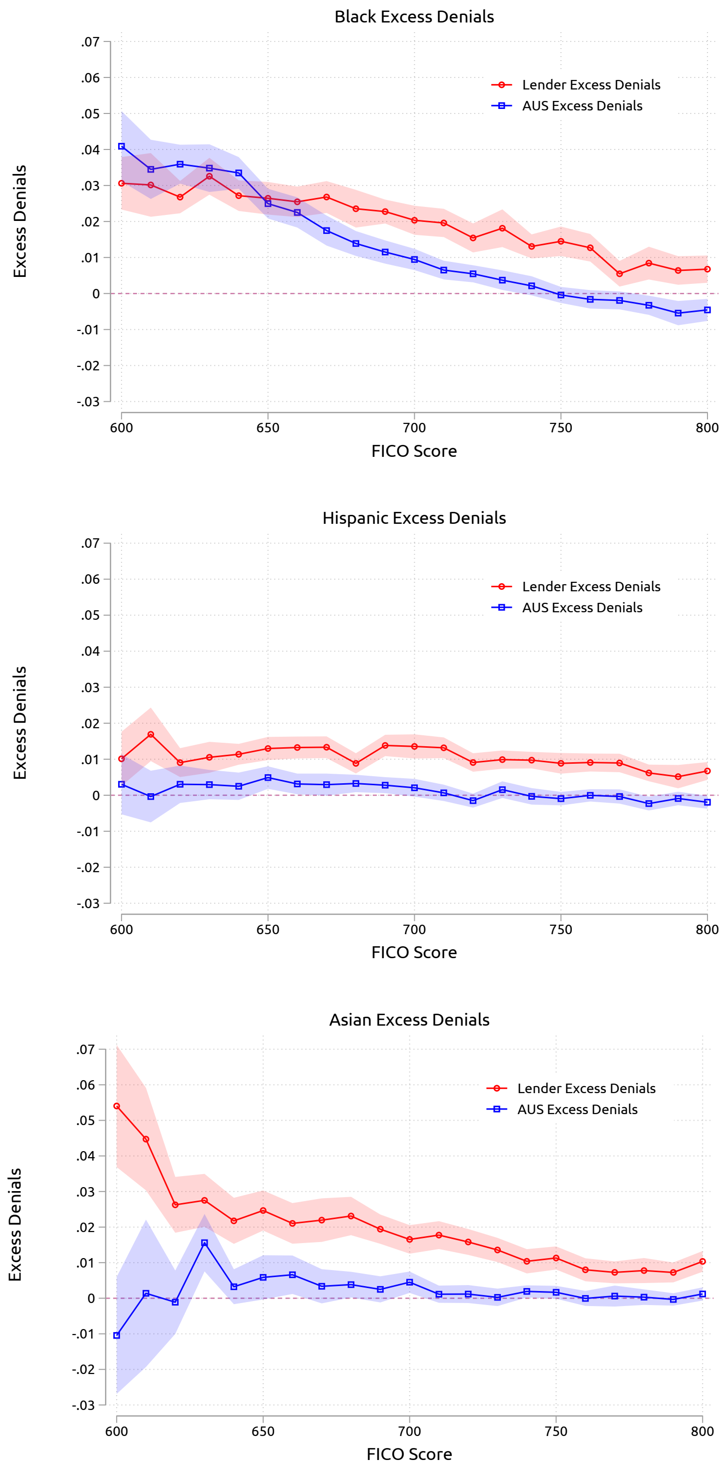


Figure 1: Lender and AUS minority exess denials relative to White by credit score

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

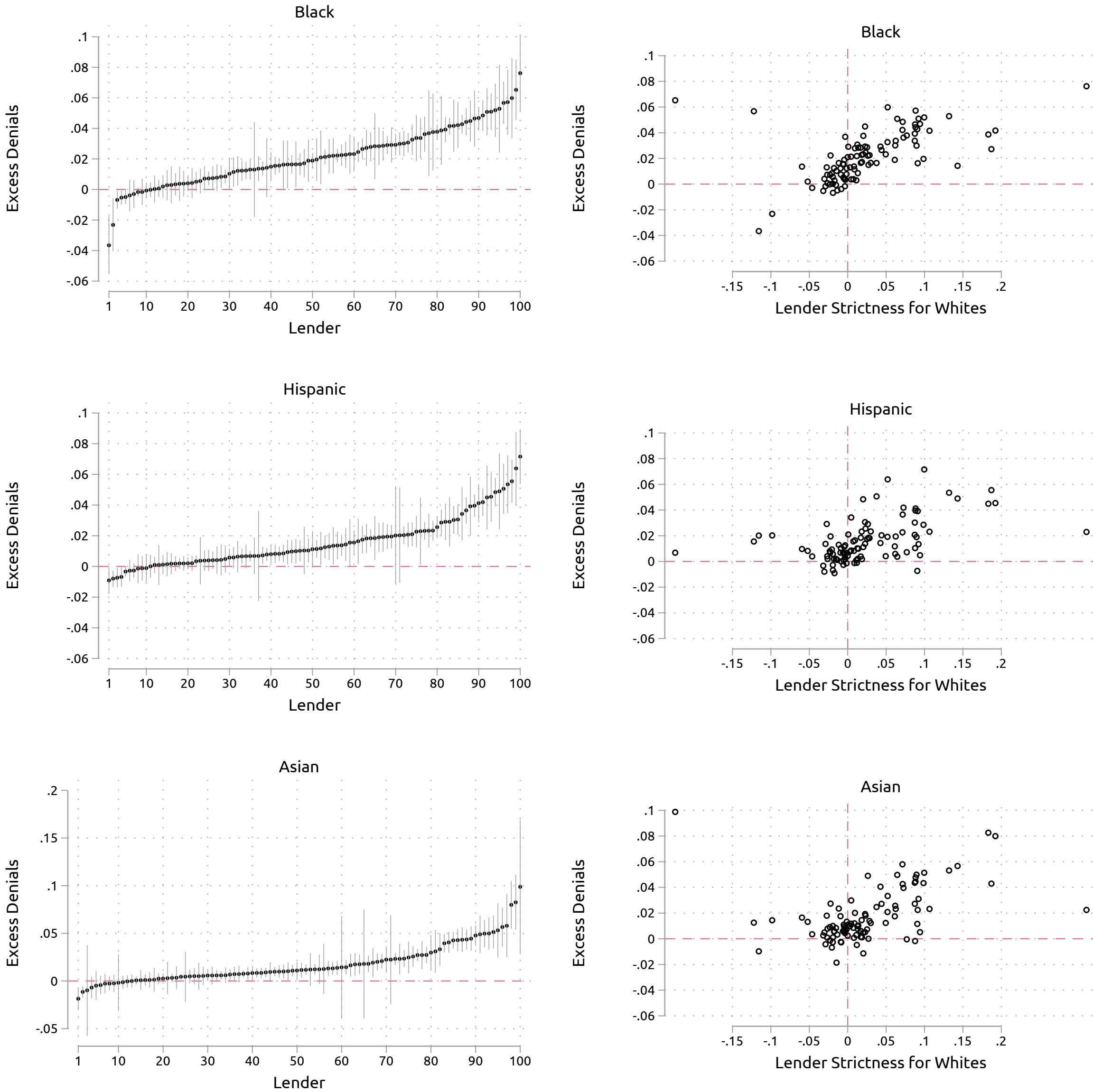
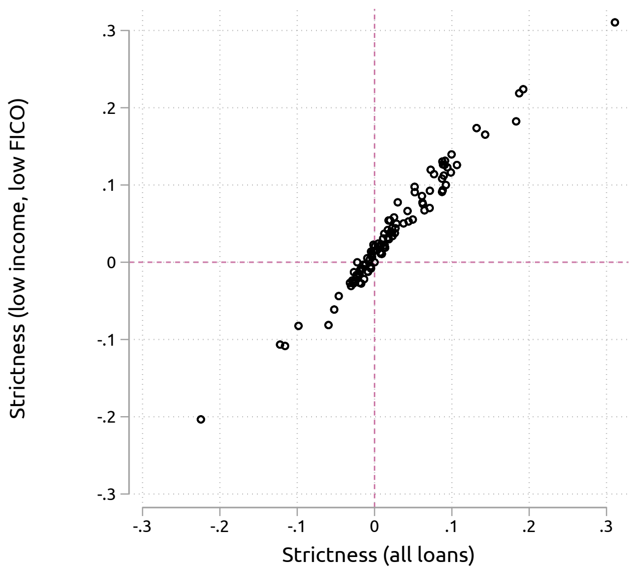


Figure 2: Excess denials for the top 100 lenders

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.

A. Lender strictness for all loans vs lower credit score and lower income



B. Lender strictness for White vs Black applicants

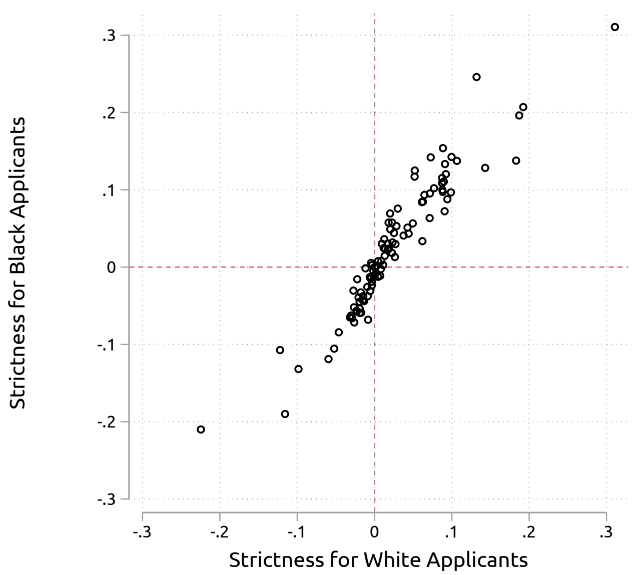
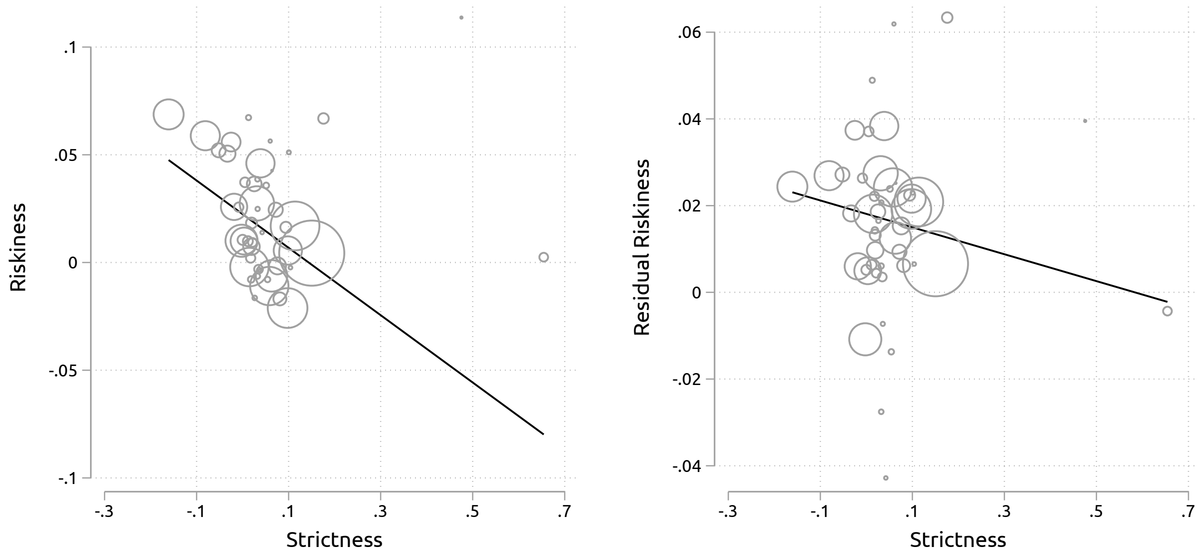


Figure 3: How does lender strictness vary across different sub-samples of the data?

Note: In panel A, lender strictness coefficients in the horizontal axis are lender fixed effects from a regression of lender denials on all controls including AUS outcomes for only White borrowers. The vertical axis shows similar coefficients in a subsample restricted to White applicants with both an income and credit score less than the 90th percentile of those fields among Black applicants who were denied credit by their lender. In panel B, lender strictness is definited similarly, but estimated in the subsample of Black and White applicants separately. Data source: HMDA.

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



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

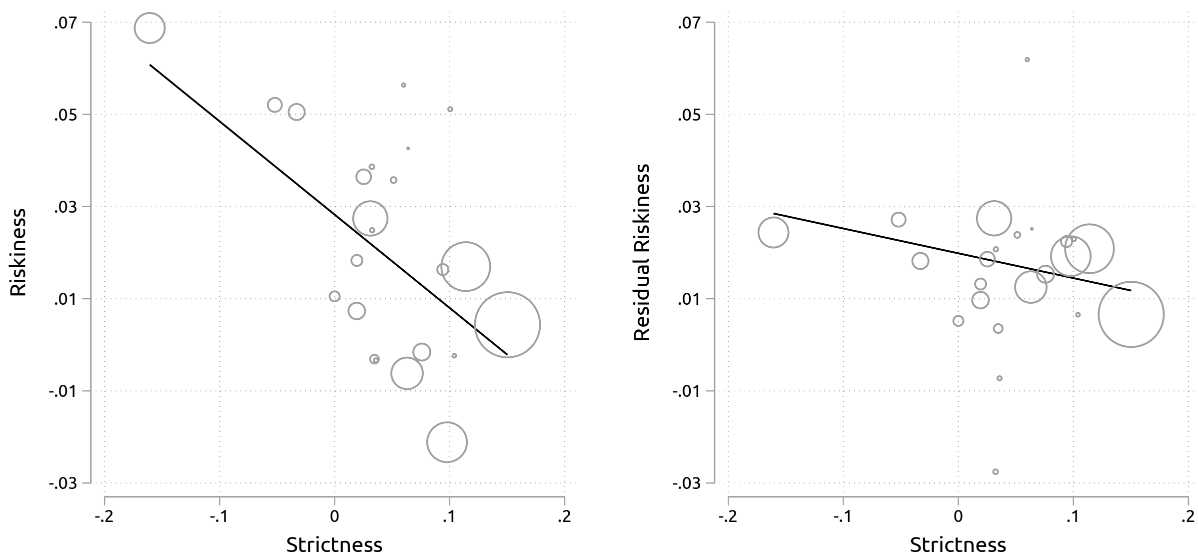


Figure 4: Lender strictness and loan performance

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.

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

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 credit score, LTV, DTI, income, etc., the estimated lender denial gaps are quite similar, suggesting a limited role for discriminatory selection into AUS evaluation. (As in the main text, we use the term "FICO" as a reference to credit scores in general, including VantageScore, FICO scores, and other credit scores.)

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 4, 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

[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.

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 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

[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.

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 identical credit score.

D Do Lenders Discourage Minorities from Applying for Mortgages? Evidence from the National Mortgage Database

As we discuss in Section 2 of the main text, before a formal application is submitted, lenders typically pre-screen prospective borrowers. If minorities are unduly discouraged from applying at the pre-screen stage, then our analysis of application rejections in the HMDA data may give us a distorted view of discrimination in the mortgage market. When lenders conduct a pre-screen, it often involves checking prospective borrowers' credit history, which generates a new credit **inquiry** in their credit record. In this section, we test for racial and ethnic differences in the probability of observing a mortgage application conditional on observing a mortgage inquiry, and racial and ethnic differences in the probability of observing an inquiry in the first place.

D.1 Mortgage Applications vs. Mortgage Inquiries

For this analysis, we draw on data from the National Mortgage Database (NMDB®). The NMDB, created by the Federal Housing Finance Agency and the Consumer Financial Protection Bureau, merges together individual credit history data with loan-level mortgage origination and servicing data for a nationally representative 5 percent sample of closed-end first-lien residential mortgages in

[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.

the United States. The NMDB is a unique source of information: We observe if and when existing mortgage borrowers *inquire* about getting a new mortgage — for example when seeking to refinance — as well as borrowers' race and ethnicity, in a large nationally representative sample (see Figure A.6 for an assessment of the representativeness of the NMDB).

In Figure A.4, we compare mortgage application counts in HMDA to inquiries from NMDB, by credit score and by race and ethnicity, in three different ways.[5] If minorities are more likely to be discouraged from applying, then we may observe fewer applications per inquiry for minorities relative to Whites. One challenge in comparing HMDA applications to NMDB inquiries is that HMDA includes (but does not identify) home purchase applications from prospective first-time home buyers, while the NMDB, by construction, only records inquiries from non-first-time buyers since this sample is composed of those who already have successfully obtained a mortgage. To minimize sample differences caused by this issue, we focus our analysis on the middle of 2020 when there was a boom in first-lien *refinance* applications relative to all other types of mortgage applications. In addition, since the NMDB provides an indicator of first-time home buyer status for the loan that caused the borrower to enter the panel, we can estimate the fraction of home purchase originations in the second half of 2020 going to first-time home buyers — by race, ethnicity and credit score — and use these estimates to deflate the number of home purchase applications in HMDA. In this way we can compare the number of mortgage inquiries by existing homeowners in NMDB to the estimated number of applications from existing homeowners (the deflated purchase applications plus all refinance applications) in HMDA, by race and credit score.

In the top panel of Figure A.4, we plot the count of 1-4 family owner-occupied first-lien home purchase applications (adjusted for first-time home buying as discussed above) and refinance applications during 2020q2 and 2020q3, from HMDA, as a fraction of the number of existing borrowers with at least one mortgage inquiry during 2020q2 and 2020q3, from NMDB. A low value of this ratio would indicate that many prospective borrowers drop out of the application process after the credit check by the lender but before submitting a full application. A value close to one would indicate little such discouragement is occurring (i.e., almost all households with an inquiry sub-

[5]In the HMDA data, credit score refers to the score used by lenders for underwriting the application. In the NMDB, we observe quarterly refreshed credit scores (VantageScore 3.0) of existing borrowers; we use scores as of 2020q1 to help ensure comparability with the scores recorded in HMDA at the time of application in 2020q2 or 2020q3.

mit an application). One observation from Figure A.4 is that low-score borrowers appear to be "discouraged" from applying relative to middle- and high-score borrowers, and this discouragement occurs to a similar extent across race and ethnic groups. A second interesting observation is that, in the bottom two score groups, Black borrowers appear somewhat *more* likely than White borrowers to apply conditional on inquiring, while having a similar likelihood within the top score group. In contrast, Hispanic borrowers appear less likely to apply than White borrowers across all three score groups, though this might partly reflect that Hispanic borrowers are relatively over-represented in the NMDB as suggested by Figure A.6.

The middle panel is similar to the top panel, but now the denominator captures the intensive margin of inquiries—the total number of inquiries—rather than just the extensive margin—the number of households with at least one inquiry. Because some borrowers have more than one inquiry (which may reflect shopping around), the ratios in the middle panel are lower than in the top panel. However, the cross-sectional patterns remain quite similar. Finally, in the bottom panel, we use a broader measure of mortgage applications from HMDA. Here we include junior lien applications and mortgage applications for second homes. To the extent these applications are reflected in the inquiry data and are skewed toward White homeowners, the patterns in the top two panels could be misleading. However, panel C is little changed relative to panel B beyond a general rise in the levels. Overall, while we acknowledge the limitations of this analysis — including that we only observe inquiries from existing borrowers — Figure A.4 fails to provide evidence of large scale disparate discouragement of minorities, particularly against Black borrowers.

D.2 Inquiry Rates, by Race and Ethnicity

The above analysis of the ratio of application to inquiry volumes may overlook racial disparities in lender discouragement if that discouragement mainly affects inquiry rates in the first place. Figure A.5 uses the NMDB data to compare inquiry rates by credit score and by race and ethnicity during 2020q2 and 2020q3 among borrowers with a mortgage outstanding as of 2020q1. This figure indicates that just over 20 percent of borrowers had at least one mortgage inquiry during the middle two quarters of 2020, with little difference across race and ethnicity.[6] Inquiry rates among low-score

[6] Research has found that minorities are less likely to refinance (e.g., Gerardi et al. (2023)), so it is somewhat surprising to see such little difference across groups in inquiry rates.

borrowers were lower, similarly so for White, Black and Hispanic borrowers.

E Do Lenders Understate Minority Borrowers' Income? Evidence from the National Survey of Mortgage Originations

In this section, we use data from the National Survey of Mortgage Originations (NSMO), which is a survey component of the NMDB program, to test whether lenders may be systematically understating the incomes of minority borrowers when underwriting loans. The NSMO is a quarterly mail survey, drawing its sample from newly originated mortgages since 2013 that are part of the NMDB. The goal of the NSMO is to gain unique information about borrowers' experiences in obtaining mortgages that cannot be ascertained from administrative data alone.[7] That said, a key feature of the NSMO data is that, as part of the NMDB program, each respondent is linked to detailed administrative mortgage and credit history data.

Table A.8 displays summary statistics from the NSMO, by race and ethnicity. The patterns across group of loan amount, credit score, and "lender-reported" income (i.e., the income used by the lender for underwriting) are qualitatively similar to those we observe in the HMDA data. Note that these variables are not self-reported, but rather utilize the linked adminstrative data as discussed above, so any differences from HMDA largely reflect sampling error and differences in time period coverage. However, the NSMO asks respondents to report their "total annual household income" in one of six categorical amounts.[8] Thus, we can compare self-reported income to lender-reported income as a test of whether lenders might be understating the income of minority applicants, with the caveat that we can only do this comparison for those who were approved and obtained a mortgage since rejected mortgage applicants are not in the NSMO sample.

Table A.8 shows that about 39 percent of White borrowers self-report an income category that *exceeds* the income category derived from lender-reported income. This fraction for White borrowers is higher than those for Black, Asian, and Hispanic borrowers — the opposite of what one might expect if lenders were more likely to understate minority incomes.

Figure A.7 plots this fraction for all borrowers at every $1,000 increment of lender-reported

[7] For more information, see https://www.fhfa.gov/nsmodata. We use a confidential version of the NSMO.

[8] The categories are (1) less than $35k; (2) $35k-$49k; (3) $50k-$74k; (4) $75k-$99k; (5) $100k-$174k; (6) $175k or more.

income. This figure shows that within each income category (delineated by the vertical dotted lines), this fraction rises between the lower and upper bounds of the category. In other words, near the lower bound, self-reported income is less likely to exceed the category lenders report, compared to near the upper bound, which is intuitive.

In Table A.9, we formally test whether the outcome shown in A.7 is shifted upward for minorities, which would be consistent with lenders being more likely to understate minority incomes on mortgage applications. In columns 2 and 4, we include a detailed set of controls (see table notes for a full list), including dummies for each $1,000 increment of lender-reported income. In all columns, we exclude loans with lender-reported income of $180k or higher since this is the top category and borrowers cannot self-report a higher income; columns 3 and 4 also exclude the second-highest income category, since this is such a wide category and borrowers toward the lower end of this category are highly unlikely to self-report higher income than what lenders report. Overall, we fail to find evidence that lenders are more likely to understate minority borrower incomes. For Black borrowers, in columns 2 and 4, there is no statistically significant difference in the probability that self-reported income exceeds lender-reported incomes relative to non-Hispanic White borrowers, while Asian and Hispanic borrowers are *less* likely than White borrowers to have self-reported income the exceeds lender-reported incomes.

F Racial Differences in Mortgage Borrowing Experiences: Evidence from the National Survey of Mortgage Originations

In Table A.10, we draw on the NSMO data, which asks respondents several questions about the mortgage process and their level of satisfaction along a variety of dimensions, to test for differences by race and ethnicity in mortgage borrowing experiences. In Table A.10, we run OLS regressions for six different outcome variables, flexibly controlling for a detailed set of loan and borrower characteristics, including loan type, loan purpose, loan amount, credit score, income, LTV, and several other variables (see table notes for more details). As noted earlier, a key feature of the NSMO data is that most of these controls come from administrative data and therefore measurement error is less of a concern than if they were self-reported. In columns 1 and 2, we focus on two yes/no questions that ask about delays that were experienced in the mortgage process. The first question asks: "In the process of getting this mortgage from your mortgage lender/broker, did you delay or postpone the closing date?" The results in column 1 show that Black borrowers and other minority borrowers (includes Native Americans and Pacific Islanders) were 4-5 percentage points more likely than non-Hispanic White borrowers to have postponed closing, which is a large difference relative to the mean of 16 percent.

The second question asks: "In the process of getting this mortgage from your mortgage lender/broker, did you redo/refile paperwork due to processing delays?" The results in column 2 again show sizable differences for minority borrowers relative to White borrowers. Black borrowers are nearly 10 percentage points more likely than similarly-situated non-Hispanic White borrowers to have experienced processing delays that led to having to redo paperwork. These processing delays could be one factor that led to closing delays for minority borrowers.

The NSMO also asks respondents about their level of satisfaction with (1) the application process, (2) the loan closing process, (3) the timeliness of disclosure documents, and (4) their mortgage lender/broker; respondents can answer "very," "somewhat," or "not at all."[9] Columns 3-6 indicate that Black borrowers were less likely to be "very satisfied" than non-Hispanic White borrowers on all four questions, as were Asian borrowers.

Overall, these data indicate that Black and other minority groups are more likely to experience delays in the mortgage process, and are less satisfied with the mortgage process and their lender. This analysis highlights other ways in which lenders may treat minorities differently that might not be reflected in standard assessments of differential treatment that focus on more easily observed outcomes such as mortgage denials and mortgage pricing (see next section).

G Do Minorities Pay More for Mortgages? Evidence from the 2018-2019 HMDA Data

In Table A.11, we use the 2018-2019 HMDA data to test for racial and ethnic differences in mortgage pricing. Following Bhutta and Hizmo (2020), we test for differences in the two dimensions of mortgage prices charged by lenders: the interest rate and upfront mortgage origination fees. Columns 1-3 of Table A.11 present estimates of racial differences in interest rates. In column 1, we

[9]The NSMO asks about borrowers' satisfaction along a few other dimensions, which we do not show here.

essentially present raw differences across groups in interest rates, controlling only for application date and county fixed effects. These results indicate that, before controlling for any risk factors, Black and Hispanic borrowers got interest rates that were on average 10 and 12 basis points higher, respectively, than non-Hispanic White borrowers, while Asian borrowers got interest rates that were 8 basis points lower than White borrowers. These unconditional differences are similar within lenders (column 2). In column 3, we flexibly add a battery of borrower and loan controls (similar to our main denial regressions). After including these detailed controls, we find a 1 basis point difference in interest rates between Black and White borrowers (statistically significant at the 10 percent level), and a 2 basis point difference between Hispanic and non-Hispanic White borrowers. On a typical $250,000 loan, these differences translate to approximately $1 and $2 increased monthly interest costs, respectively. Also, the coefficient on "other minority" borrowers (which includes Native Americans and Pacific Islanders) drops to zero in column 3. Overall, we find economically small differences on average by race and ethnicity in interest rates in the 2018-2019 HMDA data.

In columns 4-6, we present similar regression specifications with origination fees as a percent of the loan amount as the outcome variable. In the HMDA data, lenders report "total borrower-paid origination charges," which includes fees and any discount points paid by the borrower to the lender; lenders also report "lender credits," which are rebates that lenders pay to borrowers at closing.[10] We calculate origination fees as total origination charges minus lender credits.[11] After controlling for loan and borrower risk factors, we find that Black and Hispanic borrowers pay 0.02% and 0.04% more in fees, respectively, than White borrowers, or $50 and $100 given a $250,000 loan (the mean upfront net origination fees for all borrowers is 0.68% of the loan amount, or $1,700 given a $250,000 loan).

[10] For more information on these and other HMDA fields, see https://files.consumerfinance.gov/f/documents/cfpb_reportable-hmda-data_regulatory-and-reporting-overview-reference-chart_2023-02.pdf.

[11] The lender credit field in the HMDA captures "general lender credits;" but lenders may alternatively provide "specific credits" (credits that are used to pay for specific closing cost items), and these will not be captured in the HMDA data. The HMDA fields capturing upfront fees draws on certain elements on the closing cost disclosure form that is provided to borrowers at closing. General credits and specific credits are reported on different parts of the closing cost form, and HMDA only captures the former. For more information about reporting of specific credits on the closing cost form, see https://www.consumerfinance.gov/rules-policy/regulations/1026/interp-38/#38-e-3-Interp.

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 fixed-rate mortgages, 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) credit score 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-related 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**	0.084**	0.050**	0.032**	0.027**	0.023**	0.024**	0.020**
	(0.008)	(0.004)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Asian	0.012**	0.035**	0.033**	0.024**	0.018**	0.018**	0.015**	0.014**
	(0.004)	(0.003)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Hispanic	0.033**	0.025**	0.013**	0.008**	0.011**	0.010**	0.009**	0.009**
	(0.004)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Other	0.099**	0.070**	0.047**	0.033**	0.024**	0.023**	0.021**	0.019**
	(0.010)	(0.006)	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Joint Race	0.002	0.026**	0.016**	0.012**	0.006**	0.006**	0.005**	0.005**
	(0.004)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Missing Race	0.069**	0.048**	0.042**	0.034**	0.023**	0.021**	0.018**	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
Income Bins		Yes	Yes	Yes	Yes	Yes	Yes	Yes
FICO bins			Yes					
LTV bins								
DTI bins								
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	0.133	0.211	0.368	0.414	0.423	0.351	0.363
N. Obs.	9,718,280	9,515,418	9,515,417	9,489,809	9,487,558	9,402,889	8,737,868	8,653,841

Note - All the controls except for the race/ethnicity indicators are fully interacted with program by loan purpose indicators. Income bins are fixed effects that denote income deciles. 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**	0.023**	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**	0.025**	0.014**	0.012**	0.017**	0.012**
	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Hispanic	0.008**	0.009**	0.013**	0.011**	0.013**	0.010**	0.007**
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Other	0.011**	0.030**	0.032**	0.035**	0.017**	0.018**	0.013**
	(0.002)	(0.005)	(0.004)	(0.005)	(0.004)	(0.003)	(0.002)
Joint Race	0.003**	0.003	0.012**	0.005**	0.003**	0.007**	0.003**
	(0.001)	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)
Missing Race	0.012**	0.023**	0.025**	0.015**	0.017**	0.017**	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
Income Bins	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	0.310	0.465	0.421	0.395	0.414	0.399	0.384
N. Obs.	5,939,774	1,099,260	1,698,834	1,032,080	1,718,491	3,370,740	2,529,342
Average Lender Denial Rate	0.066	0.139	0.198	0.096	0.090	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**	0.024**	0.023**	0.034**
	(0.001)	(0.001)	(0.002)	(0.008)	(0.001)	(0.002)	(0.002)	(0.006)
Asian	0.011**	0.012**	0.010**	0.014**	0.020**	0.021**	0.019**	0.033**
	(0.001)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.002)	(0.007)
Hispanic	0.006**	0.008**	0.004**	0.005	0.013**	0.015**	0.012**	0.015**
	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.003)
Other	0.010**	0.012**	0.008**	0.024*	0.016**	0.014**	0.017**	0.041**
	(0.002)	(0.002)	(0.003)	(0.013)	(0.003)	(0.004)	(0.004)	(0.015)
Joint Race	0.003**	0.003**	0.002*	0.009**	0.008**	0.008**	0.008**	0.003
	(0.001)	(0.001)	(0.001)	(0.004)	(0.002)	(0.002)	(0.002)	(0.006)
Missing Race	0.009**	0.008**	0.011**	0.016**	0.021**	0.020**	0.022**	0.024**
	(0.001)	(0.001)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
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
Income Bins	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	0.297	0.310	0.289	0.264	0.331	0.360	0.308	0.283
N. Obs.	3,685,482	2,195,052	1,446,370	208,873	1,497,301	718,426	747,270	88,478
Average Lender Denial Rate	0.051	0.049	0.053	0.090	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, DTI, and income 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**	0.027**	0.030**
	(0.004)	(0.001)	(0.011)	(0.002)
Asian	0.008**	0.013**	0.117**	0.028**
	(0.003)	(0.001)	(0.028)	(0.003)
Hispanic	0.016**	0.008**	0.051**	0.020**
	(0.003)	(0.001)	(0.015)	(0.002)
Other	0.043**	0.016**	0.014	0.030**
	(0.006)	(0.002)	(0.035)	(0.004)
Joint Race	-0.001	0.004**	-0.022*	0.009*
	(0.002)	(0.001)	(0.012)	(0.005)
Missing Race	0.051**	0.015**	0.124**	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
Income Bins		Yes		Yes
FICO-LTV-DTI grid		Yes		Yes
Lender FE		Yes		Yes
R-Squared	0.006	0.239	0.009	0.566
N. Obs.	8372025	8172794	572131	450603
Average Lender Denial Rate	0.069	0.067	0.604	0.592

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, DTI, and income bins. The standard errors are clustered at the lender and county levels. Significance: * p<0.1, ** p<0.05. Data source: HMDA.

Table A.7: Regressions with a fully interacted grid using the AUS sample

	Lender Denial	AUS Denial
	(1)	(2)
Black	0.020**	0.014**
	(0.001)	(0.001)
Asian	0.012**	0.003**
	(0.001)	(0.001)
Hispanic	0.011**	0.001*
	(0.001)	(0.001)
Other	0.017**	0.009**
	(0.002)	(0.001)
Joint Race	0.002**	-0.001
	(0.000)	(0.001)
Missing Race	0.015**	0.005**
	(0.002)	(0.001)
Fully interacted grid	Yes	Yes
Lender-Program-Purpose FE	Yes	Yes
R-Squared	0.375	0.369
N. Obs.	8,401,117	8,538,153

Note - The fully interacted grid is a set of fixed effects created by interacting Credit Score Bins x DTI bins x LTV bins x Income bins x Loan Progam x Loan Purpose x Co-applicant f.e. x AUS Denial f.e.. 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. Income is binned into deciles. The standard errors are clustered at the lender and county levels. Significance: * $p<0.1$, ** $p<0.05$. Data source: HMDA.

Table A.8: NSMO summary statistics

	White	Black	Asian	Hispanic
Loan amount ($)	212,435.3	207,045.2	316,458.3	225,217.5
Credit score	721.3	686.4	735.3	701.4
Lender-reported income ($)	83,888.7	73,775.2	101,388.2	76,559.8
Self-reported income > Lender-reported income	0.39	0.38	0.33	0.35
N	27,240	2,145	1,637	2,610

Note: This table displays summary statistics of individual-level data from the National Survey of Mortgage Originations (NSMO), covering mortgages originated between 2013 and 2020. Race and ethnicity reflect the identity of the survey respondent; "White" refers to non-Hipanic White borrowers. Credit score refers to the VantageScore 3.0 measured near the time of mortgage origination. Lender-reported income refers to the income used by the lender in underwriting the loan; the next row shows the fraction of respondents NSMO whose self-reported income category (respondents could choose one of six categories) is higher than the income category derived from lender-reported income.

Source: Authors' calculations based on the National Survey of Mortgage Originations (NSMO).

Table A.9: Are lenders more likely to understate the income of minority borrowers?

	Sample: Income < \$180,000		Sample: Income < \$100k	
	(1)	(2)	(3)	(4)
Black	-0.065***	-0.021	-0.069***	-0.024
	(0.014)	(0.013)	(0.017)	(0.016)
Asian	-0.049***	-0.086***	-0.068***	-0.112***
	(0.017)	(0.016)	(0.025)	(0.024)
Hispanic	-0.066***	-0.067***	-0.069***	-0.072***
	(0.013)	(0.012)	(0.016)	(0.015)
Other minority	-0.067***	-0.026	-0.053*	-0.008
	(0.022)	(0.021)	(0.027)	(0.026)
Joint	0.060***	0.013	0.079***	0.012
	(0.014)	(0.014)	(0.019)	(0.018)
Controls		Yes		Yes
N	32,699	32,699	22,408	22,408
R-square	0.201	0.276	0.175	0.263

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table displays regression results using individual-level data from the National Survey of Mortgage Originations (NSMO). The outcome variable is an indicator variable for whether an NSMO respondent self-reports an income category that is higher than the income category derived from lender-reported income (i.e., the income lenders use to underwrite the loan). The excluded race category is non-Hispanic White. Race and ethnicity reflect the identity of the survey respondent, except when there is one White borrower and one non-White borrower on the loan, in which we case we categorize these loans as "Joint." In columns 2 and 4, controls include: loan purpose, loan type, county fixed effects, survey wave fixed effects, and dummies for self-employed, presence of a co-applicant or spouse/partner, whether respondents' reported income was higher or lower than a "normal" year, 11 LTV categories, 6 credit score categories, and 8 loan amount categories. We also include dummies for every \$1,000 bucket of lender-reported income starting at \$10,000. Sample for columns 1 and 2 includes all NSMO respondents with lender-reported income between \$10,000 and \$179,000; sample for columns 3 and 4 includes all NSMO respondents with lender-reported income between \$10,000 and \$99,000. All columns exclude construction loans, and loans from builders, and limit sample to fixed-rate mortgages for principal residences. Regressions are weighted using analytical weights provided in the NSMO; heteroskedasticity robust standard errors shown.
Source: Authors' calculations based on the National Survey of Mortgage Originations (NSMO).

Table A.10: Racial and ethnic differences in borrowing experiences

	Were there delays in...		Were you very satisfied with...			
	(1) Closing date	(2) Processing	(3) Application process	(4) Closing process	(5) Timely disclosures	(6) Your lender
Black	0.045***	0.096***	-0.048***	-0.054***	-0.061***	-0.071***
	(0.012)	(0.014)	(0.014)	(0.014)	(0.014)	(0.013)
Asian	0.013	0.013	-0.069***	-0.057***	-0.080***	-0.113***
	(0.012)	(0.013)	(0.016)	(0.016)	(0.016)	(0.015)
Hispanic	0.013	0.028**	-0.009	-0.002	-0.019	-0.018
	(0.010)	(0.011)	(0.012)	(0.012)	(0.013)	(0.011)
Other	0.049**	0.047**	-0.028	-0.025	-0.037	-0.068***
	(0.021)	(0.022)	(0.023)	(0.023)	(0.023)	(0.022)
Joint	0.003	0.021*	-0.036***	-0.036***	-0.046***	-0.036***
	(0.010)	(0.012)	(0.014)	(0.014)	(0.014)	(0.012)
Outcome mean	0.16	0.21	0.69	0.70	0.69	0.78
R-square	0.107	0.132	0.095	0.092	0.093	0.087
N	35,162	35,162	35,162	35,162	35,162	35,162

Note: This table displays regression results using individual-level data from the National Survey of Mortgage Originations (NSMO), estimating differences in self-reported borrowing experiences by race and ethnicity for five different experience outcomes. The excluded race category is non-Hispanic White. Race and ethnicity reflect the identity of the survey respondent, except when there is one White borrower and one non-White borrower on the loan, in which we case we categorize these loans as "Joint." All regressions include controls for: loan purpose, loan type, property type, county fixed effects, survey wave fixed effects, log income, and dummies for self-employed, presence of a co-applicant or spouse/partner, loan term, 11 LTV categories, 6 credit score categories, and 8 loan amount categories. Sample is limited to fixed-rate mortgages for principal residences with a term of 10, 15, 20, or 30 years. We exclude construction loans, and loans from builders. Regressions are weighted using analytical weights provided in the NSMO; heteroskedasticity robust standard errors shown.
Source: Authors' calculations based on the National Survey of Mortgage Originations (NSMO).

Table A.11: Do minorities pay more for mortgages?

	Interest Rate (%)			Origination Fees (%)		
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.10***	0.08***	0.01*	0.11***	0.05***	0.02***
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Asian	-0.08***	-0.05***	-0.04***	-0.08***	-0.04**	-0.01
	(0.01)	(0.01)	(0.00)	(0.02)	(0.02)	(0.01)
Hispanic	0.12***	0.07***	0.02***	0.11***	0.07***	0.04***
	(0.01)	(0.01)	(0.00)	(0.02)	(0.01)	(0.01)
Other minority	0.04***	0.04***	-0.00	0.12***	0.08***	0.04***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
Joint	-0.01	-0.00	-0.01***	0.14	0.03	0.02
	(0.01)	(0.01)	(0.00)	(0.11)	(0.03)	(0.02)
Missing Race	-0.02***	-0.00	-0.01***	-0.05***	-0.04***	0.01***
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)
Outcome mean	3.6	3.6	3.6	.68	.68	.68
App Date FE	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Lender FE		Yes	Yes		Yes	Yes
Loan Amount Bins			Yes			Yes
Co-Applicant			Yes			Yes
Income Bins			Yes			Yes
FICO-LTV-DTI grid			Yes			Yes
N	805,977	805,784	784,818	805,977	805,784	784,818
R-square	0.49	0.56	0.74	0.05	0.23	0.43

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table displays regression results using loan-level HMDA data for originated mortgages only, estimating differences by race and ethnicity in interest rates and origination fees (net of reported lender credits) as a percent of the loan amount. The excluded race category is non-Hispanic White. Race and ethnicity generally reflect the identity of the applicant listed first, except when there is one White borrower and one non-White borrower, in which we case we categorize these loans as "Joint." In columns 3 and 6, all controls except application date and county are interacted with loan purpose and loan type fixed effects. Sample of loans includes first-lien 30-year fixed-rate home purchase and refinance mortgages (including cash-out refinancings) for site-built single-unit properties. Standard errors are clustered by lender.

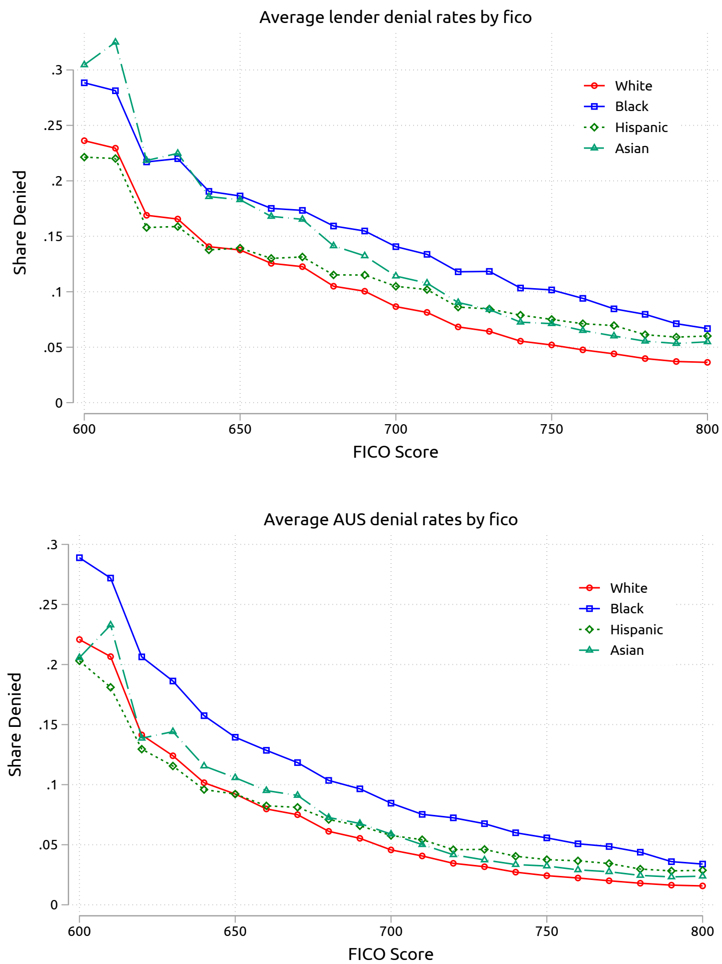


Figure A.1: Lender and AUS average denials by credit score

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

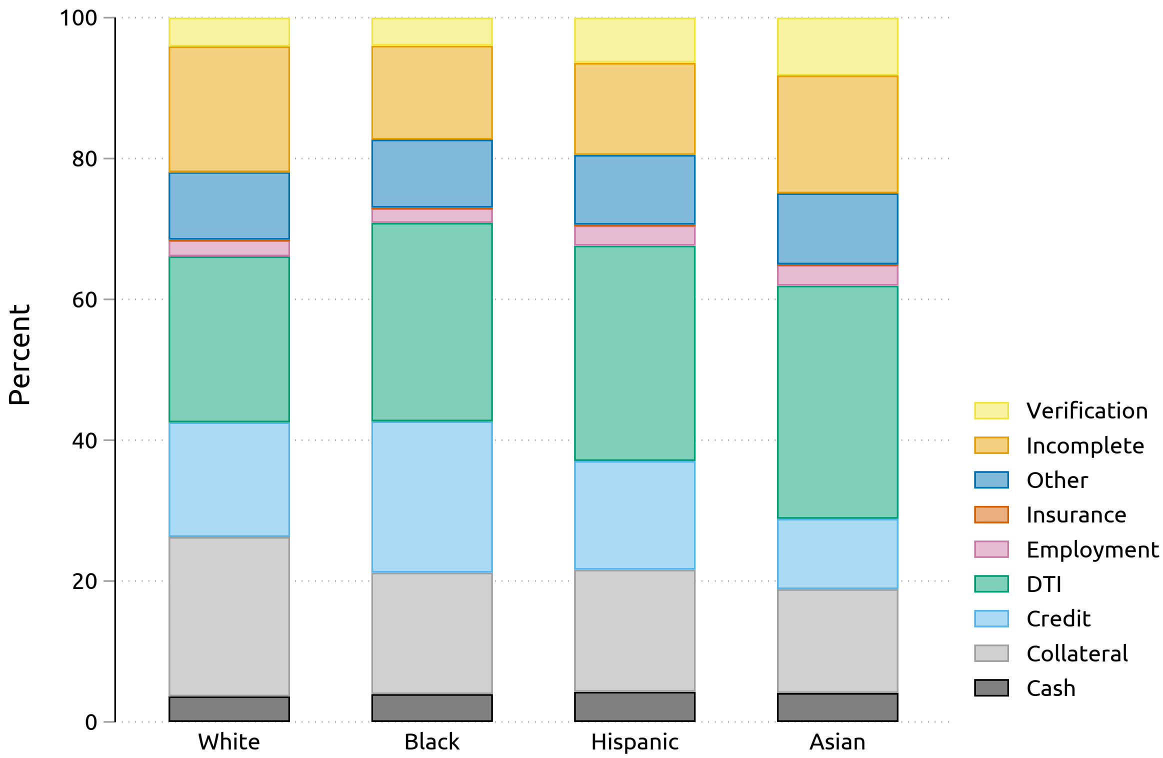


Figure A.2: Raw shares of denial reasons provided by lenders

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

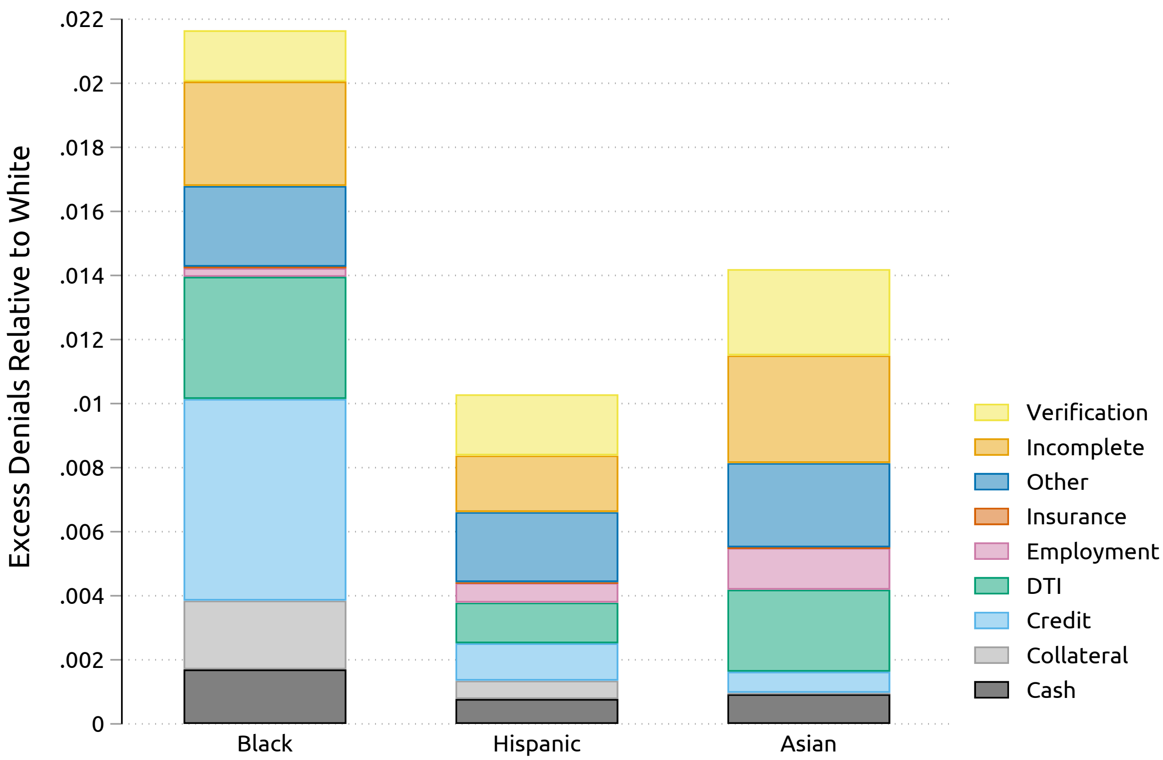
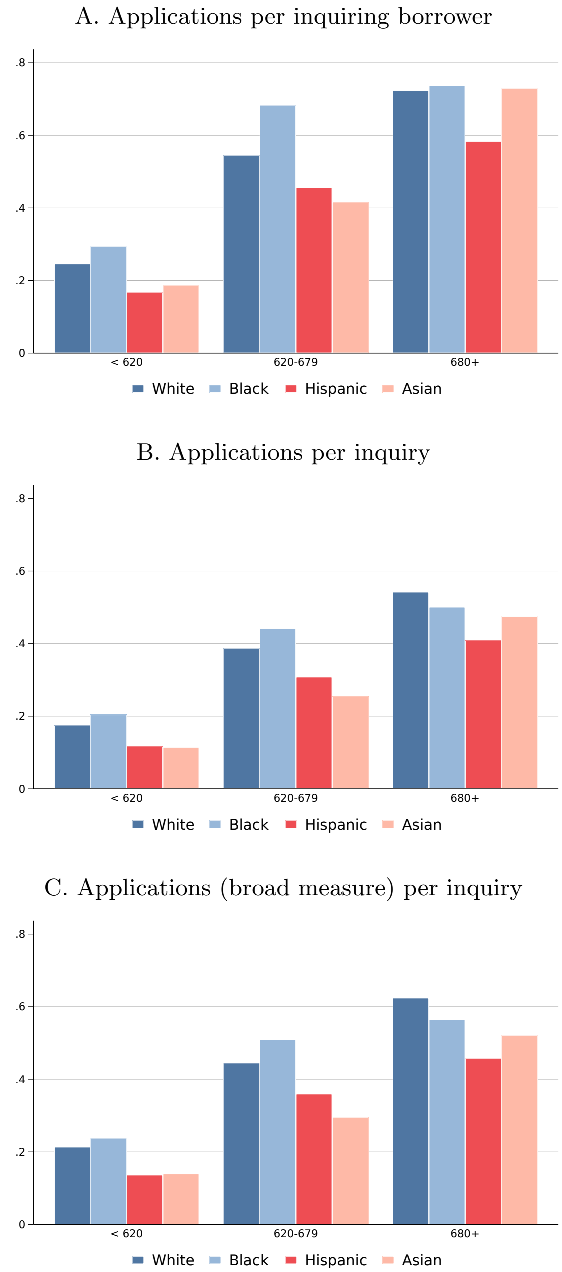


Figure A.3: Excess denial gaps broken down by lender provided denial reason

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.

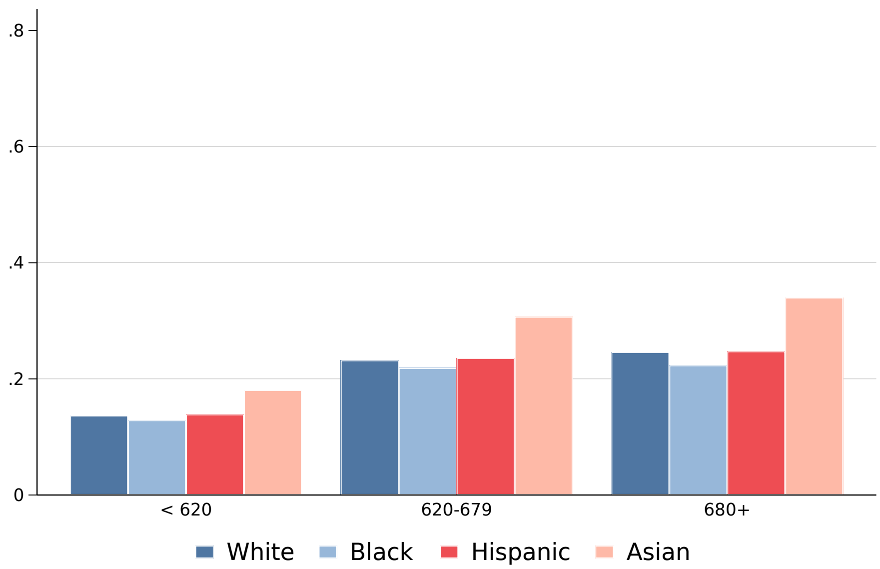
Figure A.4: Ratios of applications to inquiries



Note: Figures plot the number of mortgage applications recorded in HMDA in 2020q2 and 2020q3 as a fraction of the number of mortgage inquiries observed in the NMDB in 2020q2 and 2020q3, by credit score and by race and ethnicity. See text for more details.

Sources: Authors' calculations based on the National Mortgage Database (NMDB) and Home Mortgage Disclosure Act (HMDA).

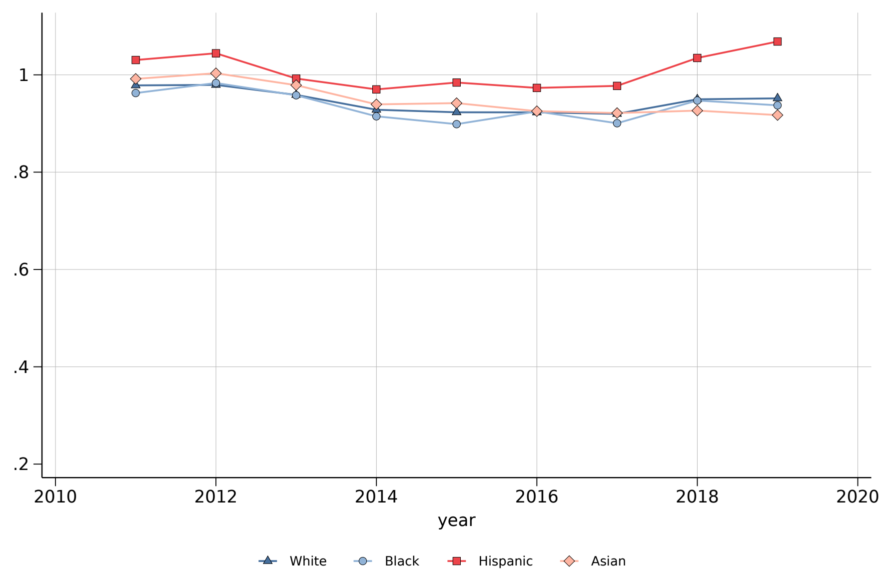
Figure A.5: Inquiry rates



Note: This figure shows the likelihood of having at least one mortgage inquiry during 2020q2 or 2020q3 among borrowers with a mortgage that was still open as of March 31, 2020.

Source: Authors' calculations based on the National Mortgage Database (NMDB).

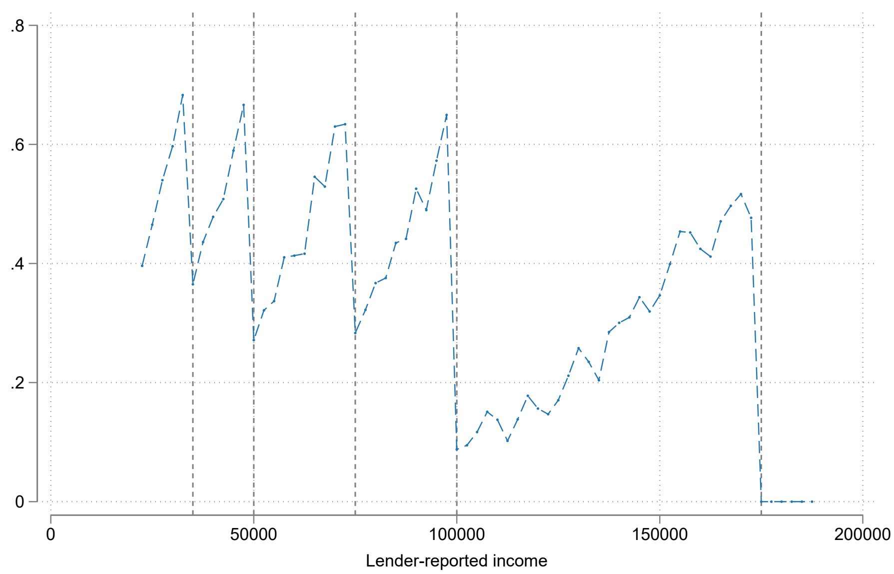
Figure A.6: Ratio of NMDB originations to HMDA originations



Note: This figure assesses the representativeness of the NMDB data by plotting the ratio of the number of first-lien home purchase and refinance mortgage originations in the NMDB data to the number of such originations in the HMDA data, by year and by race and ethnicity. In most years, NMDB origination counts are between 90 percent and 100 percent of HMDA origination counts, regardless of race or ethnicity.

Sources: Authors' calculations based on the National Mortgage Database (NMDB) and Home Mortgage Disclosure Act (HMDA).

Figure A.7: Fraction with self-reported income higher than lender-reported income



Note: This figure shows the fraction of respondents in the National Survey of Mortgage Originations (NSMO) whose self-reported income category is higher than the income category derived from lender-reported income (i.e., the income lenders use to underwrite the loan); the dashed vertical lines define the income categories.

Source: Authors' calculations based on the National Survey of Mortgage Originations (NSMO).