Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

Do Minorities Pay More for Mortgages?

Neil Bhutta and Aurel Hizmo

2020-007

Please cite this paper as:
Bhutta, Neil, and Aurel Hizmo (2020). "Do Minorities Pay More for Mortgages?," Finance and Economics Discussion Series 2020-007. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2020.007.

Do Minorities Pay More for Mortgages?

Neil Bhutta and Aurel Hizmo[1]
October, 2019

(Click here for most recent version)

Abstract: We test for racial discrimination in the prices charged by mortgage lenders. We construct a unique dataset where we observe all three dimensions of a mortgage's price: the interest rate, discount points, and fees. While we find statistically significant gaps by race and ethnicity in interest rates, these gaps are offset by differences in discount points. We trace out point-rate schedules and show that minorities and whites face identical schedules, but sort to different locations on the schedule. Such sorting may reflect systematic differences in liquidity or preferences. Finally, we find no differences in total fees by race or ethnicity.

JEL Codes: G21, G28, R51
Keywords: Discrimination, Fair Lending, Mortgage, Points, Interest Rate, FHA, Consumer Protection, High-Cost Mortgage

---

[1] Both authors are employed at the Federal Reserve Board. Spencer Perry provided excellent research assistance. Thanks to Dan Ringo for many helpful conversations; and to Elliot Anenberg, Robert Avery, Jaclene Begley, Haj Hadeishi, and Tom Mayock for their comments; to Karen Pence and Kathy Gibbons for assistance in obtaining data; and to Brian Synowiec and Andrew Holliday at HUD for compiling and providing the FHA data and for technical assistance. The views and conclusions expressed in this paper do not necessarily reflect those of Federal Reserve.

I. Introduction

In the wake of the housing crisis, the U.S. Department of Justice charged several of the largest mortgage lenders with fair lending violations for overcharging black and Hispanic borrowers during the housing boom, leading to unprecedented settlements totaling well over $500 million.[2] Although regulation of the mortgage market has ramped up dramatically since the crisis, discrimination against minority borrowers continues to be a major concern.[3] Most recently, in a letter to financial regulators, Senators Elizabeth Warren and Doug Jones raise concerns that lenders continue to discriminate, citing new research by Bartlett, Morse, Stanton and Wallace (2019) that finds interest rate gaps between minority and white borrowers.[4]

While interest rate gaps are often considered to be evidence of discrimination, in this paper we show that, because of the multi-dimensional nature of mortgage pricing, interest rate gaps alone do not provide compelling evidence. Borrowers face both upfront mortgage costs (i.e. "points" and fees) and future mortgage costs (i.e. interest payments), and can trade one for the other depending on their liquidity needs, their expected duration in the loan, or their time preferences. Indeed, using unique new data where we observe both price dimensions, we find that while black and Hispanic borrowers, on average, obtain higher interest rates, these higher rates are compensated by lower upfront costs.[5] Rather than reflecting discrimination, our results underscore that interest rate gaps can arise because minority borrowers may differentially demand lower-point/higher-rate loans.

To conduct our analysis we construct a unique loan-level dataset covering hundreds of lenders across the country, with rich details for each loan: demographics including the race, ethnicity, gender, and income of borrowers; all key underwriting

---

[2] In the three largest settlements, Bank of America (on behalf of Countrywide) settled for $335 million in 2011, Wells Fargo settled for $175 million in 2012, and JPMorgan Chase settled for $55 million in 2017.

[3] Note that both taste-based discrimination and statistical discrimination are illegal in the U.S. These forms of discrimination are often referred to generally as "disparate treatment." Another form of illegal discrimination is "disparate impact," which refers to practices that do not have a business necessity and have a disproportionate impact on protected groups.

[4] The letter can be found here: https://www.warren.senate.gov/download/2019610-letter-to-regulators-on-fintech-final

[5] Even using data on annual percentage rates (APR) – which is supposed to provide an "all-in" measure of price and is available in certain datasets – is likely to generate misleading conclusions about discrimination, as we will show.

variables; and, most notably, detailed price data on the interest rate, discount points, and fees. In contrast, the data used in previous research only contain interest rates, or are limited to pre-crisis loans in small samples or data from a single lender. We obtained administrative data from the Federal Housing Administration (FHA) covering the universe of FHA-insured home purchase loans originated in 2014 and 2015 and then merged these data with: (1) loan-level data collected under the Home Mortgage Disclosure Act (HMDA), and (2) loan-level data from Optimal Blue, a mortgage pricing platform used by many lenders.

The market for FHA-insured loans is an important one to study. In the years after the financial crisis, the FHA program has been a dominant source of credit for less wealthy, and moderate- to low- credit-score borrowers, and is heavily used by minorities. According to HMDA data, nearly 55 percent of black and Hispanic borrowers buying their home used FHA-insured loans in 2014 and 2015.[6] At the same time, over 70 percent of FHA loans are originated by smaller non-bank lenders – institutions that arguably face the least regulatory scrutiny.

Because FHA loans are fully-insured by the U.S. government and sold into Ginnie-Mae-guaranteed securities, mortgage originators are exposed to little or no credit risk and thus have limited incentives to price risk and statistically discriminate.[7] As such, unobserved credit risk variables, which is often a concern in this literature, should not pose a serious threat to identification and any within-lender differences in pricing that we might find are unlikely to reflect statistical discrimination.[8]

We begin our empirical analysis by examining minority-white differences in interest rates. Without any controls, black and Hispanic FHA borrowers on average pay interest rates that are about 8 and 10 basis points higher, respectively, than rates paid by non-Hispanic white FHA borrowers. After controlling for borrowers' FICO scores,

[6] This calculation excludes borrowers using Veterans Administration loans, which are also heavily used by minorities but require borrowers to be armed services veterans.

[7] To "statistically discriminate" in this context means to use race and ethnicity as a proxy for unobserved risk factors and to charge minorities more due to an expectation of elevated risk. See Bayer, Ferreira, and Ross (2017a) for evidence on unexplained gaps in default by race.

[8] Available rate sheets from several lenders confirm that there is only a modest amount of risk-based pricing, primarily for low FICO scores. If necessary, our matched dataset allows us to control for all key borrower risk characteristics such as credit score, loan to value (LTV) ratio, debt-to-income (DTI) ratio, liquid assets, and several others.

geographic and week fixed effects, and lender fixed effects, we continue to find statistically significant, albeit economically modest, differences in interest rates: black and Hispanic borrowers obtain rates that are 2-3 basis points more than non-Hispanic white borrowers. [9]

Next we turn to an analysis of discount points. Conditional on the same set of variables as above, we find that black and Hispanic borrowers pay *fewer* discount points, offsetting the difference in interest rates.[10] More importantly, we trace out the relationship between interest rates and discount points and find that black and Hispanic borrowers face a point-rate schedule that is virtually identical to that of white borrowers. In other words, lenders compensate minority and white borrowers equally when they select a higher interest rate, *across the entire range of interest rates*.

Our results suggest that the slightly higher average interest rates of black and Hispanic borrowers do not appear to reflect discrimination by lenders, but instead reflect some degree of sorting by race and ethnicity along the point-rate schedule – possibly due to differential preferences, or needs, to avoid upfront costs. At the other end of the spectrum, Asian borrowers tend to get the lowest interest rates and pay the most upfront points.

While all borrowers face the same point-rate price schedule, sorting to different locations along the schedule can result in lenders earning higher average profits on loans to black and Hispanic borrowers. This can happen if market imperfections enable lenders to capture some of the additional premia generated when higher-rate loans are sold in the secondary market, rather than fully passing through premia to borrowers.

Using additional data from Optimal Blue on secondary market pricing, we find that marginal premia are not entirely passed through to borrowers.[11] As a result, lenders earn slightly higher revenue – and higher profit, if origination costs are independent of point-rate choices – from black and Hispanic borrowers, since these borrowers are more likely to select into higher-rate/lower-point loans. We estimate that revenue generated

---

9 Note that throughout the paper we use the term "Hispanic" as shorthand for the group identified in the HMDA data as "Hispanic or Latino."

10 As we will see, FHA borrowers typically do not pay positive points, but rather "negative points", also referred to as lender credits. By paying a higher interest rate, borrowers can get cash credits at closing to help pay for upfront closing costs.

11 In other words, the tradeoff between points and interest rate that all borrowers face is not as favorable as it could be.

from black borrowers is about 0.08 points (i.e. 0.08 percent of the mortgage balance or \$130 on average) higher than white borrowers, and the revenue gap for Hispanic borrowers is about 0.06 points. Again, these revenue gaps arises not because the lenders in our data charge higher prices to minority borrowers, but because they are able to extract higher margins from borrowers – regardless of race and ethnicity – who want lower upfront costs.[12]

Finally, we assess differences in other types of upfront fees (underwriting fee, credit check fee, etc.). In the HMDA data, lenders must report the APR – a single measure of loan cost that is a function of the interest rate, points, fees, and insurance premiums –if the APR exceeds a threshold. We exploit a unique situation in 2014 where a change in FHA insurance premiums triggered a rise in APRs such that the APRs on a substantial fraction of FHA loans had to be reported. Given data on the APR, interest rate, and mortgage insurance premium for a loan, we can estimate total upfront fees for the loan. We find no evidence of systematic discrimination in fees by race or ethnicity.

The only paper we are aware of that, like ours, studies racial and ethnic discrimination in pricing in the post-crisis mortgage market is Bartlett et al. (2018). They focus on loans purchased and guaranteed by the government-sponsored enterprises, Fannie Mae and Freddie Mac. Like with the FHA market, lenders and investors in the GSE market have limited credit risk exposure due to the GSE guarantee. Using HMDA data matched to loan servicing data, they find interest rate gaps that are similar in magnitude to ours and conclude that lenders discriminate. But in contrast to our data, they do not directly observe discount points. Without data on discount points, it is not possible to reach firm conclusions about discrimination.

The lack of data on points and fees could be a confounding factor in other recent papers that document interest rate differentials in the pre-crisis era. Examples include Cheng, Lin, and Liu (2015) and Ghent, Hernandez-Murillo, and Owyang (2014). Haughwout, Mayer, and Tracy (2009) also do not observe discount points or fees,

[12] One potential concern might be that lenders are differentially steering minority borrowers into higher-rate/lower-point loans. While we cannot rule out this interpretation, both the small differences in interest rates across groups and our finding that Asian borrowers get the lowest interest rate/highest point loans make us lean toward the explanation that sorting arises due to different liquidity needs/preferences rather than lender steering.

although they do not find any interest rate differentials in their sample of subprime loans.

In contrast to these papers, Woodward (2008) collected complete pricing data for a sample of about 1,500 FHA loans originated in 2002. Her study is the only other paper we know of that that traces out the tradeoff between interest rates and discount points. We build on Woodward's analysis by tracing the point-rate schedule separately by race, and using a much larger sample of loans originated in the post-crisis era. Notably, Woodward (2008) finds almost no benefit to consumers, in terms of reduced upfront costs, associated with higher interest rates. This result differs considerably from the clear point-rate tradeoff that we see in our data. The more substantial point-rate tradeoff we observe may reflect new post-crisis regulations, particularly new rules that prevent mortgage brokers and loan officers from increasing their own compensation through selling high interest rate loans.

In follow-up work to Woodward (2008), Woodward and Hall (2012) find that mortgage brokers typically earned nearly 1 point (i.e. 1 percent of the loan amount) more from black and Hispanic borrowers in combined front-end and back-end fees.[13] Their analysis of brokers' revenue is similar to our revenue analysis described above, but we find far smaller differences by borrower race and ethnicity in revenue in our post-regulatory-reform data.[14]

One implication of our finding that borrowers sort into different point-rate locations is that that black and Hispanic borrowers will tend to have the highest APRs – even when treated fairly. Although APR aims to provide an "all-in" measure of loan cost, in practice the APR formula discounts upfront costs too aggressively. Researchers often use HMDA data to estimate gaps in the prevalence of "high-cost lending" (loans with APRs above the reporting threshold; see, e.g. Bhutta and Ringo 2014; Bayer, Ferreira, and Ross 2017b), but such estimates could be biased. We show that in our sample, this type of approach generates misleading results about the presence of discrimination.

---

[13] Two other studies using different pre-crisis datasets have also found that brokers earn higher fees from minority borrowers (Ambrose, Conklin, and Lopez 2018; Clarke and Rothenberg 2017).

[14] One difference between our analysis and that in Woodward and Hall (2012) is that we report within-lender estimates. In regressions where we exclude lender fixed effects, we find bigger revenue gaps, but still far smaller than what is found in Woodward and Hall (2012).

One caveat to our analysis is that discrimination can occur along several margins beyond pricing, which are beyond the scope of this paper. One important margin is the loan application accept/reject decision, which has been difficult to study due to a lack of data on rejected applicants.[15] Munnell et al. (1996) collected the necessary data from several lenders in Boston and found evidence of discrimination in the application decisions. However, given the movement towards automated underwriting decisions since the early 1990's, this is an area that is ripe for new research.

Another margin where racial bias could affect outcomes for prospective minority borrowers is in service quality. For example, Hanson et al. (2016) provide experimental evidence that response rates by loan officers to online mortgage inquiries differ by the race of the inquirer (see also Ross et al. 2008).

The rest of the paper proceeds as follows. In the next section, we provide institutional details the FHA program and discuss pricing in this market. Then we describe our data, followed by a discussion of our empirical results. Finally, we conclude.

II. Background: FHA Loan Costs and Interest Rate Dispersion

Since the onset of the recent financial crisis, the Federal Housing Administration (FHA) has been the main source of credit for less wealthy, low- and moderate-credit-score borrowers. The FHA does not directly extend credit, but insures loans funded by private lenders provided that they meet the FHA's underwriting guidelines, and are within statutory loan size limits. These loans are thereafter securitized in Ginnie Mae guaranteed pools.[16]

FHA insurance provides investors with an explicit government guarantee against credit risk. In return for this guarantee, the FHA charges borrowers upfront and annual mortgage insurance premiums, which are used to fund the FHA program. Premiums vary only slightly by LTV, loan size, and loan term, and do not vary at all by credit score

[15] The HMDA data are the only public source of information on rejected applications, but they do not provide data on the most important underwriting variables such as credit score, so they cannot be used to reliably test for discrimination in accept/reject decisions.

[16] Ginnie Mae guarantees that investors will receive pass-through payments in a timely manner.

or other risk metrics like DTI. These insurance premiums are the only component of the cost for an FHA loan that the FHA determines.

In addition to insurance premiums, FHA borrowers face other costs common to most mortgages, which are market determined. First, borrowers face fees which have to be paid at origination such as underwriting fees, appraisal fees, title fees, etc. However, the main driver of the cost of a mortgage is the interest rate and the associated "discount points" paid. Lenders often present borrowers with a menu with different combinations of interest rates and "discount points" to choose from. Borrowers can pay discount points, each point equal to one percent of the mortgage balance, in exchange for a lower interest rate. Borrowers can also choose negative points, also known as "lender credits", in return for a higher mortgage rate. As we will show later, FHA borrowers often get lender credits, which they use to help reduce the upfront closing costs to get a mortgage.

Because interest rates, discount points, and fees are determined in the market, as opposed to being set by the FHA, the potential for price dispersion and discriminatory pricing exists in the market for FHA loans just as it does in the market of other types of loans. Our data, which we describe in more detail in the next section, shows a high degree of price dispersion for FHA loans in 2014 and 2015. As an illustration, we plot the dispersion in interest rates for 30-year fixed-rate FHA home purchase mortgages with the maximum LTV, and zero discount points. Figure 1 plots the residual from a regression of interest rates on FICO score, MSA-by-week fixed effects, and day-of-rate-lock fixed effects. We find a 45 basis point difference between the 10^{th} and 90^{th} percentile interest rate, and well over a 100 basis point difference between the 1^{st} and 99^{th} percentile.[17] This observed price dispersion is unlikely to reflect risk-based pricing since all of these mortgages are fully insured by the government, and insurance premiums are paid separately by the borrower. In this paper, we assess one potential factor that could be driving this dispersion: differential treatment of borrowers on the basis of their race and ethnicity.

[17] In related work with Andreas Fuster, we conduct a more in-depth analysis of price dispersion in mortgages and assess how borrowers' shopping behavior and mortgage knowledge affect interest rates (Bhutta, Fuster, and Hizmo 2019).

III. Data

We obtained administrative loan-level data from the FHA on all FHA-insured mortgages originated in 2014 and 2015. These data provide a rich set of borrower, loan, and property attributes, including the interest rate, term of the loan, property location (census tract), and all key underwriting variables such as credit score, LTV ratio, income, DTI, liquid assets, and several others. These data also provide important dates for the transaction, including the date of application for FHA insurance, and the closing date of the loan.

That said, three key pieces of information are missing from the FHA data: borrower race and ethnicity; lender identifiers; and upfront finance charges, i.e. fees and discount points. As discussed earlier, data on both interest rates and upfront charges are needed to accurately assess whether minorities pay more. In order to obtain these additional pieces of information, we merge in two additional loan-level datasets.

III.A. Merging in HMDA Data

First, we merge the FHA data with the Home Mortgage Disclosure Act (HMDA) data. The HMDA requires most banks and credit unions, and many nonbank lenders, to annually report data on every mortgage application they receive. These data are widely considered to be a comprehensive source of data on residential mortgage lending in the U.S. Indeed, in 2016 they covered about 90 percent of all first-lien originations reported to national consumer credit bureaus (Bhutta, Laufer, and Ringo 2017).

Matching the FHA data to the HMDA data provides us with borrowers' race and ethnicity, and lender identifiers. It also provides us with the APR for a subset of loans. Given the APR, interest rate, and term of the loan, we can then estimate total upfront costs (i.e. fees plus discount points), but only for those loans where APR is reported.

Under HMDA rules, lenders must report the APR as a spread over the "average prime offer rate" (APOR), but only when the spread exceeds 150 basis points (1.5 percentage points).[18] Loans with reportable spreads are commonly referred to as

[18] APOR is an estimate by the Federal Financial Institutions Examination Council of the prime APR being offered by lenders, based on the offered interest rate and points reported in Freddie Mac's Primary Mortgage Market Survey.

"higher-priced loans." As discussed earlier, the APR on a mortgage aims to measure the total cost of the loan, and thus incorporates discount points and fees as well as mortgage insurance premiums. In mid-2013, the FHA raised its annual insurance premium to 135 basis points for the entire term of the loan.[19] In addition, the FHA charged a one-time upfront premium of 175 basis points (which raises the APR by about 15 basis points).[20] Combined, these two premiums added about 150 basis points to the APR of FHA-insured loans, helping to push many FHA loans originated in 2014 over the APR reporting threshold. Indeed, lenders reported the APR spread on about 44 percent of FHA loans in 2014, compared to less than 5 percent in previous years (Bhutta, Ringo, and Popper 2015). Then in January 2015, FHA reduced the annual premium to 85 basis points, thus sharply reducing the number of loans with a reported APR spread.

We merge the FHA and HMDA datasets together by matching *exactly* on the following variables: year of origination (i.e. 2014), loan type (i.e. FHA-insured), loan purpose (i.e. home purchase), census tract of the property, loan amount, and borrower income. Of the nearly 1.4 million home purchase loans originated in 2014 and 2015 in the FHA dataset, over 85 percent matched uniquely to a HMDA-reported loan on these variables. Details on the matching procedure and match quality are provided in the Appendix.

We drop certain types of loans from our final FHA-HMDA dataset. Almost all FHA home purchase loans are 30-year, fixed-rate, first-lien mortgages for owner-occupied, single-family, site-built (i.e. not manufactured) properties. We exclude from the final dataset the roughly 5 percent of matched loans that do not fit these criteria. In addition, we drop the 9 percent of loans with LTVs under 90 percent. The annual FHA premium for such loans is not assessed for the entire loan term, and thus few under-90-LTV loans have a reportable APR. The final dataset consists of nearly 970,000 mortgages.

---

[19] The annual premium was 5 basis points lower for loans with an LTV at or below 95 percent, but most FHA-insured loans had LTVs over 95 percent. See Bhutta and Ringo (2016) for more on the changes over time to FHA premiums.

[20] The upfront premium is a one-time fee equal to 1.75 percent of the loan amount. For loans with interest rates ranging from 3.75 percent to 4.85 percent (95 percent of the loans in our dataset in 2014 fall within this range), this fee increases the APR by 0.145 to 0.154 percentage points or 14.5 to 15.4 basis points.

III.B. Merging in Optimal Blue Data

Second, we merge in data from Optimal Blue. Optimal Blue is a lending services company that provides mortgage lenders with a software platform that can be used during the interest rate lock process. Optimal Blue retains the data entered by lenders, and these data can be purchased for research. These data include borrower FICO score, DTI and LTV ratios, FHA status, date of rate lock, loan amount, occupancy status and the ZIP code of the securing property. In addition, the Optimal Blue data include both the interest rate and the discount points paid or received for a given rate. Observing a precise measure of discount points is the key reason we merge in the Optimal Blue data.

Using the Optimal Blue data helps address two key concerns with our APR-derived measure of points and fees in the FHA-HMDA data. First, for most borrowers who get a relatively low interest rate, we generally do not observe an APR spread. Thus, in the FHA-HMDA data we do not observe the upfront costs paid by these borrowers, and cannot assess whether minority borrowers paid more points than white borrowers to obtain these low interest rates. With the Optimal Blue data, we observe discount points across the entire range of interest rates.

Second, while many FHA borrowers get lender credits (negative points), lenders generally do not include negative points in their APR calculation. Though the rules for calculating APR require that lenders include discount points *paid* by borrowers, the rule is not clear about whether to include points *received* by borrowers, and many lenders seem to not include lender credits in APR calculations.[21] Thus, if white borrowers are more likely than minority borrowers to receive credits at a given interest rate, relying only on the FHA-HMDA data may generate results that are biased against finding that minorities pay more.

We merge the Optimal Blue data to the FHA-HMDA data in two steps. First, we match exactly on county, loan amount, LTV, and interest rate.[22] Then we break ties by comparing the closing date from the FHA data to the lock date in the Optimal Blue data, and by comparing the FICO score recorded in the FHA data to that recorded in the

[21] See https://www.mtgprofessor.com/tutorial_on_annual_percentage_rate_(apr).htm

[22] We round LTV to the nearest 10th and round interest rate to the nearest 100th in each dataset prior to matching.

Optimal Blue data.[23] We estimate that Optimal Blue covers about one-quarter of the mortgage market, and we match about 158,000 loans to our FHA-HMDA analysis dataset, implying a match rate of about 65 percent (see the Appendix for more details). Although Optimal Blue does not cover a majority of the market, our matched FHA-HMDA-Optimal Blue dataset looks very similar to the broader market on a variety of observable dimensions, such as minority shares, average FICO score, and average income (see the Appendix for these results).

III.C. Summary Statistics

Table 1 provides a first glimpse at pricing and other characteristics across borrower race and ethnicity groups. We split borrowers into five mutually exclusive groups: non-Hispanic white, Hispanic white, black, Asian, and other.[24] The first row of Panel A displays average interest rates on FHA-insured loans in the larger FHA-HMDA data. We measure interest rates relative to the prime mortgage rate reported in Freddie Mac's weekly Primary Mortgage Market Survey (PMMS), using either the week of FHA insurance application, or the week of closing if the closing occurred more than 60 days after the FHA application date.[25] On average, interest rates for (non-Hispanic) white borrowers were 6 basis points over prime, while black and Hispanic borrowers paid about 14 and 17 basis points over prime, respectively.

The remainder of Panel A in Table 1 presents other characteristics of borrowers and their loans. Within the set of FHA-insured loans, risk characteristics do not vary dramatically by race. For example, the average FICO scores of white and Asian borrowers are only 10-15 points higher than of black and Hispanic borrowers. In contrast, studies of the broader population of borrowers -- including FHA and non-FHA loans – indicate that credit scores for white and Asian borrowers are significantly higher than for black and Hispanic borrowers (e.g. Bhutta and Ringo 2016).

---

[23] There can be differences in the timing of when these credit scores are obtained, leading to slightly different scores across the two datasets. To help break ties, we impose that the two FICO scores must be within 10 points of each other.

[24] To categorize borrowers by race and ethnicity, we follow the methodology in Bhutta, Laufer, and Ringo (2017), which takes into account the race and ethnicity both the borrower and co-borrower, if there is one.

[25] The FHA data includes exact day of closing and FHA insurance application. We drop a few loans where the FHA application date is reported to have occurred after closing. We also trim the data at the 1st and 99th percentiles of the interest rate spread.

Panel B shows interest rate spreads by group in the FHA-HMDA-Optimal Blue data. One advantage of the Optimal Blue data is that we observe the exact date when the rate was locked. So in Panel B, we measure the interest spread using the prime rate in the week the borrower locked in their interest rate. While the level of interest rate spreads are slightly higher than in the larger FHA-HMDA data, the differences across groups are similar: Hispanic and black borrowers pay about 10 basis points more than white borrowers and 15 bp more than Asian borrowers.[26]

IV. Estimating Racial and Ethnic Gaps in Mortgage Costs

IV.A. Interest Rate Gaps in the FHA-HMDA Data

Table 2 presents results of OLS regressions of the interest rate spread over prime on borrower race and ethnicity indicators, where the omitted reference category is non-Hispanic white. Column 1 presents the unconditional rate gap estimates, mirroring the gaps already observed in Table 1.

Column 2 adds in county and week fixed effects, which reduces the Hispanic coefficient noticeably. Next, the specification presented in column 3 adds in flexible borrower controls for FICO score, income, co-borrower status, and borrower and co-borrower gender. The rate gaps for black and Hispanic borrowers are cut in half after adding in these controls and the adjusted R-square increases. [27]

Examination of rate sheets for FHA loans suggests that lenders may adjust interest rates modestly based on FICO score and geography, consistent with the results thus far. Figure 2 shows the relationship between FICO and interest rate spreads separately for white, black and Hispanic borrowers, conditional on county fixed effects.[28] For all three groups, interest rates rise moderately as FICO scores fall below 680, and are basically flat from 680 to 800. This figure shows that the modest gap in interest rates between white and minority borrowers holds across the FICO distribution. Moreover, the

[26] Appendix Table 1 shows that the average interest rate (the level, as opposed to the spread, across all borrowers) in the FHA-HMDA-Optimal Blue data is about 4 basis points higher than in the FHA-HMDA data.
[27] Consistent with the notion that there is little, if any, risk-based pricing in the FHA market due to the FHA guarantee, rate sheets for FHA loans that we have looked at indicate modest price adjustments along just a few dimensions. Indeed, when we control for several other risk factors such as LTV, DTI, and months of reserves, our regression results are virtually unchanged both in terms of rate gap estimates and R-squared.
[28] We condition spreads on geographic location by subtracting out the average spread in the borrower's county.

minority-white rate gap widens only slightly at the 90th percentile. At any given FICO score, interest rate dispersion (i.e. the difference between the 90th and 10th percentile rate) is similarly large for all groups.

Finally, turning back to Table 2, the regression in column 4 includes lender fixed effects. The R-squared jumps, indicating considerable price differences across lenders, consistent with previous research (e.g. Bayer, Ferreira, and Ross 2017b; Bhutta, Popper, and Ringo 2014). In addition, the rate gaps for black and Hispanic borrowers narrow further to about 3 basis points for black borrowers and about 1.5 basis points for Hispanic borrowers (note that the widest gap is actually between black and Asian borrowers).

Do these statistically significant, albeit economically modest, interest rate gaps imply that lenders illegally discriminate against minority borrowers? Because originators take little, if any, credit risk and risk-based pricing is minimal for FHA loans, these gaps are unlikely to simply reflect a misspecification problem (e.g. lenders pricing a component of risk that we do not observe in the data). Nonetheless, additional analysis is necessary to reliably answer this question. In particular, we need to assess differences in points paid, which we turn to next.

IV.B. Discount Point Gaps in the FHA-HMDA-Optimal Blue Data

A key contribution of this paper is to incorporate data on upfront charges for mortgages to allow for a more complete assessment of minority-white pricing differences in the mortgage market. In this section, we use the merged FHA-HMDA-Optimal Blue dataset to test whether differences in discount points can account for minority-white differences in interest rates.

To begin, we repeat the same interest rate regressions presented in the previous section using the FHA-HMDA-Optimal Blue merged dataset. The results, shown in the first three columns of Table 3, are similar to those shown in Table 2 using the much larger FHA-HMDA dataset. In column 3, after including borrower controls and lender fixed effects, the estimated interest rate gaps for black and Hispanic borrowers relative to non-Hispanic white borrowers mirror what we found previously in the more comprehensive FHA-HMDA dataset.

Next we test for gaps in the amount of points paid. As background, Figure 3 shows the distribution of discount points for all borrowers. The modal value is zero points, and most borrowers do not pay points but instead receive lender credits (negative points). This distribution makes sense for the FHA market, where borrowers tend to be cash constrained, and thus may need to minimize upfront costs.

Columns 4-6 of Table 3 show estimates from a series of regressions of discount points paid on borrower race and ethnicity indicators, and an expanding set of controls. Jumping to column 6, which includes our full set of controls, we find that that black and Hispanic borrowers pay slightly fewer points than white borrowers, and Asian borrower pay more points, consistent with a tradeoff between the interest rate and discount points. We estimate that, on average, black borrowers pay about 0.08 points less (or $160 on a $200,000 loan) than white borrowers, and Hispanic borrowers pay about 0.04 points less. Thus, minority borrowers are, at least to some degree, compensated for paying slightly higher interest rates.

Figure 4 displays in greater detail the tradeoff between discount points and interest rates, showing average discount points, by race and ethnicity, at each decile of the interest rate spread. In the top panel, we condition on borrower characteristics, and county and week of rate lock fixed effects, while in the bottom panel we also include lender fixed effects. The reference category is white borrowers in the lowest interest rate category, and we pin down the levels for these graphs using their average amount of points (-0.4).[29]

Remarkably, the rate-point schedules for black, Hispanic, and white borrowers are virtually identical, particularly in the bottom panel where we include lender fixed effects. This finding implies that each borrower group typically faces the same price schedule. But borrowers can choose different locations along the schedule. Indeed, the results in Tables 2 and 3 suggest that black and Hispanic borrowers tend to choose slightly higher interest rates and lower points compared to non-Hispanic white borrowers, and especially compared to Asian borrowers.

Differences across borrowers in their point-rate choice may be influenced by a combination of factors, such as their amount of available liquid assets, preferences for

[29] In the Appendix we provide full regression results and standard errors.

liquidity, and their expected time horizon in the mortgage. Importantly, Figure 4 shows that even minority borrowers who get especially high interest rates of 40 to 60 basis points over prime are, on average, compensated with lender credits to the same degree as white borrowers who get those same interest rates.[30] Overall, we find little evidence of systematic discrimination in the market, with interest rate gaps offset by points gaps, and all racial/ethnic groups facing almost identical point-rate schedules on average.

Finally, we look at racial and ethnic pricing differences lender by lender, focusing on the largest 100 lenders in our data. The top panel in Figure 6 plots the black-white gaps in points, conditional on the interest rate and other controls, for each of the 100 top lenders.[31] The gap is expressed in dollars for a typical loan of $175,000, and with lenders sorted by the magnitude of their gap. The black-white gap is positive for half of the lenders and negative for the other half, with the gap for the median lender being very close to zero. Of the 100 lenders, the black-white gap is less than $250 (less than 0.15 points) for 85 of them, and in most cases not statistically different from zero. There are a few lenders with a more sizeable gap – the estimated gap is between $500 and $900 for 3 lenders – although there are also a few lenders with large negative gaps, which makes us cautious about drawing strong conclusions about the positive gaps. The bottom panel for Figure 6 plots the Hispanic-white gaps and shows similar patterns to the top panel.

IV.C. Does Lender Revenue Vary by Race and Ethnicity?

Here we conduct an analysis that is similar to that of Woodward and Hall (2012), examining lender *revenue* by race and ethnicity of the borrower. The results thus far indicate that black and Hispanic FHA borrowers face the same price schedule as white

[30] The tradeoff we find between points and rate differs considerably from the findings in Woodward (2008), which finds almost no tradeoff (i.e. no compensation to consumers from paying a higher interest rate) for FHA loans originated through mortgage brokers in the early 2000's.

[31] We regress discount points on race and ethnicity interacted with lender dummies, controlling for borrower and loan characteristics, as well as rate spread interacted with lender dummies to allow for a different point-rate schedule for each lender. To obtain a reasonable amount of precision for each lender-specific gap estimate, we focus on the largest 100 lenders in our data. The top 100 are selected from the set of lenders that have at least 10 loans to Hispanic borrowers and at least 10 loans to black borrowers. The 100 lenders in our sample account for just over 80 percent of loans in our matched FHA-HMDA-Optimal Blue dataset. The smallest lender in this top 100 group has 300 loans in the dataset.

borrowers. However, because minority and white borrowers tend to differ in their point-rate choice, the revenue and profits that lenders ultimately earn could differ by race and ethnicity.

To see why, consider an expression for revenue in equation (1). The revenue (as a percentage of the loan amount), R, that a lender earns from borrower i equals discount points paid, P, plus the market value of a loan with an interest rate spread, r :[32]

$$R_i = P_i + v(r_i) \qquad (1)$$

This market value, $v(r)$, reflects the premium that can be generated on the secondary market for a loan with a given interest rate spread. Naturally, this premium increases with the rate spread since higher spreads generate a greater stream of interest income for investors.[33]

When a borrower accepts a higher rate spread, the additional premia from the secondary market can be used to compensate the borrower by reducing the amount of points charged (or increasing the lender credits). However, if lenders do not fully pass through higher premia to borrowers, then revenue will be increasing in the rate spread. Moreover, since the cost to originate a loan should not depend on borrowers' selection of a particular point-rate combination, then profits could conceivably also be increasing in the rate spread in the absence of full pass through.

In our primary dataset we do not observe the premium at which each particular loan was sold in the secondary market. However, we are able to draw on another sample of mortgages in the Optimal Blue data where we observe the mortgage interest rate and the price at which the lender sells the mortgage to a secondary market investor.

The red line in Figure 5 shows the average "price" on the secondary market by mortgages' interest rate spread, where price on the y-axis is measured in dollars per hundred dollars of the loan amount.[34] For example, a secondary market price of 104

---

[32] Another component of revenue are various other fees that are charged. We consider these other fees in the next section.
[33] Equation (1) is analogous to equation (1) in Woodward and Hall (2012), but here we are thinking about the total revenue generated from the loan, as opposed to the revenue accruing solely to mortgage brokers. Part of the total revenue considered here would be distributed to a mortgage broker or loan officer.
[34] Price estimates for the red line are based on secondary market prices for 13,797 mortgages. These estimates are conditional FICO and fixed effects for county, week, and lender.

implies that the investor will pay $104,000 for a $100,000 loan. The extra $4,000, or 4 points, is the premium.

The black line in Figure 5 mimics Figure 4, only now with the y-axis shown in terms of price rather than points (price equals 100 minus the number of discount points paid). Together, the red and black lines tell us that, for example, loans with a rate spread of 30-40 basis points generate, on average, a premium of about 5.5 points, which is then split between borrowers and lenders: borrowers get about 2 points in lender credits, and lenders get the other 3.5 points.[35] In other words, the gap between the red and black lines represents lenders' average revenue as a function of the rate spread.

The key takeaway from Figure 5 is that the red and black lines are not quite parallel: lenders' revenue rises somewhat with the interest rate paid by the borrower. As such, because black and Hispanic borrowers sort into slightly higher rate loans compared to white and Asian borrowers, lenders earn slightly more revenue from black and Hispanic borrowers. Table 4 provides estimates of these revenue gaps. The outcome variable is our estimate of the revenue from each loan as given by equation (1), using the relationship shown by the red line in Figure 5 to estimate $v(r_i)$. In column 3, with the full set of controls, the coefficients indicate that revenue from black borrowers is about 0.075 points (i.e. 0.075 percent of the loan amount) higher than white borrowers. Given a typical loan size of $175,000, this translates into about $130.

As alluded to earlier, this revenue gap could reflect higher profits on loans to black and Hispanic borrowers if the cost to originate a loan (i.e. processing an application and doing the underwriting) does not vary with the rate spread. However, some of the lenders we study may act as loan servicers, and therefore could face some risk due to default and prepayment, and this risk likely increases with rate spreads. Overall, we view our revenue gap estimates as an upper bound on the profitability gaps between borrower groups.

Notably, the estimated revenue gaps in column (2), which exclude lender fixed effects, are somewhat bigger. For example, the coefficient for black borrowers is about 0.19 points, which is about $330 for a $175,000 loan. This finding suggests that black

[35] In addition to profits, the remaining 3.5 points would be used to compensate loan officers, and pay for underwriting costs and other origination costs.

borrowers – for some reason – tend to go to somewhat more expensive lenders than white borrowers, which is consistent with the conclusions in Bayer, Ferreira, and Ross (2017b). But within lender (column 3), this gap drops sharply by about 60 percent to \$130, as already noted.[36]

IV.E. Differences in Fees

One remaining area where differences in mortgage costs can arise is in upfront fees. Lenders often charge an assortment of fees that can add up to thousands of dollars. The list of fees is not standardized from one lender to the next, but often includes fees for things like underwriting, application, appraisal, credit check, processing, tax service, and more.[37]

Data on fees for mortgages is rarely available. Indeed, Woodward (2008) hand collected data from mortgage closing documents in order to obtain such data. In this paper, we take advantage of a unique opportunity where we observe the APR on nearly half of all FHA-insured home purchase loans originated in 2014. The APR for a mortgage is a measure of loan cost that incorporates the interest rate and most fees (as well as positive discount points, but recall that most FHA borrowers do not pay points).[38]

As discussed earlier in Section III.A., under HMDA rules, lenders must report the APR as a spread over the APOR (the average prime offer rate) when the spread exceeds 150 basis points. By 2014, the FHA had raised their insurance premiums to such a level that the premiums alone pushed APRs up by nearly 150 basis points. The first column of Table 5 shows that in 2014, nearly half of all FHA loans in our FHA-HMDA dataset had reportable APRs.

Conditional on observing the APR spread, we can estimate fees, *F*, as:

---

[36] Appendix figure A3 shows revenue gaps estimated separately for the top 100 lenders in our sample. The results are very similar to Figure 6: the revenue gaps are zero for the median lender and small or negative for the majority of lenders.

[37] Some of these fees may go to third parties such as an appraisal company. However, in many instances these third parties may be wholly or partially owned by the lender.

[38] As discussed in Section III.B., lenders are required to include positive discount points in their APR calculations, but not negative points. Since most FHA borrowers do not pay positive discount points, and lenders generally do not include negative points, we view the APRs for FHA mortgages as largely capturing the interest rate and fees.

$$F = (APR - APOR^{*}) - (i + MI^{A} + MI^{U} - APOR)$$

On the right-hand-side of this equation, $APR - APOR^{*}$ refers to the actual reported APR spread in the HMDA data, and the second expression in parentheses estimates what the APR spread would be if it excluded fees, as:

- the interest rate, i,
- plus the annual FHA insurance premium (135 basis points for most loans), MI^{A},
- plus the upfront FHA premium (15 basis points), MI^{U},
- minus APOR in the week of application for FHA insurance.[39]

This difference between the actual APR spread and the APR spread excluding fees yields an estimate of fees expressed in percentage points: the amount by which the APR increases due to fees.[40] Table 5 indicates that fees added 12 basis points to the APR, on average, across all borrowers. Using the formula to compute the APR for a 30-year fixed rate loan of $200,000 at an interest rate of 4 percent, this 12 basis point amount implies about $3,000 in fees.[41]

To estimate gaps by race and ethnicity in fees, we run regressions similar to the previous interest rate regressions, but now use our measure of fees, *F*, as the outcome. As noted just above, only about half of borrowers in our sample have a reported APR spread in HMDA and thus have an estimate for points and fees paid. However, as shown in Figure 7, for those borrowers with an interest rate of 20 basis points or higher over prime, we observe the APR spread almost universally. As such, here we test for gaps in fees within the set of borrowers who had an interest rate spread of at least 20 basis points. Since our earlier evidence indicated only small interest rate differences by

[39] $APOR^{*}$ is the prime offer rate during the week the rate was locked in by the borrower, which may not be the same as the week of FHA application or closing; we do not observe the week of rate lock in the HMDA or FHA data.

[40] One potential issue with this calculation is that, for any fixed fee (e.g. $2000), APR rises slightly more than 1 for 1 with the interest rate. Thus, our estimate of fees, *F*, will tend to be higher for higher interest rate loans. But for modest differences in interest rates this issue is negligible.

[41] To be sure, for many FHA borrowers these fees are offset to some degree by lender credits (negative discount points). As explained earlier, our understanding is that APRs generally do not include lender credits, so our estimate of *F* should, for the most part, just capture total fees before any credits were applied.

race and ethnicity, it mitigates our concern about differences by race and ethnicity in the extensive margin of being in the group of borrowers with a rate spread of at least 20 basis points.

Column 1 in Table 6 shows estimated gaps in fees paid for this subset of borrowers, controlling only for county and week fixed effects. The black-white gap is small, at just over 0.5 basis points, while the Hispanic-white gap of about 1.5 basis points implies a fee gap of about $350 for a $200,000 loan.[42] However, after adding in borrower controls in column 2, these gaps shrink and are no longer statistically significant, and the within-lender results in column 3 continue to be near zero and insignificant.

IV.F. Differences in the Probability of Getting a High-Cost Mortgage

Researchers have often used the HMDA data to document sizeable differences across groups in the probability of getting a "high-cost loan," i.e. loans with a reported APR spread (see, for example, Avery, Brevoort, and Canner 2006; Bocian, Ernst, and Li 2008; Been, Ellen, and Madar 2009; Bhutta and Ringo 2014; Bayer, Ferreira, and Ross 2017b). Keeping with this tradition, Table 7 shows regression results where the outcome variable is an indicator for whether the loan is high-cost.

Column 1 presents the raw gaps between groups in the fraction of high-cost FHA loans in 2014. Mirroring the statistics presented in Table 5, black and Hispanic borrowers exhibit large gaps of 9 and 15 percentage points, respectively, in the likelihood of having a high-cost loan. In column 2, even after controlling for borrower risk (income and credit score), and county, week, and lender fixed effects, black and Hispanic borrowers have higher probabilities of 4 and 3 percentage points, respectively. Particularly for black borrowers, much of the raw gap remains.

Given our previous findings of no differences when we looked at more granular measures of loan costs, it is clear that the gaps in probability of getting a high-cost loan that we see in column 2 of Table 6 provide misleading evidence on discrimination by lenders. A key complication with the type of regression in Table 6 stems from the way APR is calculated. Because the APR calculation assumes that a loan is held to maturity,

[42] At an interest rate in the neighborhood of 4 percent.

it tends to overweight the interest rate of the loan and underweight upfront charges, as most borrowers stay in the same loan for far less than 30 years. Thus, even when minority borrowers are treated fairly, their APRs will tend to be higher if there is a systematic preference among minorities toward higher-rate, low-upfront-cost loans – as we found earlier.[43]

This issue is exacerbated by lenders not netting out the lender credits associated with higher interest rates when calculating APRs. Moreover, small differences in APRs can generate large differences in the *fraction* of borrowers above the higher-priced threshold, if the distribution of APRs has a lot of mass near the APR reporting threshold.

To illustrate how these issues can generate biased results, we conduct a simple counterfactual. Earlier, we estimated that black borrowers receive 0.08 lender credits in exchange for a 3.6 basis points higher interest rate relative to white borrowers. If, instead, black borrowers got the same amount of lender credits as white borrowers and a 3.6 basis points lower interest rate, their APR would also drop by almost 3.6 basis points. Using the empirical distribution of APR spreads for black borrowers, we calculate that an APR drop of just 3.6 basis points would lower the share of borrowers with a higher-priced mortgage by about 4 percentage points, which is the same as the black-white difference estimated in column 2 of Table 6. Therefore, the small difference in contract selection can account for the entire estimated black-white difference in the probability of having a high APR spread loan.

To be sure, the gaps found in previous research focusing on the subprime boom of the mid-2000s may reflect more meaningful pricing gaps. Nonetheless, we illustrate here that it is difficult to get a precise reading on discriminatory pricing based solely on the HMDA APR spread data, even when researchers merge in data to control for carefully for credit risk.

---

[43] Systematic differences in preferences could reflect, for example, systematic differences across groups in available liquidity. There is modest evidence for this in Table 1, which shows that black and Hispanic borrowers are less likely than white and Asian borrowers to have at least three months of reserves.

V. Conclusion

In this paper we test for discrimination against minority borrowers in the prices charged by mortgage lenders. To do so, we construct a unique dataset of FHA-insured loans originated in 2014 and 2015, where we observe rich borrower and loan details, most notably information on all components of the cost of a mortgage: interest rates, discount points, and other fees. We find that while black and Hispanic borrowers pay slightly higher interest rates than non-Hispanic white borrowers, this rate gap is offset by a gap in discount points. Black and Hispanic borrower tend to choose slightly higher interest rates in return for lower upfront costs. In contrast, Asian borrowers tend select lower interest rates and pay more points relative to white borrowers. More generally, we find that minority and white borrowers all face the same point-rate schedule. Finally, we also find no difference by race and ethnicity in other fees charged (underwriting fee, credit check fee, appraisal fee, etc.).

A key message from our findings is that mortgage pricing is multi-dimensional, and it is important to take into account discount points and fees alongside interest rates when assessing differences in costs across consumers. Along the same lines, it is important to be aware of the limitations of using the mortgage APR spread variable recorded in the HMDA data. Finally, major regulatory changes following the financial crisis, as well as technological progress in the mortgage industry, underscore the importance of continued research on discrimination using more contemporary datasets.

REFERENCES

Ambrose, B. W., Conklin, J., & Lopez, L. (2018). Preferential Treatment in Financial Contracts: Does Broker Race Affect Mortgage Prices?.

Avery, R., Brevoort, K., & Canner, G. (2006). Higher-Priced Home Lending and the 2005 HMDA Data. Federal Reserve Bulletin A123, Vol. 92.

Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2018). Consumer lending discrimination in the FinTech era.

Bayer, P., Ferreira, F., & Ross, S. (2017a). The Vulnerability of Minority Homeowners in the Housing Boom and Bust. *American Economic Journal: Economic Policy,* 9(1), 344-45.

Bayer, P., Ferreira, F. and Ross, S. (2017b). What Drives Racial and Ethnic Differences in High-Cost Mortgages? The Role of High-Risk Lenders. *The Review of Financial Studies*, 31(1), pp.175-205.

Been, V., Ellen, I., and Madar, J. (2009). The High Cost of Segregation: Exploring Racial Disparities in High-Cost Lending. *Fordham Urban Law Journal*. 36(3).

Bhutta, N., Laufer, S., & Ringo, D. (2017). Residential Mortgage Lending in 2016: Evidence from the Home Mortgage Disclosure Act Data. Federal Reserve Bulletin, Vol. 103, No. 6.

Bhutta, N., Popper, J., & Ringo, D. (2015). The 2014 Home Mortgage Disclosure Act Data. Federal Reserve Bulletin, 101(4).

Bhutta, N., Fuster, A., & Hizmo, A. (2019). Paying Too Much? Price Dispersion in the U.S. Mortgage Market.

Bocian, D., Ernst, K., & Li, W. (2008). Race, Ethnicity and Subprime Home Loan Pricing. *Journal of Economics and Business*, 60(1-2), 110-124.

Cheng, P., Lin, Z., & Liu, Y. (2014). Racial Discrepancy in Mortgage Interest Rates. *The Journal of Real Estate Finance and Economics, 51*(1), 101-120.

Clarke, K. A., & Rothenberg, L. S. (2017). Mortgage Pricing and Race: Evidence from the Northeast. *American Law and Economics Review*, 20(1), 138-167.

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

Hanson, A., Hawley, Z., Martin, H., & Liu, B. (2016). Discrimination in mortgage lending: Evidence from a correspondence experiment. *Journal of Urban Economics*, *92*, 48-65.

Haughwout, A. F., Mayer, C. J., & Tracy, J. S. (2009). Subprime Mortgage Pricing: The Impact of Race, Ethnicity, and Gender on the Cost of Borrowing. *SSRN Electronic Journal*.

Munnell, A. H., Tootell, G. M., Browne, L. E., & McEneaney, J. (1996). Mortgage lending in Boston: Interpreting HMDA data. *The American Economic Review*, 25-53.

Ross, S. L., Turner, M. A., Godfrey, E., & Smith, R. R. (2008). Mortgage lending in Chicago and Los Angeles: A paired testing study of the pre-application process. *Journal of Urban Economics*, 63(3), 902-919.

Woodward, S. E. (2008). A study of closing costs for FHA mortgages.

Woodward, S., & Hall, R. (2012). Diagnosing Consumer Confusion and Sub-optimal Shopping Effort: Theory and Mortgage-Market Evidence. *American Economic Review,* 102(7), 3249-76.

Table 1: Summary statistics

	All	White	Black	Hispanic	Asian	Other
A. FHA-HMDA data						
Interest rate spread over prime (percentage points)	0.089	0.061	0.137	0.166	0.029	0.081
FICO score	680.7	683.5	670.5	678.2	686.8	678.1
Loan amount ($, thousands)	181.74	174.44	177.16	187.82	236.44	203.09
Income ($, thousands)	66.03	67.25	61.53	58.38	71.51	74.46
Fraction single female	0.254	0.255	0.452	0.227	0.227	0.111
Fraction with max LTV (96.5%)	0.89	0.89	0.89	0.90	0.81	0.88
Debt service to income (%)	40.6	39.8	41.6	42.5	42.5	40.7
Months reserves (share ≥ 3 months)	0.43	0.43	0.41	0.41	0.48	0.44
N	971,222	568,897	104,204	165,662	23,926	108,533
B. FHA-HMDA-Optimal Blue data						
Interest rate spread over prime (percentage points)	0.150	0.122	0.221	0.225	0.073	0.122
N	157,853	92,156	15,777	30,303	3,593	16,024

Note: FHA home purchase loans originated in 2014 and 2015. Prime rate is the average rate reported in the Freddie Mac Primary Mortgage Market Survey. In Panel A, we use the prime rate either in the week of FHA insurance application, or in the week of closing if the closing date was more than 60 days after the FHA application date. In Panel B, we use the prime rate in the week of rate lock using the lock date observed in the Optimal Blue data. Sample restricted to 30-year fixed rate loans with loans amounts under 625,000 and LTVs over 90%, for single-family, site built, owner-ccupied properties. Data are trimmed at the 1st and 99th percentiles of the interest rate spread. White refers to non-Hispanic white borrowers, while Hispanic refers to Hispanic white borrowers.

Table 2: Interest rate spread regressions

	y = interest rate spread			
	(1)	(2)	(3)	(4)
Black	0.0756**	0.0786**	0.0382**	0.0308**
	(0.0066)	(0.0031)	(0.0024)	(0.0019)
Hispanic	0.1047**	0.0645**	0.0312**	0.0156**
	(0.0078)	(0.0035)	(0.0028)	(0.0021)
Asian	-0.0321**	-0.0315**	-0.0368**	-0.0265**
	(0.0069)	(0.0039)	(0.0037)	(0.0031)
Other	0.0202**	0.0067**	-0.0054**	0.0035**
	(0.0034)	(0.0018)	(0.0016)	(0.0011)
County and week fixed effects		yes	yes	yes
Borrower controls			yes	yes
Lender fixed effects				yes
Adj. R-squared	0.016	0.171	0.248	0.412
N	971,222	971,137	971,137	971,082

Data source: Merged FHA-HMDA data for home purchase loans originated in 2014 and 2015.
Notes: * p<0.05 ** p<0.01. Standard errors, clustered at county level, in parentheses. Week fixed effects refer to week of application for FHA insurance. Borrower controls includes gender, co-borrower presence and gender, dummy variables for 9 FICO score categeories, dummy variables for income deciles. Reference group is non-Hispanic white borrowers.

Table 3: Interest rate spread and discount points regressions

	Outcome: Interest rate spread			Outcome: # discount points paid		
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.0955**	0.0440**	0.0311**	-0.0127	-0.0305	-0.0830**
	(0.0059)	(0.0049)	(0.0037)	(0.0175)	(0.0165)	(0.0115)
Hispanic	0.0770**	0.0416**	0.0226**	-0.1051**	-0.0997**	-0.0501**
	(0.0064)	(0.0045)	(0.0032)	(0.0319)	(0.0304)	(0.0131)
Asian	-0.0207**	-0.0209**	-0.0212**	0.0758**	0.0851**	0.0660**
	(0.0062)	(0.0056)	(0.0052)	(0.0273)	(0.0271)	(0.0221)
Other	0.004	-0.0100**	-0.0035	0.0478**	0.0171	-0.004
	(0.0030)	(0.0029)	(0.0023)	(0.0164)	(0.0149)	(0.0109)
County and week FE	yes	yes	yes	yes	yes	yes
Borrower controls		yes	yes		yes	yes
Lender Fixed Effects			yes			yes
Adj.R-squared	0.213	0.373	0.52	0.157	0.161	0.513
N	157,485	157,485	157,332	157,485	157,485	157,332

Data source: HMDA-FHA-Optimal Blue matched data for home purchase loans originated in 2014 and 2015.

Notes: * p<0.05, ** p<0.01. Standard errors, clustered at the county level, in parentheses. Borrower controls includes gender, co-borrower presence and gender, dummy variables for 9 FICO score categeories, dummy variables for income deciles.

Table 4: Estimating racial and ethnic gaps in lenders' revenue

	(1)	(2)	(3)
Black	0.454**	0.189**	0.075**
	(0.043)	(0.035)	(0.022)
Hispanic	0.259**	0.096**	0.057**
	(0.034)	(0.026)	(0.018)
Asian	-0.019	-0.009	-0.031
	(0.031)	(0.030)	(0.024)
Other	0.069**	-0.028	-0.019
	(0.025)	(0.023)	(0.014)
County and week FE	yes	yes	yes
Borrower controls		yes	yes
Lender Fixed Effects			yes
Adj.R-squared	0.263	0.408	0.695
N	157,485	157,485	157,332

Data source: HMDA-FHA-Optimal Blue matched data for home purchase loans originated in 2014 and 2015.
Notes: * $p<0.05$, ** $p<0.01$. Standard errors, clustered at the county level, in parentheses. The dependent variable is mortgage revenue from discount points plus expected YSP. See text for details. Borrower controls includes gender, co-borrower presence and gender, dummy variables for 9 FICO score categeories, dummy variables for income deciles.

Table 5: Estimated total points and fees for loans with reported APR

	All	White	Black	Hispanic	Asian	Other
Fraction with APR spread reported	0.48	0.44	0.54	0.59	0.41	0.45
Point and fees (percentage point contribution to APR)	0.12	0.12	0.13	0.14	0.12	0.10
N	428,613	248,736	47,154	74,930	10,781	47,012

Source: Merged FHA-HMDA data for home purchase loans originated in 2014.

Note: Prime rate is the average rate reported in the Freddie Mac Primary Mortgage Market Survey in the week of FHA insurance application.

Table 6: Fees regressions

	y = Fees (percentage point contribution to APR)		
	(1)	(2)	(3)
Black	0.0055*	-0.0009	0.0021
	(0.0027)	(0.0026)	(0.0016)
Hispanic	0.0153**	0.0017	0.0018
	(0.0018)	(0.0017)	(0.0014)
Asian	-0.0054	-0.0088*	-0.0048
	(0.0043)	(0.0042)	(0.0038)
Other	-0.0172**	-0.0158**	-0.0011
	(0.0020)	(0.0021)	(0.0014)
County and week fixed effects	yes	yes	yes
Borrower controls		yes	yes
Lender fixed effects			yes
Adj. R-squared	0.133	0.155	0.528
N	100,423	100,423	100,322

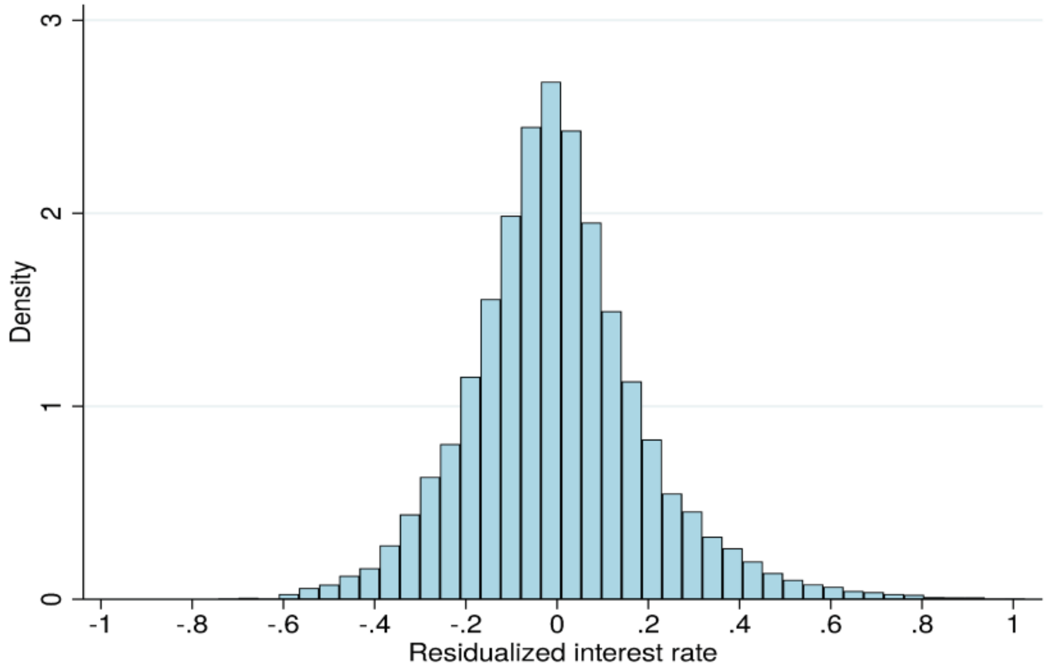
Data source: Merged FHA-HMDA data for home purchase loans originated in 2014.
Notes: * p<0.05 ** p<0.01. Standard errors, clustered at county level, in parentheses. Sample restricted to borrowers with interest rate spread of at least 20 basis points. Points and fees estimated using data on APR spread, interest rate, APOR during week of application, and mortgage insurance premiums. See text for details. Week fixed effects refer to week of application for FHA insurance. Borrower controls includes gender, co-borrower presence and gender, dummy variables for 9 FICO score categeories, dummy variables for income deciles.

Table 7: Probability of loan being higher-priced regressions

	y = **1**(higher-priced)	
	(1)	(2)
Black	0.0940**	0.0367**
	(0.0083)	(0.0032)
Hispanic	0.1519**	0.0266**
	(0.0097)	(0.0033)
Asian	-0.0332**	-0.0346**
	(0.0090)	(0.0051)
Other	0.006	-0.0038
	(0.0052)	(0.0026)
County and week fixed effects		yes
Borrower controls		yes
Lender fixed effects		yes
Adj. R-squared	0.015	0.323
N	428,613	428,387

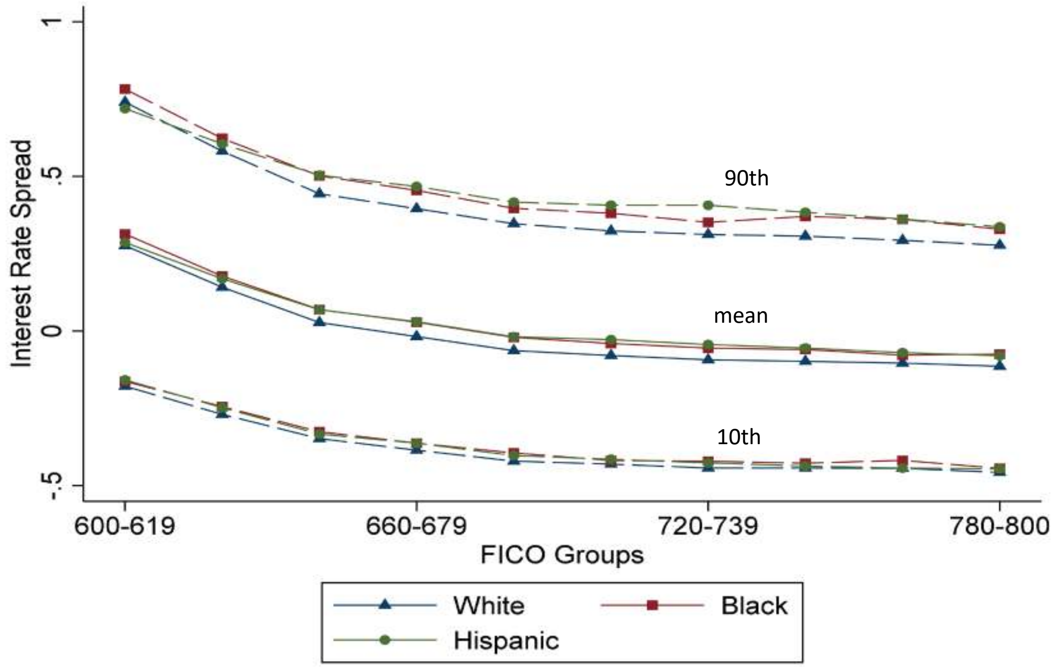
Data source: Merged FHA-HMDA data for home purchase loans originated in 2014.
Notes: * p<0.05 ** p<0.01. Standard errors, clustered at county level, in parentheses. Higher priced loans are those with reported APR spreads in the HMDA data because the spread exceeds 150 basis points. Week fixed effects refer to week of application for FHA insurance. Borrower controls includes gender, co-borrower presence and gender, dummy variables for 9 FICO score categeories, dummy variables for income deciles.

Figure 1: Dispersion in interest rates for FHA loans



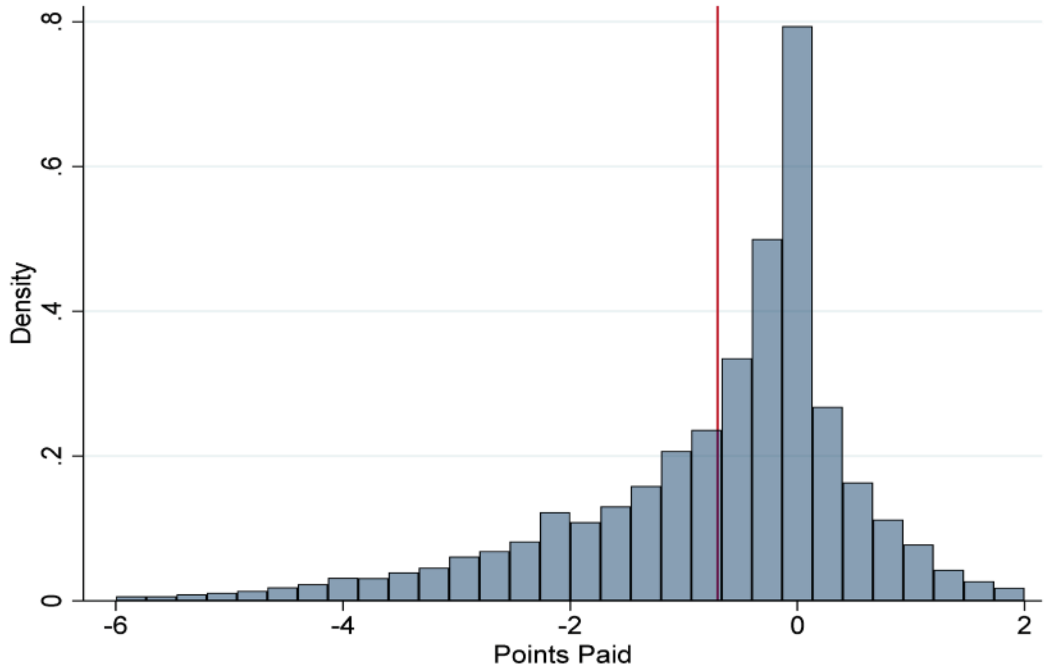
Source: Optimal Blue data for home purchase loans originated in 2014 and 2015.
Note: Sample restricted to loans with zero discount points. Figure plots residuals from regression of interest rates on 9 FICO bin dummies, MSA-by-week fixed effects, and day of rate lock fixed effects.

Figure 2: Interest rate distribution, by FICO score and race or ethnicity



Source: Merged FHA-HMDA data for home purchase loans originated in 2014 and 2015.
Note: Interest rate spread measures difference between interest rate on loan and prime rate during the week the application for FHA insurance was submitted. conditional on county fixed effects.

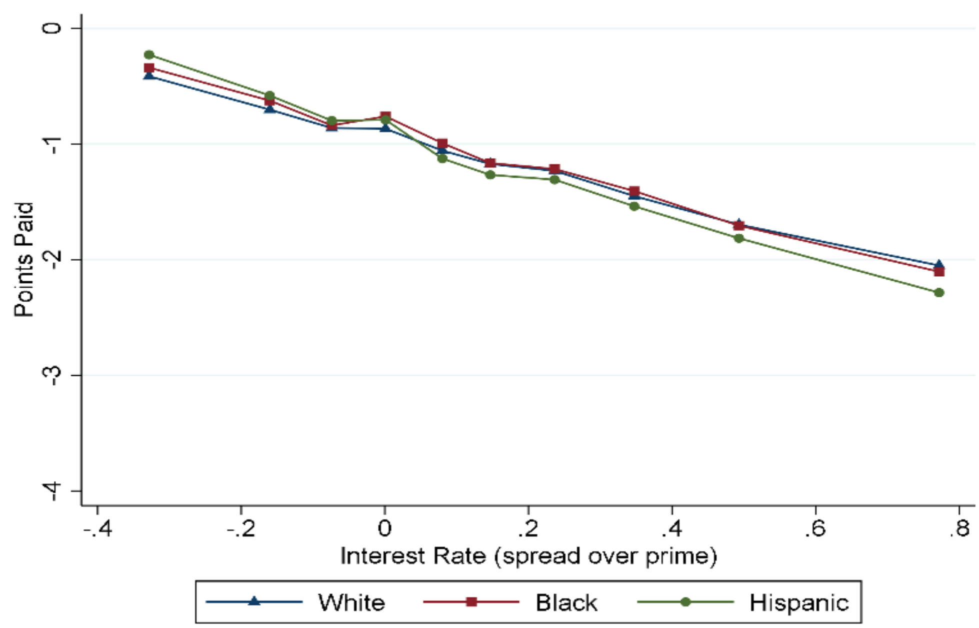
Figure 3: Distribution of discount points



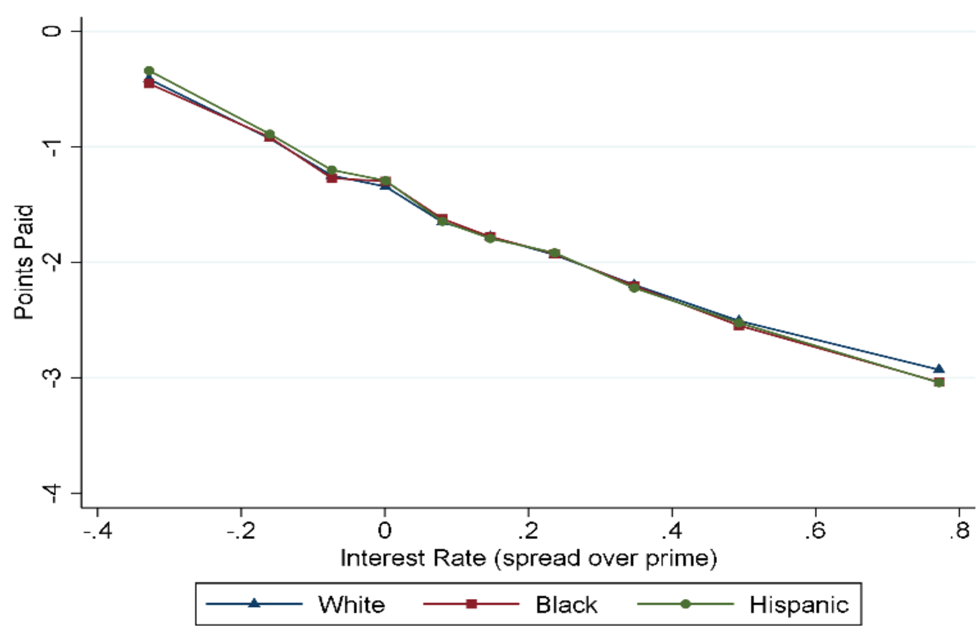
Source: Merged FHA-HMDA-Optimal Blue data for home purchase loans originated in 2014 and 2015.
Note: Negative points indicate that the borrower received cash at closing, which can be used to help pay for closing costs. Red line show the mean of the distribution.

Figure 4: The tradeoff between discount points and interest rates, by race

A. No lender fixed effects

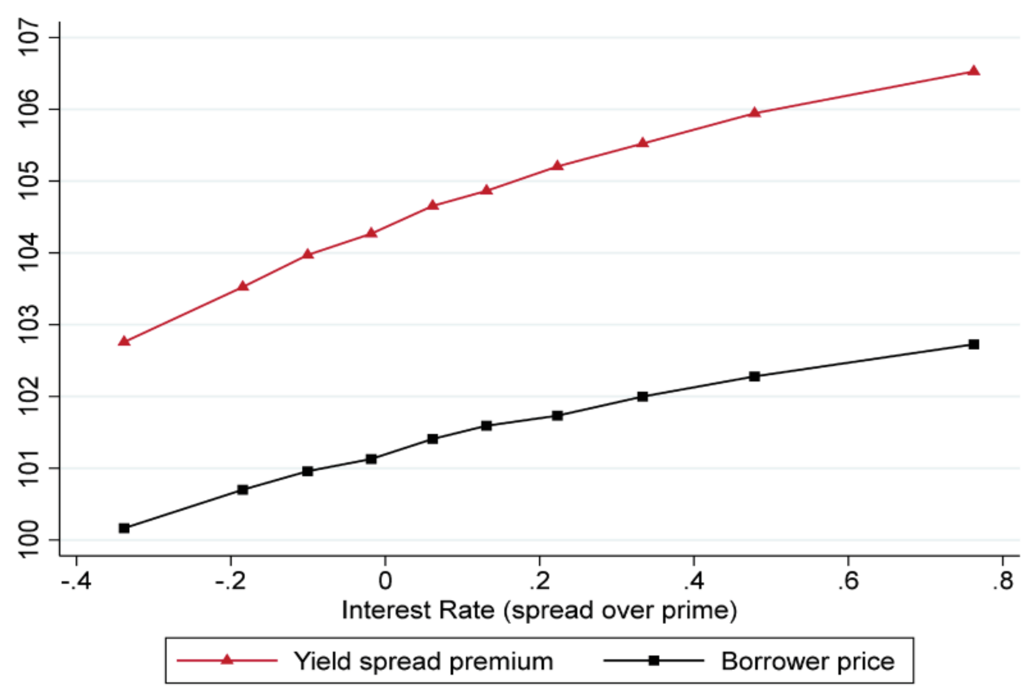


B. With lender fixed effects



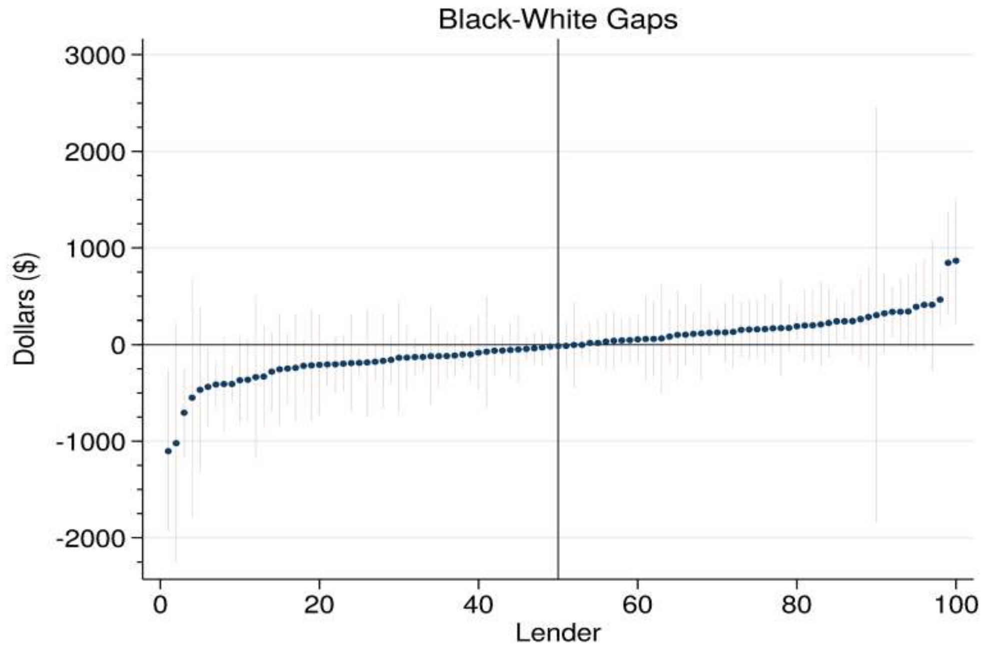
Source: Merged HMDA-FHA-Optimal Blue dataset
Figures plot coefficients from regressions of discount points on interest rate spread deciles interacted with race and ethnicity. Only white, black and Hispanic borrowers, and those with interest spreads over -50 bps and under 110 bps, are included in the regression. Controls include gender, dummies for 9 FICO buckets, dummies for income decile, and county and week of rate lock fixed effects. The only difference between panels A and B is that B shows coefficients from a regression that also controls for lender fixed effects. For regression output and standard errors, see Appendix Table A2. Negative values for points paid implies borrowers recieved money from lenders at closing.

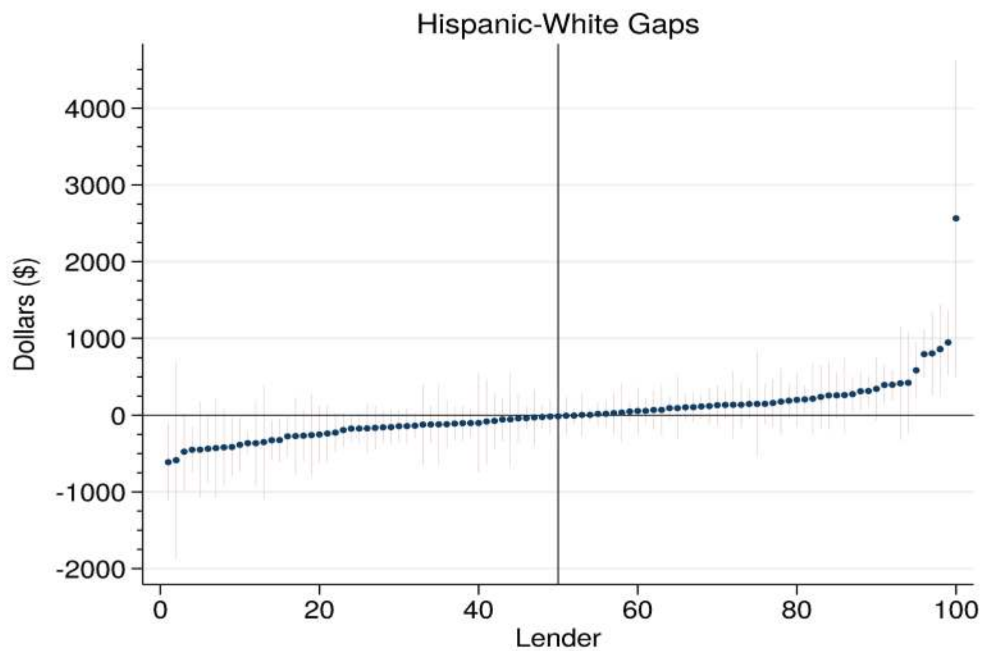
Figure 5: Borrower price and YSP, by interest rate spread



Source: Optimal Blue data for home purchase loans locked in 2014 and 2015
Notes: Black line shows the average borrower price, defined as 100 minus the number of discount points, based on 221,787 home purchase mortgage rate locks. The red line shows the average yield-spread premium (YSP), or secondary market price, based on different set of 13,797 mortgage rate locks where we observe the YSP instead of the borrower price. Prices shown are conditional on FICO, LTV, DTI, loan amount, and fixed effects for lender, MSA, and lock date.

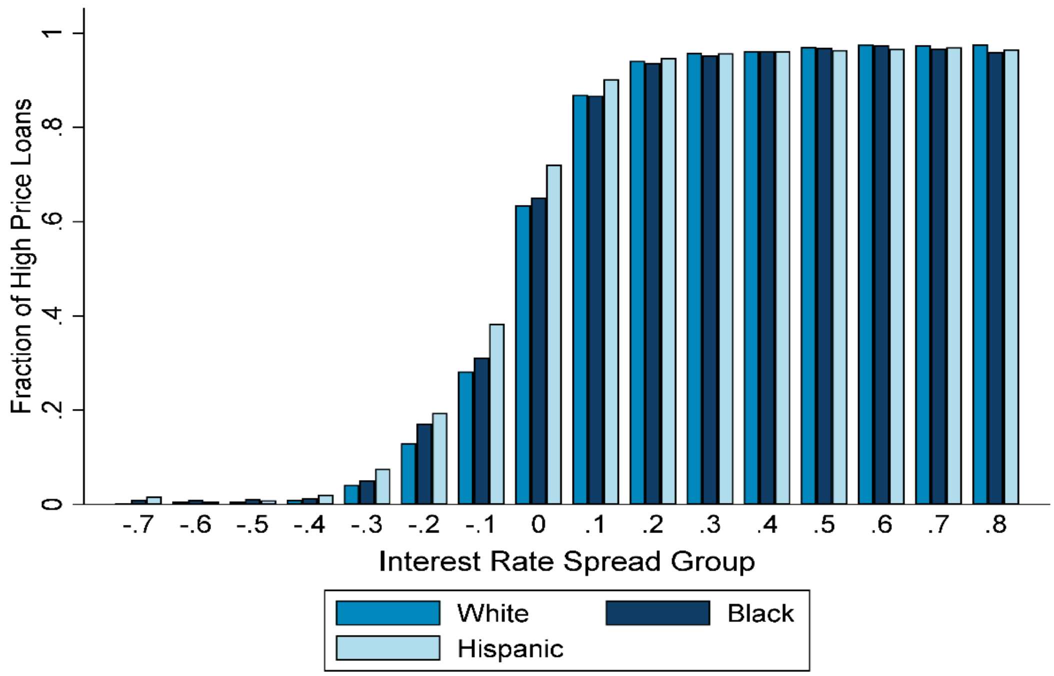
Figure 6: Lender-specific pricing gaps, largest 100 lenders





Source: Merged FHA-HMDA-Optimal Blue data for home purchase loans originated in 2014 and 2015.
Notes: Sample restricted to the largest 100 lenders that have at least 10 loans to black borrowers and 10 loans to Hispanic borrowers. Each data point represents the racial/ethnic gap in upfront points charged for a loan of $175k for a single lender. The vertical lines represent 95 percent confidence intervals. Gaps estimated from a regression of discount points on race and ethnicity dummies interacted with lender dummies, controlling for rate spread bins interacted with lender, and the full set of borrower and loan controls as described in column 4 of Table 2.

Figure 7: Share of loans with APR spread reported, by interest rate spread and by race or ethnicity



Source: Merged FHA-HMDA data for home purchase loans originated in 2014.
Note: Interest rate spread measures difference between interest rate on loan and prime rate during the week the application for FHA insurance was submitted. High price loans are those with a reported APR spread in the HMDA data because the APR spread exceeds 150 basis points.

Appendix

A1: Summary statistics for merged datasets vs. full HMDA and full FHA datasets

	(1)	(2)	(3)	(4)	(5)
	HMDA Data	FHA Data	FHA-HMDA Merged Data	FHA-HMDA Merged Data (Analysis Sample)	FHA-HMDA - Optimal Blue Merged Data (Analysis Sample)
N	1,411,978	1,359,978	1,160,429	971,222	157,853
Loan Amount ($, 000's)	185.82	184.48	182.61	181.74	186.82
Income ($, 000's)	67.07	67.14	66.60	66.03	69.14
Home value ($, 000's)		190.61	188.62	185.63	194.03
Fraction with max LTV		0.80	0.80	0.89	0.92
Interest Rate		4.09	4.09	4.09	4.13
Fraction High-Cost in 2014 (i.e. has reported APR spread)	0.45		0.46	0.48	0.56
FICO		680.78	680.82	680.66	679.30
DTI		40.53	40.41	40.60	41.07
Fraction with reserves ≥ 3 months		0.45	0.45	0.43	0.43
First Time Buyer Share		0.82	0.82	0.83	0.81
White	0.58		0.59	0.59	0.58
Black	0.11		0.10	0.11	0.10
Hispanic	0.17		0.17	0.17	0.19
Asian	0.03		0.03	0.02	0.02
Single Female	0.26		0.26	0.25	0.25

Table A1 displays summary statistics for FHA-insured home purchase loans originated in 2014 before and after matching the HMDA and administrative HUD/FHA datasets. Column 1 shows that 1.4M loans were reported in HMDA, slightly above the number in the FHA dataset (column 2) because the FHA data we obtained contains only the loans endorsed by the FHA by the end of 2015. Because endorsements can occur several weeks after closing, our FHA data is missing some FHA loans that were closed toward the end of 2015. For 2014 and most of 2015, the FHA data has slightly more loans than reported in HMDA. We match the two sets of FHA purchase loans on exact loan amount, census tract, and borrower income, after dropping non-unique observations from each dataset. Column 3 indicates that nearly 1.2M loans were matched in this manner, or about 85 percent of the number in column 2. Column 3 also shows that the summary statistics for the matched sample are nearly identical to those in the original datasets. Column 4 shows summary statistics for our main analysis sample, which is restricted to 30-year fixed rates loans with LTVs over 90 percent, for owner-occupied, 1-4 family, non-manufactured housing. We also drop outliers, trimming at the 1st and 99th percentile of the interest rate spread, and we drop borrowers with FICO scores below 600 and above 800. Finally, we match the HMDA-FHA analysis sample to the Optimal Blue data on interest rate locks. We first merge on exact county, loan amount, LTV, and interest rate, and then break ties by requiring that the closing date reported in the HUD data falls within the interest rate lock period, and that the FICO score reported in Optimal Blue is within 10 points of the score reported to HUD. We believe that the Optimal Blue data covers about 25 percent of all originations. Thus, the roughly 158 thousand matched loans in column 5 is about 65 percent of the maximum number of matches we could have expected.

Table A2: Regression of points paid on race and interest rate spread

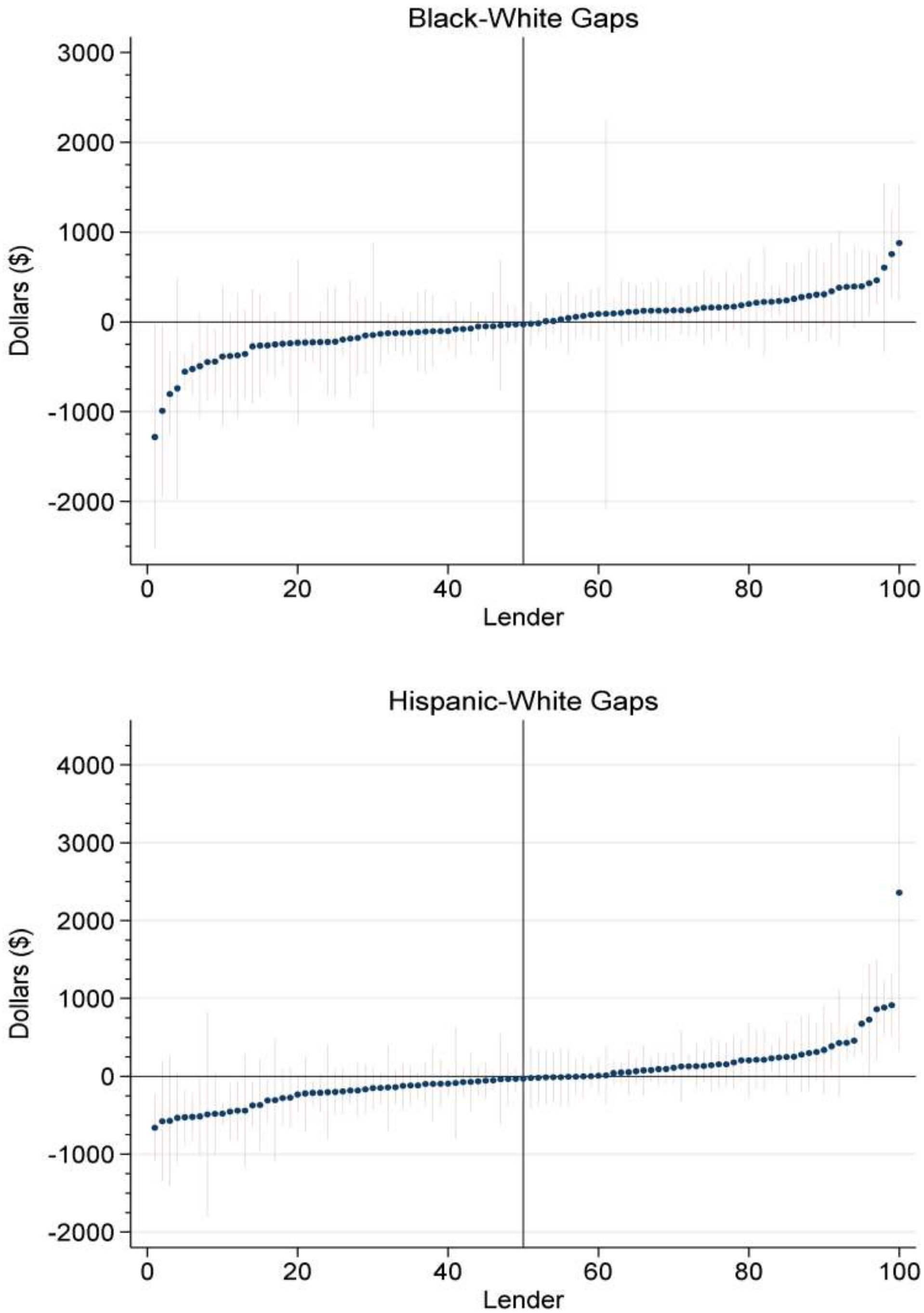
	(1)	(2)
Black	0.0748	-0.0389
	(0.04273)	(0.03220)
Hispanic	0.1860**	0.0738**
	(0.04656)	(0.02318)
Deciles of interest rate spread		
2nd decile	-0.2891**	-0.5148**
	(0.02828)	(0.01489)
3rd decile	-0.4469**	-0.8353**
	(0.03494)	(0.01868)
4th decile	-0.4536**	-0.9303**
	(0.03558)	(0.02088)
5th decile	-0.6431**	-1.2386**
	(0.04041)	(0.02485)
6th decile	-0.7585**	-1.3649**
	(0.04355)	(0.02566)
7th decile	-0.8179**	-1.5215**
	(0.04308)	(0.02562)
8th decile	-1.0360**	-1.7818**
	(0.04781)	(0.03029)
9th decile	-1.2826**	-2.0932**
	(0.05218)	(0.03535)
10th decile	-1.6362**	-2.5166**
	(0.06616)	(0.04417)
(Black)*(Interest spread decile)		
2nd decile	0.0035	0.0498
	(0.04871)	(0.03343)
3rd decile	-0.0529	0.0156
	(0.05472)	(0.03831)
4th decile	0.0320	0.0849*
	(0.05331)	(0.03991)
5th decile	-0.0103	0.0677
	(0.05675)	(0.04051)
6th decile	-0.0671	0.0383
	(0.05372)	(0.04148)
7th decile	-0.0586	0.0474
	(0.05762)	(0.04200)
8th decile	-0.0311	0.0309
	(0.05130)	(0.03993)
9th decile	-0.0839	-0.0021
	(0.05310)	(0.03973)
10th decile	-0.1282	-0.0686
	(0.06714)	(0.05065)

(Hispanic)*(Interest spread decile)		
2nd decile	-0.0638	-0.0350
	(0.04265)	(0.02510)
3rd decile	-0.1227**	-0.0256
	(0.04331)	(0.02363)
4th decile	-0.1090	-0.0222
	(0.05764)	(0.02699)
5th decile	-0.2552**	-0.0665*
	(0.07560)	(0.03329)
6th decile	-0.2801**	-0.0888*
	(0.07333)	(0.03447)
7th decile	-0.2621**	-0.0565
	(0.06469)	(0.03365)
8th decile	-0.2738**	-0.1000**
	(0.07125)	(0.03803)
9th decile	-0.3042**	-0.0913*
	(0.07380)	(0.03587)
10th decile	-0.4206**	-0.1854**
	(0.08731)	(0.04730)
Lender Fixed Effects		Yes
Adj.R-squared	0.261	0.671
N	135099	134954

* p<0.05, ** p<0.01; standard errors in parentheses, clustered at county level

Table A2 shows regression results corresponding to **Figure 4**, using the merged HMDA-FHA-Optimal Blue dataset (see Table A1 for details) on FHA home purchase loans originated in 2014. The coefficients trace out how average points paid changes with interest rates by race. Only white, black and Hispanic borrowers, and those with interest spreads over -55 bps and under 90 bps, are included in the regression. White borrowers are the reference race category. Controls include gender, dummy variables for 9 FICO score buckets, dummy variables for income decile, and county and week of rate lock fixed effects. The main interest rate decile coefficients show that points paid declines steadily as interest rates increase. The black borrower main and interaction coefficients indicate that black borrowers do not pay a statistically different amount of points than white borrowers, regardless of including lender fixed effects. The Hispanic main and intereraction coefficients in column 1 indicate that Hispanic borrowers pay slightly more in points relative to white borrowers at the lowest interest rate decile, but pay somewhat less in points than white borrowers at higher interest rate deciles. However, these results attenuate noticeably after including lender fixed effects in column 2.

Figure A3: Lender-specific revenue gaps, largest 100 lenders



Source: Merged FHA-HMDA-Optimal Blue data for home purchase loans originated in 2014 and 2015

Notes: Each data point represents gap estimate for a single lender, and vertical lines represent 95 percent confidence intervals. Gaps estimated from regression of loan revenue, based on equation (1) (see text), on race and ethnicity dummies interacted with lender dummies, controlling for the full set of controls as described in column 4 of Table 2.