Does Borrower and Broker Race Affect the Cost of Mortgage Credit?*

Brent W. Ambrose,† James N. Conklin,‡ and Luis A. Lopez§

February 17, 2020

ABSTRACT

We test for pricing disparities in mortgage contracts using a novel dataset that allows us to observe the race and ethnicity of both parties to the loan. We find that minorities pay more in fees than similarly qualified whites when obtaining a loan through the same white broker. Critically, we find that the premium paid by minorities depends on the race of the broker. We also examine recent policy changes regarding broker compensation rules that may reduce these price disparities, but may also limit access to credit for minorities.

JEL Classification: J15, L85, R20.

Keywords: Discrimination, Credit Rationing, Regulations, Dodd-Frank Act, Mortgages

*We thank Sumit Agarwal, Wayne Archer, Patrick Bayer, Neil Bhuta, Larry Cordell, Ed Coulson, Stuart Gabriel, Kris Gerardi, Andra Ghent, Aurel Hizmo, Lawrence Katz, David Ling, Rich Martin, Gonzalo Maturana, Andy Naranjo, Ruchi Singh, Will Strange, Joe Tracy, Susan Woodward, Tyler Yang, Tony Yezer, and the seminar participants at Baruch College, the Federal Reserve Bank of Philadelphia, the University of Florida, Georgia State University, Laval University, the National University of Singapore, the Pennsylvania State University, the 2018 Singapore Management University Conference on Urban and Regional Economics, the 2018 AREUEA National Meeting, the 2018 American Real Estate Society Meeting, and the 2020 Allied Social Science Association meeting for their helpful comments and suggestions. We also thank the Penn State Institute for Real Estate Studies for providing access to the New Century Mortgage database. James Conklin received research support from the University of Georgia's Terry-Sanford research award. We received the 2018 ARES Best Paper Award in Real Estate Finance, sponsored by Real Capital Analytics. We affirm that we have no material financial interests related to this research.

†Smeal College of Business, Pennsylvania State University, 381 Business Building, University Park, PA 16802 (e-mail: bwa10@psu.edu; office: 814-867-0066) – Corresponding author.

‡Terry College of Business, University of Georgia (e-mail: jnc152@uga.edu).

§College of Business Administration, University of Illinois at Chicago (e-mail: lal@uic.edu).

I. Introduction

Economists and policy makers have long observed that economic outcomes are correlated with race and ethnicity in many countries and societies. In the United States, these observations created pressure to pass legislation to remove discriminatory practices, such as the 1964 Civil Rights Act that ended segregation and sought to create equal opportunities in labor markets and the 1968 Fair Housing Act that explicitly targeted discriminatory practices in the housing market. Yet, over half a century later, evidence suggests that discriminatory practices in housing and financial markets may continue. For example, several large mortgage lenders were recently subject to litigation involving charges of disparate treatment of minorities (New York, Office of the Attorney General, Civil Rights Bureau, 2008, 2007, 2006; United States Dept. of Justice, 2010, 2011) and academic research continues to find evidence of loan pricing disparities (Bhutta and Hizmo, 2019; Woodward and Hall, 2012; Ghent, Hernández-Murillo, and Owyang, 2014; Cheng, Lin, and Liu, 2015; Bartlett et al., 2018) suggesting that discrimination in the mortgage industry remains a concern.[1]

Earlier studies often relied on observing differences in transaction outcomes (e.g., accept/reject decisions, pricing, or mortgage performance) across race as evidence for discrimination. The seminal study in this line is Munnell et al. (1996), who reported that minorities in Boston had mortgage loan denial rates that were twice as high as white applicants. In addition, Berkovec et al. (1994, 1998) found evidence consistent with discrimination using FHA loan performance (i.e. mortgage default) data.[2] However, using loan denial rates and ex post loan performance to measure potential

[1]In other areas, recent studies have found evidence for discrimination or preferential treatment in areas such as online markets for iPods (Doleac and Stein, 2013), apartment rentals on Airbnb (Edelman, Luca, and Svirsky, 2017), employment practices (Fadlon, 2015), single-family housing markets (Haughwout, Mayer, and Tracy, 2009; Woodward and Hall, 2012; Bayer et al., 2012; Ewens, Tomlin, and Wang, 2014; Bartlett et al., 2018; Fuster et al., 2018), policing and criminal justice (Anwar and Fang, 2006; Antonovics and Knight, 2009; Abrams, Bertrand, and Mullainathan, 2012; Anwar, Bayer, and Hjalmarsson, 2012), and medical treatment (Anwar and Fang, 2012) to name just a few. Furthermore, evidence of discrimination exists internationally as well with Zussman (2013) reporting evidence of discriminatory practices in used car sales in Israel, Fisman, Paravisini, and Vig (2017) finding evidence for discrimination based on cultural affinity in the personal loan market in India, Glover, Pallais, and Pariente (2017) identifying evidence for statistical discrimination in hiring based on performance of cashiers in French grocery stores, Hjort (2014) reporting evidence for discrimination based on ethnicity using data from Kenya, and Hedegaard and Tyran (2018) using a novel experiment from Denmark to estimate the price elasticity of ethnic discrimination.

[2]Yezer (2006) and Ladd (1998) provide summaries of the earlier literature on mortgage discrimination while

discrimination is controversial (Horne, 1997; Ross and Yinger, 2002; Brueckner, 1996; Ross, 1996; Yinger, 1996) with such comparisons suffering from two distinct issues. First, they lack information on the individual on the other side of the transaction (the loan officer or broker); to draw conclusions about discrimination, it is important to know not only the race and ethnicity of the borrower, but also the race and ethnicity of the individual loan officer. Yet, such information is unavailable in most administrative datasets, and knowing the source of the disparate treatment (whether it is the originating institution or individual) is crucial to designing public policy remedies. Second, previous studies focusing on contract interest rate differentials (Avery, Canner, and Cook, 2005; Bhutta and Ringo, 2015; Courchane and Nickerson, 1997; Crawford and Rosenblatt, 1999; Black, Boehm, and DeGennaro, 2003; Boehm and Schlottmann, 2007; Haughwout, Mayer, and Tracy, 2009; Ghent, Hernández-Murillo, and Owyang, 2014; Bayer, Ferreira, and Ross, 2017; Bartlett et al., 2018), application rejection rates (Munnell et al., 1996; Ross and Yinger, 2002; Horne, 1997; Ross and Yinger, 2002), or performance differences (Bayer, Ferreira, and Ross, 2016; Deng and Gabriel, 2006; Berkovec et al., 1998, 1994; Brueckner, 1996; Ross, 1996; Yinger, 1996) often found it challenging to control for important borrower characteristics observed by the lender but unobserved by the econometrician (e.g., credit quality), making it difficult to reject alternative explanations (Horne, 1997; Ross and Yinger, 2002). To overcome these challenges, we use a novel administrative dataset containing all underwriting information collected for over 300,000 mortgages originated by mortgage brokers between 2003 and 2007. In doing so, we make four key contributions.

First, whereas previous studies only observe the race/ethnicity of the borrower, our dataset allows us to identify the race and ethnicity of the individual lender (i.e. broker) by using the Bayesian Improved First Name Surname Geocoding (BIFSG) method developed in Voicu (2018). Thus, we provide additional insights into the literature examining within and across ethnic and racial group interactions (Agarwal et al., 2019; Li, 2014; Wong, 2013; Zhang and Zheng, 2015; Bertrand, Luttmer, and Mullainathan, 2000; Bayer, McMillan, and Rueben, 2004).These studies

Courchane and Ross (2019) summarize more recent research and court cases covering discrimination and disparate treatment.

provide evidence regarding the heterogeneous discounts and premiums that ethnic/racial groups charge each other. Thus, our study complements the existing literature by documenting a new channel by which pricing differentials in housing markets occur.

Second, to overcome issues associated with omitted credit risk characteristics (Horne, 1997; Ross and Yinger, 2002), we focus on a measure of the cost of credit that should not be vulnerable to this criticism – mortgage broker fees.[3] Since mortgage brokers are purely middlemen that arrange loans, they do not bear credit or interest rate risk and thus standard theories of pricing suggest that their compensation should be independent of borrower credit or interest rate risk (Woodward and Hall, 2012).[4] Thus, our analysis builds on the work of Woodward and Hall (2012) and Woodward (2008), who focus on fees for loans insured by the Federal Housing Administration (FHA), to circumvent a concern that plagues existing studies: that omitted credit risk attributes explain observed differences across borrower racial/ethnic groups (Horne, 1997).

Third, our analysis is related to recent studies showing that the inclusion of lender (institution) fixed effects reduces the magnitude of earlier findings of racial disparities in mortgage outcomes (Bayer, Ferreira, and Ross, 2017; Bhutta and Hizmo, 2019; Avery, Canner, and Cook, 2005; Avery, Brevoort, and Canner, 2007). The inclusion of broker fixed effects has a similar interpretation, but at a more granular (individual broker) level, allowing us to focus on the differences in fees attributable to the individual mortgage broker – a previously unexplored area.

Finally, we contribute to the debate surrounding the efficiency and effects of financial regulations designed to protect consumers (Campbell et al., 2011). We design a test to illustrate and quantify

[3] Mortgage brokers receive compensation from two sources: origination fees paid by borrowers and lender rebates (yield spread premiums). The former refers to the numerous potential expenses such as points, application fees, underwriting fees, and other miscellaneous fees borrowers pay the broker at mortgage closing. The latter refers to the rebate the lender pays the broker for negotiating a contract interest rate above the minimum market rate the borrower qualifies to receive. Because borrowers enter the market infrequently and mortgages are heterogeneous products, consumers are at an informational disadvantage relative to market specialists (mortgage brokers) that have considerable discretion over pricing. Thus, mortgage markets are conducive to price dispersion.

[4] Although brokers may face reputation risk for delivering low quality (high credit risk) loans, we provide empirical evidence that broker compensation is not directly tied to credit risk in Section III.B. Moreover, our study uses mortgages originated during an economic expansionary period characterized by rising house prices and low early termination events (a key trigger for lender mortgage rescission). As a result, during this period broker concerns regarding reputation risks arising from credit risks are most likely minimal.

the consequences (intended and unintended) of recent consumer protection regulations that were designed to restrict broker compensation.

To preview our results, we observe that Hispanic, black, and Asian borrowers pay a significant premium relative to white borrowers when obtaining a loan through a white broker after conditioning on a rich and extensive set of borrower, loan, property, and area characteristics. Minority premiums remain, but are smaller in magnitude, after the inclusion of individual broker fixed effects, indicating that: (i) a minority borrower pays more than a comparable white borrower when using *the same broker* and (ii) minority borrowers tend to systematically select into high-fee brokers, consistent with the hypothesis originated by Yezer, Phillips, and Trost (1994). Our results confirm and complement those of Woodward and Hall (2012) for black and Hispanic borrowers, which were based on a different loan product (FHA mortgages) originated during a different time (a six-week period in 2001). However, our ability to include broker fixed effects results in more modest minority fee premiums, but we note that the premiums paid by minorities are within the range of pricing differences that triggered legal action against lenders for disparate impact. Interestingly, we also observe that white borrowers pay more on average when originating a loan from a minority broker. Finally, we note that variation exists in the premiums across broker race or ethnic groups, with Hispanic brokers charging Hispanic borrowers a premium relative to white borrowers while black brokers do not appear to charge different fees to white and black borrowers. In contrast, we find some evidence that Asian borrowers pay lower fees than comparable white borrowers when originating loans from Asian brokers. We report on a battery of robustness tests in the online appendix to alleviate concerns regarding endogeneity and selection biases. While we remain agnostic about the exact underlying causal mechanism, the results from our tests are robust and point away from typical explanations such as unobservable credit risk, broker effort, borrower contract selection, non-random borrower-broker matching, language, or ethnic enclaves.

We also study how recent regulatory changes arising from the Great Recession and financial crisis may reduce pricing disparities across racial/ethnic groups. For example, Title XIV (the Mortgage

Reform and Anti-predatory Lending Act) of the Dodd-Frank Act places severe restrictions on how mortgage brokers may be compensated.[5] Since we find significant fee disparity across race, our study shows the importance for continued evaluation of the effectiveness of regulatory oversight versus the reliance on enforcement of existing laws to combat disparate treatment in mortgage markets. We first focus on the Dodd-Frank regulation meant to increase pricing transparency: the proposed elimination of dual compensation (broker compensation from both the borrower and the lender as discussed in footnote 3). We continue to find fee differences on transparently priced loans, i.e., mortgages without dual compensation. This suggests that a regulation banning dual compensation, per se, is unlikely to eliminate racial price disparities. Next, we consider whether differences in fees arise from borrower heterogeneity with respect to broker loan production costs, and whether the Dodd-Frank regulations may result in credit rationing disparities. Based on a quantile regression framework, we estimate at the 30^{th} quantile that over 25 percent of Hispanic and black borrowers (and six percent of Asian borrowers) with loans originated by white brokers would have been at risk of being credit rationed as a result of fee caps imposed by Dodd-Frank. In contrast, only about 16 percent of white borrowers with loans originated by white loan officers would be at risk of credit rationing. Thus, although the restrictions may reduce pricing disparities, they may also result in credit rationing to borrowers needing extra effort by mortgage brokers to originate loans as suggested by Yezer, Phillips, and Trost (1994) and Yezer (2017).

II. Data

We use data on loan applications for brokered, first-lien, residential mortgages that were approved and funded by New Century Financial Corporation between January 2003 and March 2007.[6]

[5] The Dodd-Frank Act is available online at `https://www.congress.gov/bill/111th-congress/house-bill/4173/text`. See Section 1403, Prohibition on Steering Incentives, which amends Section 129B of the Truth in Lending Act.

[6] New Century began originating loans in 1997 and stopped in March 2007 when it filed for bankruptcy. We restrict the sample to the period after 2002 due to incomplete data on broker surname prior to 2003. Ambrose, Conklin, and Yoshida (2016) discuss the data in greater detail.

Although the loans were funded by a single lender, our analysis focuses on mortgages that were originated by 124,736 independent mortgage brokers who had access to a variety of lenders, thus reducing concerns that our results are idiosyncratic to one particular lender. Ambrose, Conklin, and Yoshida (2016) provide comparisons to mortgages in other studies that indicate that the New Century loans are representative of the overall subprime market. Nonetheless, we also discuss below that our sample is representative of the subprime market before the Great Recession.

We use the New Century data because each loan file contains elements central to our analysis: the borrower's Home Mortgage Disclosure Act (HMDA) race code and the broker's name and office location.[7] The dataset also contains borrower, property, and loan characteristics as well as broker fees. Based on property location, we merge the New Century data with Census 2000 data to gain geographic controls. The Census variables are similar to those used in Bayer, Ferreira, and Ross (2017). Table 1 in the online appendix lists and describes the variables.

A. Sample Specification and Representativeness

Following Ambrose, Conklin, and Yoshida (2016) and Conklin (2017), we exclude loan applications with missing data or when (1) the borrower's and co-borrower's combined monthly income is negative or greater than \$26,900; (2) the combined loan-to-value ratio is negative or larger than 125 percent; (3) the borrower's FICO credit score is less than 450; (4) the debt-to-income ratio is negative or larger than 60 percent; and (5) the borrower's age is reported as less than 18 years or older than 99 years. We also winsorize the one percent tails of the combined monthly income and broker fees. Furthermore, we keep loans originated by white, Hispanic, black, or Asian/Pacific

[7] During the application stage, the applicant (or the loan officer) fills out a form that asks the applicant to identify her race and ethnicity. The ethnicity question allows the applicant to self-identify as either "Hispanic or Latino" or "Not Hispanic or Latino," while the race question allows the applicant to self-identify as "American Indian or Alaska Native," "Asian," "African American," "Native Hawaiian or Other Pacific Islander," or "white." The categories follow the classification standards of federal data on race and ethnicity (62 Fed. Reg. 131 (9 July 1997)). Hence, to be consistent with the federal classification standards, we categorize borrowers as American Indian or Alaska Native, Asian or Pacific Islander, African American, Hispanic, or white. The Hispanic category applies to all borrowers who self-identify as Hispanic or Latino. The other categories apply to borrowers who self-identify as the corresponding race but not Hispanic or Latino.

Islander brokers to borrowers in those same racial/ethnic groups. The final sample includes 323,846 originated loans.

As noted in section A.1 of the online appendix, the typical principal borrower is a 40-year old, married male with a credit score of 619 and an annual income of approximately \$68,500.[8] The average loan is an adjustable rate mortgage with a loan amount of \$172,800 on a 30-year term with a prepayment penalty.[9] Forty-two percent of these loans were originated to purchase a residential property, and the rest to refinance an existing mortgage. Among refinances, 85 percent are cash-out mortgages having loan amounts that exceed the outstanding balance of debt being refinanced.

To ensure that the sample is representative of the subprime market from 2003 to 2007, we provide an analysis in the online appendix that compares the New Century data with the subprime loan sample in Demyanyk and Van Hemert (2009), which is comprised of loans across many subprime lenders and covers roughly half of the subprime mortgage market (85 percent of the securitized subprime market). As we detail in section A.1, the descriptive statistics across the two samples are quite similar. In section A.1, we also compare the New Century data to the Home Mortgage Disclosure Act (HMDA) loan application register data. We note that the minority share of subprime originations in HMDA for New Century (51 percent) is nearly identical to the share in the rest of the subprime market (52 percent); thus alleviating concerns that the New Century data suffers from selection issues based on borrower minority status.

B. Observable Race and Ethnicity

While we observe borrower race and ethnicity (due to HMDA reporting requirements), we do not directly observe the race and ethnicity of brokers. However, we are able to infer their race and ethnicity using a Bayesian-based classifier approach, which is similar in spirit to the methodology used

[8]In cases where there are multiple borrowers on the loan, the income represents the combined income of these borrowers. Since approximately 41 percent of the loans are low-doc (stated income) loans, the average income reported in the data is likely inflated (Ambrose, Conklin, and Yoshida, 2016).

[9]Here we report the exponential of the average log loan amount ($\exp^{12.06} = 172,800$). The average loan amount of \$206,000 is reported in Table A.1.

by regulators to determine consumer race and ethnicity (Consumer Financial Protection Bureau, 2014b). In addition, various courts have relied on Bayesian-based classification methods in cases where it was necessary to infer an individual's race or ethnicity (e.g., Guardians Ass'n of N.Y.C. Police Dep't v. Civil Serv. Comm'n, 1977, ¶32).[10]

We infer mortgage broker race using the Bayesian Improved First Name Geocoding (BIFSG) method developed in Voicu (2018) (see section A.2 in the online appendix for a detailed discussion). The intuition of the approach is to calculate the probability (Bayesian score) that a person falls into a certain race/ethnicity based on the individual's last name, first name, and location. A Bayesian score for each race is calculated for every broker in the sample.[11] To categorize a broker's race discretely, we apply a "maximum a posteriori" (MAP) classification scheme that sets an individual's race to that of the group with the highest Bayesian score.[12] Relative to other classification schemes, MAP is more accurate, minimizes bias, and maximizes data coverage (Voicu, 2018).

Table 2 reports the number of unique brokers in the sample by the number of loans they originated. We identify the race/ethnicity of 124,736 individual brokers. Sixteen percent are identified as Hispanic, 8 percent as African-American, 4 percent as Asian or Pacific Islander, and the rest as white. To our knowledge, the only other source of demographic information on mortgage loan officers is Hanson et al. (2016). Whereas our sample covers 2003 to 2007, the Hanson et al. (2016) sample is from 2012, a period when subprime mortgage lending was virtually non-existent. But, consistent with Hanson et al. (2016), we find that the overwhelming majority of loan officers are white. The similarity between our loan officer demographics and those reported in Hanson

[10]The name matching method employed in the Guardians Association case was devised in a study conducted by the Rand Institute that inferred the racial profile of subjects in the case by comparing their names to 8,000 surnames obtained from the US Bureau of Census (Guardians Ass'n of N.Y.C. Police Dep't v. Civil Serv. Comm'n, 1977, ¶33). Guardians Ass'n of N.Y.C. Police Dep't v. Civil Serv. Comm'n (1977) was ultimately upheld on appeal (Guardians Asso. of N.Y.C. Police Dep't, Inc. v Civil Serv. Com., 1980; Civil Serv. Com. v. Guardians Ass'n., 1983) and has been cited in subsequent rulings (e.g., United States v. Brown, 2018).

[11]Bayesian scores are calculated for the six race/ethnicity groups defined by the U.S. Census: white, Hispanic, African-American, Asian or Pacific Islander, American Indian or Alaskan Native, and two or more races.

[12]The sample includes very few brokers categorized as American Indian/Alaskan native or Two or More Races. Thus, we exclude these groups from our analysis. In section A.2 of the online appendix we explore other discrete Bayesian classification systems and find that the results are materially unaffected by the choice of classification scheme. We also provide accuracy tests for the BIFSG methodology using publicly available Florida voter data.

et al. (2016) suggest that brokers that worked with New Century are representative of loan officers in the broader mortgage market.

We partition the data into three subsamples: i) Hispanic and white brokers and borrowers (HW); ii) black and white brokers and borrowers (BW); and iii) Asian/Pacific Islander and white brokers and borrowers (AW). Note that the white borrower/white broker observations are the same in each subsample and serve as the reference group. Performing the analysis separately by minority group, with whites as the reference group, may shed more light on the channels through which "minority premiums" emerge.[13]

Table 3 provides observation counts by broker and borrower race for each subsample. We observe that brokers tend to originate loans to borrowers who share the same race or ethnicity. Nineteen percent of loans arranged by white brokers in the HW sample were to Hispanic borrowers. In contrast, 81 percent of those originated by Hispanic brokers were to Hispanic borrowers. We observe similar patterns in the BW and AW subsamples.

III. Analysis of Broker Fees

We examine differences in log broker fees, calculated as the natural logarithm of the sum of front- and back-end fees. Front-end fees include the application fee, underwriter fee, mortgage brokerage firm fee and points.[14] Borrowers generally incur these fees during the loan origination process and pay them at closing.[15] Back-end fees include the yield spread premium and correspondence premium. The yield spread premium is the total rebate that the lender provides to the broker at closing for locking a contract rate above the minimum rate the borrower qualifies to receive (par).[16]

[13] We thank an anonymous referee for this suggestion.

[14] In the mortgage literature points generally refer to fees paid by the borrower directly to the lender to "buy down" the interest rate on the loan. In contrast, in our sample points represent compensation the broker negotiates from the borrower. One point is equivalent to 1 percent of the loan balance at origination.

[15] For refinance loans, origination fees are often rolled into the loan amount. In other words, borrowers do not pay fees out of pocket directly at closing, but rather obtain a larger loan amount to cover the fees.

[16] In theory, the broker can use this rebate to offset origination fees. However, evidence suggests that increases in yield spread premiums are associated with relatively small decreases in origination fees (Woodward and Hall, 2010; Ambrose and Conklin, 2014).

The correspondence premium is analogous to a yield spread premium; it compensates the broker for originating a loan at an interest rate above par.

A. Distribution of Broker Fees

Table 3 displays the mean value of broker fees and several key underwriting factors by borrower-broker race/ethnicity across the three race subsamples. Broker fees vary considerably across broker and borrower race groups. Minorities pay more, on average, than white borrowers within a given broker race in each subsample. In panels A and C, there is also evidence that white borrowers pay higher fees when obtaining a loan through a minority broker. However, underwriting factors also differ across groups with large differences existing in the share of stated income loans and average annual income across groups, which suggests that borrowers and loan products may vary systematically with broker and borrower race.[17] Thus, unconditional mean fee differences may be uninformative. We address this more formally in Section III.B.

Figure 1 shows the kernel density of log broker fees by borrower-broker race/ethnicity for each subsample. For Hispanic borrowers, the distribution of fees sits to the right of white borrowers regardless of the race of the mortgage broker. In the BW subsample, the picture is less clear as the right tail of the black borrower fee distribution for both white and black brokers appears to have an additional mass. Finally, in the AW subsample, the minority fee distribution sits to the right of the white fee distribution when the broker is white, but not when the broker is Asian/Pacific Islander. The API broker distribution looks similar to the black broker fee distribution. Overall, the unconditional fee distributions in Figure 1 suggest that minority premiums exist, regardless of the broker's race or ethnicity.

[17]Since large difference exist in stated income share across groups, and these loans were often used to inflate income (Jiang, Nelson, and Vytlacil (2014) and Ambrose, Conklin, and Yoshida (2016)), in robustness checks we exclude stated income loans from the analysis. The main results of the paper are materially unaffected.

B. Empirical Model

Table 3 indicates that significant mean differences exist in underwriting factors across broker and borrower race combinations. These differences in borrower characteristics and loan products across groups could drive the observable variation in broker fees. Indeed, Bayer, Ferreira, and Ross (2017) note that a limitation of recent studies on mortgage pricing is that some of the key loan attributes associated with high cost loans are unobservable in standard data sets, making it impossible to determine whether demand for these product types explains the minority premiums. In contrast, our administrative data set contains all information collected by the lender at origination and the characteristics of the originated mortgage. This allows us to account for observable differences across borrowers, brokers, and product features. Thus, we test the impact of borrower's and broker's minority status on broker fees with the following ordinary least squares (OLS) regression:

$$P_{imt} = \delta_1 B_i^M + \delta_2 L_i^M + \delta_3 B_i^M \times L_i^M + X'_{imt}\beta + \tau_t + \kappa_m + \varepsilon_{imt} \tag{1}$$

where P_{imt} is the natural logarithm of broker fees paid by borrower i, in metropolitan statistical area (MSA) m, at time t. B_i^M is a dummy variable that equals one when the borrower is a minority, and zero otherwise. L_i^M is a dummy variable that equals one when the broker is a minority, and zero otherwise. X_{imt} denotes the matrix of control variables (described in Table 1), τ_t denotes origination year-quarter fixed effects, and κ_m denotes MSA fixed effects. The origination year-quarter fixed effects account for variation in broker fees that arise from temporal changes in the economic environment. The MSA fixed effects account for geographic-specific differences. The error term ε_{imt} is clustered at the MSA level.

We classify control variables into four broad categories: borrower, loan, property type, and area/geography. As noted above, these variables represent virtually all information collected at the time of origination thereby allowing the regression framework to estimate the effect of differences in borrower race or ethnicity holding constant all observable factors that might affect origination fees. Borrower controls include variables that describe demographic attributes (i.e., gender, age, and marital status), and underwriting risk factors (i.e., credit score, a subprime indicator if the FICO score is less than 620, income, debt-to-income, and employment status). Property type controls indicate whether the collateral is owner-occupied, a second home, an investment property, a condominium, a two-to-four unit multifamily, or a single-family residence. Loan controls include variables that describe features specific to the loan contract such as the loan purpose (i.e., purchase, refinance, or cash-out refinance), loan type (i.e., adjustable-rate, interest only, or fixed-rate), loan amount, combined loan-to-value ratio (CLTV), loan term, spread between the contract interest rate and the two-year Constant Maturity Treasury, prepayment penalty presence, stated-income documentation, and loan arrangement settings (i.e., co-borrower presence and face-to-face meeting). As in Haughwout, Mayer, and Tracy (2009), we allow the loan-to-value to affect the cost of credit non-linearly by using dummy variable bins.[18]

Finally, area controls include variables that influence the competitive setting and economic environment at the property location. This category includes the MSA/quarter Broker Herfindahl-Hirschman (HHI) index that acts as a proxy for market competition among brokers. Ambrose and Conklin (2014) show that broker HHI affects the costs of obtaining a mortgage. The area controls also include the Pahl index that provides a measure of mortgage broker regulations and occupational licensing requirements across states (Pahl, 2007). The effect of regulation on fees is ambiguous as increased monitoring of broker activities could decrease fees, while increased costs of broker compliance could increase fees. We include the share of college educated adults in the county to control for the effect observed by Woodward and Hall (2012) that borrower education (proxied by area education level) affects the cost of credit.[19] We capture variation in area wealth levels by including the per capita income at the zip code level, county level median income, and the county poverty share in the year of loan origination. We also include the share of the county

[18]Specifically, the loan-to-value ratio we use five CLTV categories: $CLTV < 80\%$, $80\% \leq CLTV < 85\%$, $85\% \leq CLTV < 90\%$, $90\% \leq CLTV \leq 95\%$, and $CLTV \geq 95\%$.

[19]The county share of college educated adults also controls for the broker's education level, which we do not observe directly.

adult population that is unmarried. To capture geographic differences in housing markets, we include the county rent to price ratio and the county share of housing that is owner-occupied. To control for the possibility that brokers and borrowers are operating within ethnic enclaves, we include additional county demographic controls measured as a fraction of county population: percent Hispanic, percent black, percent Asian or Pacific Islander, percent foreign born, only English speaking share, and Spanish speaking share.[20] Finally, we include the monthly MSA unemployment rate from the Bureau of Labor Statistics and the log distance in miles between the borrower's and broker's zip codes reported by New Century.

We estimate equation (1) separately for the HW, BW, and AW subsamples. The white borrowers that work with white brokers are the same in each subsample and serve as the reference group. The parameters δ_j, where $j \in \{1, 2, 3\}$, represent the coefficients of interest as they reveal whether minority premiums exists and to what extent they vary with broker race.

Since brokers had significant discretion over the fees negotiated on each loan, broker heterogeneity may explain the observed pricing differentials. For example, if minority borrowers select into "high-fee" white brokers, while white borrowers select into "low-fee" white brokers, then the observed differences may simply reflect that the two borrower groups use different mortgage brokers. To address this issue, we expand equation (1) to include individual mortgage broker fixed effects, α_k.[21]

The broker fixed effects models exploit within broker variation in borrower race to identify minority pricing premiums. The intent is to isolate variation in fees and borrower minority status from variation in unobserved broker attributes. These models are similar in spirit to Bayer, Ferreira, and Ross (2017) and Munnell et al. (1996), however, we control for potential unobserved heterogeneity at a more granular (individual) level. Our broker fixed effects models also closely approximate the identification strategy used in experimental paired-audit studies (e.g. Ayers and

[20] We thank an anonymous referee for this suggestion. The set of demographic control variables mirrors those included in Bayer, Ferreira, and Ross (2017).

[21] Note that the stand-alone broker race term (L^M) is absorbed by the broker fixed effects when α_k is included in the regression model.

Siegelman, 1995). By including broker fixed effects along with a rich set of control variables, we ask whether a minority borrower pays more than a comparable white borrower when obtaining a loan from the same mortgage broker. Additionally, we observe whether within broker minority premiums vary across broker race.[22]

C. Are Mortgage Brokers Compensated for Borrower Credit Risk?

As mentioned above, standard theories of pricing suggest that mortgage broker compensation should not vary systematically with borrower credit risk because brokers do not bear default risk on the loans they originate. We provide empirical evidence in support of this hypothesis by comparing the estimated coefficients from a linear probability model of mortgage default with the estimated coefficients on the log fee model.[23] If broker compensation is directly related to risk, then we would expect the coefficient estimates to follow the same pattern. Figure 2 shows the coefficient estimates (with 95% confidence intervals) for a set of risk characteristics that are commonly used in the mortgage default literature.[24] In the right panels, we plot the corresponding coefficient estimates from the broker fee model.[25] No clear relationship between default risk and broker compensation emerges. For example, although high CLTV loans (>95%) increase the likelihood of default, they are not associated with greater broker compensation.[26] FICO score is inversely related to default,

[22]Exploiting within-broker variation comes at a cost, however, as many individual brokers in the sample originate only a few loans. For example, approximately 60 percent of the unique loan officers originated only one loan. Forty three percent of the white loan officers originated loans to both minority and white borrowers while 36 percent of the Hispanic, black, and API loan officers originated loans to both white and minority borrowers. Thus, identification in the broker fixed effects regression relies on variation in fees and minority status within the subset of brokers that originated loans to both minorities and whites. We note that over 50% of the mortgages in the sample are originated by brokers that meet this criteria.

[23]For the default model, the dependent variable takes a value of one if the mortgage becomes 60 or more days delinquent within two years of origination and zero otherwise, and the control variables are those from equation (1).

[24]In the interest of brevity, we do not report coefficient estimates for all controls used in the regression. Tabulated results are available upon request and we note that the results are consistent with the extant literature.

[25]The full set of coefficient estimates are available upon request. The default regression model and fee model used to create Figure (2) include the entire sample (HW, BW, and AW). For ease of interpretation, we use credit score bins. However, our primary results do not use credit score bins, but rather let credit (FICO) score and a subprime indicator (FICO<620) enter directly into the model. The results are materially unaffected by this change.

[26]Broker compensation may also be inversely related to risk. For example, lenders may offer greater yield spread premiums to brokers on lower risk loans. If this is the case, then the coefficient estimates in the right panels should mirror those in the left panels. Again, this is not borne out in the figure.

but the same relationship does not hold with respect to broker fees. Finally, although second homes and stated income loans increase the likelihood of default, broker compensation is actually lower on these loans. Taken together, Figure 2 provides compelling evidence that broker fees are not directly tied to credit risk, thus alleviating concerns that minority premiums, if they exist, are due to credit risk factors that are unobservable to the econometrician.

D. *Main Results*

We now turn to our main results. Table 4 contains the OLS estimates from equation (1). Columns (1) - (3) correspond to the Hispanic/white (HW) subsample. In addition to the borrower and broker race variables, we include the natural logarithm of loan amount as a control in all models. Column (1) shows that Hispanic borrowers that obtain a loan through a white broker pay 9% more than white borrowers that also use a white broker.[27] In dollar terms, this premium translates into an additional $500 in fees on the average loan, which is nearly identical to the Latino premium ($489) estimated by Woodward (2008) in a study of 7,560 Federal Housing Administration (FHA) insured loans originated in 2001 (Table 3a). Interestingly, the Hispanic premium exists even when the broker is Hispanic; a Hispanic borrower pays 11% more than a white borrower that receives a loan through a Hispanic broker.[28] The minority/minority premium is not significantly different from the minority premium with a white mortgage broker. In other words, Hispanic borrowers pay a significant premium relative to white borrowers regardless of the broker's race. Column (1) also shows that white borrowers pay a small premium (2%) when obtaining a loan through a Hispanic broker.

The results in column (1) focus on market-level disparities in the cost of mortgage credit.

[27]In a log-linear model with a dummy variable ($ln(y) = \alpha + \beta D + \epsilon$), the percentage increase in y when the dummy changes from zero to one is $100 \times (exp(\beta) - 1)$. However, when β is relatively small, as is the case in our study, the percentage change in y can be approximated by $100 \times \beta$. For ease of interpretation, we will use this approximation throughout the paper.

[28]The last two rows of Table 4 report the minority premium charged by minority brokers ("Minority/Minority Premium") along with the corresponding p-value from a test of the null hypothesis that the minority/minority premium is zero. The minority/minority premium is calculated as $\delta_1 + \delta_2 + \delta_3 - \delta_2$.

In column (2) we introduce mortgage broker fixed effects.[29] The large increase in the adjusted R-squared moving from column (1) to (2) is consistent with individual mortgage brokers having considerable discretion in pricing. After accounting for broker heterogeneity, Hispanic borrowers pay a 5% premium relative to white borrowers when obtaining a loan through the same white broker. The minority/minority premium is 6%, indicating that Hispanic borrowers pay 6% more than whites when they obtain a loan through the same Hispanic broker, providing evidence that a within broker minority premium exists regardless of the race of the loan officer.[30]

The inclusion of broker fixed effects significantly reduces the magnitude of the minority premium from 9% (11%) to 5% (6%) for white (Hispanic) brokers. The fact that broker fixed effects account for a large portion of the minority premium documented in column (1) suggests that Hispanic borrowers systematically select into high-fee brokers as hypothesized by Yezer, Phillips, and Trost (1994). This is closely related to, and consistent with, the findings of Bayer, Ferreira, and Ross (2017) that lender fixed effects reduce racial differences in the likelihood of receiving high-cost loans. Finally, in column (3) we add the full set of control variables and fixed effects. The magnitude of the minority premium declines slightly in column (3), however, it remains statistically and economically significant.

Columns (4)-(6) report the estimates using the black/white (BW) subsample. Column (4) shows that black borrowers that obtain a loan through a white mortgage broker pay 14% more, on average, than white borrowers that work with a white broker. This translates into a $785 premium, which is statistically and economically reduced when using a black broker. The minority/minority premium is 6% when the mortgage broker is black. White borrowers obtaining a loan through a black broker pay 6% more than white borrowers working with a white broker.

[29]We use the reghdfe package (Correia, 2014, 2016) to estimate the broker fixed effects models in Stata. This package iteratively eliminates singleton groups (e.g., loans by brokers which originated only one loan), which explains the reduction in sample size. To test for differences in coefficient estimates across models, we follow the procedure outlined in the regdhfe Stata help file. The minority premium coefficients across columns (1) and (2) are statistically significantly different from one another.

[30]Broker race drops from the model in the broker fixed effects specifications. Thus, the minority/minority premium is calculated as $\delta_1 + \delta_3$ in these models.

Consistent with the results in the HW subsample, the inclusion of broker fixed effects in column (5) significantly reduces the magnitude of the minority premium. After accounting for broker heterogeneity, black borrowers pay an 8% premium relative to white borrowers when obtaining a mortgage through the same white broker. A key departure from the HW results, however, is that we find no evidence that the same black broker treats white and black borrowers differently. To see this, notice that the minority/minority premium is economically small (2%) and not significantly different from zero in column (5).

Although the magnitude of the minority borrower coefficient declines from column (5) to column (6), it is still economically and statistically significant.[31] A black borrower pays 5% more than a comparable white borrower when obtaining a loan through the same white broker, which translates to a $281 premium. This estimate is somewhat smaller than the African American premium of $563 reported in Woodward (2008). Note that the minority/minority premium is estimated as zero with a p-value of 0.79. Thus, we find no evidence that the same black broker treats a black borrower differently from a comparable white borrower.

We repeat the analysis in columns (7)-(9) using the Asian/white (AW) subsample. Here again, we see in column (7) that minorities pay a premium (4%) relative to white borrowers when obtaining a loan through a white broker. However, the minority/minority premium results are quite different in the AW subsample relative to the HW and BW subsamples. The minority/minority premium of -10% indicates that Asian borrowers actually pay less than their white counterparts when obtaining a loan through an Asian broker. Consistent with the HW and BW results, white borrowers on average pay more (7%) to obtain a loan through an Asian broker. Column (8) shows that Asian borrowers still pay a premium of 3% relative to white borrowers who receive a loan through the same white broker. Also, we still see evidence that the same Asian broker treats Asian borrowers differently from white borrowers (minority/minority premium of -5%). A similar pattern appears in the saturated regression model in column (9), however, the minority/minority premium of -4%

[31]The coefficients (8% versus 5%) are statistically significantly different from each other.

is slightly outside traditional statistical significance thresholds.

To summarize, we first note that minority borrowers pay a premium relative to white borrowers when they obtain loans through white brokers. This holds even after controlling for an extensive set of control variables and broker fixed effects, suggesting the minority borrowers receive different treatment than comparable white borrowers when obtaining loans through the same white broker. Second, a significant portion of the racial mortgage pricing disparity is explained by broker fixed effects, which suggests that minorities tend to systematically select into high-fee brokers. Third, there is some evidence that white borrowers pay more on average when obtaining a loan through a minority broker. Fourth, minority/minority premiums vary across minority groups with Hispanic brokers charging a premium to Hispanic borrowers, while black brokers do not appear to treat white and black borrowers differently. For Asian brokers, there is evidence that Asian borrowers receive more favorable treatment relative to white borrowers.

Finally, with respect to whether the statistically significant premiums are economically meaningful, we note that the magnitude of the observed premium paid by minority borrowers is above the threshold cutoff established in recent consent decrees agreed to by various financial institutions accused of disparate treatment of minorities in residential mortgage origination fees (New York, Office of the Attorney General, Civil Rights Bureau, 2007, 2006, 2008).[32] For example, using estimates from our saturated regression models, along with the 20 basis point threshold cutoff agreed to by HCI Mortgage in its settlement with the New York Attorney General for determining incidences of disparate treatment, we identified 9,290 minority borrowers that paid premiums (conditional) above the level that indicates potential disparate treatment.[33] Thus, although one might argue

[32]For example, the consent decree between the New York Attorney General and GreenPoint Financial (New York, Office of the Attorney General, Civil Rights Bureau, 2007) establishes a 25 basis point threshold in fee disparity that would require GreenPoint to "implement appropriate remedial measures to minimize the potential for future pricing disparities by the Broker, including mandatory fair lending training and oral and/or written counseling (p. 6, Section 5.2(a))." In conducting the analysis, the consent decree stipulates that GreenPoint may only control for "race-and-ethnicity-neutral factors" such as credit score, loan product type characteristics, and property type and location, which is similar to our regression specification.

[33]See New York, Office of the Attorney General, Civil Rights Bureau (2008). Our analysis indicates that 2.8%, 24.1%, and 0.2% of Hispanic, black, and Asian borrowers, respectively, paid premiums above the cutoff used to indicate possible disparate treatment.

that the minority premiums in our study are small in magnitude, they are clearly within the range that drew the attention of regulators in recent years.

E. Robustness Checks

We perform a number of additional empirical exercises to confirm that our primary results are robust to different specifications and methodologies. We briefly describe these tests here, with more detailed discussion in the online appendix. First, we examine whether our analysis is sensitive to the method used in inferring broker race. In the majority of our analysis, we infer the broker's race using the MAP BIFSG classfication scheme discussed in Sections II.B. However, we also used different classification schemes based on the BIFSG scores (see online appendix section A.2) and report the baseline regression results for different threshold classification schemes in Table A.7. Regardless of the classification threshold, results are similar in sign, significance, and magnitude to those reported in the saturated regression models of Table 4. We also used BIFSG scores directly, as opposed to a binary classification scheme, to estimate the effect of loan officer race on minority premiums. Again, this methodology produced estimates similar to those using the MAP BIFSG scheme (see Table A.8). Thus, we conclude that the results are not sensitive to the method or criteria used to infer broker race.

A second concern is whether the fee differential across borrower race reflects differences in broker effort required to generate a successful loan application. Brokers bear the risk that a loan application does not result in a funded loan (funding uncertainty). Generally, brokers are only compensated on applications that result in funded loans. If funding uncertainty co-moves with race and broker fees, then the coefficient estimates in the previous section may be biased. To address this concern, we report the results for the estimation of a Heckman model that accounts for funding uncertainty in online appendix section A.3 and find no evidence that the minority pricing premiums are driven by funding uncertainty.

A third concern is that observed pricing differentials reflect borrower contract selection rather than disparate treatment. To fix ideas, suppose a white broker offers two distinct contracts to every applicant, regardless of borrower race: i) a high front-end fee/low rate contract and ii) a low front-end fee/high rate contract. Also assume, consistent with existing evidence, that the broker earns greater revenue on the latter (Woodward and Hall, 2012; Ambrose and Conklin, 2014). If minority borrowers tend to select the contract that generates greater revenue for the broker, while white borrowers select into the lower revenue contract, then we would observe minority fee premiums even though there is no differential treatment by the mortgage broker.[34] We investigate this possibility by examining the tradeoff between front-end and back-end fees across borrower and broker race. The results, discussed in detail in online appendix section A.4, point to differential treatment by brokers rather than borrower contract selection.

In online appendix section A.5, we formally examine the borrower's choice of mortgage broker race to ensure that pricing disparities are not driven by borrower selection into broker race. Results using a propensity score matching technique – reported in Tables A.10 and A.11 – are consistent with our primary results presented in Table 4.

Finally, we examine whether the premiums uncovered in our main analysis are driven by language differences (see online appendix section A.6). For example, perhaps Hispanic brokers charge a premium for providing bilingual services. Although we do not observe the language in which the mortgage application was taken, we do observe whether the borrower is a US citizen, and use this information as a proxy for the use of a foreign language. We exclude non-US citizens in online appendix Table A.12 and find that the results reported in Table 4 are unchanged.

F. Does credit risk explain minority premiums?

In Section III.C, we provided evidence that broker compensation is not directly tied to standard measures of credit risk. However, the potential remains that minority pricing premiums reflect broker compensation for credit risk. To investigate this possibility, we examine whether minority

[34] We thank Aurel Hizmo for pointing out this possibility in his discussion of our paper at the 2020 Allied Social Science Association meeting.

pricing premiums vary with an observable measure of credit quality (FICO score), after conditioning on the rich set of control variables and fixed effects used in our saturated regression models. We interact borrower minority status with credit score quartiles to see if the minority pricing premium varies across credit scores. Figure 3 shows the estimated minority fee premiums (and 95% confidence intervals) across credit quartiles. Each panel represents a separate regression.[35] For example, the regression used to create the top left graph includes loans originated to white and Hispanic borrowers by white brokers. The top right panel, on the other hand, includes loans originated to Hispanic and white borrowers by Hispanic brokers.

A clear pattern emerges across the panels in Figure 3. The minority pricing premium is generally larger at higher credit scores. For example, in the top left panel, Hispanics with low credit scores (Bin 1) pay a 2% premium relative to comparable white borrowers with low credit scores. However, this minority premium increases significantly to 7% when comparing high credit score Hispanics to comparable high credit score whites (Bin 4). This pattern holds in five out of the six panels in Figure 3.[36]

If minority fee premiums reflect additional credit risk, then we should observe higher default rates on loans to minorities. Additionally, given the positive relationship between credit score and minority premium documented in Figure 3, we would expect that the difference in default rates between minorities and whites would also be positively related to credit score. To test these two predictions, we estimate default regression models with the same controls as the fee regressions in Figure 3. We plot the marginal effect of minority status on the likelihood of default across credit score quartiles in Figure 4.[37] Across all panels, there is no evidence that minorities are more likely to default at any point in the credit score distribution. Thus, borrower credit risk does not explain the minority pricing premiums or the positive relationship between these premiums and credit

[35] Coefficient estimates from these models are reported in tabular form in online appendix Table A.13.

[36] Note that the minority premium estimates across credit score bins are not always statistically distinguishable from one another. However, in all graphs the minority premium point estimate is largest for the high credit score bin.

[37] Coefficient estimates for from these models are reported in tabular form in appendix Table A.14

scores. Ultimately, we are left with a puzzle calling for future research on why minority premiums are generally positively correlated with borrower credit quality.

IV. Policy Implications

In the wake of the 2007-2008 financial crisis, regulators and lawmakers focused on curbing perceived abuses in mortgage lending, with mortgage brokers garnering significant attention. Although policymakers recognized the importance of brokers in helping consumers choose loans, they held concerns that the way brokers were being paid motivated them to steer borrowers into risky and expensive loan products (Consumer Financial Protection Bureau, 2014a). In response to these concerns, the Board of Governors of the Federal Reserve System (the Board) proposed rules in 2009 governing mortgage broker compensation that were later incorporated into the Dodd-Frank Act. The Consumer Financial Protection Bureau (CFPB) subsequently issued regulations (effective January 1, 2014) reconciling the Board's broker compensation rules (Kider and Kamensky, 2015). Thus, in this section, we analyze how the new broker compensation rules could have impacted broker fees and borrower access to credit.

A. Dual Compensation

In the run-up to the financial crisis, mortgage brokers could be compensated through either direct fees from the borrower, rebates (YSP) from the lender, or a combination of the two (dual compensation). In response to the crisis, regulators proposed a rule that would ban dual compensation.[38] By eliminating (or restricting) dual compensation, policy makers aimed to increase transparency in the loan origination process. Supporters of this restriction argue that dual compensation leads to borrower confusion and suboptimal shopping behavior. Indeed, two recent studies document that borrowers pay significantly higher fees on dual compensation loans (Woodward and

[38]The Board's 2010 Loan Originator Final Rule, which amended Regulation Z of the Truth-in-Lending Act (TILA), prohibits dual compensation (Consumer Financial Protection Bureau, 2012). After the CFPB inherited responsibility for the Regulation Z, the rule was republished at 12 CFR 1026.36(d) (Consumer Financial Protection Bureau, 2012).

Hall, 2012; Ambrose and Conklin, 2014). Proponents of dual compensation, on the other hand, argue that it provides valuable flexibility for consumers by allowing borrowers to choose lower out-of-pocket fees in exchange for a higher interest rate. At this time, dual compensation is not prohibited, but the CFPB has indicated an interest in further investigating dual compensation to determine how it affects borrower confusion and ultimately mortgage choice.[39] Thus, the results of this investigation will aid the CFPB in determining whether it should proceed with its initial proposal to ban dual compensation.

We re-estimate our baseline regression with two subsamples that exclude dual compensation within each set of HW, BW, and AW loans: (i) loans where brokers were compensated entirely through up-front fees, and (ii) loans where brokers received all compensation from the lender (yield spread premiums). We report the results using up-front fees as the dependent variable in columns (1) through (3) of Table 5 and columns (4) through (6) report the results using back-end fees as the dependent variable. Note that sample sizes (and power) are significantly reduced because dual compensation loans (excluded in this analysis) make up a large share (65%) of our overall sample. Under front-end fee only compensation schemes, Hispanic and black borrowers paid significantly higher premiums relative to their white counterparts (columns 1 and 2); Asian and Pacific Islander borrowers did not (column 3). With back-end only fees, black borrowers (column 5) paid a premium to obtain a loan while borrowers in other racial or ethnic groups did not. Taken together, the results suggest that the elimination of dual compensation to increase price transparency, per se, is unlikely to completely eliminate racial price disparities.

B. Broker Costs, Fee Caps, and Credit Rationing

In the post-crisis period, a residential mortgage loan can be categorized as "qualified" or "non-qualified." Broadly speaking, the CFPB deems a loan as a Qualified Mortgage (QM) if it has

[39] Details available at the CFPB's website: `www.consumerfinance.gov/policy-compliance/rulemaking/final-rules/`.

features that make it affordable and safe to the typical borrower.[40] To acquire QM status, the lender must follow certain underwriting criteria and the loan must exclude prohibited contract features that are deemed risky (e.g., negative amortization, interest-only payments, balloon loans, term greater than 30 years).[41] Among the restrictions, loan originator points and fees are capped at 3 percent of the loan amount (see Section 1026.43(e) of Regulation Z).[42]

The fee caps are meant to make loans affordable and to reduce broker discretion in pricing. But, the caps may also have an unintended consequence of causing mortgage brokers to withdraw from providing services to loan applicants that require higher levels of effort or service.[43] To consider how fees vary by loan applicant, we model the broker's revenue (fees) as follows:

$$P_{ijmt} = k_{ijmt} + \pi_{ijmt}, \tag{2}$$

where P_{ijmt} is the total dollar amount of revenue generated from loan applicant i (including fees and yield spread premium), by broker j, in market m, at time t. k_{ijmt} is the broker's production cost for originating the loan and π_{ijmt} is the excess profit generated on the loan applicant. The

[40] See `www.consumerfinance.gov/ask-cfpb/what-is-a-qualified-mortgage-en-1789/` for details. A key incentive to originate QM loans is that lenders are afforded certain legal protections against borrower-initiated lawsuits (Bhutta and Ringo, 2015). Additionally, lenders (or sponsors) do not face risk retention requirements on securitizations of qualified residential mortgages (QRM), which have the same definition as QM loans (see 24 CFR Part 267 – available at `www.gpo.gov/fdsys/pkg/FR-2014-12-24/pdf/2014-29256.pdf`). As a result of the regulatory benefits afforded on QM loans, an overwhelming majority of newly issued mortgages are classified as QM. A recent survey conducted by the American Bankers Association suggests that 91 percent of the typical bank's mortgage originations were QM in 2016 (American Bankers Association, 2017).

[41] We focus only on regulations directly related to mortgage broker compensation; the potential impacts of other regulations affecting mortgage brokers are outside the scope of our analysis. However, we note that these regulations do cover some of the contract features that are prevalent in our data (e.g., interest only loans) and thus would preclude them from obtaining QM status.

[42] For loans less than $100,000 the fees can exceed 3 percent. The fee caps for these loans are:

- $60,000 to $100,000: $3,000.
- $20,000 to $60,000: 5 percent of the loan amount.
- $12,500 to $20,000: $1,000 or less.
- $12,500 or less: 8 percent of the loan amount.

[43] For example, a mortgage broker quoted in the New York Times complained, "I will now get paid the same amount to process a plain-vanilla loan as I will a complex loan of equal size that requires more work," while the director at the National Association of Mortgage Brokers expressed concerns that the new rules will drive small, independent brokerages out of business (Browning, 2011).

implicit assumption is that brokers charge borrowers fees to cover their production costs. Although there are multiple sources of costs (e.g., marketing, overhead, and so on), a large portion of this cost compensates the mortgage broker for time, effort and search costs. In a perfectly competitive market, π_{ijmt} would be driven to zero. However, the mortgage market is not perfectly competitive.[44] Thus, the model allows both cost and excess profit to vary across individuals (e.g., loan and borrower characteristics), brokers, markets, and time to address the heterogeneity in the provision of brokerage services.

Empirically, we estimate the production cost for each loan applicant ($\hat{k}_{ijmt}$) and then compare it to the fee cap imposed by Regulation Z. If the estimated production cost exceeds the fee cap, then we assume that the borrower would be credit rationed under current regulations because the broker cannot recover the associated costs. We follow an approach similar to Berndt, Hollifield, and Sandås (2017) by fitting a quantile regression model of broker fees:[45]

$$q_\alpha(Fees|\mathbf{\Gamma}) = \mathbf{\Gamma}'\beta_\alpha, \tag{3}$$

where α is the quantile of interest and $\mathbf{\Gamma}$ includes the conditioning variables from equation 1.[46] The selected value of α fits a regression line where $(1-\alpha)$ of the observations lie above the regression line. The predicted values from this regression provide an estimate of the minimum (conditional) fee required for the broker to originate a loan, which is an estimate of the loan production cost, $\hat{k}_{ijmt}$ (Liu, Laporte, and Ferguson, 2008). If this cost estimate exceeds the fee caps imposed on

[44] Mortgage markets contain a high degree of information asymmetry (Agarwal, Chang, and Yavas, 2012; Albertazzi et al., 2015; Adelino, Gerardi, and Willen, 2013; Keys et al., 2009). In the context of our paper, mortgage brokers, who participate in the market frequently, enjoy an informational advantage over borrowers that enter the market infrequently. This information asymmetry means that broker revenue may frequently deviate significantly from loan production costs.

[45] An alternative is a stochastic frontier model similar to the approach used in Woodward (2008), where k is symmetrically distributed around a mean and π is distributed non-negative with an asymmetric distribution. However, this approach is intractable in our context due to the large number of covariates and fixed effects that are likely to affect costs and profits.

[46] The Γ vector excludes property type controls since they do not directly relate to the loan production costs as well as the log loan amount since we express fees on the left-hand-side in nominal terms. Furthermore, to determine whether a loan would have been credit rationed, we divided the predicted broker compensation by the loan amount and compare the ratio to the fee caps. We suppress the subscripts for ease of interpretation.

QM loans, then we assume that the borrower would be unable to obtain a loan. Put differently, the borrower would be credit rationed under current regulations.

To generate a baseline of possible credit rationing, we compare the actual fees observed to the fee caps outlined above to determine which borrowers would be credit rationed under current regulations assuming that the actual fees equal loan production costs and that brokers earned zero excess profit. The zero profit condition is, of course, a heroic assumption, which we relax after reporting our baseline results. The top left panel in Figures A.3, A.4, and A.5 in the online appendix report the results for the HW, BW, and AW loans, respectively. We note that nearly half of Hispanic borrowers that worked with Hispanic brokers and about 44 percent of Hispanic borrowers that obtained a loan from a white broker would have been credit rationed. Meanwhile, around 40 percent of white borrowers would be credit rationed regardless of the broker's race. In addition, more than half of the black borrowers would have been rationed while Asian/Pacific Islander borrowers appear to be less likely to face rationing risk, especially Asians who obtain loans from Asian brokers. Overall, these results provide an upper bound on credit rationing created by the current regulations and suggest that racial credit rationing disparities would exist under a zero profit assumption.[47]

Next, we turn to our quantile regression results reported in Table 6. We present results for different values of α due to the mechanical relationship between the choice of α and the fraction of borrowers that are classified as credit rationed in our data; larger values of α shift the estimated cost function upwards. Columns (1), (2), and (3) provide the point estimates using the 10^{th}, 20^{th}, and 30^{th} quantiles, respectively. Panels A, B, and C, show the quantile estimates for the HW, BW, and AW subsamples. The minority status dummy variables and interaction term are strongly significant for each quantile regression. The estimates in column (1) of Panel A, for example, indicate that Hispanic borrowers pay \$305 more than whites at the 10^{th} quantile. At higher quantiles, the racial

[47]The final rule implementing Regulation Z adopted an amendment allowing the exclusion of up to two "bona fide" discount points from the points and fees calculation for qualified mortgages. However, this exclusion is unlikely to affect our analysis because true discount points – paid to the lender to buy the interest rate down – are virtually non-existent in our sample of primarily subprime hybrid ARMs.

price disparities are greater.

Using the point estimates from Table 6, we infer the cost (conditional) for each loan applicant, and the proportion of borrowers that would be at risk of credit rationing under current regulations. For example, Figure 5 reports the results for the 30^{th} quantile.[48] In general, the HW and BW subsamples show that Hispanic and black borrowers represent the groups most at risk of losing access to credit, with 20 percent of Hispanic and 29 percent of black borrowers who obtained financing from white brokers at risk of credit rationing. In contrast, 17 and 14 percent of white and Asian/Pacific Islander borrowers who obtain financing from white brokers would encounter credit rationing risk, respectively. Thus, our results indicate that racial disparities in credit rationing risk are likely to exist after imposition of the regulation. In particular, Hispanic and black borrowers – who account for 45 percent of the loans in our sample – encounter the highest risk of credit rationing regardless of how we estimate loan production costs for each loan applicant.

C. Subprime Resurgence

The qualified mortgage (QM) designation requires that loan origination fees adhere to the caps outlined in the previous section. The benefit of these caps is that they may reduce the scope for brokers to price discriminate, however, they also increase the likelihood of credit rationing. But, borrowers that would be credit rationed in the QM market may still be able to obtain mortgage credit in the non-QM market because non-QM loans are not subject to the same fee limitations. In the non-QM market, mortgage brokers retain considerable discretion over mortgage pricing. These mortgages are typically extended to borrowers with a blemished credit history. In other words, non-QM loans are subprime mortgages. But, because the term subprime carries such a negative stigma in the wake of the recent financial crisis, the industry has rebranded these loans as "nonprime" mortgages (Grind, 2017; Olick, 2018). Non-QM loans carry significantly higher interest rates and downpayment requirements relative to their QM counterparts, and thus, even

[48] Figures A.3 through A.5 in the online appendix report comparisons for the 10^{th}, 20^{th}, 30^{th} quantiles.

if high loan production costs do not fully eliminate access to credit, a steep penalty exists for obtaining a mortgage in the nonprime market.

Although the current mortgage lending environment is dominated by QM loans, non-QM loans are gaining market share. In the years following the financial crisis, subprime mortgage lending virtually disappeared. However, in 2017, $4.1 billion in securities backed by non-prime mortgages were issued and the first quarter of 2018 saw $1.3 billion in non-prime issuances – more than double the amount issued in the same quarter a year earlier (McLannahan and Rennison, 2018). Clearly there is growing demand for nonprime securities from secondary mortgage market investors.

Existing research shows that subprime lending – which focuses on borrowers with blemished credit histories – is concentrated in minority neighborhoods, and subprime mortgages are originated disproportionately to minority borrowers (Mayer and Pence, 2009; Faber, 2013; Calem, Gillen, and Wachter, 2004; Pennington-Cross, Yezer, and Nichols, 2000)). This is driven, at least in part, by the fact that the two largest minority groups (Hispanics and blacks) have lower credit scores than non-minorities, on average (Board of Governors of the Federal Reserve System, 2007). Moving forward, these differences in credit scores across racial groups suggest that nonprime lending will continue to focus disproportionately on minority borrowers. Given that nonprime lending (i) is more likely to be a source of credit for minorities, (ii) is gaining market share, and (iii) is not subject to QM fee caps that limit broker pricing discretion, the racial pricing disparities we identified in the pre-crisis period are likely to persist despite recent regulatory changes.[49]

D. *Summary*

Although the new rules on broker fees limit how mortgage brokers may collect fees on loan originations, potentially reducing pricing disparities, our analysis indicates that these rules alone are unlikely to eradicate minority premiums and instead may place borrowers at risk of credit rationing.

[49] In a contemporaneous paper, Bhutta and Hizmo (2019) provide evidence of post-regulation racial pricing disparities in a sample of Federal Housing Administration (FHA) loans originated in 2014. Note that although the Bhutta and Hizmo (2019) paper is similar in spirit to our own, a key difference is that we control for the minority status of the loan officer.

However, even if borrowers can avoid credit rationing by obtaining credit in the nonprime market, significant pricing disparities are likely to exist in that market because brokers have considerable pricing discretion on non-QM loans.

V. Conclusion

This paper uses a nationwide dataset of loans originated between 2003 and mid-2007 by over 124,000 unique mortgage brokers to test for differential treatment in financial contracts by examining both front- and back-end fees that borrowers pay at origination. Focusing on fees paid to mortgage brokers, rather than interest rates, helps us overcome the challenge of potentially omitted risk characteristics that plague previous studies. Our unique dataset also allows us to infer the broker's race, providing the opportunity to observe the race of both sides to the contract.

We find that minority pricing premiums exist when the mortgage broker is white after conditioning on an extensive set of borrower, loan, property, and area characteristics. Premiums are smaller, but remain significant, after including individual broker fixed effects, which indicates that a minority borrower pays more than a comparable white borrower when obtaining a loan from the same mortgage broker. The results also suggest that minorities tend to select into high-fee brokers. Importantly, we find that the premium a minority pays depends critically on the race of the mortgage broker. Hispanic brokers charge Hispanic borrowers a premium relative to comparable white borrowers, but observably similar white and black borrowers pay the same fees when obtaining a loan from the same black broker. We also find some evidence that Asian borrowers pay lower fees than comparable white borrowers when obtaining loans from Asian brokers.

The Dodd-Frank Act of 2010 enacted a number of regulations designed to curb perceived abuses in the mortgage industry. These regulations severely restrict broker discretion in setting mortgage origination fees. For example, the yield spread premium rebate can no longer be paid unless the borrower also reviews a similar loan without it, and mortgage brokers are limited in the ability to collect compensation on the basis of the loan terms other than the loan balance at origination.

We document a possible negative or adverse effect of these regulations in the form of potential credit rationing. Assuming that loan fees reflect broker production costs, we estimate that the restrictions on broker fees could result in a large percentage of minority borrowers being at risk of credit rationing. As a result, our study fills the need articulated by Campbell et al. (2011) for rigorous analysis of the effectiveness of regulatory interventions following the Great Recession and it demonstrates the trade-offs policy makers face when designing new regulations.

REFERENCES

Abrams, D. S., M. Bertrand, and S. Mullainathan. 2012. Do judges vary in their treatment of race? *Journal of Legal Studies* 41:347–84. ISSN 0047-2530. doi:10.1086/666006.

Adelino, M., K. Gerardi, and P. S. Willen. 2013. Why don't lenders renegotiate more home mortgages? redefaults, self-cures and securitization. *Journal of Monetary Economics* 60:835–53. ISSN 0304-3932. doi:10.1016/j.jmoneco.2013.08.002.

Agarwal, S., Y. Chang, and A. Yavas. 2012. Adverse selection in mortgage securitization. *Journal of Financial Economics* 105:640–60. ISSN 0304-405X. doi:10.1016/j.jfineco.2012.05.004.

Agarwal, S., H.-S. Choi, J. He, and T. F. Sing. 2019. Matching in housing markets: The role of ethnic social networks. *Review of Financial Studies* 32:3958–4004. ISSN 0893-9454. doi: {10.1093/rfs/hhz006}.

Albertazzi, U., G. Eramo, L. Gambacorta, and C. Salleo. 2015. Asymmetric information in securitization: An empirical assessment. *Journal of Monetary Economics* 71:33–49. ISSN 0304-3932. doi:10.1016/j.jmoneco.2014.11.002.

Ambrose, B. W., and J. N. Conklin. 2014. Mortgage brokers, origination fees, price transparency and competition. *Real Estate Economics* 42:363–421.

Ambrose, B. W., J. N. Conklin, and J. Yoshida. 2016. Credit rationing, income exaggeration, and adverse selection in the mortgage market. *Journal of Finance* 71:2637–86.

American Bankers Association. 2017. Aba survey: Regulatory burdend continues to restrict mortgage lending. Available at `https://www.aba.com/Press/Pages/033017RealEstateLendingSurvey.aspx`.

Angrist, J. D., and J.-S. Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion.* Princeton University Press.

Antonovics, K., and B. G. Knight. 2009. A new look at racial profiling: Evidence from the boston police department. *Review of Economics and Statistics* 91:163–77. ISSN 0034-6535. doi:10.1162/rest.91.1.163.

Anwar, S., P. Bayer, and R. Hjalmarsson. 2012. The impact of jury race in criminal trials. *Quarterly Journal of Economics* 127:1017–55. ISSN 0033-5533. doi:10.1093/qje/qjs014.

Anwar, S., and H. Fang. 2012. Testing for the role of prejudice in emergency departments using bounceback rates. *The BE Journal of Economic Analysis & Policy* 12. ISSN 1935-1682. doi:10.1515/1935-1682.3247.

Anwar, S., and H. M. Fang. 2006. An alternative test of racial prejudice in motor vehicle searches: Theory and evidence. *American Economic Review* 96:127–51. ISSN 0002-8282. doi:10.1257/000282806776157579.

Avery, R. B., K. P. Brevoort, and G. B. Canner. 2007. The 2006 HMDA data. *Fed. Res. Bull. A73* 93.

Avery, R. B., G. B. Canner, and R. E. Cook. 2005. New information reported under HMDA and its application in fair lending enforcement. *Fed. Res. Bull.* 91:344–.

Ayers, I., and P. Siegelman. 1995. Race and gender discrimination in bargaining for a new car. *American Economic Review* 85:304–21.

Bartlett, R., A. Morse, R. Stanton, and N. Wallace. 2018. Consumer-lending discrimination in the fintech era. *UC Berkeley Public Law Research Paper* Available at SSRN: https://ssrn.com/abstract=3063448.

Bayer, P., F. Ferreira, and S. L. Ross. 2016. The vulnerability of minority homeowners in the housing boom and bust. *American Economic Journal-Economic Policy* 8:1–27. ISSN 1945-7731. doi:{10.1257/pol.20140074}.

———. 2017. What drives racial and ethnic differences in high-cost mortgages? the role of high-risk lenders. *Review of Financial Studies* 31:175–205.

Bayer, P., R. McMillan, and K. Rueben. 2004. What drives racial segregation? New evidence using Census microdata. *Journal of Urban Economics* 56:514–535. ISSN 0094-1190. doi:{10.1016/j.jue.2004.06.002}.

Bayer, P. J., M. D. Casey, F. V. Ferreira, and R. McMillan. 2012. Price discrimination in the housing market. *Economic Research Initiatives at Duke (ERID) Working Paper No. 127* .

Berkovec, J. A., G. B. Canner, S. A. Gabriel, and T. H. Hannan. 1994. Race, redlining, and residential mortgage loan performance. *Journal of Real Estate Finance and Economics* 9:263–94. ISSN 0895-5638. doi:10.1007/BF01099279.

———. 1998. Discrimination, competition, and loan performance in fha mortgage lending. *Review Of Economics And Statistics* 80:241–50. ISSN 0034-6535. doi:10.1162/003465398557483.

Berndt, A., B. Hollifield, and P. Sandås. 2017. What broker charges reveal about mortgage credit risk. *Sveriges Riksbank Working Paper Series No. 336.* Available at SSRN: `ssrn.com/abstract=2918688`.

Bertrand, M., E. Luttmer, and S. Mullainathan. 2000. Network effects and welfare cultures. *Quarterly Journal of Economics* 115:1019–1055. ISSN 0033-5533. doi:{10.1162/003355300554971}.

Bhutta, N., and A. Hizmo. 2019. Do minorities pay more for mortgages. *Available at SSRN: https://ssrn.com/abstract=3352876 or http://dx.doi.org/10.2139/ssrn.3352876* .

Bhutta, N., and D. Ringo. 2015. Effects of the ability to repay and qualified mortgage rules on the mortgage market. *FEDS Notes* .

Black, H. A., T. P. Boehm, and R. P. DeGennaro. 2003. Is there discrimination in mortgage pricing? the case of overages. *Journal of Banking & Finance* 27:1139–65.

Board of Governors of the Federal Reserve System. 2007. Report to the congress on credit scoring and its effects on the availability and affordability of credit.

Boehm, T. P., and A. Schlottmann. 2007. Mortgage pricing differentials across hispanic, african-american, and white households: Evidence from the american housing survey. *Cityscape* 93–136.

Bohren, J. A., A. Imas, and M. Rosenberg. 2019. The dynamics of discrimination: Theory and evidence. *American Economic Review* 109:3395–436.

Browning, L. 2011. New fed rule for mortgage brokers. *The New York Times* Available at `https://nyti.ms/2jKQm7M`.

Brueckner, J. K. 1996. Default rates and mortgage discrimination: A view of the controversy. *Cityscape: A Journal of Policy Development and Research* 2:65–8.

Calem, P. S., K. Gillen, and S. Wachter. 2004. The neighborhood distribution of subprime mortgage lending. *Journal of Real Estate Finance and Economics* 29:393–410.

Campbell, J. Y., H. E. Jackson, B. C. Madrian, and P. Tufano. 2011. Consumer financial protection. *Journal of Economic Perspectives* 25:91–113. ISSN 0895-3309. doi:10.1257/jep.25.1.91.

Cheng, P., Z. Lin, and Y. Liu. 2015. Racial discrepancy in mortgage interest rates. *Journal of Real Estate Finance and Economics* 51:101–20. ISSN 1573-045X. doi:10.1007/s11146-014-9473-0.

Civil Serv. Com. v. Guardians Ass'n. 1983. *463 U.S. 1228, 103 S. Ct. 3568* .

Conklin, J. N. 2017. Financial literacy, broker–borrower interaction and mortgage default. *Real Estate Economics* 2:376–414.

Consumer Financial Protection Bureau. 2012. Loan originator compensation requirements under the truth in lending act. Available at `files.consumerfinance.gov/f/201301_cfpb_final-rule_loan-originator-compensation.pdf`.

———. 2014a. 2013 loan originator rule: Small entity compliance guide. Available at `files.consumerfinance.gov/f/201401_cfpb_complaince-guide_loan-originator.pdf`.

———. 2014b. Using publicly available information to proxy for unidentified race and ethnicity: A methodology and assessment. Available at `files.consumerfinance.gov/f/201409_cfpb_report_proxy-methodology.pdf`.

———. 2015. Consumer's mortgage shopping experience: A first look at results from the national survey of mortgage borrowers. Available at `files.consumerfinance.gov/f/201501_cfpb_consumers-mortgage-shopping-experience.pdf`.

Correia, S. 2014. REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects. Statistical Software Components, Boston College Department of Economics.

———. 2016. Estimating Multi-Way Fixed Effect Models with reghdfe.

Courchane, M., and D. Nickerson. 1997. Discrimination resulting from overage practices. In *A Special Issue of the Journal of Financial Services Research: Discrimination in Financial Services*, 133–51. Springer.

Courchane, M. J., and S. L. Ross. 2019. Evidence and actions on mortgage market disparities: Research, fair lending enforcement, and consumer protection. *Housing Policy Debate* 29:769–94. doi:10.1080/10511482.2018.1524446.

Crawford, G. W., and E. Rosenblatt. 1999. Differences in the cost of mortgage credit implications for discrimination. *Journal of Real Estate Finance and Economics* 19:147–59.

Dell'Ariccia, G., D. Igan, and L. U. Laeven. 2012. Credit booms and lending standards: Evidence from the subprime mortgage market. *Journal of Money, Credit and Banking* 44:367–84.

Demyanyk, Y., and O. Van Hemert. 2009. Understanding the subprime mortgage crisis. *Review of Financial Studies* 24:1848–80.

Deng, Y., and S. Gabriel. 2006. Risk-based pricing and the enhancement of mortgage credit availability among underserved and higher credit-risk populations. *Journal of Money Credit and Banking* 38:1431–1460. ISSN 0022-2879. doi:{10.1353/mcb.2006.0079}.

Doleac, J. L., and L. C. Stein. 2013. The visible hand: Race and online market outcomes. *The Economic Journal* 123:F469–92. ISSN 1468-0297. doi:10.1111/ecoj.12082.

Domingos, P., and M. Pazzani. 1996. Beyond independence: Conditions for the optimality of the simple bayesian classifiers. In *Proc. 13th Intl. Conf. Machine Learning*, 105–12.

Edelman, B., M. Luca, and D. Svirsky. 2017. Racial discrimination in the sharing economy: Evidence from a field experiment. *American Economic Journal: Applied Economics* 9:1–22.

Elliott, M. N., P. A. Morrison, A. Fremont, D. F. McCaffrey, P. Pantoja, and N. Lurie. 2009. Using the census bureau's surname list to improve estimates of race/ethnicity and associated disparities. *Health Services and Outcomes Research Methodology* 9:69–.

Ewens, M., B. Tomlin, and L. C. Wang. 2014. Statistical discrimination or prejudice? a large sample field experiment. *The Review of Economics and Statistics* 96:119–34. doi:10.1162/REST_a_00365.

Faber, J. W. 2013. Racial dynamics of subprime mortgage lending at the peak. *Housing Policy Debate* 23:328–49. doi:10.1080/10511482.2013.771788.

Fadlon, Y. D. 2015. Statistical discrimination and the implication of employer-employee racial matches. *Journal of Labor Research* 36:232–48. ISSN 0195-3613. doi:10.1007/s12122-015-9203-2.

Fisman, R., D. Paravisini, and V. Vig. 2017. Cultural proximity and loan outcomes. *American Economic Review* 107:457–92.

Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther. 2018. Predictably unequal? the effects of machine learning on credit markets. *Working Paper* .

Ghent, A. C., R. Hernández-Murillo, and M. T. Owyang. 2014. Differences in subprime loan pricing across races and neighborhoods. *Regional Science and Urban Economics* 48:199–215.

Glover, D., A. Pallais, and W. Pariente. 2017. Discrimination as a self-fulfilling prophecy: Evidence from french grocery stores. *Quarterly Journal of Economics* 132:1219–60. ISSN 0033-5533. doi: 10.1093/qje/qjx006.

Grind, K. 2017. Does anyone remember how to make a subprime mortgage. *Wall Street Journal* Https://www.wsj.com/articles/does-anyone-remember-how-to-make-a-subprime-mortgage-1497259803.

Guardians Ass'n of N.Y.C. Police Dep't v. Civil Serv. Comm'n. 1977. *431 F. Supp. 526 (S.D.N.Y.)* .

Guardians Asso. of N.Y.C. Police Dep't, Inc. v Civil Serv. Com. 1980. *633 F.2d 232* .

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

Haughwout, A., C. Mayer, and J. Tracy. 2009. Subprime mortgage pricing: The impact of race, ethnicity, and gender on the cost of borrowing. *Brookings-Wharton Papers on Urban Affairs* 2009:33–63.

Heckman, J. J. 1979. Sample selection bias as a specification error. *Econometrica* 47:153–62.

———. 1990. Varieties of selection bias. *The American Economic Review* 80:313–8.

Hedegaard, M. S., and J.-R. Tyran. 2018. The price of prejudice. *American Economic Journal – Applied Economics* 10:40–63. ISSN 1945-7782. doi:10.1257/app.20150241.

Hjort, J. 2014. Ethnic divisions and production in firms. *Quarterly Journal of Economics* 129:1899–946. ISSN 0033-5533. doi:10.1093/qje/qju028.

Horne, D. 1997. Mortgage lending, race, and model specification. *Journal of Financial Services Research* 11:43–68.

Jiang, W., A. A. Nelson, and E. Vytlacil. 2014. Liar's loan? effects of origination channel and information falsification on mortgage delinquency. *Review of Economics and Statistics* 96:1–18.

Keys, B. J., T. Mukherjee, A. Seru, and V. Vig. 2009. Financial regulation and securitization: Evidence from subprime loans. *Journal of Monetary Economics* 56:700–20. ISSN 0304-3932. doi:10.1016/j.jmoneco.2009.04.005.

Kider, M., and F. Kamensky. 2015. Subtitle A: Loan originator compensation restrictions and enforcement. *Mortgage Compliance Magazine* .

Ladd, H. 1998. Evidence on discrimination in mortgage lending. *Journal of Economic Perspectives* 12:41–62. ISSN 0895-3309.

Li, Q. 2014. Ethnic diversity and neighborhood house prices. *Regional Science and Urban Economics* 48:21–38. ISSN 0166-0462. doi:{10.1016/j.regsciurbeco.2014.04.007}.

Liu, C., A. Laporte, and B. S. Ferguson. 2008. The quantile regression approach to efficiency measurement: insights from monte carlo simulations. *Health economics* 17:1073–87.

Mayer, C. J., and K. Pence. 2008. Subprime mortgages: what, where, and to whom? Working Paper, National Bureau of Economic Research.

———. 2009. Subprime mortgages: What, where, and to whom? In E. Glaeser and J. Quigley, eds., *Housing markets and the economy: risk, regulation, and policy: essays in honor of Karl E. Case*. Cambridge, MA: Lincoln Land Institute.

McLannahan, B., and J. Rennison. 2018. Us subprime mortgage bonds back in fashion with $1.3bn of deals in first quarter of 2018. *Financial Times* .

Munnell, A. H., G. M. Tootell, L. E. Browne, and J. McEneaney. 1996. Mortgage lending in boston: Interpreting hmda data. *American Economic Review* 86:25–53.

Nadauld, T. D., and S. M. Sherlund. 2013. The impact of securitization on the expansion of subprime credit. *Journal of Financial Economics* 107:454–76.

New York, Office of the Attorney General, Civil Rights Bureau. 2006. Assurance of discontinuance pursuant to executive law 63(15) in the matter of Countrywide Home Loans, Inc. Available at `https://ag.ny.gov/sites/default/files/press-releases/archived/Countrywide%20Assurance%20Final%20Signed%20PDF.pdf`.

———. 2007. Assurance of discontinuance pursuant to executive law 63(15) in the matter of: Greenpoint Mortgage Funding, Inc. Available at `https://ag.ny.gov/sites/default/files/pdfs/bureaus/civil_rights/Greenpoint%20AOD%20Final%20PDF.pdf`.

———. 2008. Assurance of discontinuance pursuant to executive law 63(15) aod no. 08-160. in the matter of Home Consultants, Inc. D/B/A HCI Mortgage. Available at `https://ag.ny.gov/sites/default/files/pdfs/bureaus/civil_rights/HCI%20Settlement%20Agreement%201-5-08.pdf`.

Olick, D. 2018. Subprime mortgages make a comeback – with a new name and soaring demand. *CNBC* Https://www.cnbc.com/2018/04/12/sub-prime-mortgages-morph-into-non-prime-loans-and-demand-soars.html.

Pahl, C. J. 2007. A compilation of state mortgage broker laws and regulations, 1996-2006. Working Paper, Federal Reserve Bank of Minneapolis.

Pennington-Cross, A., A. Yezer, and J. Nichols. 2000. Credit risk and mortgage lending: Who uses subprime and why?

Piskorski, T., A. Seru, and J. Witkin. 2015. Asset quality misrepresentation by financial intermediaries: evidence from the rmbs market. *Journal of Finance* 70:2635–78.

Puhani, P. 2000. The heckman correction for sample selection and its critique. *Journal of economic surveys* 14:53–68.

Rosenbaum, P. R., and D. B. Rubin. 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* 39:33–8.

Ross, S. L. 1996. Flaws in the use of loan defaults to test for mortgage lending discrimination. *Cityscape: A Journal of Policy Development and Research* 2:41–8.

Ross, S. L., and J. Yinger. 2002. The color of credit: Mortgage discrimination, research methodology, and fair-lending enforcement. *MIT Press Books* 1.

Tzioumis, K. 2018. Data for: Demographic aspects of first names. doi:10.7910/DVN/TYJKEZ.

United States Dept. of Justice. 2010. Consent order, United States v. PrimeLending. Available at `https://www.justice.gov/sites/default/files/crt/legacy/2011/02/08/primelendsettle.pdf`.

———. 2011. United States of America v. Countrywide Financial Corporation, U.S. District Court Central District of California. Available at `https://www.justice.gov/sites/default/files/crt/legacy/2011/12/21/countrywidecomp.pdf`.

United States v. Brown. 2018. *U.S. Dist. LEXIS 40326, 2018 WL 1278577 (N.D. Ill. March 12, 2018)* .

Voicu, I. 2018. Using first name information to improve race and ethnicity classification. *Statistics and Public Policy* 5:1–13. doi:10.1080/2330443X.2018.1427012.

Wong, M. 2013. Estimating ethnic preferences using ethnic housing quotas in Singapore. *Review of Economic Studies* 80:1178–1214. ISSN 0034-6527. doi:{10.1093/restud/rdt002}.

Woodward, S. E. 2008. A study of closing costs for fha mortgages. *Prepared for the U.S. Department of Housing and Urban Development. Washington, D.C., The Urban Institute* .

Woodward, S. E., and R. E. Hall. 2010. Consumer confusion in the mortgage market: Evidence of less than a perfectly transparent and competitive market. *The American Economic Review* 100:511–5.

———. 2012. Diagnosing consumer confusion and sub-optimal shopping effort: Theory and mortgage-market evidence. *The American Economic Review* 102:3249–76.

Word, D. L., C. D. Coleman, R. Nunziata, and R. Kominski. 2008. Demographic aspects of surnames from census 2000. *Unpublished manuscript* Available at `citeseerx.ist.psu.edu/viewdoc/download`.

Yezer, A. 2017. Commentary: What can we learn from government attempts to modify the allocation of mortgage and consumer credit in the united states. *Cityscape: A Journal of Policy Development and Research* 19:199–205.

Yezer, A., R. Phillips, and R. Trost. 1994. Bias in estimates of discrimination and default in mortgage lending: The effects of simultaneity and self-selection. *Federal Reserve Bank of Philadelphia, Proceedings* 197–222.

Yezer, A. M. 2006. Discrimination in mortgage lending. *A Companion to Urban Economics* 197–210.

Yinger, J. 1996. Why default rates cannot shed light on mortgage discrimination. *Cityscape: A Journal of Policy Development and Research* 2:25–32.

Zhang, J., and L. Zheng. 2015. Are people willing to pay for less segregation? Evidence from US internal migration. *Regional Science and Urban Economics* 53:97–112. ISSN 0166-0462. doi:{10.1016/j.regsciurbeco.2015.05.002}.

Zuehlke, T. W., and A. R. Zeman. 1991. A comparison of two-stage estimators of censored regression models. *The Review of Economics and Statistics* 185–8.

Zussman, A. 2013. Ethnic discrimination: Lessons from the israeli online market for used cars. *Economic Journal* 123:F433–68. ISSN 0013-0133. doi:10.1111/ecoj.12059.

Table 1. Variables and Definitions

Variables	Definitions
Dependent Variable	
Log broker fees	Natural logarithm of one plus the yield spread premium rebate, correspondence premium, application fee, underwriter fee, brokerage firm fee and points.
Borrower Controls	
Log age	Natural logarithm of the primary borrower's age in years.
Gender	An indicator equal to one if the primary borrower is female.
Gender unknown	An indicator equal to one if the primary borrower is unknown.
Marital status	An indicator equal to one if the primary borrower is not married.
Marital status unknown	An indicator equal to one if the marital status of the primary borrower is unknown.
Log combined income	Natural logarithm of the combined monthly income of all borrowers on the loan.
FICO score	Primary borrower's FICO credit score.
Subprime	An indicator equal to one if the FICO score is below 620.
Debt-to-income ratio	Borrower's debt-to-income ratio at origination.
Self-employed	An indicator equal to one if the primary borrower is self-employed.
Property Type	
Owner-occupied[a]	An indicator equal to one if the subject property is owner-occupied.
Investment property	An indicator equal to one if the subject property is an investment property.
Second home	An indicator equal to one if the subject property is a second home.
Single-family[a]	An indicator equal to one if the property is a single-family residence.
2-4 Unit	An indicator equal to one if the property is a 2-4 unit property.
Condominium	An indicator equal to one if the property is a condominium.
Loan Controls	
Face	An indicator equal to one if the borrower and broker meet face-to-face.
Purchase[a]	An indicator equal to one if the mortgage is for a home purchase.
ARM	An indicator equal to one if the mortgage is an adjustable rate mortgage.
Cash	An indicator equal to one if the mortgage is a cash-out refinance loan.
Co-borrower	An indicator equal to one if there is a co-borrower on the loan.
CLTV below 80%[a]	An indicator equal to one if the combined loan to value ratio is in [0%,80%).
CLTV between 80 and 85%	An indicator equal to one if the combined loan to value ratio is in [80%,85%).
CLTV between 85 and 90%	An indicator equal to one if the combined loan to value ratio is in [85%,90%).
CLTV between 90 and 95%	An indicator equal to one if the combined loan to value ratio is in [90%,95%).
CLTV greater than 95%	An indicator equal to one if the combined loan to value ratio is in [95%,100%).
Interest-only	An indicator equalt to one if the loan payments are interest only.
Log loan amount	Natural logarithm of the loan amount.
Log loan term	Natural logarithm of the loan term in years.
Spread	Contract rate minus the 2-year constant maturity treasury at origination.
Prepay	An indicator equal to one if the loan contains a prepayment penalty.
Stated Income	An indicator equal to one if the loan is not a full income documentation loan.
Area Controls	
Log distance	Log distance in miles between property location and broker's zip code.
Broker HHI	MSA/year level Herfindah-Herschman index of mortgage broker competition.
MSA unemployment	MSA unemployment rate in the month of loan origination.
Pahl-index	State level Pahl-index of mortgage broker regulations.
Zip per capita income	Per capita income of the the zip code in the year of origination.
Household income	Median county household income in 2000.
Poverty share	Percentage of county households in 2000 who have income below poverty level.
Rent-Price ratio	County-level rent-to-price ratio times 200 in 2000.
College educated	Percentage of county residents in 2000 that had a bachelors degree or higher.
Occupancy share	Percentage of county housing units in 2000 that are owner occupied.
Single share	Percentage of county residents in 2000 that were never married and at least 15 years old.
African American share	percentage of county residents in 2000 that self-identify as African American (non-Hispanic).
Hispanic share	Percentage of county residents in 2000 that self-identify as Hispanic.
Asian/Pacific Islander share	Percentage of county residents in 2000 that self-identify as Asian or Pacific Islander.
Foreign share	Percentage of county residents in 2000 that were born abroad.
English share	Percentage of county households in 2000 who speak only English at home.
Spanish share	Percentage of county households in 2000 who speak only Spanish at home.

[a] the base for the corresponding categorical variable.

Table 2. Unique Brokers by Race

Loans per broker	White	Hispanic	African American	Asian/Pacific Islander
1	53,941	11,946	5,682	3,130
2	15,057	3,294	1,486	719
3	6,959	1,580	740	368
4	3,888	870	391	181
5	2,333	582	255	124
6	1,687	424	176	74
7	1,163	332	132	58
8	900	217	108	39
9	698	180	60	26
10+	3,409	979	394	154
Total	90,035	20,404	9,424	4,873
Sample Share	72%	16%	8%	4%

This table reports the number of unique brokers by race and the number of loan originations they arranged in our sample.

Table 3. Summary Statistics of Broker Fees and Underwriting Factors

Panel A: HW	White Broker		Hispanic Broker	
	White Borrower	Hispanic Borrower	White Borrower	Hispanic Borrower
Broker Fees	5,116	6,184	5,796	6,435
Stated Income	36	49	41	59
Debt-to-income	39	41	40	41
CLTV	85	86	84	86
Credit Score	616	624	624	636
Annual Income	82,305	80,979	89,019	80,365
Age	42	40	43	40
Obs	142,539	33,415	11,117	47,342

Panel B: BW	White Broker		Black Broker	
	White Borrower	Black Borrower	White Borrower	Black Borrower
Broker Fees	5,116	5,578	5,086	5,255
Stated Income	36	33	37	36
Debt-to-income	39	40	39	40
CLTV	85	86	86	87
Credit Score	616	602	614	610
Annual Income	82,305	71,549	78,741	70,767
Age	42	44	43	43
Obs	142,539	46,709	5,788	18,539

Panel C: AW	White Broker		API Broker	
	White Borrower	API Borrower	White Borrower	API Borrower
Broker Fees	5,116	6,619	6,777	7,106
Stated Income	36	51	42	59
Debt-to-income	39	41	40	41
CLTV	85	88	85	88
Credit Score	616	638	630	653
Annual Income	82,305	102,131	101,385	115,475
Age	42	41	43	41
Obs	142,539	7,394	4,017	6,986

Panel A reports the mean values of broker fees and underwriting factors of loans originated by white and Hispanic brokers for white and Hispanic borrowers. Panel B reports the mean values of broker fees and underwriting factors of loans originated by white and black brokers for white and black borrowers. Panel C reports the mean values of broker fees and underwriting factors of loans originated by white and API brokers for white and API borrowers. The variable combined loan-to-value is the nominal combined loan amount to collateral value ratio. The variable credit score is the borrower's nominal FICO score. The definitions of the other variables are available in Table 1.

Table 4. OLS Regressions of ln(Broker Fees)

Dep. Var.: ln(Broker Fees)	(1) HW	(2) HW	(3) HW	(4) BW	(5) BW	(6) BW	(7) AW	(8) AW	(9) AW
Minority Borrower	0.09***	0.05***	0.04***	0.14***	0.08***	0.05***	0.04***	0.03***	0.03***
	(0.01)	(0.01)	(0.00)	(0.02)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Minority Broker	0.02*			0.06***			0.07***		
	(0.01)			(0.01)			(0.02)		
Minority Borrower ×	0.01	0.00	0.01	-0.08***	-0.06***	-0.05***	-0.14***	-0.08***	-0.08**
Minority Broker	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Observations	234,413	175,670	175,660	213,575	158,543	158,535	160,936	114,303	114,293
Adjusted R-squared	0.32	0.51	0.56	0.33	0.54	0.58	0.30	0.51	0.56
Log Loan Amount	Y	Y	Y	Y	Y	Y	Y	Y	Y
Broker FE	N	Y	Y	N	Y	Y	N	Y	Y
Other Controls	N	N	Y	N	N	Y	N	N	Y
Year-Quarter FE	N	N	Y	N	N	Y	N	N	Y
MSA FE	N	N	Y	N	N	Y	N	N	Y
Minority/Minority Premium	0.11	0.06	0.05	0.06	0.02	0.00	-0.10	-0.05	-0.04
P-value	0.00	0.00	0.00	0.00	0.13	0.79	0.00	0.08	0.14

This table reports OLS estimates. The dependent variable is the natural log of broker fees. Each reported covariate is a dummy variable that equals one when true, and zero otherwise. Borrower controls include age, gender, marital status, income, credit (FICO) score, a subprime mortgage indicator (FICO<620), debt-to-income ratio, and employment status. Property type controls include owner-occupancy versus investor status and single-family versus condominium or multiple unit structure. Loan controls include indicators for broker face-to-face interaction, purchase versus refinance, co-borrowing, combined debt loan-to-value, loan type (adjustable-rate, fixed-rate, interest only), interest rate spread, and borrower documentation. Area controls include distance between borrower and broker, broker competition, area unemployment rate, regulatory environment, education, income, share of housing that is owner occupied, price to rent ratio, and county population share that is: married, foreign born, Hispanic, black, Asian or Pacific Islander, English speaking, and Hispanic speaking. See Table 1 for a complete description of the variables. Robust standard errors clustered by MSA are in parentheses. The stars ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

Table 5. Proposed Ban on Dual Compensation

Y Dep. Var.: ln(Y + 1)	(1) Front End Fees HW	(2) Front End Fees BW	(3) Front End Fees AW	(4) Back End Fees HW	(5) Back End Fees BW	(6) Back End Fees AW
Minority Borrower	0.03***	0.05***	0.01	0.01	0.04**	-0.02
	(0.01)	(0.01)	(0.03)	(0.02)	(0.02)	(0.03)
Minority Borrower ×	0.02	-0.04	-0.07*	0.05	0.05	-0.09
Minority Broker	(0.02)	(0.03)	(0.04)	(0.05)	(0.08)	(0.07)
Observations	39,697	32,998	21,108	7,981	7,209	6,191
Adjusted R-squared	0.59	0.63	0.62	0.69	0.69	0.68
Log Loan Amount	Y	Y	Y	Y	Y	Y
Broker FE	Y	Y	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y
MSA FE	Y	Y	Y	Y	Y	Y

This table reports OLS estimates. The dependent variable is the natural log of broker fees. Each reported covariate is a dummy variable that equals one when true, and zero otherwise. Each column restricts the observations to the subgroup specified in the header. Borrower controls include age, gender, marital status, income, credit (FICO) score, a subprime mortgage indicator (FICO<620), debt-to-income ratio, and employment status. Property type controls include owner-occupancy versus investor status and single-family versus condominium or multiple unit structure. Loan controls include indicators for broker face-to-face interaction, purchase versus refinance, co-borrowing, combined debt loan-to-value, loan type (adjustable-rate, fixed-rate, interest only), interest rate spread, and borrower documentation. Area controls include distance between borrower and broker, broker competition, area unemployment rate, regulatory environment, education, income, share of housing that is owner occupied, price to rent ratio, and county population share that is: married, foreign born, Hispanic, black, Asian or Pacific Islander, English speaking, and Hispanic speaking. See Table 1 for a complete description of the variables. Robust standard errors clustered by MSA are in parentheses. The stars ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

Table 6. Quantile Regressions of Broker Fees

Panel A: HW	(1)	(2)	(3)
Dep. Var.: Broker Fees$_{\alpha}$	Quantile 10	Quantile 20	Quantile 30
Minority Borrower	304.96***	359.34***	411.31***
	(21.25)	(16.74)	(16.64)
Minority Broker	91.53***	174.05***	186.13***
	(32.98)	(25.97)	(25.83)
Minority Borrower × Minority Broker	159.99***	139.61***	155.54***
	(40.32)	(31.75)	(31.57)
Observations	234,413	234,413	234,413
Panel B: BW			
Minority Borrower	236.27***	292.95***	352.03***
	(15.98)	(14.39)	(13.89)
Minority Broker	124.16***	176.30***	239.73***
	(37.95)	(34.17)	(32.99)
Minority Borrower × Minority Broker	-99.08**	-116.93***	-177.77***
	(44.91)	(40.44)	(39.04)
Observations	213,575	213,575	213,575
Panel C: AW			
Minority Borrower	308.79***	392.76***	447.51***
	(37.58)	(31.82)	(30.34)
Minority Broker	271.86***	318.00***	346.93***
	(49.62)	(42.03)	(40.07)
Minority Borrower × Minority Broker	-458.24***	-471.99***	-454.88***
	(70.71)	(59.89)	(57.10)
Observations	160,936	160,936	160,936
Borrower Controls	Y	Y	Y
Property Type	N	N	N
Loan Controls	Y	Y	Y
Area Controls	N	N	N
Year-Quarter Fixed Effects	Y	Y	Y
State Fixed Effects	Y	Y	Y

This table reports the coefficient estimates at the 10^{th}, 20^{th}, and 30^{th} quantile. The dependent variable is the nominal value of broker fees. Each reported covariate is a dummy variable that equals one when true, and zero otherwise. Borrower controls include age, gender, marital status, income, credit (FICO) score, debt-to-income ratio, and employment status. Property type controls include owner-occupancy versus investor status and single-family versus condominium or multiple unit structure. Loan controls include indicators for broker face-to-face interaction, purchase versus refinance, co-borrowing, combined debt loan-to-value, loan type (adjustable-rate, fixed-rate, interest only), interest rate spread, and borrower documentation. Area controls include distance between borrower and broker, broker competition, area unemployment rate, regulatory environment, education, and income. See Table 1 for a complete description of the variables. Standard errors are in parentheses. The stars ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively. The last row provides the mean predicted cost using the point estimates. As reference, consider that the average nominal broker fees (or cost) is $5,565.

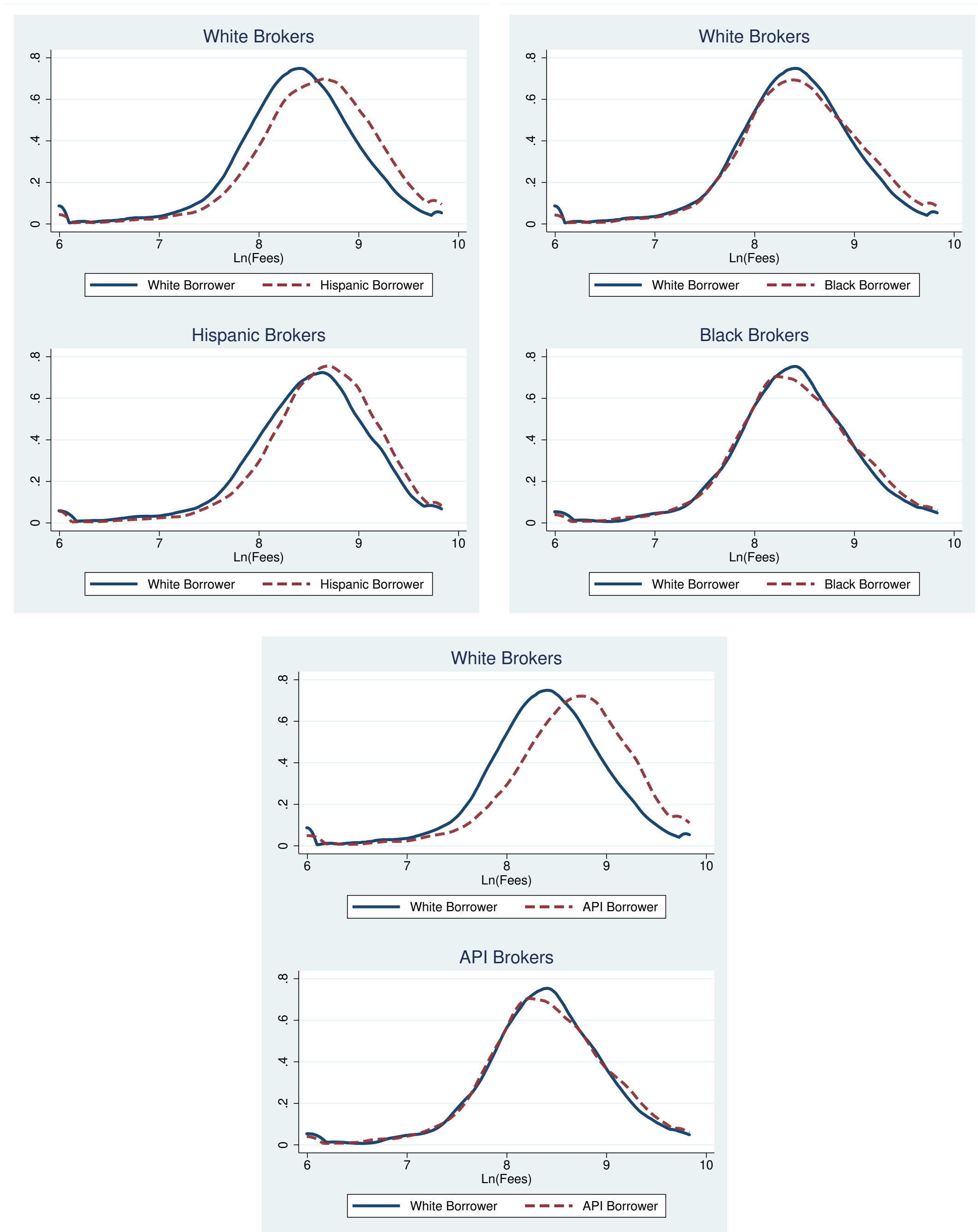

Figure 1. Kernel Density of Log Broker Fees by Group

This figure displays the distribution of log broker fees in the three main samples used in our analysis (Hispanice/White, Black/White, API/White). Distributions are separated by borrower and mortgage broker race.

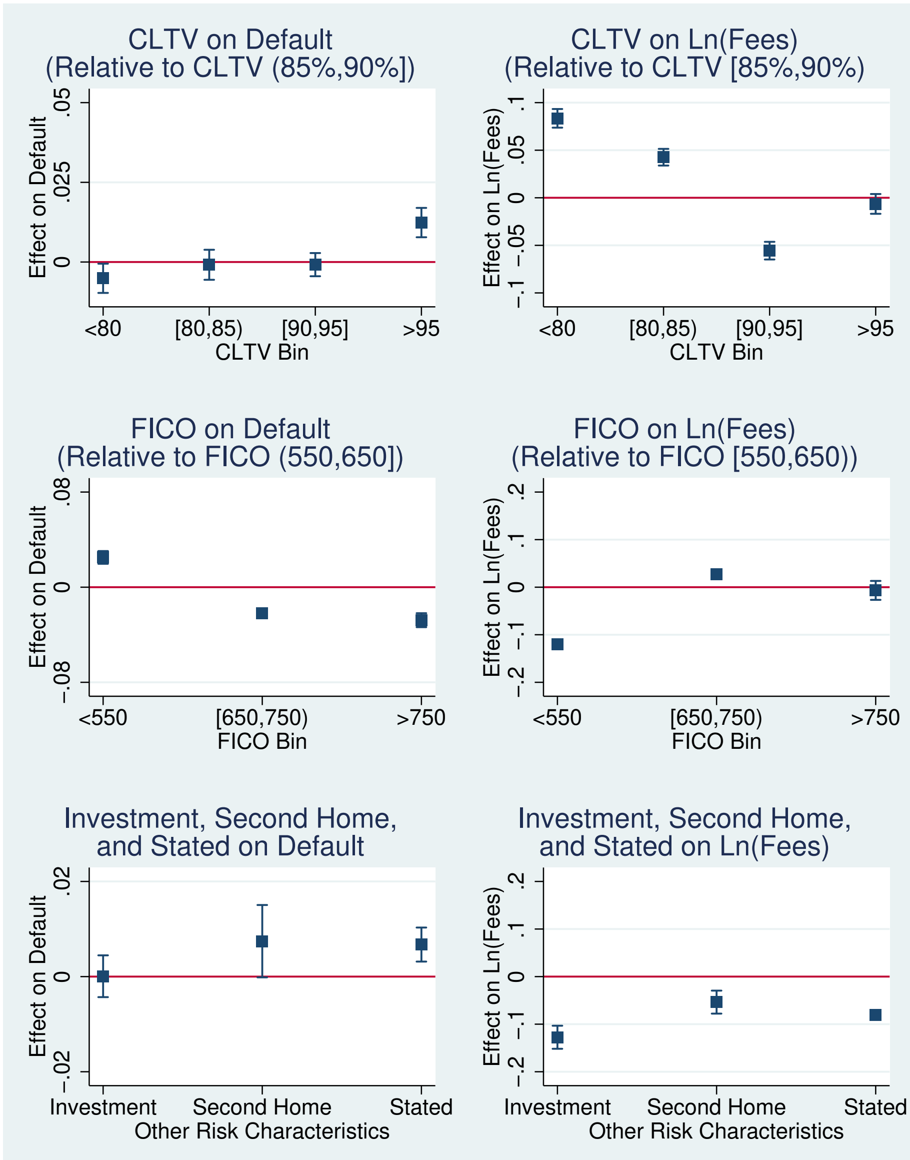

Figure 2. Default Risk Characteristics and Broker Fees
This figure shows that no clear relationship exists between mortgage risk characteristics and mortgage broker compensation. The left panels plots coefficient estimates from the mortgage default regression reported in column (1) of Table A.5 in the online appendix. The right panels plot coefficient estimates from a regression with log broker fees as the dependent variable as reported in column (2) of Table A.5.

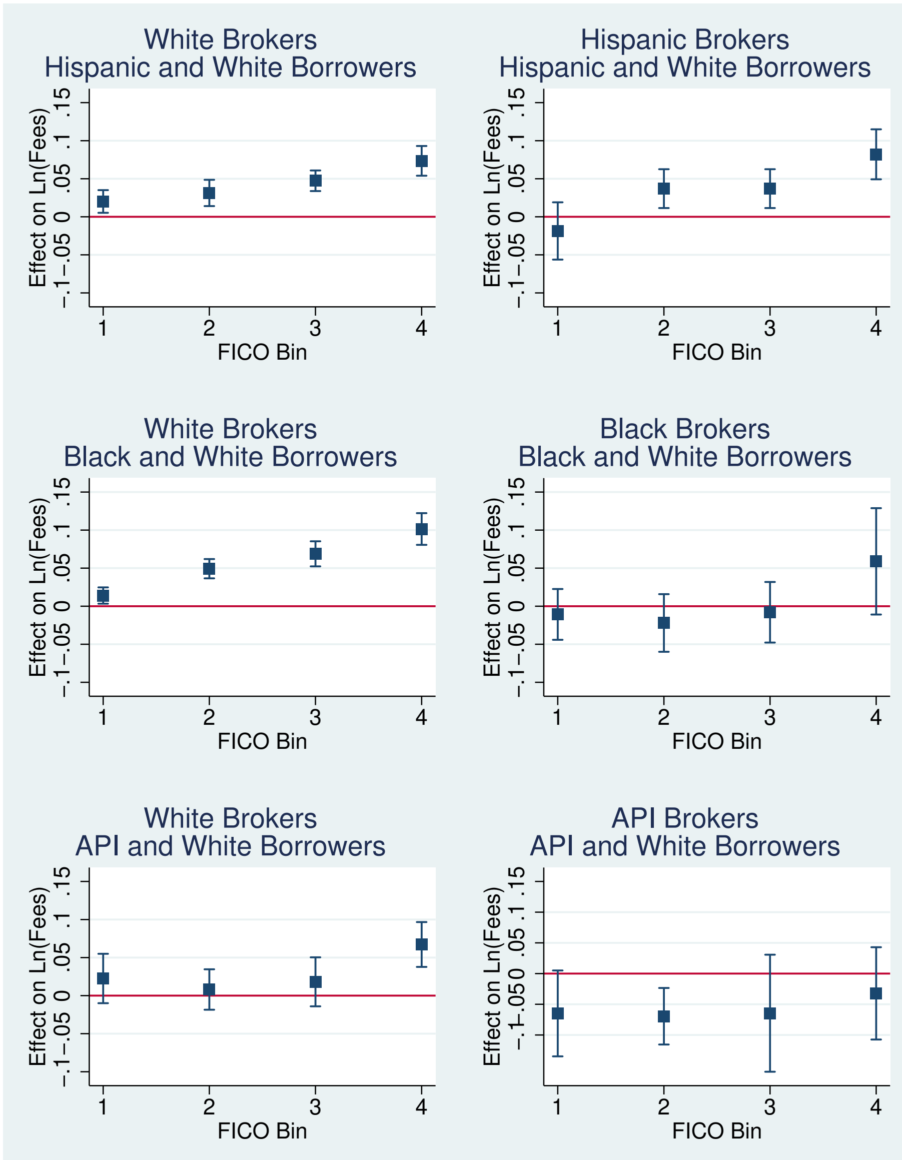

Figure 3. Minority Premium Across Credit Scores
This figure displays the minority premium estimates across FICO score quartiles separated by borrower and mortgage broker race. Each graph represents a separate regression model and corresponds to a single column in Table A.13 in the online appendix.

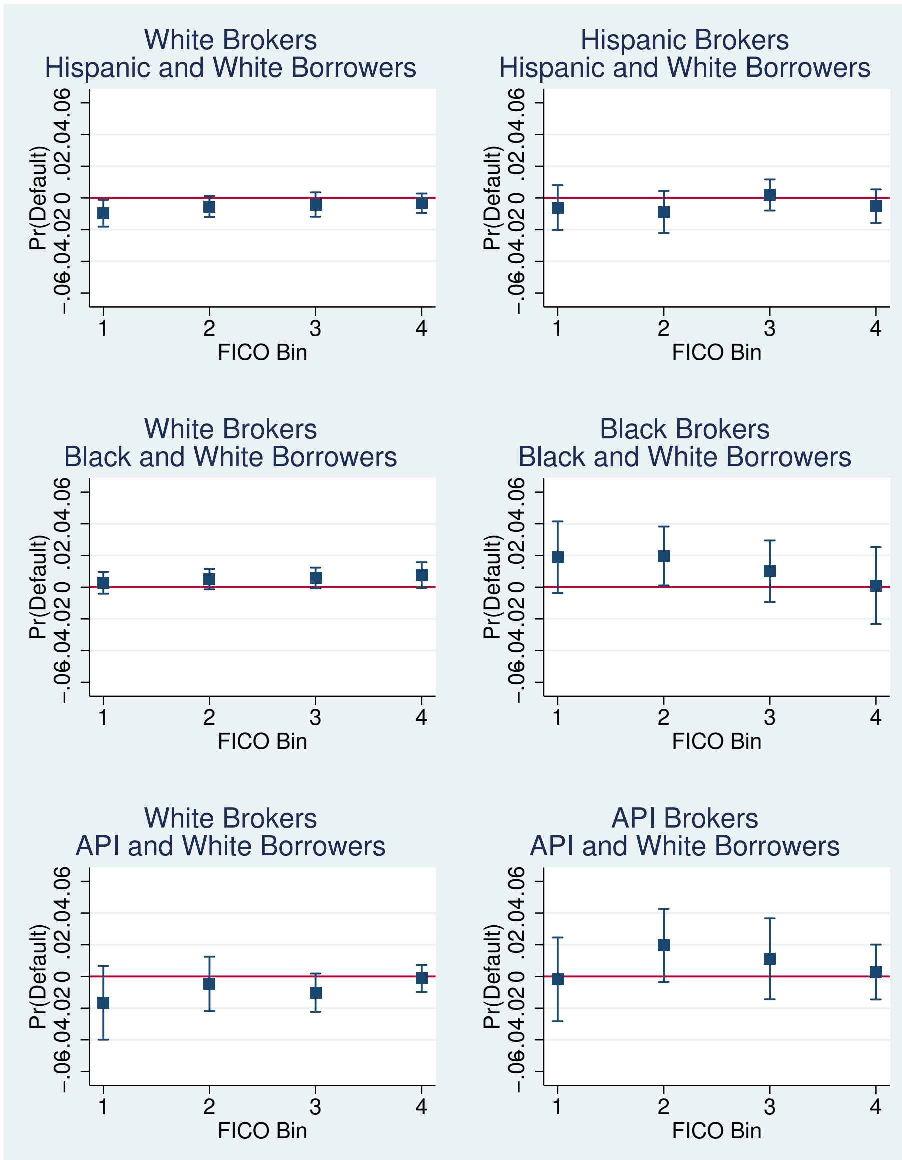

Figure 4. Minority Effect on Probability of Default Across Credit Scores
This figure displays minority effect on mortgage default across FICO score quartiles separated by borrower and mortgage broker race. Each graph represents a separate regression model and corresponds to a single column in Table A.14 of the online appendix.

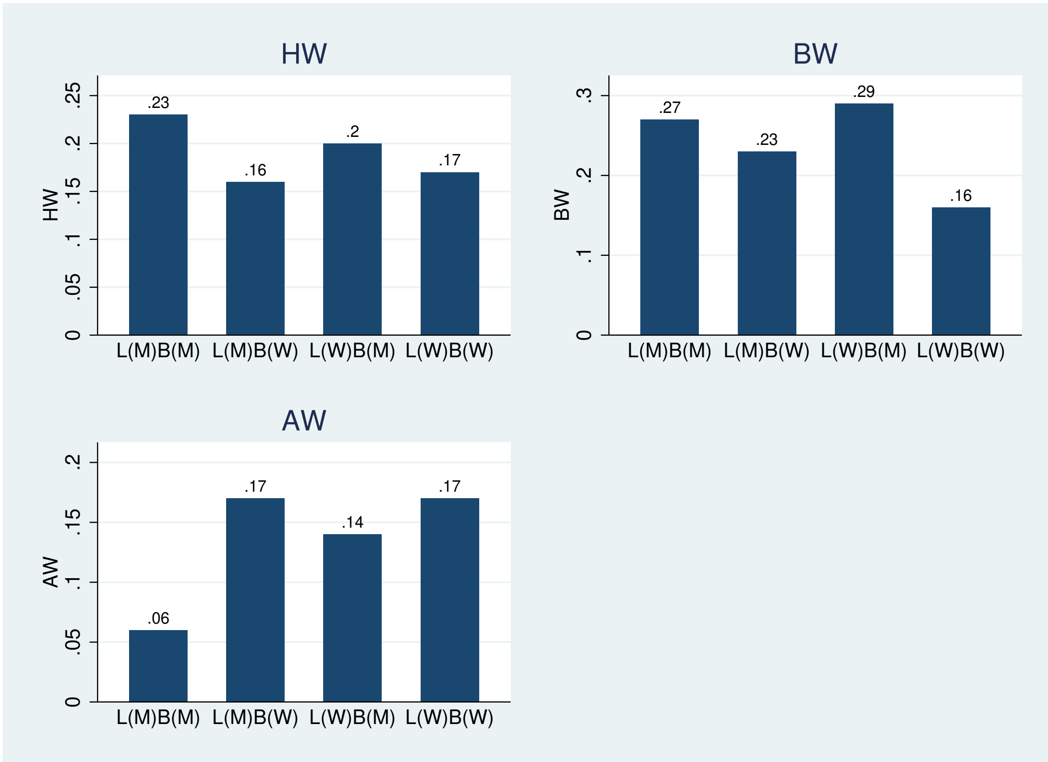

Figure 5. Borrowers At Risk of Credit Rationing
This figure displays the proportion of borrowers at risk of credit rationing. The underlying sample comprises of only Hispanic and white borrowers or brokers (HW), only black and white borrowers or brokers (BW), or only Asian and white borrowers or brokers (AW). Each graph replaces the actual broker fees with predicted fees from the regression at the 30^{th} quantile in Table 6. B(M) stands for minority borrower, B(W) stands for white borrower, L(M) stands for minority broker, and L(W) stands for white broker.

"Does Borrower and Broker Race Affect Mortgage Prices?"

Online Appendix

(Not for Publication)

A.1. Comparison of New Century Sample with Broader Market

To ensure that our sample is representative of the subprime market from 2003 to 2007, in Table A.1 we compare the New Century loan sample with the subprime loan sample in Demyanyk and Van Hemert (2009) – hereafter referred to as DVH – a highly cited paper on the subprime mortgage crisis. The DVH sample is comprised of loans across many subprime lenders and covers roughly half of the subprime mortgage market (85 percent of the securitized subprime market). Descriptive statistics across the two samples are quite similar, with two exceptions. First, the average loan size is slightly higher in the New Century sample. Second, the combined loan to value ratio (CLTV) is significantly higher in the New Century sample. The difference in CLTV ratios across the two samples, however, is likely due to unreported second liens ("silent seconds") in the DVH data. These "silent seconds" cause the true CLTV at origination to be underestimated in the DVH sample. In fact, Piskorski, Seru, and Witkin (2015) use information contained in the New Century data to show that CLTV at origination is biased downwards in data sets like the one employed in DVH. The main takeaway of Panel A is that our New Century sample is representative of subprime mortgage lending over the period covered in our study.

In Panel B of Table A.1, we calculate the minority share of subprime originations in the Home Mortgage Disclosure Act (HMDA) loan application register data. Although the HMDA data does not include information on many of the loan characteristics reported in Panel A, it does provide broad coverage of the entire mortgage market, and it includes information on applicant race and ethnicity. We identify subprime originations using the subprime lender lists compiled by the Department of Housing and Urban Development (HUD) from 2003 to 2005.[1] The minority share of subprime originations in HMDA for New Century (51 percent) is nearly identical to the share in the rest of the subprime market (52 percent).[2] This alleviates concerns that the New Century data suffers from selection issues based on borrower minority status.

A.2. Identifying the Broker's Race and Ethnicity

The New Century data contains the mortgage broker's last name, first name, and office ZIP code location. We use this information to infer the race/ethnicity of the mortgage broker using the Bayesian Improved First Name Surname Geogcoding (BIFSG) method developed in Voicu (2018). By including first name race/ethnicity information, this new methodology extends the well-known Bayesian Improved Surname Geocoding (BISG) method that relies on surname and location alone.

We begin by matching the broker's last name to a list of frequently occurring surnames from the 2000 U.S. Census. The list includes self-reported racial/ethnic distributions associated with surnames used by at least 100 individuals in the 2000 Census.[3] Next we match the broker's location to ZIP code level race/ethnicity distributions obtained from the 2011 American Community Survey 5-year estimates.[4] Finally, we match the broker's first name to a recently developed database that includes the race/ethnicity distributions associated with first names based on 2.7 million mortgage applications (Tzioumis, 2018).

[1] Using HUD's subprime lender lists to identify subprime loans is common practice in the literature (see Nadauld and Sherlund (2013), Mayer and Pence (2008) and Dell'Ariccia, Igan, and Laeven (2012) for examples). The list includes lenders that HUD identifies as specializing in originating subprime loans. This method is not perfect, however, as some subprime loans are originated by lenders not on the list, while some non-subprime loans are originated by lenders on the list.

[2] Our method of classifying borrowers as minorities is discussed in detail below.

[3] The list is publicly available at `www.census.gov/topics/population/genealogy/data` and is discussed in detail in Word et al. (2008).

[4] From https://data.census.gov/cedsci/, we downloaded the data for "Hispanic or Latino Origin by Race" while setting the geographies to "Zip Code Tabulation Area (Five-Digit)."

Following the BIFSG methodology, the conditional probability of interest is defined as:

$$p(r|f, s, z) \quad = \quad p(r|s) \times \frac{p(f|r) \times p(z|r)}{\sum_{r=1}^{6} v(r, f, s)}, \tag{4}$$

where $p(r|f, s, z)$ is the posterior probability of being race r, given first name f, surname s, and location z. We refer to this probability as the BIFSG score. $p(r|s)$ is the proportion of all people with surname s who report being of race r. This probability is then updated by the second term, where $v(r, f, s) = p(r|s) \times p(f|r) \times p(z|r)$, to create the posterior race/ethnicity distributions (BIFSG scores).[5] Each of the conditional distributions on the right hand side of equation 4 is obtained from publicly available information, as described above.

The BIFSG methodology relies on two assumptions. First, it assumes that the probability of a first name conditional on race does not vary by surname ($p(f|r) = p(f|r, s)$). The second assumption is that the probability of a location conditional on race does not vary by first name or surname ($p(z|r) = p(z|r, f, s)$). Other Bayesian race/ethnicity classification systems, including the commonly used BISG approach, require similar assumptions (Consumer Financial Protection Bureau, 2014b; Elliott et al., 2009; Tzioumis, 2018). Although the BIFSG assumptions are not directly testable with public data, previous research shows that simple Bayesian classifiers perform well, even when there are clear dependencies between attributes (e.g., r, f, s, z) (Domingos and Pazzani, 1996).[6]

"Threshold" classification schemes are commonly used to create a discrete categorical variable for an individual's race. For example, in our context an individual is classified into a race if its BIFSG score for that race is above a certain threshold, say 85%. Individuals with BIFSG scores

[5]For each broker we calculate this probability for each of the six race/ethnicity groups defined by the U.S. Census (white, Hispanic, African American, Asian or Pacific Islander, American Indian or Alaskan Native, and Two or More Races). Whereas Voicu (2018) observes an individual's location at the Census Tract level, we observe broker location at the ZIP code level.

[6]For observations that have a missing first or last name, we use the available distribution to infer the race or ethnicity of the broker. For observations that have both the first and last names but lack $p(f|r)$ estimates, we infer the race or ethnicity of the broker using the maximum likelihood in the two distributions. For over 97% of our sample, we employ the Bayesian approach described above to define race and ethnicity.

below the threshold remain unclassified and are excluded from the analysis. As the threshold is increased, the chances of race/ethnicity classification generally decrease, but at a significant cost – a greater share of the observations are dropped from the analysis. Alternatively, discrete categorization of race/ethnicity can be achieved using the "maximum a posteriori" (MAP) classification scheme, which is common when using Bayesian-based classifiers (Voicu, 2018). The MAP scheme sets race to be that of the highest Bayesian score for the individual. An obvious advantage to the MAP approach is that all observations are included in the analysis because race/ethnicity is predicted for every individual.

Table A.3 provides examples of the MAP and threshold classification schemes. The top panel provides the BIFSG scores for several hypothetical mortgage brokers. The bottom panel shows the categorization for each individual at various thresholds and under the MAP scheme. For example, under an 85% threshold scheme, the race/ethnicity of Edward Lewis in Gainesville, FL (zip-code 32608) is classified as unknown, and any mortgages originated by this broker would be excluded from our analysis. In contrast, under the MAP classification scheme, this mortgage broker is classified as white, and all loans associated with this broker would be included in our regressions. Likewise, the MAP scheme would classify both Calvin Dawson (32305) and Frank Robinson (34946) as black. However, we would drop loans by the latter broker if using an 85 percent threshold.

We report results using the MAP classification in the main text of the paper for several reasons. First, it has the distinct advantage of giving us the broadest data coverage – we retain all observations since each loan officer can be classified into a single race/ethnicity. In contrast, under an 85% threshold classification scheme, 24% of the observations in the New Century data are unusable since broker race/ethnicity is not identified for brokers associated with those loans. Second, MAP helps identify whites and blacks that are often classified as unknown under simple threshold schemes.[7] Third, and most important, using a large sample of mortgage applications, Voicu (2018)

[7] Although the MAP classification scheme has the advantage of using all available data, it does come at a modest cost in terms of accuracy when compared to simple threshold schemes. In this context, accuracy is defined as the total number of individuals whose race is correctly classified divided by the total number of individuals that are classified. For example, using Florida voter data, described in more detail below, we find that a simple 85% threshold

shows that coefficient estimates of borrower race/ethnicity in mortgage denial regressions are less biased using MAP relative to bias using an 80% threshold system. In APR regressions, he finds that bias is similar across the two methodologies. For these reasons, we use MAP classification in the body of the paper, but we stress that our results remain unchanged when we use different threshold classifications (50%, 80%, 85%, or 95%).[8]

Although Voicu (2018) extensively examines the validity of the BIFSG MAP approach, we observe location at a different level (ZIP code) than in his study (Census Tract). Thus, we examine the accuracy of the MAP BIFSG classification approach with ZIP code as the geographic identifier. To do so, we use publicly available voter registration data from the state of Florida. The data includes 13.3 million voter records, representing nearly 63 percent of Florida's population. For each voter we observe first name, last name, home ZIP code, and self-reported race and ethnicity.

Table A.4 reports the accuracy rates for each of the following groups: whites, Hispanics, African Americans, and Asians or Pacific Islanders. The accuracy rates are measured for each group as the number of voters classified correctly divided by the total number of voters classified into that group. We find that the accuracy rate for white and Hispanic voters is 86 and 83 percent, respectively. For blacks and Asian or Pacific Islanders, the accuracy rate is 74 and 70 percent, respectively.[9] Comparing our accuracy rates to those from North Carolina voter data reported in Voicu (2018), we see that our accuracy rate in Florida is significantly better for Hispanics and blacks, but less accurate for whites and Asians or Pacific Islanders. Notice, however, that when Voicu (2018) uses HMDA data, which is nationally representative, the accuracy rate of MAP BIFSG improves

---

rule results in an **86%** accuracy rate while the MAP accuracy rate is slightly lower at **83%**.

[8]We also note that our results remain the same when using the BIFSG scores as independent variables in the regression models. Voicu (2018) shows that discrete classification schemes based on BIFSG scores produce less bias in race/ethnicity coefficients than the BIFSG scores themselves in OLS models of mortgage outcomes.

[9]A potential concern with using MAP BIFSG is that accuracy rates vary across race/ethnicity. In particular, the accuracy rates are significantly lower for blacks and Asian or Pacific Islanders. To alleviate this concern, we also experimented with using different thresholds for each race/ethnicity that ensure equal accuracy rates across race/ethnicity. Using Florida voter data, we found thresholds for each race/ethnicity that gave a certain accuracy rate. For example, the thresholds to ensure 85% accuracy for white, Hispanic, and black classifications are 50%, 66%, and 72%, respectively. Results using this approach at different accuracy rates (80%, 85%, and 90%) are materially unchanged from those reported in the paper.

significantly over the rates in the North Carolina data. Thus, it is quite likely that the accuracy rates for mortgage brokers in the New Century data, which is nationally representative, are higher than those obtained in the Florida voter data.

Although we obviously cannot directly test the accuracy rate of different classification schemes for brokers in the New Century data, we again stress that our results are insensitive to the choice of classification scheme. This suggests that our results are not driven by broker race/ethnicity misclassification or sample selection issues.

A.3. Rejection and Funding Uncertainty

Mortgage brokers do not bear credit risk on the loans they arrange, so their compensation should not depend on credit risk factors directly. Thus, unobserved credit risk factors are unlikely to explain the pricing differentials. However, brokers do face production costs and rejection risk that may vary across loan applicants. For example, as Yezer (2017) points out, well organized applicants may require little effort from the broker in shepherding the loan from application through funding. Meanwhile, other applicants will require more effort, hand-holding, and financial counseling without offering the broker certainty that the application converts into a funded loan. After spending time compiling an application, some applicants will be rejected, while others will decide to withdraw their applications. In the case of a non-funded application (rejected or withdrawn), the broker receives no compensation. Thus, broker revenue on funded loans likely reflects funding uncertainty and loan production cost differences across applications.[10]

If funding uncertainty or loan production costs correlate with both broker compensation and race, our estimates in earlier sections will be biased. However, the rich set of borrower, property, loan, and area controls likely serve as suitable proxies for differences in loan production costs. With

[10] Bohren, Imas, and Rosenberg (2019) theorize that an agent's prior beliefs about preferential treatment held by other agents drive differential outcomes even if the agents do not hold the same beliefs. Hence, another source of the observed differences in fees by borrower race could be the result of mortgage brokers "believing" that minority applicants are more likely to be rejected by lenders.

respect to rejection risk and funding uncertainty, our unique data gives us the ability to account for these factors.

To this point, our analysis has focused exclusively on funded loans because mortgage broker fees are reliably recorded on these loans. But, the data also includes information on loan applications that did not convert to funded loans with New Century. We observe over 480,000 brokered loan applications that meet the requirements stated in Section II.A with 65 percent funded, 19 percent rejected, and 13 percent approved but not funded (withdrawn after approval). The remainder consist of loan applications that were withdrawn prior to approval or rejection.[11]

Variation in application outcomes enables us to incorporate rejection risk and funding uncertainty into our analysis by incorporating the probability of origination for each observation. Specifically, we estimate a two-step Heckman model that corrects for the likelihood of loan origination (Heckman, 1979, 1990). In the first step, we use a Probit selection model with a dummy for origination as the dependent variable that takes a value of one if the loan application is approved and funded, and zero otherwise. In the second step, we calculate the causal model of broker fees, which is similar to equation (1) in using the log broker fees as the dependent variable and our rich set of control. We replace the quarter-year origination fixed effects with quarter-year application fixed effects and includes an additional variable: the inverse mills ratio.[12]

To achieve identification in the causal model, the estimation approach requires an instrument for our selection model that meets two requirements. First, the instrument must be highly correlate with the application's likelihood of origination, conditional on other covariates. Second, the instrument must meet the exclusion restriction that it only affects fees through its impact on the

[11] An application can be withdrawn by the broker or the borrower. We are not able to determine what ultimately happens in the case of a withdrawn or rejected application. For example, the application may be converted by the same broker into a funded loan with another lender. Alternatively, the borrower could obtain a loan through a different broker or directly through another lender. Finally, the applicant may receive no loan. Despite the inability to determine the ultimate outcomes, since an originated loan results in a funded loan with 100 percent certainty, we can say with confidence that a rejected or withdrawn application is less likely to result in a funded loan.

[12] The inverse mills ratio, imputed using the first stage point estimates, is the likelihood that a loan application is originated over the cumulative likelihood of the loan application's outcomes. Its coefficient estimate in the causal model can be interpreted as the covariance between the loan's origination likelihood and the fees paid, relative to the variation in the loan application's outcome.

likelihood of origination.[13] Hence, we use the Non-New Century Subprime Rejection Rate as an instrument for our selection model. The Non-New Century Subprime Rejection Rate is the annual county-level rejection rate on loan applications by subprime lenders other than New Century. The intuition of the instrument is as follows: rejection rates across subprime lenders *within* the same geographic area are likely to be highly correlated since the lenders face the same applicant pool and underlying property market fundamentals within those areas. But, the rejection rate of other subprime lenders within that location should not directly affect the fees a broker charges when originating a loan through New Century. We construct the subprime county-level rejection rate using HMDA loan application data. We identify subprime loans using the subprime lender lists available on the the Department of Housing and Urban Development's website.[14] We anticipate that the subprime rejection rate (excluding New Century) is inversely related to New Century's propensity to approve and fund a loan application. We also believe that it is reasonable to assume that it does not directly influence fees on brokered loans in our sample.[15]

Columns (1), (4), and (7) of Table A.9 report results for the selection (rejection) models for the HW, BW, and AW samples, respectively. In each of these columns, the area subprime rejection rate (the instrument) is significantly negatively related to the likelihood that an application results in a funded loan. Also, the race coefficients are significant in the selection models as well. The racial price disparities remain even after accounting for the risk of an application not funding.

Next we turn to the causal model of broker fees. The racial price disparities in columns (3), (5) and (8) remain even after accounting for the risk of an application not funding.[16] Note, though,

[13]Without an instrument that meets the exclusion restriction, identification in the causal model derives solely from the nonlinearity of the inverse Mills ratio. However, as Puhani (2000) points out, relying on this nonlinearity for identification is often problematic. Thus, we use an instrument to achieve identification.

[14]See `https://www.huduser.gov/portal/datasets/manu.html`. As discussed in Section II.A, using the HUD lists to identify subprime loans is common in the literature even though the method is not perfect. The HUD lists are only available through 2005, so we use the 2005 subprime lender list to identify subprime loans in 2005 and 2006.

[15]It is possible, however, that the instrument does not meet the exclusion restriction. For example, in relatively constrained credit markets (e.g., where rejection rates are high), local brokers may charge additional fees to account for lower volume. We thank an anonymous referee for pointing this out. We proceed with the analysis noting this potential limitation.

[16]For comparison purposes, we report OLS estimates corresponding to the casual models in Columns (3), (6), and (9). The results are nearly identical to those in the Heckman fee models.

that the inverse mills ratio in the second stage is statistically insignificant for two of the subsamples (HW and AW), which suggests that our rich set of control variables in the fee regression adequately accounts for funding uncertainty and rejection risk. Taken together, the results in Table A.9 suggest that the minority pricing premiums observed in this study are not due to funding uncertainty.

A.4. Contract Selection

It is not clear whether the observed pricing disparities are due to differential treatment by the mortgage broker or contract selection by the borrower.[17] For example, suppose a white mortgage broker offers the same menu of contracts to all borrowers, regardless of race, but some of those contracts generate higher revenue for the broker. If minorities systematically select into the high revenue contracts, while white borrowers select into the low revenue contracts, then observed fee differentials are due to borrower selection rather than disparate treatment by the mortgage broker. In a contemporaneous paper, Bhutta and Hizmo (2019) provide evidence that minorities and whites tend to select into different contracts in terms of points and interest rates.

To rule out borrower contract selection, we follow an approach similar to Bhutta and Hizmo (2019) to investigate whether brokers treat minorities differently. We examine whether minority and white borrowers face the same trade-off between front-end fees and back-end (YSP) fees. In theory, an increase in back-end fees should be exactly offset by a decrease in front-end fees. However, evidence suggests that YSP increases are only partially offset by reductions in front-end fees (Woodward and Hall, 2012; Ambrose and Conklin, 2014). Our primary interest is not the tradeoff itself, but whether it varies with borrower and broker race. If minorities face a different front-end/back-end fee tradeoff, this suggests that brokers are treating minorities differently.

We examine the tradeoff between front-end fees and back-end fees after conditioning on the rich set of control variables and fixed effects used in our saturated regression models. The dollar

[17] In this section, we define a contract as a unique pairing of front-end fees and mortgage contract rate (and hence back-end fees).

amount of front-end fees is the dependent variable. We create indicator variables for the dollar amount of back-end fees and interact borrower minority status with these indicators.[18]

We estimate six separate regression models and plot the tradeoff coefficient estimates for minorities and non-minorities in Figure A.1. The top two panels show that Hispanic borrowers face a different tradeoff than white borrowers, regardless of the race of the mortgage broker. Black borrowers also face a different tradeoff than white borrowers when working with white brokers. This provides some evidence that brokers in these subsamples are treating minorities differently from whites. The tradeoffs between minorities and whites are not significantly different in the remaining panels.

While figure A.1 provides some evidence that minorities and white face different front-end/back-end fee tradeoffs, we note that even if minorities and whites face the same tradeoff it would not preclude differential treatment by mortgage brokers. Suppose, for example, that with no yield spread premium, a white broker charges a white borrower \$4,000, but charges a comparable minority \$5,000. Even if both minorities and whites face the same YSP/front-end fee tradeoff, minorities still are treated differently. Figure A.2 plots the adjusted means of front-end fees at different levels of back-end fees by borrower and broker race.[19] This figure depicts not only differences in front-end/back-end fee tradeoffs, but also differences in front-end fee levels. Thus, even on loans with no back end fees, Hispanics are predicted to pay higher fees, regardless of broker race. For black borrowers working with white brokers, differential treatment is purely driven by differences in the front-end/back-end fee tradeoff. The bottom left panel shows that API and white borrowers face the exact same tradeoff when working with white brokers, but at every level of yield spread API borrowers are expected to pay significantly higher fees. For black brokers and API brokers, there is no evidence of differential treatment.

[18] The buckets for back-end fees are as follows: [0,0], (0,1000], (1000,2000], (2000,3000], (3000,4000], (4000,5000], (5000,6000], (6000,7000]. We exclude observations with back-end fees greater than \$7000 as 98.5% of the observations fall into the smaller buckets. 90% of the observations have \$4000 or less in back end fees, and 31% have no back-end fees.

[19] These adjusted means are obtained in Stata using the regression models of Figure A.1 along with the margins comand.

Taken together, the results in Figures A.1 and A.2 suggest that minority premiums are not merely the result of differences in contract selection across groups. Rather, these figures imply that pricing disparities are driven by differential treatment.

A.5. Borrower/Broker Selection

Our analysis thus far is based on an implicit assumption that either the borrower's choice of broker race is random, or that the propensity to self-select into brokers of the same race is not driven by the dependent variable, broker fees. We believe the latter is a reasonable assumption since survey evidence suggests that many borrowers consider only a single broker (Woodward and Hall, 2012). For example, a recent report from the Consumer Financial Protection Bureau shows that nearly 50 percent of consumers who take out a purchase mortgage only consider a single lender or broker prior to application. Moreover, the report indicates that 77 percent of mortgage borrowers apply to only one lender or broker (Consumer Financial Protection Bureau, 2015). This provides strong support for our assumption that price is not driving self-selection in our data. Additionally, the extensive set of borrower, loan, property, and area characteristics reduces concerns of omitted variable bias.

However, to further assuage selection concerns, we explicitly model the borrower's choice to obtain a loan through a minority broker using a propensity score matching technique, a semi-parametric approach that obtains balanced treatment and control subgroups.[20] The benefits to this approach are twofold. First, Rosenbaum and Rubin (1985) show that matching on propensity scores mimics random sampling and thus, mitigates self-selection bias. Second, and more importantly, this approach permits us to model the borrower's propensity to choose a white or minority broker

[20]A standard approach to deal with selection issues is a Heckman correction model. Ultimately we do not use this method for two reasons. First, we were unable to find an instrument for the broker choice model that is likely to meet the exclusion restriction in the causal fee model. In such cases, OLS results are often more reliable (Puhani, 2000). Second, there is a high degree of censoring in our data. In a selection model of the choice to use a minority broker, over 95 percent of the observations for white borrowers would be censored in the fee regression. When there is a high degree of censoring, OLS is often preferable to a Heckman model (Puhani, 2000; Zuehlke and Zeman, 1991).

and thus, construct balanced subgroups. For each white borrower that goes to a minority broker (the treatment group), we find a similar white borrower in the same zip code that goes to a white broker (the control group) using propensity score matching. Likewise, for each minority borrower that goes to a minority broker, we find a comparable minority borrower that obtains a loan from a white broker.

To estimate the propensity scores, we first fit for each borrower group a probit model of the broker's minority status on a subset of the baseline controls that includes year-quarter fixed effects.[21] The covariates in this model arise prior to the borrower's broker choice and thus, we avoid the "bad control" problem (Angrist and Pischke, 2008).[22] We then add the broker's zip code multiplied by ten to the fitted propensity score estimates to create a modified propensity score.[23] This approach forces matching within zip codes and adjusts for geographic differences in lending practices that possibly drive differences in broker fees.

We next estimate

$$P_{imt} = \delta L_i^M + X'_{imt}\beta + \tau_t + \kappa_m + \varepsilon_{imt} \tag{5}$$

for each of the borrower race groups. These regressions will show whether minority brokers charge more than white brokers to a comparable borrower of the same race. Table A.10 reports the results for white borrowers. Consistent with our results in Table 4, there is evidence in all three columns that white borrowers pay more, on average, when using a minority broker. For minority borrowers, the evidence varies across races as show in Table A.11. Hispanic borrowers pay more, on average, to get a loan from a Hispanic broker. Black borrowers, on the other hand, pay similar fees regardless of the race of the loan officer. Finally, Asian borrowers pay less in broker fees when obtaining a

[21] For white borrowers, we estimate a separate probit regression for each potential race of the broker (Hispanic, black, an Asian). For each of the other borrower minority groups, we fit one probit model.

[22] Although we can use a propensity score technique to model the choice of broker race, using the same methodology to model borrower race is infeasible, as treatment assignment (borrower race) occurs before the other covariates are determined. Thus, we cannot determine the treatment effect of borrower race on fees using propensity score matching.

[23] The added zip code times ten forces exact matching at the zip code level when using nearest neighbor propensity score matches with a caliper width of 0.02.

loan from an Asian broker. These findings are consistent with the results of columns (1), (4), and (7) of Table 4.[24]

The parameter estimates are similar to the previous results. Hence, propensity score results confirm the baseline estimates of the price disparities borrowers encounter between white and minority brokers.

A.6. Differences in Language

We examine whether the premiums uncovered in our main analysis are driven by language differences. For example, perhaps Hispanic brokers charge a premium for providing bilingual services. Although we do not observe the language in which the mortgage application was taken, we do observe whether the borrower is a US citizen, and use this information as a proxy for the use of a foreign language. Thus, we exclude non-US citizens from the sample and re-estimate the regression. The results are reported in Table A.12. The results are consistent with those reported in Table 4.

[24] In the other columns of Table 4, the minority broker dummy drops from the model, making it infeasible to determine whether a white or Minority broker charges more within borrower race.

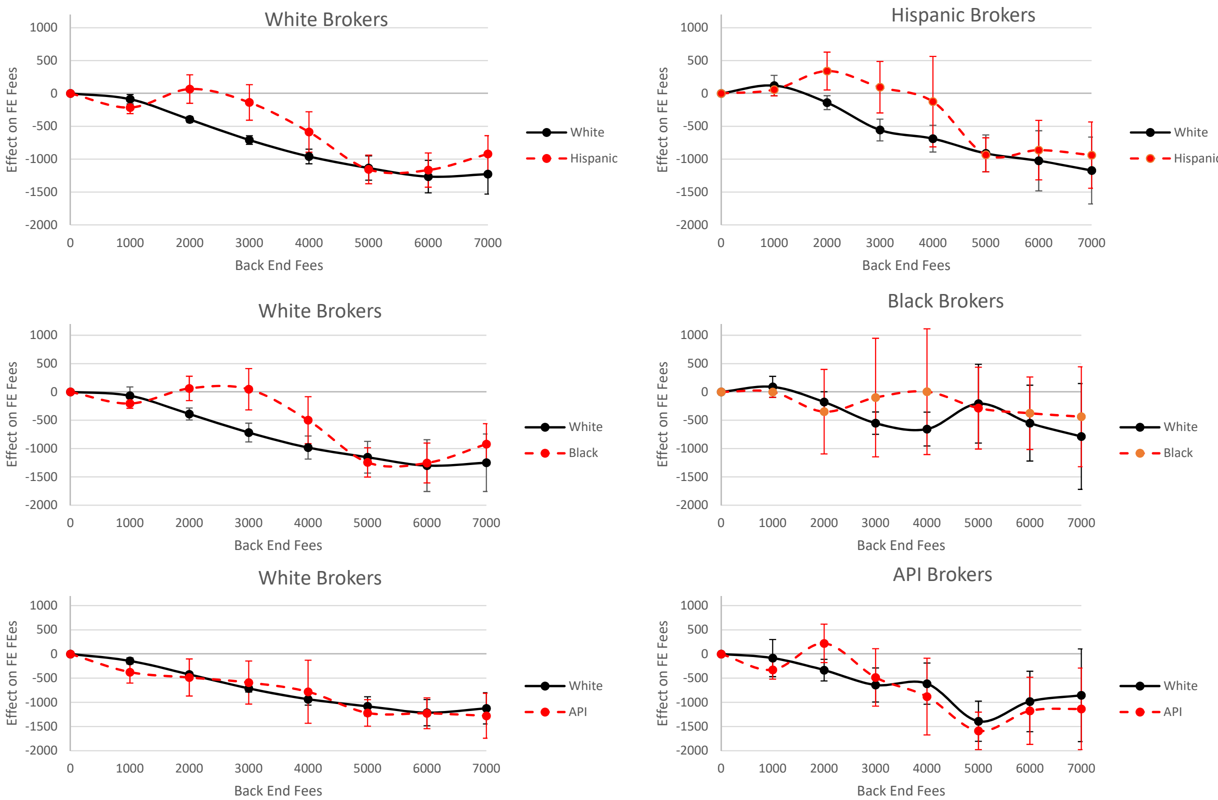

Figure A.1. Tradeoff Between Front- and Back-end Fees

This figure displays coefficient estimates from the fee tradeoff regression in Section A.4 of the Online Appendix. Each graph represents a separate regression model.

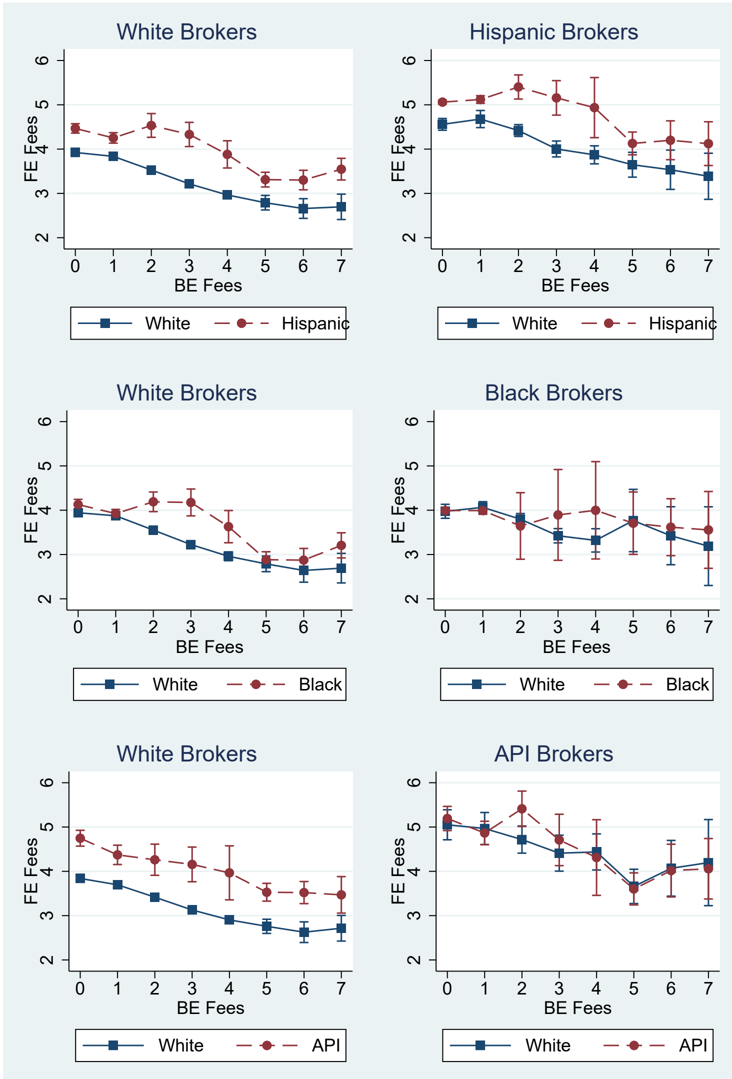

Figure A.2. Adjusted Conditional Mean of Front-end Fees

This figure displays adjusted means calculated from the fee tradeoff regression in Section A.4 of the Online Appendix. Each graph represents a separate regression model.

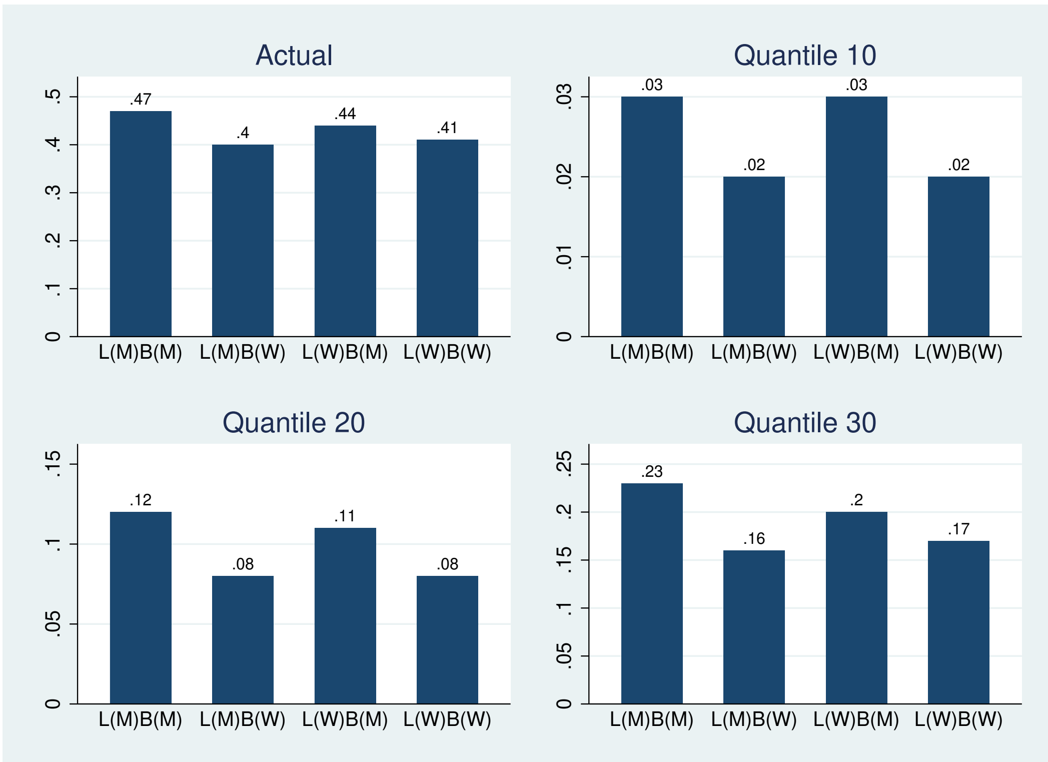

Figure A.3. Borrowers At Risk of Credit Rationing: Hispanics and Whites
This figure displays the proportion of borrowers at risk of credit rationing. The underlying sample comprises of only Hispanic and White borrowers or brokers. The graph titled "Actual" displays the share of loans that have broker fees exceeding their cap. The other graphs replace the actual broker fees with predicted fees from the regression at the 10^{th}, 20^{th}, or 30^{th} quantile in Table 6. B(M) stands for minority borrower, B(W) stands for white borrower, L(M) stands for minority broker, and L(W) stands for white broker.

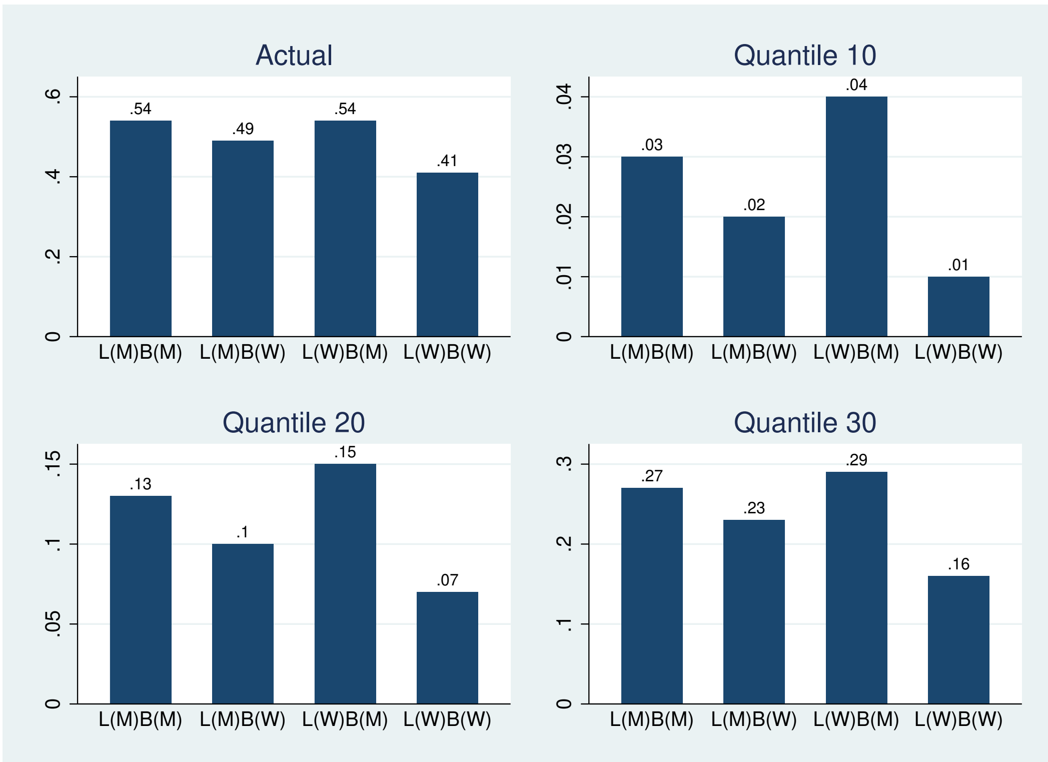

Figure A.4. Borrowers At Risk of Credit Rationing: African Americans and Whites
This figure displays the proportion of borrowers at risk of credit rationing. The underlying sample comprises of only African American and White borrowers or brokers. The graph titled "Actual" displays the share of loans that have broker fees exceeding their cap. The other graphs replace the actual broker fees with predicted fees from the regression at the 10^{th}, 20^{th}, or 30^{th} quantile in Table 6. B(M) stands for minority borrower, B(W) stands for white borrower, L(M) stands for minority broker, and L(W) stands for white broker.

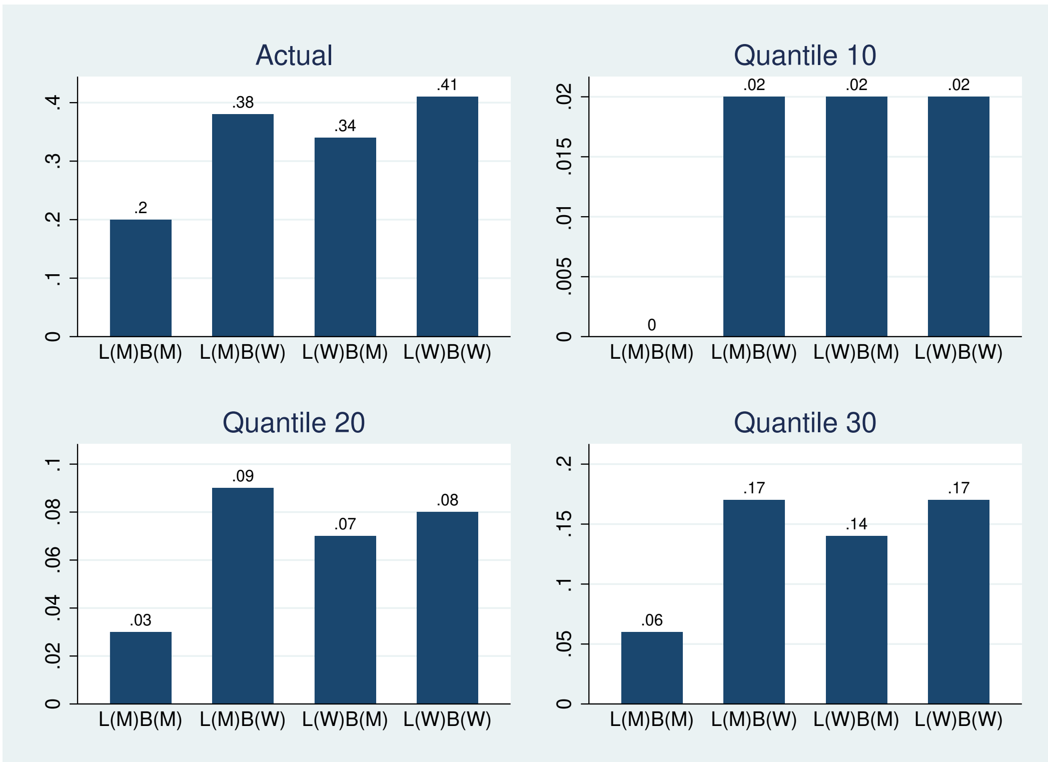

Figure A.5. Borrowers At Risk of Credit Rationing: Asians/Pacific Islanders and Whites
This figure displays the proportion of borrowers at risk of credit rationing. The underlying sample comprises of only Asian/Pacific Islander and White borrowers or brokers. The graph titled "Actual" displays the share of loans that have broker fees exceeding their cap. The other graphs replace the actual broker fees with predicted fees from the regression at the 10^{th}, 20^{th}, or 30^{th} quantile in Table 6. B(M) stands for minority borrower, B(W) stands for white borrower, L(M) stands for minority broker, and L(W) stands for white broker.

Table A.1. New Century Comparison with Overall Subprime Market

Panel A	Demanyanyk and Van Hemert Sample	New Century Sample
Average Loan Size (*1000)	191	206
FRM(%)	23	21
Purchase(%)	38	42
Refinancing (cash out) (%)	54	49
Refinancing(no cash out)(%)	8	8
Fico Score	619	619
CLTV (%)	75	86
Debt-to-Income Ratio(%)	40	40
Investor Dummy (%)	8	9
Low-Doc (%)	35	41
Prepayment Penalty Dummy(%)	73	74
Mortgage Rate(%)	8	8
Panel B	Subprime (excl. New Century) HMDA	New Century HMDA
Minority (%)	52	51

Note: Panel A reports a comparison of summary statistics in Demanyanyk and Van Hemert (2011) with the New Century data sample used in this study. Weighted (by number of loans) averages of summary statistics for loans from 2003 to 2006 are reported in the Demanyanyk column. Panel B reports the share of subprime loan originations to minority borrowers in HMDA. HUD subprime lender lists are used to identify subprime loans. Since the lists are only available through 2005, Panel B includes originations reported in HDMA from 2003 to 2005.

Table A.2. Summary Statistics of Variables

Variable	ALL	HW	BW	AW
Log age	3.70	3.69	3.72	3.70
	(0.27)	(0.27)	(0.28)	(0.28)
Gender	0.39	0.35	0.40	0.36
	(0.49)	(0.48)	(0.49)	(0.48)
Gender unknown	0.00	0.00	0.01	0.01
	(0.07)	(0.07)	(0.07)	(0.07)
Marital status	0.57	0.60	0.56	0.61
	(0.50)	(0.49)	(0.50)	(0.49)
Marital status unknown	0.42	0.39	0.44	0.39
	(0.49)	(0.49)	(0.50)	(0.49)
Log combined income	8.65	8.66	8.61	8.68
	(0.57)	(0.57)	(0.58)	(0.59)
FICO score	619.07	621.62	612.22	618.93
	(60.47)	(60.44)	(59.29)	(60.49)
Subprime	0.50	0.48	0.55	0.50
	(0.50)	(0.50)	(0.50)	(0.50)
Debt-to-income ratio	0.40	0.40	0.40	0.40
	(0.09)	(0.09)	(0.09)	(0.09)
Investor	0.09	0.08	0.10	0.08
	(0.28)	(0.26)	(0.30)	(0.28)
Second home	0.02	0.02	0.02	0.02
	(0.13)	(0.13)	(0.13)	(0.13)
Self-employed	0.24	0.26	0.21	0.25
	(0.43)	(0.44)	(0.41)	(0.43)
2-4 Unit	0.07	0.06	0.06	0.05
	(0.26)	(0.24)	(0.24)	(0.21)
Condominium	0.07	0.07	0.06	0.07
	(0.25)	(0.25)	(0.24)	(0.26)
Face	0.41	0.40	0.36	0.36
	(0.49)	(0.49)	(0.48)	(0.48)
Purchase	0.42	0.41	0.40	0.40
	(0.49)	(0.49)	(0.49)	(0.49)
ARM	0.79	0.79	0.78	0.79
	(0.41)	(0.41)	(0.41)	(0.41)
Cash	0.49	0.50	0.51	0.50
	(0.50)	(0.50)	(0.50)	(0.50)
Co-borrower	0.30	0.33	0.31	0.35
	(0.46)	(0.47)	(0.46)	(0.48)
CLTV below 80%	0.25	0.26	0.25	0.25
	(0.43)	(0.44)	(0.43)	(0.43)
CLTV between 80 and 95%	0.13	0.13	0.13	0.13
	(0.33)	(0.33)	(0.34)	(0.34)
CLTV between 85 and 90%	0.12	0.11	0.12	0.12
	(0.32)	(0.32)	(0.33)	(0.32)
CLTV between 90 and 95	0.16	0.15	0.17	0.16
	(0.37)	(0.36)	(0.38)	(0.37)
CLTV greater than 95%	0.35	0.35	0.33	0.34
	(0.48)	(0.48)	(0.47)	(0.47)

This table reports the means of the variables in this study. Standard deviations are reported in parentheses.

Table A.2. Summary Statistics of Variables (Continued)

Variable	ALL	HW	BW	AW
Interest-only	0.18	0.19	0.14	0.17
	(0.38)	(0.39)	(0.35)	(0.37)
Log loan amount	12.06	12.08	11.96	12.06
	(0.61)	(0.59)	(0.60)	(0.60)
Log loan term.	5.87	5.87	5.87	5.87
	(0.10)	(0.10)	(0.09)	(0.09)
Spread	1.66	1.59	1.78	1.64
	(1.16)	(1.14)	(1.17)	(1.16)
Prepay	0.74	0.75	0.70	0.72
	(0.44)	(0.43)	(0.46)	(0.45)
Stated Income	0.41	0.43	0.35	0.38
	(0.49)	(0.49)	(0.48)	(0.48)
Log distance	3.03	3.04	3.06	3.10
	(1.47)	(1.49)	(1.50)	(1.54)
Broker HHI	0.05	0.05	0.05	0.05
	(0.07)	(0.07)	(0.07)	(0.07)
MSA unemployment	5.14	5.18	5.04	5.05
	(1.45)	(1.50)	(1.28)	(1.34)
Pahl-index	7.85	7.98	7.22	7.30
	(3.75)	(3.79)	(3.78)	(3.74)
Zip per capita income (in $1,000s)	25.97	26.61	27.21	29.19
	(14.28)	(14.62)	(14.46)	(15.31)
College educated	0.16	0.16	0.16	0.17
	(0.05)	(0.05)	(0.05)	(0.05)
Single share	0.28	0.27	0.27	0.27
	(0.05)	(0.05)	(0.05)	(0.05)
Foreigh share	0.15	0.15	0.12	0.12
	(0.11)	(0.11)	(0.10)	(0.09)
Household income (in $1000s)	45.61	45.52	45.74	46.54
	(9.60)	(9.55)	(9.50)	(9.74)
African American share	0.12	0.09	0.14	0.09
	(0.12)	(0.09)	(0.13)	(0.09)
API share	0.05	0.05	0.04	0.05
	(0.07)	(0.05)	(0.05)	(0.08)
Hispanic share	0.18	0.20	0.13	0.13
	(0.17)	(0.18)	(0.13)	(0.13)
Occupancy share	0.64	0.65	0.66	0.67
	(0.10)	(0.10)	(0.10)	(0.10)
Poverty share	0.12	0.12	0.11	0.11
	(0.05)	(0.05)	(0.05)	(0.04)
Rent-Price ratio	1.04	1.05	1.02	1.00
	(1.10)	(1.24)	(0.65)	(0.78)
English share	0.75	0.74	0.80	0.80
	(0.16)	(0.16)	(0.13)	(0.12)
Spanish share	0.15	0.16	0.11	0.11
	(0.14)	(0.14)	(0.10)	(0.09)
Observations	323,846	234,413	213,575	160,936

This table reports the means of the variables in this study. Standard deviations are reported in parentheses.

Table A.3. Example of Broker's Race Assignment

First Name	Last Name	Zip	Posterior Likelihood Distribution: White	Hispanic	African American	Asian or Pacific Islander	AIAN
Edward	Lewis	32608	0.65	0.01	0.32	0.00	0.00
Charles	Clark	34103	0.93	0.00	0.07	0.00	0.00
Amy	Ramos	33405	0.12	0.86	0.00	0.01	0.00
Luis	Moreno	33012	0.00	1.00	0.00	0.00	0.00
Frank	Robinson	34946	0.17	0.00	0.83	0.00	0.00
Calvin	Dawson	32305	0.04	0.00	0.96	0.00	0.00

First Name	Last Name	Zip	MAP	Classification at Threshold: At 85%	At 95%	At 97%
Edward	Lewis	32608	White	Unknown	Unknown	Unknown
Charles	Clark	34103	White	White	Unknown	Unknown
Amy	Ramos	33405	Hispanic	Hispanic	Unknown	Hispanic
Luis	Moreno	33012	Hispanic	Hispanic	Hispanic	Unknown
Frank	Robinson	34946	Black	Unknown	Unknown	Unknown
Calvin	Dawson	32305	Black	Black	Black	Unknown

This table depicts the threshold rule to infer the broker's race. Panel A shows the proportion of Census 2000 respondents with the specified surname by race. Panel B reports the resulting broker race classification by threshold (51, 85, 95 or 97 percent). The threshold is the proportion of the Census 2000 respondents that must self-identify as the same race to gain the classification of that race. Failing to meet the threshold results in an unknown race classification.

Table A.4. Bayesian Classifier Accuracy Rates

Method	Group	FL Voter Data (ZIP)	Accuracy rate NC Voter Data (Tract)	HMDA Data (Tract)
BIFSG	NH White	86%	93%	95%
BIFSG	Hispanic	83%	75%	87%
BIFSG	NH Black	74%	65%	74%
BIFSG	NH API	70%	76%	88%

This table reports the accuracy rate of the MAP classification scheme for race that rely on BIFSG methods. The FL Voter accuracy rates reflect the classification of FL voter data using our algorithm for race. The HMDA accuracy rates reflect that of the BIFSG approach used in **?**. The accuracy rate is measured as the number of observations correctly categorized in the specified group divided by the total number of observations classified into that same group, excluding unclassified observations.

Table A.5. Default Risk Characteristics and Ln(Broker Fees)

	(1)	(2)
Dep. Var.:	Default	Ln(Fees)
CLTV less than 80%	-0.01**	0.08***
	(0.00)	(0.01)
CLTV between 80 and 85%	-0.00	0.04***
	(0.00)	(0.00)
CLTV between 90 and 95%	-0.00	-0.06***
	(0.00)	(0.00)
CLTV greater than 95%	0.01***	-0.01
	(0.00)	(0.01)
FICO between 450 and 550	0.03***	-0.12***
	(0.00)	(0.00)
FICO between 650 and 750	-0.02***	0.03***
	(0.00)	(0.00)
FICO between 750 and 850	-0.03***	-0.01
	(0.00)	(0.01)
Investment property	0.00	-0.13***
	(0.00)	(0.01)
Second home	0.01*	-0.05***
	(0.00)	(0.01)
Stated Income/Low Documentation	0.01***	-0.08***
	(0.00)	(0.00)
Observations	247,395	249,138
Adjusted R-squared	0.08	0.57
Log Loan Amount	Y	Y
Broker FE	Y	Y
Other Controls	Y	Y
Year-Quarter FE	Y	Y
MSA FE	Y	Y

This figure shows that no clear relationship exists between mortgage risk characteristics and mortgage broker compensation. The dependent variable in column (1) is an indicator variable that takes a value of one if the mortgage defaults within 24 months of the origination date. The dependent variable in column (2) is the natural log of broker fees. Borrower controls include age, gender, marital status, income, debt-to-income ratio, and employment status. Loan controls include indicators for broker face-to-face interaction, purchase versus refinance, co-borrowing, combined debt loan-to-value, loan type (adjustable-rate, fixed-rate, interest only), and interest rate spread. Area controls include distance between borrower and broker, broker competition, area unemployment rate, regulatory environment, education, income, share of housing that is owner occupied, price to rent ratio, and county population share that is: married, foreign born, Hispanic, Black, Asian or Pacific Islander, English speaking, and Hispanic speaking. See Table 1 for a complete description of the variables. Robust standard errors clustered by MSA are in parentheses. The stars ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

Table A.6. Distribution of Loans by Broker's Race and Threshold

	At 51%		At 85%		At 95%		At 97%	
	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
API	10,564	3.26	7,693	2.38	5,392	1.66	4,366	1.35
Black	22,520	6.95	9,645	2.98	4,282	1.32	2,836	0.88
Hispanic	57,500	17.76	48,460	14.96	40,214	12.42	36,504	11.27
White	225,610	69.67	179,339	55.38	125,679	38.81	99,437	30.71
Unknown	7,652	2.36	78,709	24.3	148,279	45.79	180,703	55.8
Total	323,846	100	323,846	100	323,846	100	323,846	100

This table displays the distribution of originated loans by the broker's race classified using different frequency thresholds (51, 85, 95 or 97 percent).

Table A.7. OLS Regressions of ln(Broker Fees) Using Different Threshold Classification Schemes

Dep. Var.: ln(Broker Fees)	(1) HW (51)	(2) HW (85)	(3) HW (95)	(4) BW (51)	(5) BW (85)	(6) BW (95)	(7) AW (51)	(8) AW (85)	(9) AW (95)
Minority Borrower	0.04***	0.05***	0.05***	0.05***	0.06***	0.06***	0.03***	0.04***	0.05***
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Minority Borrower ×	0.01	0.00	0.00	-0.05***	-0.04**	-0.02	-0.08**	-0.07**	-0.10***
Minority Broker	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Observations	172,984	145,036	108,790	154,932	118,987	82,700	112,667	94,555	69,027
Adjusted R-squared	0.56	0.56	0.56	0.58	0.58	0.58	0.56	0.56	0.57
Log Loan Amount	Y	Y	Y	Y	Y	Y	Y	Y	Y
Broker FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
MSA FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Minority/Minority Premium	0.05	0.05	0.05	0.01	0.01	0.03	-0.04	-0.03	-0.05
P-value	0.00	0.00	0.00	0.69	0.55	0.36	0.16	0.26	0.09

This table reports OLS estimates. Broker race is inferred using different threshold schemes (51%, 85%, 95%). The dependent variable is the natural log of broker fees. Each reported covariate is a dummy variable that equals one when true, and zero otherwise. Borrower controls include age, gender, marital status, income, credit (FICO) score, a subprime mortgage indicator (FICO<620), debt-to-income ratio, and employment status. Property type controls include owner-occupancy versus investor status and single-family versus condominium or multiple unit structure. Loan controls include indicators for broker face-to-face interaction, purchase versus refinance, co-borrowing, combined debt loan-to-value, loan type (adjustable-rate, fixed-rate, interest only), interest rate spread, and borrower documentation. Area controls include distance between borrower and broker, broker competition, area unemployment rate, regulatory environment, education, income, share of housing that is owner occupied, price to rent ratio, and county population share that is: married, foreign born, Hispanic, Black, Asian or Pacific Islander, English speaking, and Hispanic speaking. See Table 1 for a complete description of the variables. Robust standard errors clustered by MSA are in parentheses. The stars ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

Table A.8. OLS Regressions of ln(Broker Fees) Using BIFSG Scores Directly

Dep. Var.: ln(Broker Fees)	(1) HW	(2) BW	(3) AW
Minority Borrower	0.04***	0.05***	0.03***
	(0.00)	(0.01)	(0.01)
Minority Borrower ×	0.01	-0.06***	-0.07**
Minority Broker	(0.01)	(0.02)	(0.03)
Observations	188,265	172,783	126,794
Adjusted R-squared	0.56	0.58	0.56
Log Loan Amount	Y	Y	Y
Broker FE	Y	Y	Y
Other Controls	Y	Y	Y
Year-Quarter FE	Y	Y	Y
MSA FE	Y	Y	Y
Minority/Minority Premium	0.05	-0.01	-0.05
P-value	0.00	0.63	0.15

This table reports OLS estimates. BIFSG scores, rather than a binary classification system, are included directly in the regression models. The dependent variable is the natural log of broker fees. Each reported covariate is a dummy variable that equals one when true, and zero otherwise. Borrower controls include age, gender, marital status, income, credit (FICO) score, a subprime mortgage indicator (FICO<620), debt-to-income ratio, and employment status. Property type controls include owner-occupancy versus investor status and single-family versus condominium or multiple unit structure. Loan controls include indicators for broker face-to-face interaction, purchase versus refinance, co-borrowing, combined debt loan-to-value, loan type (adjustable-rate, fixed-rate, interest only), interest rate spread, and borrower documentation. Area controls include distance between borrower and broker, broker competition, area unemployment rate, regulatory environment, education, income, share of housing that is owner occupied, price to rent ratio, and county population share that is: married, foreign born, Hispanic, Black, Asian or Pacific Islander, English speaking, and Hispanic speaking. See Table 1 for a complete description of the variables. Robust standard errors clustered by MSA are in parentheses. The stars ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

Table A.9. Two-Step Heckman: Correcting for Origination Uncertainty

	Heckman		OLS	Heckman		OLS	Heckman		OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	HW	HW	HW	BW	BW	BW	AW	AW	AW
Dep. Var.:	1[Originated]	Ln(Fees)	Ln(Fees)	1[Originated]	Ln(Fees)	Ln(Fees)	1[Originated]	Ln(Fees)	Ln(Fees)
Minority Borrower	-0.04***	0.07***	0.07***	-0.08***	0.08***	0.08***	-0.03**	0.04***	0.04***
	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Minority Broker	-0.12***	0.03***	0.03***	-0.16***	0.07***	0.06***	-0.12***	0.04***	0.04***
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
Minority Borrower	0.08***	0.02***	0.02***	0.06***	-0.04***	-0.04***	0.04*	-0.09***	-0.09***
× Minority Broker	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
Subprime rejection rate	-0.52***			-0.51***			-0.58***		
(county/year excluding NCEN)	(0.09)			(0.10)			(0.11)		
Inverse Mills Ratio		0.03			-0.21***			0.00	
		(0.07)			-0.07			(0.08)	
Observations	341,868	341,883	234,405	315,456	315,456	213,566	232,080	232,080	160,930
Censored Obs.	113,701	113,708	N/A	107,958	107,958	N/A	75,478	75,478	N/A
Log Loan Amount	Y	Y	Y	Y	Y	Y	Y	Y	Y
Broker FE	N	N	N	N	N	N	N	N	N
Other Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
MSA FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

This table reports two-step Heckman estimates. The selection models (columns (1), (4), and (7)) set as the dependent variable 1[Originated], a dummy variable that indicates whether the loan is approved and funded. The causal models (columns (2), (5), and (8)) set the natural log of broker fees as the dependent variable and reports the inverse mills ratio inferred from the selection model. Columns (3), (6), and (9) present OLS results for comparison. Borrower controls include age, gender, marital status, income, credit (FICO) score, a subprime mortgage indicator (FICO<620), debt-to-income ratio, and employment status. Property type controls include owner-occupancy versus investor status and single-family versus condominium or multiple unit structure. Loan controls include indicators for broker face-to-face interaction, purchase versus refinance, co-borrowing, combined debt loan-to-value, loan type (adjustable-rate, fixed-rate, interest only), interest rate spread, and borrower documentation. Area controls include distance between borrower and broker, broker competition, area unemployment rate, regulatory environment, education, income, share of housing that is owner occupied, price to rent ratio, and county population share that is: married, foreign born, Hispanic, Black, Asian or Pacific Islander, English speaking, and Hispanic speaking. See Table 1 for a complete description of the variables. Robust standard errors clustered by MSA are in parentheses. The stars ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

Table A.10. OLS of ln(Broker Fees) Post-matching for White Borrowers

Dep. Var.: ln(Broker Fees)	(1) HW White Borrower	(2) BW White Borrower	(3) AW White Borrower
Minority Broker	0.03***	0.05***	0.04**
	(0.01)	(0.01)	(0.02)
Observations	19,396	10,471	7,271
Adjusted R-squared	0.36	0.41	0.35
Log Loan Amount	Y	Y	Y
Broker FE	N	N	N
Other Controls	Y	Y	Y
Year-Quarter FE	Y	Y	Y
MSA FE	Y	Y	Y
Treatment Count	10,426	5,468	3,777
Control Count	8,970	5,003	3,494

This table reports OLS estimates that use ex-post observations from propensity score matches. The dependent variable for each regression is the natural log of broker fees. Each reported covariate is a dummy variable that equals one when true, and zero otherwise. Column (1) compares White borrowers who went to Hispanic brokers with observably similar White borrowers who went to White brokers. Column (2) compares White borrowers who went to Black brokers with observably similar White borrowers who went to White brokers. Column (3) compares White borrowers who went to API brokers with observably similar White borrowers who went to White brokers. Treatment Count reports the number of observations in the group of interest, and Control Count reports the number of counterfactual observations. Borrower controls include age, gender, marital status, income, credit (FICO) score, a subprime mortgage indicator (FICO<620), debt-to-income ratio, and employment status. Property type controls include owner-occupancy versus investor status and single-family versus condominium or multiple unit structure. Loan controls include indicators for broker face-to-face interaction, purchase versus refinance, co-borrowing, combined debt loan-to-value, loan type (adjustable-rate, fixed-rate, interest only), interest rate spread, and borrower documentation. Area controls include distance between borrower and broker, broker competition, area unemployment rate, regulatory environment, education, income, share of housing that is owner occupied, price to rent ratio, and county population share that is: married, foreign born, Hispanic, Black, Asian or Pacific Islander, English speaking, and Hispanic speaking. See Table 1 for a complete description of the variables. Robust standard errors clustered by MSA are in parentheses. The stars ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

Table A.11. OLS of ln(Broker Fees) Post-matching for Minority Borrowers

	(1) HW	(2) BW	(3) AW
Dep. Var.: ln(Broker Fees)	Hispanic Borrower	Black Borrower	API Borrower
Minority Broker	0.03**	-0.00	-0.03*
	(0.01)	(0.01)	(0.02)
Observations	49,163	22,307	4,835
Adjusted R-squared	0.38	0.49	0.31
Log Loan Amount	Y	Y	Y
Broker FE	N	N	N
Other Controls	Y	Y	Y
Year-Quarter FE	Y	Y	Y
MSA FE	Y	Y	Y
Treatment Count	34991	13250	3113
Control Count	14172	9057	1722

This table reports OLS estimates that use ex-post observations from propensity score matches. The dependent variable for each regression is the natural log of broker fees. Each reported covariate is a dummy variable that equals one when true, and zero otherwise. Column (1) compares Hispanic borrowers who went to Hispanic brokers with observably similar Hispanic borrowers who went to white brokers. Column (2) compares Black borrowers who went to Black brokers with observably similar Black borrowers who went to white brokers. Column (3) compares API borrowers who went to API brokers with observably similar API borrowers who went to white brokers. Treatment Count reports the number of observations in the group of interest, and Control Count reports the number of counterfactual observations. Borrower controls include age, gender, marital status, income, credit (FICO) score, a subprime mortgage indicator (FICO<620), debt-to-income ratio, and employment status. Property type controls include owner-occupancy versus investor status and single-family versus condominium or multiple unit structure. Loan controls include indicators for broker face-to-face interaction, purchase versus refinance, co-borrowing, combined debt loan-to-value, loan type (adjustable-rate, fixed-rate, interest only), interest rate spread, and borrower documentation. Area controls include distance between borrower and broker, broker competition, area unemployment rate, regulatory environment, education, income, share of housing that is owner occupied, price to rent ratio, and county population share that is: married, foreign born, Hispanic, Black, Asian or Pacific Islander, English speaking, and Hispanic speaking. See Table 1 for a complete description of the variables. Robust standard errors clustered by MSA are in parentheses. The stars ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

Table A.12. OLS Regressions of ln(Broker Fees) Excluding Non-U.S. Citizens

Dep. Var.: ln(Broker Fees)	(1) HW	(2) BW	(3) AW
Minority Borrower	0.04***	0.05***	0.04***
	(0.00)	(0.00)	(0.01)
Minority Borrower ×	0.01	-0.05***	-0.08**
Minority Broker	(0.01)	(0.02)	(0.03)
Observations	165,549	154,889	111,213
Adjusted R-squared	0.56	0.58	0.56
Log Loan Amount	Y	Y	Y
Broker FE	Y	Y	Y
Other Controls	Y	Y	Y
Year-Quarter FE	Y	Y	Y
MSA FE	Y	Y	Y
Minority/Minority Premium	0.05	0.000	-0.05
P-value	0	.76	.15

This table reports OLS estimates from the sample that excludes borrowers that are not U.S. citizens. The dependent variable is the natural log of broker fees. Each reported covariate is a dummy variable that equals one when true, and zero otherwise. Borrower controls include age, gender, marital status, income, credit (FICO) score, a subprime mortgage indicator (FICO<620), debt-to-income ratio, and employment status. Property type controls include owner-occupancy versus investor status and single-family versus condominium or multiple unit structure. Loan controls include indicators for broker face-to-face interaction, purchase versus refinance, co-borrowing, combined debt loan-to-value, loan type (adjustable-rate, fixed-rate, interest only), interest rate spread, and borrower documentation. Area controls include distance between borrower and broker, broker competition, area unemployment rate, regulatory environment, education, income, share of housing that is owner occupied, price to rent ratio, and county population share that is: married, foreign born, Hispanic, Black, Asian or Pacific Islander, English speaking, and Hispanic speaking. See Table 1 for a complete description of the variables. Robust standard errors clustered by MSA are in parentheses. The stars ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

Table A.13. Minority Premium Across Credit Scores

Dep. Var.: ln(Broker Fees)	(1) White Broker HW	(2) Hisp. Broker HW	(3) White Broker BW	(4) Black Broker BW	(5) White Broker AW	(6) API Broker AW
Minority Borrower	0.02***	-0.02	0.01**	-0.01	0.02	-0.06*
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.04)
Minority Borrower						
× FICO Q2	0.01	0.06***	0.04***	-0.01	-0.01	-0.00
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)
× FICO Q3	0.03***	0.06**	0.05***	0.00	-0.00	0.00
	(0.01)	(0.02)	(0.01)	(0.02)	(0.03)	(0.05)
× FICO Q4	0.05***	0.10***	0.09***	0.07*	0.04**	0.03
	(0.01)	(0.03)	(0.01)	(0.04)	(0.02)	(0.05)
FICO Q2	0.10***	0.05***	0.09***	0.12***	0.10***	0.13***
	(0.01)	(0.02)	(0.00)	(0.02)	(0.00)	(0.02)
FICO Q3	0.12***	0.09***	0.11***	0.14***	0.12***	0.13***
	(0.01)	(0.03)	(0.01)	(0.03)	(0.01)	(0.04)
FICO Q4	0.11***	0.08***	0.10***	0.08**	0.11***	0.12***
	(0.01)	(0.03)	(0.01)	(0.04)	(0.01)	(0.04)
Observations	129,147	46,443	139,890	18,597	106,420	7,811
Adjusted R-squared	0.57	0.51	0.58	0.57	0.56	0.48
Log Loan Amount	Y	Y	Y	Y	Y	Y
Broker FE	Y	Y	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y
MSA FE	Y	Y	Y	Y	Y	Y

This table reports OLS estimates where the dependent variable is the natural log of broker fees. The minority borrower covariate is a dummy variable that equals one when the borrower is a minority, and zero when white. Borrower controls include age, gender, marital status, income, credit (FICO) score quantile dummies, debt-to-income ratio, and employment status. Property type controls include owner-occupancy versus investor status and single-family versus condominium or multiple unit structure. Loan controls include indicators for broker face-to-face interaction, purchase versus refinance, co-borrowing, combined debt loan-to-value, loan type (adjustable-rate, fixed-rate, interest only), interest rate spread, and borrower documentation. Area controls include distance between borrower and broker, broker competition, area unemployment rate, regulatory environment, education, income, share of housing that is owner occupied, price to rent ratio, and county population share that is: married, foreign born, Hispanic, Black, Asian or Pacific Islander, English speaking, and Hispanic speaking. See Table 1 for a complete description of the variables. Robust standard errors clustered by MSA are in parentheses. The stars ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

Table A.14. Minority Effect on Probability of Default Across Credit Scores

Dep. Var.: 1[Default]	(1) White Broker HW	(2) Hispanic Broker HW	(3) White Broker BW	(4) Black Broker BW	(5) White Broker AW	(6) API Broker AW
Minority Borrower	-0.01**	-0.01	0.00	0.02	-0.02	-0.00
	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Minority Borrower						
× FICO Q2	0.00	-0.00	0.00	0.00	0.01	0.02
	(0.00)	(0.01)	(0.00)	(0.01)	(0.02)	(0.02)
× FICO Q3	0.01	0.01	0.00	-0.01	0.01	0.01
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.02)
× FICO Q4	0.01	0.00	0.00	-0.02	0.02	0.00
	(0.00)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
FICO Q2	-0.02***	-0.01**	-0.02***	-0.03***	-0.02***	-0.02*
	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)
FICO Q3	-0.04***	-0.03***	-0.04***	-0.04***	-0.04***	-0.03*
	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)
FICO Q4	-0.05***	-0.04***	-0.06***	-0.05***	-0.05***	-0.03***
	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)
Observations	128,265	46,153	138,792	18,461	105,651	7,766
Adjusted R-squared	0.07	0.04	0.07	0.11	0.07	0.05
Log Loan Amount	Y	Y	Y	Y	Y	Y
Broker FE	Y	Y	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y
MSA FE	Y	Y	Y	Y	Y	Y

This table reports OLS estimates where the dependent variable is an indicator variable that takes a value of one if the mortgage defaults within 24 months of the origination date. The minority borrower covariate is a dummy variable that equals one when the borrower is a minority, and zero when white. Borrower controls include age, gender, marital status, income, credit (FICO) score quantile dummies, debt-to-income ratio, and employment status. Property type controls include owner-occupancy versus investor status and single-family versus condominium or multiple unit structure. Loan controls include indicators for broker face-to-face interaction, purchase versus refinance, co-borrowing, combined debt loan-to-value, loan type (adjustable-rate, fixed-rate, interest only), interest rate spread, and borrower documentation. Area controls include distance between borrower and broker, broker competition, area unemployment rate, regulatory environment, education, income, share of housing that is owner occupied, price to rent ratio, and county population share that is: married, foreign born, Hispanic, Black, Asian or Pacific Islander, English speaking, and Hispanic speaking. See Table 1 for a complete description of the variables. Robust standard errors clustered by MSA are in parentheses. The stars ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.