J Real Estate Finan Econ (2019) 59:549–588
https://doi.org/10.1007/s11146-018-9683-y

An Unintended Consequence of Mortgage Financing Regulation – a Racial Disparity

James B. Kau[1] · **Lu Fang**[2] · **Henry J. Munneke**[1]

Published online: 12 November 2018

Abstract
This study investigates whether mortgage financing regulation unintentionally leads to minorities paying a higher loan contract rate under a risk-based pricing system. We provide evidence that minority borrowers prepay less frequently than comparable non-minority borrowers and thus have lower termination risk. Racially neutral lending policies prohibit the lender from considering this reduced termination risk, resulting in a disparate impact from the overstatement of a minority borrower's termination risk. While we find little evidence of a rate differential among borrowers under the current regulatory structure, results show minorities pay a higher rate when the variation in termination risk is recognized.

Keywords Mortgage discrimination · Prepayment · Default · And regulation

JEL Classifications G21 · G28 · J15 · R20

Introduction

This study focuses on the loan pricing behavior of lenders under racially neutral lending policies and further investigates if mortgage financing regulation may unintentionally lead to minorities paying more for a mortgage loan. The issue of racial discrimination in mortgage lending has consistently drawn the attention of researchers since the enactment of the Equal Credit Opportunity Act (ECOA) in 1974. Ross and Yinger (2002) and Turner

✉ Lu Fang
fangl@longwood.edu

James B. Kau
jkau@uga.edu

Henry J. Munneke
hmunneke@uga.edu

[1] Terry College of Business, University of Georgia, Athens, GA 30602, USA

[2] College of Business & Economics, Longwood University, Farmville, VA 23909, USA

and Skidmore (1999) note that racial discrimination in mortgage markets manifests itself in several forms. Differential treatment discrimination occurs when equally qualified individuals are treated differently due to explicitly considering prohibited factors such as a borrower's race or ethnicity. For example, lenders may deny loan applications of minority applicants with a higher frequency, steer minority borrowers into the less favorable instruments (e.g., subprime mortgage market) or price loans differently given the borrower's level of risk. Simply put, differential treatment discrimination means that applicants with equivalent credit-related characteristics are treated differently. Disparate-impact discrimination occurs when lenders apply a racially neutral policy to all borrowers, however the policy itself disproportionately excludes or burdens borrowers in a certain group. An extensive literature exists that has examined these forms of discrimination and specifically whether fair-lending laws and regulations have successfully eliminated race-based disparity in credit allocations and in loan pricing.[1] Overall, the results are mixed. The focus of this study on the loan pricing behavior of lenders is consistent with the lending industry's shift towards a risk-based pricing system. Under this system, borrowers with different risk levels have access to credit, but at rates relative to their risk (Turner and Skidmore 1999; Ghent et al. 2014). This implies borrowers with equivalent levels of risk should pay equivalent prices.

Using a sample of 30-year first-lien fixed-rate subprime mortgage loans for home purchase, this study evaluates the existence of contract rate disparity between racial/ethnic groups within a framework which mimics lenders' loan pricing behavior. We first employ a competing-risks hazard model to simultaneously assess the default and prepayment risks of each loan in the sample based on the loan performance data. In the loan hazard model, the effects of a borrower's race as well as the effects of the racial composition of a borrower's neighborhood are assessed. For each loan in the sample, the probability of a borrower defaulting upon or prepaying a loan is predicted based on the loan hazard model. The loan-level predicted default and prepayment probabilities are then incorporated into a loan contract rate determination model to investigate whether race-based disparity in mortgage loan pricing exists after explicitly controlling for termination risk.

The results of the competing-risks loan hazard model indicate African American borrowers and Hispanic borrowers tend to be less likely to exercise their prepayment option than similar non-Hispanic white borrowers. However, we fail to reject the null of no impact on default risk of the race and ethnicity variables. In a completely competitive market, lower prepayment risk of loans reduces borrowing cost and enhances returns to lenders, thus it should be reflected in loan pricing and result in more favorable loan terms. However, the current fair-lending laws and regulations prohibit lenders from using a borrower's race and ethnicity to assess a borrower's risk profile, leading lenders to over-assess the risk level of a loan to a minority borrower. Overstating a borrower's termination risk leads to a higher contract rate being paid. The estimation results indicate African American borrowers pay an additional 15–45 basis points and Hispanic borrowers pay roughly 10–24 extra basis points for their contract rate.

[1] Black et al. (1978) and Munnell et al. (1996) represent early studies focusing on the issue of credit allocation. Holmes and Horvitz (1994), Tootell (1996), and Ross and Tootell (2004) are examples of studies emphasizing the issue of redlining, analyzing whether the racial composition of an applicant's neighborhood affects loan credit allocations. A few recent studies investigate the existence of disparate treatment by mortgage loan originators (Ross et al. 2008; Hanson et al. 2016) using experimental methods. Turner and Skidmore (1996), Ladd (1998), LaCour-Little (1999), and Ross and Yinger (2002) provide thorough and detailed reviews of previous studies.

The methods applied in this study make it noticeably different from prior studies. In most of the prior racial studies in mortgage lending, the existence of racial disparity in the contract rate is evaluated through a reduced-form contract rate determination model. In a reduced-form model, a loan's contract rate is regressed against a borrower's race or the racial composition of a borrower's neighborhood, as well as a set of covariates that are believed to be associated with the risk level of a loan. One potential problem with this approach is that a reduced-form model does not enable us to explicitly explore the effects of race and ethnicity on loan termination patterns and to separately derive mortgage pricing implications of these termination patterns. By contrast, the analysis utilized in this study allows us to mimic lenders' pricing behavior by predicting the probability of a borrower defaulting upon or prepaying a loan and explicitly account for loan termination risk on the contract rate by including predicted probabilities of termination. More importantly, while prior studies using the reduced-form model normally focus on default risk factors while ignoring the impact of prepayment risk on the loan's contract rate, this study simultaneously assesses the competing risks of mortgage default and prepayment. Kau et al. (1992) note that default and prepayment are substitutes for one another and one cannot accurately value either option without the other. In addition, numerous authors have emphasized the importance of incorporating prepayment risk in the valuation of mortgage-backed security (e.g., see Dunn and McConnell 1981a & 1981b; Schwartz and Torous 1989, Schwartz and Torous 1992; Chernov et al. 2017). Firestone et al. (2007) note that in practice default is a relatively rare event relative to mortgage prepayment (0.6% vs 92% in their study). Using this framework, we are able to explore whether borrowers of different racial and ethnic groups have different prepayment risk, and thus enabling us to determine whether fair-lending laws and regulations lead to minority borrowers paying a significantly greater contract rate.

Literature Review

Discrimination against minority borrowers in loan pricing may be directed at the individual borrower or based on the racial and ethnic composition of a borrower's neighborhood (redlining).[2]

The Survey of Consumer Finance (SCF) data and the American Housing Survey (AHS) data are widely used in prior studies to examine mortgage discrimination in loan pricing as both data sets have information on a borrower's race. Using the 1983 SCF data and focusing on conventional fixed-rate mortgage loans, Duca and Rosenthal (1994) did not find evidence of racial discrimination in the conventional mortgage market. Getter (2006), using more recent 1998 and 2001 SCF data which better differentiates households with severe delinquency problems from households with minor repayment problems, arrived at the same conclusion - a borrower's race does not affect a loan's contract rate. However, Cheng et al. (2015) using the 2001, 2004, and 2007 SCF data consisting of various types of mortgage products, found that racial disparity in the mortgage contract rate is both statistically and economically significant. The newer SCF data used by Cheng et al. (2015) contain detailed information on a

[2] The focus of our review of the literature is to examine racial and ethnic disparity in loan pricing. For literature on racial discrimination in credit allocations, please see Turner and Skidmore (1996) and Ross and Yinger (2002) for the summary.

borrower's race as well as the borrower's characteristics including income, wealth, debts, credit quality, age, and education. Based on the 1989–2001 AHS data, Boehm et al. (2006) employed the Blinder's decomposition approach commonly used in empirical studies on labor discrimination and found empirical evidence of racial discrimination in the mortgage market.

Although both the SCF data and the AHS data appear to provide rich data sources for analysis of racial disparity in mortgage lending, they have some drawbacks. The SCF data include detailed information on the borrower, but do not contain detailed information on the attributes of the loan or the collateral. The AHS data, which is a household-level data set, fails to provide information on a household's financial condition (e.g., a borrower's credit score), but includes detailed information on a borrower's personal traits, the loan terms, and the collateral. These limitations have led to the use of proprietary data sets to investigate the issue of racial disparity in mortgage lending, with the hope these data will contain information to better account for risk.

Based on a proprietary data set of conventional loans and FHA/VA loans originated by a national home mortgage lender from 1988 to 1989, Crawford and Rosenblatt (1999) focused on differences in yield premiums across racial groups. They conclude that conventional loan interest rates are race-neutral, and although African American borrowers are shown to pay a smaller premium on average, the difference is not economically significant. Black et al. (2003) utilize another proprietary data set from a major mortgage lending institution to investigate whether racial differences in overage pricing exist in either the home purchase mortgage market or refinancing market.[3] Their results indicate minority borrowers are significantly more likely to pay a positive overage than similarly situated white borrowers for home purchase mortgage loans. However, they did not find any significant difference in overage pricing across racial groups in the refinancing mortgage market. Courchane (2007), using a sample of conventional home purchase loans and refinancing loans, fail to provide any evidence of racial discrimination in either of the two mortgage markets. By contrast, Zhang (2013), by matching proprietary loan data from a national bank to HMDA data to obtain information on a borrower's race, finds non-Hispanic Asians and non-Hispanic African Americans tend to receive a lower price for their first-lien conventional loans for home purchase.

Other studies have investigated the issue of racial discrimination in mortgage lending by emphasizing the issue of redlining. Nothaft and Perry (2002) match their loan origination data to census survey data to obtain the racial composition of a borrower's neighborhood. They find that borrowers in predominately Hispanic or Asian neighborhoods pay slightly higher interest rates, while borrowers in predominately African American neighborhoods occasionally pay slightly lower interest rates. By contrast, also using census survey data to measure the racial composition of a borrower's neighborhood, Kau et al. (2012) reveal that borrowers in predominantly African American neighborhoods or Hispanic neighborhoods tend to prepay less frequently, and after the predicted loan default and prepayment probabilities are controlled for, only borrowers in predominately African American neighborhoods pay significantly higher contract rates.

[3] An overage is defined as "a difference between the price at which a loan closes and the minimum price acceptable to the lending institution as quoted on the lender's rate sheet". For details, see Black et al. (2003).

A recent study by Ghent et al. (2014), utilizing loan level data combined with Home Mortgage Disclosure Act (HMDA) data to obtain a borrower's race and ethnicity as well as the racial and ethnic composition of a borrower's neighborhood, simultaneously investigate the impact of both a borrower's race and the neighborhood's traits on loan contract rate after accounting for the predicted loan default and prepayment probabilities.[4] Their results provide empirical evidence of adverse pricing against African Americans and Hispanics. In addition, they also find that the evidence of adverse pricing is strongest for home purchase mortgage loans and loans originated by non-depository institutions. Bayer et al. (2017) find that minority borrowers tend to be more likely to obtain a high cost loan, and racial and ethnic disparities can be largely explained by the lender's foreclosure risk.

The current study utilizes similar technique of Kau et al. (2012), but is able to focus on the impact of the individual borrower's race and ethnicity on loan pricing. The technique employed allows the current study to relax the orthogonality assumption of Ghent et al. (2014) so that the contract rate of a loan is allowed to impact loan performance. Additionally, a semi-parametric matching approach is utilized in the current study as a robust test. The results show minorities may pay higher mortgage rates despite regulation that prohibits discrimination, but the results are weak in that the discrimination result is not present in a matched sample; however, the study unambiguously shows minorities pay higher rates than would be the case if regulation allowed lenders to compensate minorities that prepay less frequently by offering those minorities lower mortgage interest rates.

Model

The framework of the modeling is established from a lender's perspective, as a loan's contract rate is determined at the time of loan origination. In setting the contract rate of a loan, the risk level of the loan at origination is assessed, including the likelihood of a borrower to default upon and to prepay a loan over the loan's life. It is assumed lenders have an accurate model to evaluate loan termination risk. In our model, the borrower's loan termination behavior is derived from the borrower's default and prepayment behavior observed in the sample.

Borrower's Loan Termination Behavior Model

The borrower's loan termination behavior is modeled using the Cox discrete-time competing-risks loan hazard model (Deng et al. 2000). The model takes into consideration that default and prepayment are substitutes (Kau et al. 1992). The borrower's loan termination behavior is observed and modeled at monthly intervals (in other words, for each loan, the information is restructured to have one observation for each month in which the loan is current). For each month of each loan subsequent to loan origination, loan performance data is used to determine whether a borrower continued, prepaid, or defaulted upon a loan. This discrete-time model solves the issue of left truncation and right censoring which are common issues in the mortgage literature, and

[4] The default and prepayment hazards are estimated using two separate models.

allows the loan default and prepayment probabilities to be estimated appropriately.[5] Specifically, a multinomial logit model is utilized to model a borrower's loan termination behavior as[6]:

$$\ln\left(p_{jt}/p_{0t}\right) = \delta_j \alpha_{jt} + \varphi_j{}' \mathbf{q}_{jt} + \beta_j{}' \mathbf{x}_{jt} + \varepsilon_{jt} \qquad j = 1, 2 \tag{1}$$

where p_{0t} is the probability of a loan being continued in period t; p_{jt} is the probability of a loan being terminated in period t given this loan had been continued by the beginning of period t, where $j = 1$ is the probability of default at time t, and $j = 2$ is the probability of prepayment at time t. In this study, t refers to mortgage time. The variable α_{jt} is the baseline hazard rate for default and prepayment respectively, and is allowed to vary by mortgage time. The vector $\mathbf{q}_{jt}$ includes a set of variables indicating a borrower's race and ethnic group, as well as variables measuring the racial and ethnic composition of a borrower's neighborhood at mortgage time t. The vector of non-racial covariates ($\mathbf{x}_{jt}$) describes the observed characteristics of the loan, the borrower, the collateral, the neighborhood, and the economic conditions. These covariates may or may not be time varying.

It is assumed in the model that the possible future mortgage interest rate path is of particular interest to the lender when predicting loan termination probabilities and is among the time-varying covariates in x_{jt}.[7] In this study, the 10-year treasury constant maturity yield is adopted as the benchmark for the mortgage interest rate of 30-year fixed-rate residential mortgage loans.[8] In order to predict the possible future 10-year yield path seen from loan origination $y_t(y_0)$, the commonly used Cox, Ingersoll and Ross (CIR) term structure model is employed.[9] In this term structure model, the whole term structure is assumed to be driven by a spot interest rate ($r(t)$). This spot interest rate is assumed to follow a mean-reverting stochastic process with volatility affected by the level of the spot rate as shown in Eq. (2):

$$dr(t) = \gamma(\theta - r(t))dt + \sigma\sqrt{r(t)}dz(t) \tag{2}$$

[5] See Fang and Munneke (2016) for a discussion on why the discrete-time competing-risks loan hazard model can solve the issue of left truncation and right censoring.

[6] The multinomial logit model is chosen with the independence of irrelevant alternatives (IIA) assumption that the odds ratio for any pair of choices is assumed to be independent of any third alternative (one event is not informative to the other conditional on all of the covariates in the model), and choices at any point in time are independent of those at any other point in time. Based on the IIA assumption, a widely adopted approach was utilized here to estimate this multinomial logit model in which we estimated default hazard and prepay hazard separately. Hence, it is not necessary to estimate default and prepay hazard models within a simultaneous equation framework, especially given the findings that studies show separate models perform well for most of the data (Allison 2010). The advantage of estimating default and prepay hazard models separately also includes the flexibility in having different specifications for different hazard models.

[7] For other time-varying covariates, we either used the actual values if we were able to observe them, or extrapolated values from the known values, whichever seemed more reasonable and appropriate.

[8] This 10-year treasury constant maturity yield is widely used as the benchmark for the mortgage interest rate of 30-year fixed-rate residential mortgage loans in practice.

[9] Though several studies in asset pricing argue other interest rate models perform better than the CIR term structure model with respect to out-of-sample prediction, those models could only be employed to forecast the mean, not the density of the spot interest rate needed here. In addition, the CIR term structure model is the standard model used in mortgage literature.

where the first term on the right side of the equation is the deterministic part with θ as the long-term mean of the spot interest rate and γ as the reversion rate, whereas the second part describes the stochastic movements.

Using the estimated parameters in Equation (2),[10] the density of future spot interest rate for any forecast interval conditional on the spot interest rate at origination $dF(r(t)|\, r(0))$ is forecasted, based on the transition density of the spot interest rate implied by the CIR term structure model.[11] As a change in the spot interest rate in the future ($r(t)$) is believed to result in a change in the 10-year yield in the future ($y_t(r(t))$), and the latter impacts the probability of a borrower defaulting upon or prepaying a loan, the forecasted conditional density of future spot interest rate $dF(r(t)|\, r(0))$ is employed to predict default probability and prepayment probability for each loan in each period t as in Equations (3) and (4)[12]:

$$\hat{p}_{1t}(y_0) = \int p_{1t}[y_t(r(t))]\, dF(r(t)|r(0)) \tag{3}$$

$$\hat{p}_{2t}(y_0) = \int p_{2t}[y_t(r(t))]\ \ dF(r(t)|r(0)) \tag{4}$$

where $\hat{p}_{1t}(y_0)$ is the predicted default probability in period t seen from loan origination given that this loan had been continued by the beginning of period t and $\hat{p}_{2t}(y_0)$ is the predicted prepayment probability in period t. Here, as the forecasted conditional future spot interest rate $(r(t)|\, r(0))$ is a continuous stochastic variable, in Equations (3) and (4) the integrated expectations are numerically approximated through a discretization approach in which the spot interest rate domain was divided into numerous but finite intervals.

The predicted probability of each termination event in any particular period t, calculated based on Equations (3) and (4), is aggregated over a 10-year span to generate a total predicted probability of each event $\hat{P}_k$ ($k = 1, 2$) seen from origination as in Equation (5).[13]

[10] The parameters in Equation (2) were estimated with the use of 4 time series of yields with different maturities from 1987 to 2007 within the framework of the single-factor CIR term structure model. Those 4 time series are 6-month T-bill yield, 1-year Fama-Bliss bond yield, 3-year Fama-Bliss bond yield, and 5-year Fama-Bliss bond yield. Data were obtained from CRSP. The reason we chose this estimation period (from 1987 to 2007) is many studies have found there was a shift in Federal Reserve monetary policy in the early 1980s (Duan and Simonato 1999) and the loan data in this study ends in 2007 based on loan origination year. We used the GAUSS code offered by Jin-Chuan Duan on his website to implement the estimation, the one he used to yield the results in Duan and Simonato (1999). We would like to acknowledge this help from him.

[11] Notice here, this study forecasts the conditional density of the future spot interest rate rather than the simple conditional mean to account for all of the possible path of future interest rate. For the transition density of the spot interest rate, see Cox, Ingersoll, Ross (1985). Here, a normal distribution was used to closely approximate the true transition density.

[12] Note here, as we use the forecasted conditional density not the forecasted conditional mean to calculate the predicted probability of each event at time t ($\hat{p}_{kt}, k = 1, 2$), the future market mortgage interest rate at time t (the future 10-year yield here, $y_t(r(t))$) could be any positive value in the spectrum. For each specific value of $y_t(r(t))$, the predicted probability of each event at time t based on that specific value could be calculated as $p_{kt}[y_t(r(t))]$. As in the single-factor CIR model, the whole term structure is driven by the spot interest rate $r(t)$, the distribution of the future 10-year yield at time t is determined by the distribution of the future spot interest rate at time t specified as $dF(r(t)|\, r(0))$. Therefore, considering all of the possible values of $y_t(r(t))$, the predicted probability of each event at time t is calculated as in Equation (3) and (4).

[13] In this study, we chose a 10-year span instead of a 30-year span because in reality, most of the 30-year fixed rate mortgage loans are prepaid within the first ten years if they were not defaulted upon. Notice here, a capital P is used to distinguish total loan termination probabilities from time-specific loan termination probabilities. The subscript k tells the type of the event, 1 for default and 2 for prepayment.

This reflects the lenders' concern on the total predicted probability of each event over the life of a loan rather than that of a time-specific predicted probability for a particular period *t*.

$$\hat{P}_k = \sum_{t=1}^{T}\left(1+\frac{y_0}{12}\right)^{-t}\hat{p}_{kt}\prod_{s=1}^{t-1}\left(1-\sum_{k=1}^{2}\hat{p}_{ks}\right) \qquad k=1,2 \tag{5}$$

Here, $\hat{p}_{kt}$ is the predicted probability of event *k* in period *t* given that the loan had been continued by the beginning of period *t* with the probability as $\prod_{s=1}^{t-1}\left(1-\sum_{k=1}^{2}\hat{p}_{ks}\right)$. Hence, $\hat{p}_{kt}\prod_{s=1}^{t-1}\left(1-\sum_{k=1}^{2}\hat{p}_{ks}\right)$ is the unconditional predicted probability of event *k* in period *t*. These probabilities are discounted by the 10-year yield at origination (y_0) with the assumption that loan termination at earlier stages of a loan is more severe to the lenders. Summation of discounted unconditional time-specific predicted probability of event *k* over a 10-year window results in a total predicted probability of each event ($\hat{P}_k$).

Loan Contract Rate Determination Model

Risk-based pricing by lenders depends on the assessment of a loan's risk level. To capture termination risk, the total predicted probabilities ($\hat{P}_k$) based on borrower's termination model are included in a linear loan contract rate determination model, Eq. (6). Such a model can be written as:

$$C_0 = \alpha_0 y_0 + \beta_1\hat{P}_1 + \beta_2\hat{P}_2 + \lambda^{'}\mathbf{b} + \xi^{'}\mathbf{z} + \varepsilon \tag{6}$$

In addition to the predicted termination probabilities, the model includes: 1) the 10-year yield at origination (y_0) which is believed to be the benchmark used to set the interest rate for 30-year fixed-rate mortgage loans; 2) a set of race and ethnicity variables, **b,** including variables indicating a borrower's race and ethnicity group as well as measures of the racial and ethnic composition of a borrower's neighborhood at origination; and 3) a set of non-racial variables at origination (**z**) describing the traits of the loan, the borrower, the collateral, and the collateral's neighborhood. The estimation of this model allows one to test whether a borrower's race and ethnicity have an effect on the contract rate after controlling for termination risk as well as the benchmark interest rate.

Data and Specifications

Data

The data used in this study contain 30-year first-lien fixed-rate home-purchase residential mortgage loans serviced by GMAC Residential Capital Company, LLC (GMAC ResCap). GMAC ResCap was a finance firm that provided home financing, loan servicing, and mortgage-back securities (MBS) issuance in the U.S. before the recent financial crisis. Loans in this data were all packaged into private-label mortgage-backed securities and traded in the secondary mortgage market. As the loan servicer, GMAC ResCap collected detailed information on the loan, borrower, and collateral at the time of loan origination. In addition, they also tracked the performance of each loan they serviced on a monthly basis. The monthly loan performance data provides information on the current interest rate and balance of a loan, as well as prepayment and delinquency status.

GMAC ResCap data do not provide information on a borrower's race or ethnicity. This information is available in the Home Mortgage Disclosure Act Loan/Application Register (HMDA-LAR) data. The Home Mortgage Disclosure Act (HMDA) is a federal law that was enacted by the Congress in 1975 and implemented by the Federal Financial Institutions Examination Council (FFIEC). Starting in 1989, amendments to HDMA, resulting from the Financial Institutions Reform, Recovery, and Enforcement Act (FIRREA), required financial institutions to collect and disclose data on race, gender, and income of the applicant and the co-applicant (if applicable) for each loan application (the LAR data). It is estimated that in 2007, more than 8800 lending institutions reported their data to FFIEC, accounting for approximately 80% of all home lending nationwide (Avery et al. 2007).[14]

To identify a borrower's race and ethnicity, the loan data from GMAC ResCap were matched with the HMDA-LAR data. One difficulty in matching the data is the two data sets define the location of the property underlying the loan differently. Loans in the GMAC ResCap data are reported based on the zip code, while loan applications in the HMDA-LAR data are reported on the basis of the census tract. In addition, the definition of a census tract adopted in HMDA-LAR data varies across time.[15] To address this issue, we first matched the GMAC ResCap loan data to a source of readily available property transactions data from Miami-Dade County, FL, in order to identify the geographic coordinates of the property securing each loan. To make this matching feasible, the GMAC ResCap loan data were restricted to 30-year first-lien fixed-rate home-purchase mortgage loans with underlying properties located in Miami. This loan-property-sales matching is conducted based on the value of the underlying property, property sale month (loan origination month), property type, and zip code, leading to a matched loan-property-sales sample.[16] Using geographic coordinates from the property sales, the location of the loans within the 1980, 1990, and 2000 census tracts are identified. Each matched loan-property-sale combination was then matched to approved conventional loan applications in the HMDA-LAR data based on the loan amount (in thousand dollars), loan origination year (loan action year), property type, property occupancy status, loan purpose, lien status (if applicable), and census tract.[17]

[14] The FIRREA amendments in 1989 expanded the coverage of HMDA to many independent non-depository lending institutions, in addition to the previously covered savings associations, banks, and credit unions. For detailed information on who is required to report HMDA data, see the descriptions on the Federal Financial Institutions Examination Council (FFIEC) website https://www.ffiec.gov/hmda/reporter.htm.

[15] HMDA-LAR data from 1989 to 1991 used the 1980 census tract boundaries, data from 1992 to 2002 followed the 1990 census tract definitions, and data from 2003 to 2012 adopted 2000 census tract definitions.

[16] Each mortgage loan was matched to property sales in the pool with replacements requiring that the gap between the appraised value of the property in the loan data and the transaction price of the property in the property sale data to be the minimum one in the pool. If multiple property sales were matched to a loan and those multiple properties were located in the same census tract, this loan is treated as having a unique property sale match, and its census tract identification number is identified as the common one for those multiple properties.

[17] As the loan amount in HMDA-LAR data is in thousand dollars, we allowed loan amount to differ by up to $1000. Lien status is a new field added to the HMDA-LAR data beginning January 1, 2004.

Only loans with a unique HMDA-LAR match are included for analysis in this study.[18]

One issue with the use of a borrower's race and ethnic information in the HMDA-LAR data is the reporting rules changed significantly in 2004 with regard to the classification of race and ethnicity. Prior to 2004, an applicant (co-applicant) was required to self-identify his/her race from among the five racial classifications: American Indian, Asian, African American, Hispanic, and White. However, since 2004, applicants were asked to separately report their ethnicity (Hispanic or non-Hispanic) and race (American Indian, Asian, African American, Native Hawaiian, and White). In order to align the race and ethnicity data pre- and post-2004, a hierarchy among all of the race and ethnic classifications is adopted.[19] Specifically, if an applicant post-2004 self-identified as an African American/American Indian/Asian/Native Hawaiian, no matter whether this applicant self-identified as a Hispanic or not, this applicant is treated as an African American/American Indian/Asian/Native Hawaiian. If an applicant reported Hispanic for ethnicity and White for race, he/she is treated as a Hispanic. Under this hierarchy, white borrowers are non-Hispanic whites. If the applicant and the co-applicant chose different racial and ethnic groups, a hierarchy ordering similar to the one described above is used. If either of the two applicants self-identified as an African American, that loan is considered to be taken by an African American borrower. If one applicant self-identified as a Hispanic and the other self-reported as a non-Hispanic white, that loan is treated as one to a Hispanic. If both of the two borrowers self-reported as non-Hispanic whites, the mortgage loan is considered to be given to a non-Hispanic white.

The matched loan-property-HMDA data were augmented with other data sources used to describe the characteristics of the underlying property's neighborhood, based on the property's location. The decennial census survey data, normalized to the 1990 census tract boundaries, was used to generate time-varying variables describing a neighborhood's traits (including housing occupancy rate and poverty rate) on a monthly basis by using a linear time-trend between the decennial data from 1990 to 2000 and from 2000 to 2010. In addition, the yearly HMDA data was aggregated on the census-tract level to create measures of the demographic characteristics of each census tract, specifically, measures of median applicant income as well as the racial and ethnic composition of a neighborhood.[20] These aggregated HMDA neighborhood variables are based on a three-year window (the previous year, the current year, and the next year). To generate these variables, we use all of the loan applications including applications that were approved or denied for 1-to-4 family dwelling purchases in a

[18] In this study, if there are multiple HMDA-LAR matches for a loan that meet those matching criteria above, and if those multiple matches have exactly the same race and ethnicity information on the borrower and the co-borrower, this loan is allowed to be identified as having a unique HMDA-LAR match, since the goal of the data matching is to obtain the race and ethnicity information on the borrower(s).

[19] Avery et al. (2007) discussed several hierarchies to solve this issue, and this study adopted one hierarchy that is reasonable here given the demographic characteristics of the population in Miami-Dade County, FL.

[20] In this study, the census-tract level aggregated HMDA data is preferred rather than the decennial census survey data to generate the time-varying variables describing the demographic characteristics of a census tract, because the aggregated HMDA data is updated every year and is believed to be more accurate than census survey data which is updated every ten years. Those measures include the median applicant income, the proportion of African American applicants, the proportion of non-Hispanic white applicants, and the proportion of Hispanic applicants. The applicant income is inflation adjusted by a GDP per capita deflator. All income is defined in 2009 dollars. The racial and ethnic group of a loan application is identified using the hierarchy described above.

census tract across all lenders. The resulting yearly census-tract level measures were further linearly time-trended to generate monthly measures.

Characteristics of the local housing market are measured by the change of and variation in housing prices within a neighborhood (census tract) using the property transactions data.[21] A median housing price index was generated to measure the changes in house price over the life of a loan relative to the time of origination. The inflation-adjusted median house price index was calculated for each census tract based on the 1990 census tract definitions, and for each month of analysis by creating a three-year window of sale, eighteen months before and eighteen months after that month.[22] The resulting index is a unique median house price index for each census tract on a monthly basis. The standard deviation of the housing sale prices is also calculated monthly for each census tract using the same three-year window to measure the heterogeneity in housing sale price in a neighborhood. Recent house price appreciation rate is calculated for each loan at origination using the change in the median house price on a census tract in the two 3-year periods prior to origination.[23]

There were initially 1994 loans originated in Miami-Dade County, FL that could be matched to the property transaction and HMDA-LAR data sets. Loans with missing values on the loan, the borrower, the collateral or the neighborhood characteristics, or loans without completed loan performance records were deleted. The sample is restricted to loans originated from Jan. 1997 to Dec. 2006. The window could not be extended beyond 2006, due to very few loans being originated beyond this point. In addition, since only a few loans were originated to American Indian, or Asian, or Native Hawaiian borrowers, they were removed from the data. The final sample consists of 1404 observations of 30-year first-lien, fixed-rate subprime residential mortgage loans for home purchase. Their performance was tracked from Jan. 2000 to Oct. 2010, a period covering the recent mortgage crisis.

Model Specifications and Endogeneity Issue

On the basis of the option-theoretic model of mortgage loan “financial” termination (Kau et al. 1992), default and prepayment are driven by different financial incentives. Therefore, different baseline hazard rates (α_{jt}) are applied to the default and prepayment hazards. A scaled Standard Default Assumption schedule (SDA) is used for default hazard and mortgage year fixed effects are employed to measure the prepayment baseline hazard rate.[24] Kau et al. (1992) note that changes in market interest rates and changes in the value of the collateral are the two most prominent time-varying factors affecting a borrower’s decision to default or prepay. In this discrete-time model, the market interest rate change at time t is measured by the gap between the 10-year

[21] The data on the pool of property transactions are sales from 1990 to 2013 in Miami-Dade County, FL.

[22] The housing sale price is inflation adjusted by a GDP per capita deflator. All prices are defined in 2009 dollars.

[23] Recent housing price appreciation rate at origination is defined as the ratio of the median housing sale price in a neighborhood in a three-year period prior to the month of loan origination to the median housing sale price in the same neighborhood in another three-year period prior to the three-year pre-origination period, then minus 1.

[24] The traditional Public Securities Association (PSA) schedule is not used because previous studies argued this schedule did not describe the pattern of actual prepayments well, for more details, see Kau et al. (2004).

Treasury constant maturity yield at loan origination and the 10-year yield at time t lagged by 2 periods ($y_0 - y_{t-2}$).[25] To measure the change in the value of the underlying house, a monthly census-tract-level median housing price index is employed with the assumption that the value of a house changes at the same pace as the median housing price index in its census tract. For each mortgage time t, in order to reflect the changes in the house price since loan origination, the ratio of the median house price index level at time t to the one at origination is calculated, and this ratio is named the relative house price at time t (RHP_t). Additionally, an interaction term of the market interest rate change ($y_0 - y_{t-2}$) and relative house price at time t (RHP_t) is also included in this competing-risks model to account for any correlation between interest rate change and house price change.

Variables describing the traits of the loan, borrower, property, and neighborhood are also included in the hazard models. Loan traits include the loan contract rate spread at origination ($C_0 - y_0$), original LTV, FICO score of the borrower, original loan amount, loan origination season fixed effects, and dichotomous indicators for whether a borrower provided full income documentation at origination, whether the loan is encumbered by a prepayment penalty at time t, whether the underlying property is owner occupied, and whether the property is a single family detached house or a condo.[26] Also, several time-varying variables are included to measure the characteristics and evolution of an underlying property's neighborhood, including the heterogeneity in housing sale price, housing occupancy rate, poverty rate, and the median income of loan applicants at time t.[27]

The specification of the default and prepayment models differs. In the default equation, the prepayment penalty variable at time t is excluded, whereas in the prepayment equation, the original LTV ratio, the income documentation status variable, and the variable of neighborhood-level heterogeneity in housing price at time t are excluded, because these variables are not believed to directly affect the corresponding hazard, and also because of identification purpose.[28]

The contract rate C_0 appears in the multinomial logit Equation (1) as part of the loan contract rate spread and will be endogenous if there are omitted variables in the estimation of the contract rate equation that are correlated with the errors in the estimation of Equation (1). In other words, if a lender believes a borrower is more

[25] A yield at time t lagged by 2 periods is used because in practice there is usually a gap between a borrower's decision and actual termination, and borrowers typically rely on past information to make their decisions. Notice here, for the first mortgage month and second mortgage month, the 10-year yield at loan origination (at time 0) is used as the 10-year yield at time t lagged by 2 periods.

[26] Contract rate spread at origination is defined as the gap between the contract rate and the 10-year yield at origination. We chose the contract rate spread instead of contract rate itself because the 10-year yield as the benchmark mortgage interest rate varied considerably within our study period, and this spread allows us to make comparisons across mortgage vintages.

[27] Heterogeneity in housing sale price at time t was measured by the standard deviation of housing sale price within a three-year window prior to time t in a neighborhood. Housing occupancy rate and poverty rate in a neighborhood at time t are from the census survey data. Median income of loan applicants in a neighborhood at time t is generated by the aggregated HMDA data described in the data subsection. In order to reflect the rank of each census tract in terms of median applicant income in Miami-Dade County, FL at time t, the ratio of the median applicant income in a census tract at time t to the median applicant income in Miami-Dade County, FL at time t was calculated and included in the model.

[28] We also tested whether those excluded variables affect the corresponding hazard respectively. None of the coefficient estimates of those variables are statistically significant.

likely to default or prepay, the lender would charge a higher contract rate (spread), however being charged a higher contract rate (spread) makes the borrower more likely to terminate a loan. Since the usual two-stage least square procedure is inappropriate of a non-linear logit model, we follow Petrin and Train (2003) and use a control function method to address this endogeneity issue. To implement the CF method, a reduced-form contract rate model is estimated in which the contract rate is regressed against all of the exogenous variables in the system. The residual from the reduced-form contract rate model is then included in the estimation of the multinomial logit model to solve the potential issue of endogeneity. Recall that the termination probabilities (p_{jt}) are estimated based on the monthly observations of each loan and then combined with the estimated CIR term structure model according to Eqs. (3) and (4), and the results aggregated over time, according to Eq. (5), to obtain the expected probabilities ($\hat{P}_k$) at origination.

Endogeneity is also an issue in the contract rate model (Eq. (6)) due to each of the predicted probabilities ($\hat{P}_k$) in the final contract rate equation being a function of the contract rate (spread). To address this issue, a set of generated IVs ($\tilde{P}_k$) are used in the estimation of the final contract rate in place of the corresponding predicted probabilities ($\hat{P}_k$). Each generated IV ($\tilde{P}_k$) is calculated as its corresponding generated variable ($\hat{P}_k$), but with the actual contract rate spread ($C_0 - y_0$) being replaced with a predicted contract rate spread ($\hat{C}_0 - y_0$). The predicted contract rate ($\hat{C}_0$) is obtained from the estimation of a reduced-form contract rate equation (the same reduced-form contract rate model used in the CF method, results not reported in the paper). The predicted contract rate ($\hat{C}_0$) is a function of all of the exogenous variables. Thus, each generated IV ($\tilde{P}_k$) is exogenous and serves as a valid IV for its corresponding generated variable ($\hat{P}_k$) in the final loan contract rate equation. With those generated IVs, 2SLS is used to estimate the contract rate model, since it is a linear in the parameters.

Other covariates included in the loan contract rate determination model are the 10-year yield at origination (y_0), as well as a set of covariates ($\mathbf{z}$) that theory suggests should affect loan pricing.[29] These additional covariates include the original LTV, FICO score of the borrower, original loan amount, loan origination season fixed effects, prepayment penalty fixed effects, borrower's income documentation status fixed effects, underlying property type fixed effects, property occupancy status fixed effects, and a list of neighborhood characteristics at loan origination including recent house price appreciation rate, heterogeneity in housing sale price, housing occupancy rate, poverty rate, and median income of loan applicants. Individual- and neighborhood-level race and ethnicity variables are included to test whether race and ethnicity affect loan pricing. Additionally, a trend term in calendar time is also included. Table 11 in Appendix A lists all of the variables used in this study and provides a detailed description of each variable.

In both the loan hazard model and loan contract rate model, some variables take a nonlinear function form based on prior studies and/or theoretical reasons. The original LTV ratio is transformed into categorical variables: loans with LTV ratio less than or

[29] This set of covariates are included in this loan contract rate determination model mainly for the identification purpose. We tested whether the main results change if those covariates are excluded. The main results remain the same.

equal to 80%, loans with LTV ratio greater than 80% but less than or equal to 90%, loans with LTV ratio greater than 90% but less than or equal to 100%, and loans with LTV ratio exceeding 100%. Furthermore, FICO score is entered as a continuous linear spline function with a knot point at 700 based on the assumption that once the FICO score is above a threshold (e.g., 700), an additional increase in FICO score would have relatively small marginal effects on loan termination probabilities/loan contract rate.[30] Finally, following prior studies, a quadratic function form of the original loan amount is utilized to allow for a non-linear relationship.

Statistical Descriptions

Table 1 presents a brief description of the characteristics of the loans at the time of loan origination in the pooled sample, including the traits of the loan, borrower, collateral, and neighborhood. The average original LTV ratio in the sample is approximately 85%, and 88 loans (6.27%) have a LTV ratio greater than 100%. More than half of the borrowers (52.35%) in the sample failed to provide full income documentation and 18.52% of the borrowers (260 loans) have a FICO score below 650. Approximately 33% of the loans are encumbered by a prepayment penalty. The average contract rate of 7.96% is relatively high, and is approximately 304 basis points higher than the 10-year treasury yield at origination. During the observation period of loan performance in this study, 122 loans (8.69%) of the 1404 loans were defaulted upon and ended up with a foreclosure/short sale/deed in lieu of foreclosure, while 1129 loans (80.41%) were prepaid. In this study, default is defined as the occurrence of a borrower being 90-days delinquent, and that occurrence eventually results in a foreclosure, short sale, or deed in lieu of foreclosure.

Table 2 provides descriptive statistics of the loans by the racial and ethnic group of the primary borrower. Around 71% of the loans (995 loans) in the sample were extended to Hispanic borrowers, about 22% of the loans (309 loans) were taken by non-Hispanic white borrowers, and approximately 7% of the loans (100 loans) were to African American borrowers. These proportions are close to the proportions from the census survey describing the racial and ethnic composition of all the residents in Miami-Dade County, FL. This indicates the working sample represents the overall Miami-Dade population.[31]

Based on the information in Table 2, borrowers of different racial and ethnic groups appear to have different loan termination patterns. The observed average default rate within the study period by African American borrowers is around 13.00%, approximately twice as high as the observed default rate by non-Hispanic white borrowers

[30] The FICO linear spline function was specified as follows: $FICO_{(FICO \leq 700)}$ = minimum (FICO, 700); and $FICO_{(FICO > 700)}$ = maximum(FICO, 700)-700. Therefore, the coefficient on $FICO_{(FICO \leq 700)}$ measures the effects of FICO score on dependent variable when FICO≤700; while coefficient on $FICO_{(FICO > 700)}$ measures the marginal effects of FICO score when FICO>700. We tested whether the results are robust to the specification of the FICO score knot point by conducting the same analysis with a knot point at 720 or 750, and results are robust.

[31] The proportion of African American borrowers in the sample of this study is slightly lower than the proportion of African American residents in Miami-Dade County, FL indicated by the census survey (approximately 20% in 2000 census survey data). This leaves a research question for future study on whether African Americans face more difficulties having access to credit than borrowers in other racial and ethnic groups. As we do not have data on loan applications, this research question is beyond the scope of this study.

Table 1 Descriptive statistics of mortgage loans at loan origination

Variable Name	Mean	Std. Dev.	Min.	Max.
Default (0,1)	0.0869	0.2818	0.0000	1.0000
Prepay (0,1)	0.8041	0.3970	0.0000	1.0000
Loan Characteristics				
Contract rate at origination (C_0)	7.9622	1.1282	5.2500	12.5000
10-year treasury yield at origination (y_0)	4.9201	0.8163	3.3300	6.8900
Contract rate spread at origination (C_0-y_0)	3.0422	0.9476	1.0100	6.4650
Original LTV	84.8048	11.8555	36.0000	107.0000
FICO at origination	700.0370	53.2957	483.0000	822.0000
Original loan amount (in $10,000)	18.1243	12.1792	2.0000	80.0000
Full income documentation (0,1)	0.4765	0.4996	0.0000	1.0000
Without prepayment penalty (0,1)	0.6702	0.4703	0.0000	1.0000
Prepayment penalty for 1 to 3 years (0,1)	0.2236	0.4168	0.0000	1.0000
Prepayment penalty for 5 years (0,1)	0.1061	0.3081	0.0000	1.0000
Property Characteristics				
Property owner occupied (0,1)	0.7778	0.4159	0.0000	1.0000
Property condo (0,1)	0.3376	0.4731	0.0000	1.0000
Neighborhood-Level Characteristics at Origination [a]				
Recent housing price appreciation rate [b]	0.1949	0.2227	−0.4519	2.8150
Heterogeneity in housing price (in $10,000) [c]	11.2656	7.8752	1.1012	43.3243
Housing occupancy rate (from Census Survey) [d]	0.9094	0.0858	0.5373	0.9883
Poverty rate (from Census Survey) [d]	0.1441	0.0826	0.0216	0.5988
Median applicant income (from HMDA) [e]	1.2188	0.5786	0.5483	4.8597
Proportion of non-Hispanic white applicants (from HMDA) [f]	0.2069	0.1729	0.0171	0.7069
Proportion of Hispanic applicants (from HMDA) [f]	0.6902	0.2161	0.0742	0.9778
Proportion of African American applicants (from HMDA) [f]	0.0775	0.1439	0.0000	0.8399
Sample Size	1404			

[a] The neighborhood of a loan's underlying property is defined based on the property's 1990 census tract boundaries

[b] Recent housing price appreciation rate at origination is defined as the ratio of the median housing sale price in a census tract in a three-year period prior to the month of loan origination to the median housing sale price in the same census tract in another three-year period prior to the three-year pre-origination period, then minus 1

[c] Heterogeneity in housing price at origination is defined as the standard deviation of the housing sale price in a census tract over a three-year period prior to the month of loan origination

[d] Housing occupancy rate and poverty rate were generated from the decennial census survey data in 1990, 2000, and 2010

[e] Median applicant income was generated from the HMDA data aggregated on the census tract level on a yearly basis. It is defined as the ratio of the median applicant income in a census tract at origination to the median applicant income in Miami-Dade County, FL at origination

[f] Variables on the racial and ethnic composition of a census tract were generated from the HMDA data aggregated on the census tract level on a yearly basis

(6.47%). Meanwhile, the observed prepayment rate by African Americas (77.00%) or by Hispanics (78.39%) is lower than that by non-Hispanic whites (88.03%). The observed differences in loan termination patterns among the three groups might be explained by the differences in the credit risk of a borrower(s). African Americans

Table 2 Descriptive statistics of mortgage loans at loan origination by race

	Race Group		
	Non-Hispanic white	Hispanic	African American
Variable Name	Mean	Mean	Mean
Default (0,1)	0.0647	0.0894	0.1300
Prepay (0,1)	0.8803	0.7839	0.7700
Loan Characteristics			
Contract rate at origination (C_0)	7.9198	7.9140	8.5731
10-year treasury yield at origination (y_0)	5.1425	4.8526	4.9040
Contract rate spread at origination (C_0-y_0)	2.7773	3.0614	3.6691
Original LTV	81.2395	85.4995	88.9100
FICO at origination	712.7605	698.0683	680.3100
Original loan amount (in $10,000)	22.0816	17.4520	12.5855
Full income documentation (0,1)	0.4822	0.4633	0.5900
Without prepayment penalty (0,1)	0.2201	0.3497	0.4700
Prepayment penalty for 1 to 3 years (0,1)	0.0939	0.2492	0.3700
Prepayment penalty for 5 years (0,1)	0.1262	0.1005	0.1000
Property Characteristics			
Property owner occupied (0,1)	0.7767	0.7799	0.7600
Property condo (0,1)	0.3495	0.3508	0.1700
Neighborhood-Level Characteristics at Origination [a]			
Recent housing price appreciation rate [b]	0.1815	0.2016	0.1693
Heterogeneity in housing price (in $10,000) [c]	15.8298	10.2828	6.9411
Housing occupancy rate (from Census Survey) [d]	0.8748	0.9196	0.9148
Poverty rate (from Census Survey) [d]	0.1314	0.1410	0.2142
Median applicant income (from HMDA) [e]	1.5494	1.1506	0.8755
Proportion of non-Hispanic white applicants (from HMDA) [f]	0.3545	0.1676	0.1415
Proportion of Hispanic applicants (from HMDA) [f]	0.5469	0.7567	0.4708
Proportion of African American applicants (from HMDA) [f]	0.0646	0.0533	0.3587
Number of loans	309	995	100

[a] The neighborhood of a loan's underlying property is defined based on the property's 1990 census tract boundaries

[b] Recent housing price appreciation rate at origination is defined as the ratio of the median housing sale price in a census tract in a three-year period prior to the month of loan origination to the median housing sale price in the same census tract in another three-year period prior to the three-year pre-origination period, then minus 1

[c] Heterogeneity in housing price at origination is defined as the standard deviation of the housing sale price in a census tract over a three-year period prior to the month of loan origination

[d] Housing occupancy rate and poverty rate were generated from the decennial census survey data in 1990, 2000, and 2010

[e] Median applicant income was generated from the HMDA data aggregated on the census tract level on a yearly basis. It is defined as the ratio of the median applicant income in a census tract at origination to the median applicant income in Miami-Dade County, FL at origination

[f] Variables on the racial and ethnic composition of a census tract were generated from the HMDA data aggregated on the census tract level on a yearly basis

appear to have the highest credit risk simply based on the average original LTV and the FICO score, while non-Hispanic whites appear to have the lowest risk based on these measures. Additionally, African Americans seem to be more likely to be encumbered by a prepayment penalty. African American borrowers also tend to live in a neighborhood with a relatively lower recent housing price appreciation rate, a higher poverty rate, and lower median income. The observed differences in the characteristics of the loan, the borrower, and the neighborhood among the three groups indicate the importance of accounting for all of those traits when examining the relationship between a borrower's race and loan hazard probability/ loan contract rate.

Results

Modeling Discrimination

The individual- and neighborhood-level race and ethnicity variables at time t are included in the estimation of the loan hazard probability to test if they are associated with the default and prepayment hazards after a set of loan risk factors in $\mathbf{x}_{jt}$ are accounted for.[32] Ross and Yinger (2002) note that under fair-lending laws, it is illegal for lenders to use a borrower's race and ethnicity as a proxy for unobserved risk factors by including race and ethnicity in the calculation of the predicted termination probabilities. They further note that all of the race and ethnicity variables should be included in the estimation of the loan hazard model, otherwise parameters of other variables in this hazard model would be biased if they are correlated with those race and ethnicity variables. Our base case strictly follows this guidance and thus fair-lending laws and regulations by including race and ethnicity variables in the estimation of the loan hazard model, but excluding (setting their coefficients to zero) them from predicting loan termination probabilities. This avoids what Ross and Yinger (2002) call statistical discrimination. We refer to these predicted probabilities as *color-neutral*. Separately, we build a *regulation-free* model where the termination probabilities are forecasted including the impact of race and ethnicity variables. This approach allows one to measure the impact of the over-assessment of termination risk in the contract rate.

Do Minorities Have Different Loan Termination Patterns?

The loan default hazard estimates, reported in Table 3, do not provide any empirical evidence that a borrower's race and ethnicity are associated with loan default probability, as the estimates on the individual-level race and ethnicity variables are not statistically significant. In addition, the neighborhood-level race and ethnicity variables are not shown to affect loan default probability. These findings indicate that the differences in the observed average default rate across the three racial and ethnic groups reported in Table 2 are fully explained by the differences in the covariates incorporated in the default hazard model.

[32] The variables describing the racial and ethnic composition of loan applicants in a census tract at time t are generated by the aggregated HMDA data described in the data subsection.

Table 3 Default hazard model estimates

Variables	Coef.	*P value*
Intercept	6.4472	0.1871
Market Characteristics		
Market interest rate change at time t (y_0-y_{t-2}) [a]	−0.3274	0.5332
Relative house price at time t (RHP_t) [a]	−2.9636	<0.0001
(y_0-y_{t-2}) × (RHP_t) [a]	0.6122	0.2020
Loan Characteristics		
Contract rate spread at origination (C_0-y_0)	0.0565	0.8660
Original LTV categories (Base group: Original LTV < =80)		
80 < Original LTV < =90	0.0231	0.9356
90 < Original LTV < =100	0.5379	0.1627
100 < Original LTV	1.2272	0.0423
FICO at origination continuous linear splines		
Minimum (FICO, 700)	−0.0126	0.0114
Maximum (FICO, 700) -700	−0.0005	0.8960
Original loan amount (in $10,000)	0.1378	<0.0001
Square term of original loan amount (in $10,000)	−0.0012	0.0128
Full income documentation (0,1)	−0.6791	0.0049
Property Characteristics		
Property owner occupied (0,1)	−0.5632	0.0265
Property condo (0,1)	0.3315	0.2197
Neighborhood-Level Characteristics		
Housing price heterogeneity at time t (in $10,000) [a]	0.0285	0.2367
Housing occupancy rate at time t [a]	−1.9855	0.1832
Poverty rate at time t [a]	1.5569	0.2772
Median applicant income at time t (HMDA) [a]	−1.7441	0.0002
Proportion of African Americans at time t (HMDA) [a]	−0.0234	0.9864
Proportion of Hispanics at time t (HMDA) [a]	−0.2288	0.8156
Borrower Race (Base group: Non-Hispanic white)		
African Americans	0.2831	0.5033
Hispanics	−0.0603	0.8228
Control Function Variable		
Residual [b]	0.6561	0.0611
Time Characteristics		
SDA [a]	1.0497	0.0043
Origination season fixed effects [c]	YES	
Likelihood Ratio	230.8994	

[a] Denotes time-varying variables

[b] The residual comes from the contract-rate reduced-form estimation

[c] The seasonal loan origination fixed effects estimates are omitted here, but are available upon request

Overall, the results indicate a borrower's decision to default is mainly driven by financial incentives. The change in the value of the collateral affects the default probability. The coefficient on the relative house price at time t (RHP_t) variable is statistically significant and negative. This is consistent with the option theory of mortgage loan termination (Kau et al. 1992) – as house price decreases, borrowers lose equity on their houses and tend to be more likely to default. In addition, estimation results on the variables describing the traits of the loan, the borrower, and the collateral, demonstrate that borrowers with an original LTV ratio above 100%, a lower FICO score, a larger original loan size, limited income documentation, or borrowers using a loan to purchase a house as investment rather than as their primary residence tend to be more likely to default on the loan. Note that the estimation results on the two FICO score variables indicate that when a borrower's FICO score is above a threshold, 700, an additional increase in the FICO score does not have significant marginal effects on default likelihood. Meanwhile, the original loan size has a positive impact on default probability but at a decreasing rate as the original loan size increases. Among the neighborhood-level variables, only the estimate on the median applicant income variable is significant and negative. The results indicate borrowers living in a relatively poorer neighborhood tend to be more likely to default. These results are consistent with most of the prior studies.

Table 4 demonstrates the estimation results on prepayment hazard. The estimation results on the individual-level race and ethnicity variables indicate both African American borrowers and Hispanic borrowers are less likely to prepay than non-Hispanic white borrowers after other covariates are controlled for. This result is consistent with prior studies focusing on prepayment patterns of minority borrowers (Kelly 1995; Clapp et al. 2001; Deng and Gabriel 2006; Firestone et al. 2007). It is worth noting that all of those prior studies have found that minority borrowers are less likely to prepay. Possible explanations for this finding are those minority borrowers might face more obstacles gaining access to credit in the refinance market, have limited knowledge or information of mortgage refinance opportunities, or are less mobile. Because of the limitation of the data, we leave this question to future research. However, the results clearly demonstrate minority borrowers prepay less frequently and therefore have a relatively lower lending cost than similar non-Hispanic white counterparts.

Overall, the results on other covariates are in accordance with theoretical expectations and prior studies. The results indicate changes in market interest rates play an important role in a borrower's decision to prepay. As the market interest rate drops, borrowers are more likely to prepay, as rational borrowers always have financial incentives to take advantage of a lower interest rate through prepayment (Kau et al. 1992). In addition, the likelihood of a borrower prepaying a loan is shown to also be affected by the change in the value of the collateral. In a market seeing great house price appreciation, borrowers would have more incentives to prepay their loans to cash out the equity on their house.

In addition to market interest rate change and house price change, other factors are shown to affect the probability of loan prepayment. The contract rate spread at origination ($C_0 - y_0$) appears to be positively associated with loan prepayment probability. The contract rate spread ($C_0 - y_0$) measures the extent to which the contract rate at closing (C_0) deviates from the baseline interest rate (y_0), a measure of risk premium. The results indicate that a borrower with a higher risk premium tends to be more likely to prepay. This result may also reflect the impact of points used to buy down a contract rate. As theory suggests, a borrower anticipating a lower likelihood of prepaying a loan

Table 4 Prepay hazard model estimates

Variables	Coef.	*P value*
Intercept	−10.8479	<0.0001
Market Characteristics		
Market interest rate change at time t (y_0-y_{t-2}) [a]	0.5741	0.0002
Relative house price at time t (RHP_t) [a]	1.3042	<0.0001
(y_0-y_{t-2}) × (RHP_t) [a]	−0.0734	0.5298
Loan Characteristics		
Contract rate spread at origination (C_0-y_0)	0.4834	<0.0001
FICO at origination continuous linear splines		
Minimum (FICO, 700)	0.0044	0.0020
Maximum (FICO, 700)-700	−0.0021	0.1001
Original loan amount (in \$10,000)	0.0301	0.0046
Square term of original loan amount (in \$10,000)	−0.0005	0.0111
Within prepayment penalty period at time t (0,1) [a]	−0.5057	<0.0001
Property Characteristics		
Property owner occupied (0,1)	0.0206	0.8047
Property condo (0,1)	−0.2132	0.0183
Neighborhood-Level Characteristics		
Housing occupancy rate at time t [a]	0.3369	0.4778
Poverty rate at time t [a]	−0.3579	0.4657
Median applicant income at time t (HMDA) [a]	0.1426	0.0964
Proportion of African Americans at time t (HMDA) [a]	0.1357	0.7342
Proportion of Hispanics at time t (HMDA) [a]	0.0402	0.8874
Borrower Race (Base group: Non-Hispanic white)		
African Americans	−0.3291	0.0400
Hispanics	−0.2465	0.0036
Control Function Variable		
Residual [b]	−0.0479	0.6555
Time Characteristics		
Mortgage year fixed effects [c]	YES	
Origination season fixed effects [c]	YES	
Likelihood Ratio	620.2642	

[a] Denotes time-varying variables

[b] The residual comes from the contract-rate reduced-form estimation

[c] Mortgage year fixed effects estimates and the seasonal loan origination fixed effects estimates are omitted here, but are available upon request

in the near future self-selects more points at loan closing in exchange for a reduced contract rate. Thus, the likelihood of a loan being prepaid is positively associated with loan contract rate (spread).[33] Estimation results on other covariates demonstrate

[33] Points are not included in the data set.

borrowers with a higher FICO score or a larger loan size appear to be more likely to prepay. Meanwhile, borrowers encumbered by a prepayment penalty, or borrowers using a loan to buy a condo versus a single-family detached house tend to be less likely to prepay, ceteris paribus.

Do Minorities Pay Higher Contract Rates?

The results in Table 5 indicate that when a lender sets a loan's contract rate, the lender takes into account how likely a borrower is to default or prepay in the future, as the coefficients on the predicted loan termination probability variables are significant and positive. Specifically, a 10-percentage-point increase in the default probability leads to a 15 basis point increase in the contract rate. A 10-percentage-point increase in the prepayment probability would lead to an increase in the contract rate by 26 basis points.

The estimation results on other covariates in the loan contract rate determination equation are consistent with either theoretical expectations and/or prior studies. The estimate on the 10-year yield at origination is significantly positive and close to 1. The estimates on the three category variables of original LTV ratio are all significant and positive. The results show as the original LTV ratio increases, the loan contract rate rises monotonically. A borrower's FICO score also plays an important role in determining a loan's contract rate. Lenders are shown to charge a higher contract rate for borrowers with a lower FICO score. However, when a borrower's FICO score is above 700, we fail to reject that an additional increase in FICO score does not marginally impact loan contract rate. In addition, lenders tend to charge a lower contract rate to borrowers who provide full income documentation at origination, as well as borrowers who use a loan to purchase a house as the primary residence rather than as an investment. Borrowers using a loan to purchase a condo are charged a higher contract rate on average than borrowers using a loan to purchase a single-family detached house. Among the neighborhood traits variables, the recent house price appreciation rate and the housing occupancy rate at origination are shown to affect loan contract rate in a negative way.

The estimation results on the original loan size variables and two prepayment penalty category variables require additional explanations. The results on the original loan size indicate as the original loan size rises, the contract rate decreases, but at a decreasing rate. This finding is consistent with most of the previous related studies. One possible explanation for this finding is that the original loan size is believed to be positively associated with a borrower's income and wealth, and the latter is anticipated to be negatively correlated with loan contract rate. The estimates on the two category prepayment penalty variables are shown to be significant and positive. This finding may initially be somewhat puzzling given that the prepayment penalty transfers a portion of the prepayment risk from the lender to the borrower, and thus should reduce the contract rate. The results show that as the length of the penalty period increases, the contract rate decreases, indicating the penalty itself reduces the loan contract rate. The observed positive relationship between the prepayment penalty and loan contract rate reflects that borrowers selecting a prepayment penalty tend to be riskier, other things equal.[34]

Turning our attention to the estimates on the individual- and neighborhood-level race and ethnicity variables. The results indicate an African American borrower pays a contract rate approximately 20 basis points higher than a non-Hispanic white borrower, ceteris

[34] A prior study by Mayer et al. (2013) provided evidence for this explanation.

Table 5 Loan contract rate estimates (Color neutral) -2SLS [a]

Variables	Coef.	*P value* [b]
Intercept	9.5309	<0.0001
Termination Risk		
IV[c] - Predicted default probability	1.5370	<0.0001
IV[c] - Predicted prepayment probability	2.6425	<0.0001
Loan Characteristics		
10-year treasury yield at origination (y_0)	0.8194	<0.0001
Original LTV categories (Base group: Original LTV < =80)		
80 < Original LTV < =90	0.1646	<0.0001
90 < Original LTV < =100	0.3892	<0.0001
100 < Original LTV	0.5986	<0.0001
FICO at origination continuous linear splines		
Minimum (FICO, 700)	−0.0105	<0.0001
Maximum (FICO, 700)-700	0.0002	0.7061
Original loan amount (in $10,000)	−0.0359	<0.0001
Square term of original loan amount (in $10,000)	0.0005	<0.0001
Full income documentation (0,1)	−0.1160	0.0021
Prepayment penalty categories (Base group: No prepayment penalty)		
Prepayment penalty for 1–3 years	0.6069	<0.0001
Prepayment penalty for 5 years	0.4308	<0.0001
Property Characteristics	−0.1046	0.0084
Property owner occupied (0,1)	−0.1046	0.0084
Property condo (0,1)	0.1291	0.0040
Neighborhood-Level Characteristics		
Recent housing price appreciation rate at origination	−0.1834	0.0447
Housing price heterogeneity at origination (in $10,000)	−0.0011	0.7941
Housing occupancy rate at origination	−0.4359	0.0464
Poverty rate at origination	0.1659	0.4740
Median applicant income at origination (HMDA)	−0.0333	0.4761
Proportion of African Americans at origination (HMDA)	0.0173	0.9269
Proportion of Hispanics at origination (HMDA)	0.1691	0.1921
Borrower Race (Base group: Non-Hispanic Whites)		
African Americans	0.2011	0.0111
Hispanics	−0.0197	0.5942
Market/Time Characteristics		
Time Trend	0.0004	0.9698
Loan origination season fixed effects [d]		
Number of loans	1404	
Adjusted R^2	0.7599	

[a] The contract rate determination model was estimated with *color-neutral* predicted loan termination probabilities incorporated, ignoring the impact of race and ethnicity on loan performance

[b] The method outlined in Appendix 6A of Wooldridge (2010) was employed to correct the standard errors and to calculate the *p values* to account for the presence of generated regressors

[c.] The predicted default and prepayment probabilities ($\hat{P}_k$) are generated variables and are endogenous given they are a function of the contract rate (spread). To address this issue, a set of generated IVs ($\tilde{P}_k$) are used, each serves as a valid IV for its corresponding generated variable ($\hat{P}_k$). Each generated IV ($\tilde{P}_k$) is calculated as its corresponding generated variable ($\hat{P}_k$), but with the actual contract rate spread ($C_0 - y_0$) being replaced with a predicted contract rate spread ($\hat{C}_0 - y_0$). The predicted contract rate ($\hat{C}_0$) is obtained from the estimation of a reduced-form contract rate equation

[d] The seasonal loan origination fixed effects estimates are omitted here, but are available upon request

paribus. Meanwhile, we fail to reject the null that the racial and ethnic composition of a borrower's neighborhood, or being a Hispanic borrower, does not have a significant impact on a loan's contract rate. Overall, we find empirical evidence of adverse pricing against African American borrowers in the subprime mortgage market over this study period, after controlling for the predicted loan termination probabilities as well as a set of covariates in this contract rate determination model. This result may be driven by aggressive lenders or high-risk lenders operating in minority markets (Ross et al. 2008) or members of minority groups may be less likely to comparison shop for mortgage products. Whatever the reason, African American borrowers are treated unfavorably in the mortgage market even with the presence of fair-lending laws and regulations.

Do Fair Lending Laws and Regulations Help or Hurt Minorities?

Recall that the results on the prepayment hazard model (Table 4) show that after controlling for a full set of covariates, minority borrowers including African American borrowers and Hispanic borrowers are less likely to exercise the prepayment option than non-Hispanic white borrowers. Therefore, those minorities have a relatively lower lending cost than their non-Hispanic white counterparts, ceteris paribus. In a competitive lending market, loans with a relatively lower lending cost should be originated at a relatively lower contract rate.[35] However, fair lending laws and regulations prohibit lenders from taking a borrower's race and ethnicity into account in assessing a loan's risk level. In other words, because lenders are unable to take a borrower's race and ethnicity into account in assessing a loan's risk level, lenders are overcompensated for loan termination risk when dealing with loans to minorities and thus disadvantaging minority borrowers.

In order to explore the extent to which minority borrowers are disadvantaged by those fair lending regulations, we estimate the contract rate determination model (Equation (6)) after allowing the termination probabilities to vary by a borrower's race and ethnicity. Specifically, we re-calculate the termination probabilities based on the estimates from Eq. (1) allowing the probabilities to vary by race and ethnicity. We name these predicted probabilities *regulation-free* to distinguish them from the *color-neutral* probabilities used in the prior analysis. The *regulation-free* scenario allows lenders to recognize the lower termination risk of minority borrowers. The primary hypothesis of the *regulation-free* model is that the coefficients on the individual race and ethnicity variables will *increase* to reflect the lower rate these borrowers should be charged due to their lower level of prepayment risk. The results of the estimation of the contract rate model with *regulation-free* probabilities are reported in Table 6. The results show that both of the individual-level race and ethnicity fixed effects variables are significant and positive. Specifically, the coefficient estimate on the variable of African American is approximately 35 basis points, while the coefficient estimate on the variable of Hispanic is nearly 10 basis points. Recall that in the *color-neutral* base case with fair-lending laws and regulations, only one

[35] This argument is supported in a few prior studies analyzing loan termination patterns of minority borrowers. Deng and Gabriel (2006) using a competing-risks loan hazard model found minority borrowers prepay their mortgage loans more slowly, but defaulted more. However, considering both default and prepayment risks, they revealed that the elevated default risks of loans by minority borrowers are more than offset by the damped prepayment speeds of those loans. Therefore, they argued that those damped termination risks of loans by minority borrowers should be reflected in the pricing of those loans, and the efficient risk-based pricing of loans should serve to enhance mortgage and housing affordability among those minority populations.

Table 6 Loan contract rate estimates (Regulation free) -2SLS [a]

Variables	Coef.	*P value* [b]
Intercept	9.6391	<0.0001
Termination Risk		
IV[c] - Predicted default probability	1.2754	<0.0001
IV[c] - Predicted prepayment probability	2.4349	<0.0001
Loan Characteristics		
10-year treasury yield at origination (y_0)	0.8091	<0.0001
Original LTV categories (Base group: Original LTV < =80)		
80 < Original LTV < =90	0.1633	<0.0001
90 < Original LTV < =100	0.3851	<0.0001
100 < Original LTV	0.5902	<0.0001
FICO at origination continuous linear splines		
Minimum (FICO, 700)	−0.0104	<0.0001
Maximum (FICO, 700)-700	0.0002	0.6596
Original loan amount (in $10,000)	−0.0354	<0.0001
Square term of original loan amount (in $10,000)	0.0005	<0.0001
Full income documentation (0,1)	−0.1147	0.0018
Prepayment penalty categories (Base group: No prepayment penalty)		
Prepayment penalty for 1–3 years	0.6000	<0.0001
Prepayment penalty for 5 years	0.4296	<0.0001
Property Characteristics		
Property owner occupied (0,1)	−0.1068	0.0064
Property condo (0,1)	0.1396	0.0017
Neighborhood-Level Characteristics at Origination		
Recent housing price appreciation rate at origination	−0.1660	0.0654
Housing price heterogeneity at origination (in $10,000)	−0.0009	0.8245
Housing occupancy rate at origination	−0.4092	0.0587
Poverty rate at origination	0.1684	0.4635
Median applicant income at origination (HMDA)	−0.0332	0.4728
Proportion of African Americans at origination (HMDA)	−0.0253	0.8912
Proportion of Hispanics at origination (HMDA)	0.1659	0.1914
Borrower Race (Base group: Non-Hispanic Whites)		
African American	0.3519	<0.0001
Hispanics	0.0962	0.0154
Market/Time Characteristics		
Time Trend	0.0016	0.8910
Loan origination season fixed effects [d]	YES	
Number of loans	1404	
Adjusted R^2	0.7660	

[a] The contract rate determination model was estimated with *regulation-free* predicted loan termination probabilities included, considering the impact of race and ethnicity on loan performance

[b] The method outlined in Appendix 6A of Wooldridge (2010) was employed to correct the standard errors and to calculate the *p values* to account for the presence of generated regressors

[c] The *regulation-free* predicted default and prepayment probabilities ($\hat{P}_k$) are generated variables and are endogenous in this contract rate equation being a function of the contract rate (spread). To address this issue, a set of generated IVs ($\tilde{P}_k$) are used, each serves as a valid IV for its corresponding generated variable ($\hat{P}_k$). Each generated IV ($\tilde{P}_k$) is calculated as its corresponding generated variable ($\hat{P}_k$), but with the actual contract rate spread ($C_0 - y_0$) being replaced with a predicted contract rate spread ($\hat{C}_0 - y_0$). The predicted contract rate ($\hat{C}_0$) is obtained from the estimation of a reduced-form contract rate equation

[d] The seasonal loan origination fixed effects estimates are omitted here, but are available upon request

individual-level race and ethnicity variable, *African Americans*, is significant and its coefficient estimate is around 20 basis points, while the other one, *Hispanics,* is not (Table 5). Combining these results indicates that both African Americans and Hispanics are disadvantaged by fair-lending laws and regulations. African American borrowers pay an additional 15 basis points and Hispanic borrowers pay an additional 10 basis points for their mortgage loans. This unintended consequence of regulation is equivalent to an increase in monthly mortgage payment of approximately \$19 for African American borrowers or of roughly \$13 for Hispanic borrowers for a 30-year fixed-rate mortgage loan of \$180,000 with an annual interest rate of 8%.[36] Note that the coefficient estimates on other covariates are similar between *color-neutral* and *regulation-free* scenarios.

Robustness Test - a Matching Approach

To explore the robustness of the main result of this study, a matching approach is employed. The matching approach mimics risk-based pricing; borrowers with the same level of loan termination risk should be charged the same contract rate regardless of their race and ethnicity. The matching approach applies a quasi-experimental design to construct a control (racial) group of loans that are as similar as possible to a target (racial) group of loans in terms of loan termination risk. This matching technique should mitigate systematic differences in loan termination risk between the two racial and ethnic groups and allow for systematic differences in loan contract rate between the two groups to be examined. Note that in this study, the risk level of a loan is represented by the predicted probability of loan termination ($\hat{P}_k$).[37]

To generate a balanced matched sample, we use a nearest 1-to-1 matching process with replacement based on Mahalanobis distance along with calipers restrictions. Details on this matching approach have been provided in Appendix B. The resulting matched sample is examined to ensure the target and control group borrowers have the same distributions of loan termination risk, as well as the matching variables, to determine if a balanced sample was found. A standard measure used to determine balance in the matching literature (Rubin and Thomas 2000; Austin 2011) is a mean difference *t*-test. However, the mean difference *t* test has been criticized by many scholars (Imai et al. 2008; Austin 2008 & 2009) because the significance levels are confounded with sample size. Therefore, in addition to the mean difference *t* test, the standardized difference of the mean and the ratio of the variance of each variable is used to check if a matched sample is well balanced.[38] Following Rubin (2001), a variable is well balanced if and only if the standardized difference of the mean falls in the

[36] The average original loan size and the average contract rate in the sample of this study are approximately \$180,000 and 8% respectively.

[37] The matching conducted here is similar to propensity score matching that is conducted based on a predicted propensity score (Rosenbaum and Rubin 1983). However, it is not exactly the same as the typical propensity score matching, as the loans are matched based on the predicted termination probabilities instead of the probability of a borrower being in a particular racial and ethnic group.

[38] For a single continuous variable x, the standardized difference of the mean is defined as $SD = \left(\overline{x}_{group1} - \overline{x}_{group2}\right) / \sqrt{\left(s^2_{group1} + s^2_{group2}\right) / 2}$, where $\overline{x}_{group1}$ and $\overline{x}_{group2}$ denote the sample mean of this variable in racial and ethnic group 1 and group 2, respectively, whereas s^2_{group1} and s^2_{group2} denote the sample variance of the variable in racial and ethnic group 1 and group 2, respectively. For a dichotomous variable, the standardized difference of the mean is defined as $SD = \left(\overline{x}_{group1} - \overline{x}_{group2}\right) / \sqrt{\left(\overline{x}_{group1} \times \left(1 - \overline{x}_{group1}\right) + \overline{x}_{group2} \times \left(1 - \overline{x}_{group2}\right)\right) / 2}$.

range of (−0.25, 0.25) and the ratio of the variance falls in the range of (0.5, 2). If the matched sample is determined not to be balanced for a specific caliper radius, the sampling process described in Appendix B is repeated with other values of caliper radius until a balanced matching sample is arrived at and in which there are sufficient loans for comparison across the two groups.[39] Following prior studies analyzing the optimal caliper radius ratio (ρ) for matching (Cochran and Rubin 1973; Rosenbaum and Rubin 1985; Austin 2011), an appropriate caliper radius ratio is adopted if the caliper radius ratio could significantly reduce the mean difference between matched pairs without losing many unsuccessful matching pairs. For the matching between African American borrowers and non-Hispanic white borrowers and the matching between Hispanic borrowers and non-Hispanic white borrowers, caliper radius ratios of 0.8 and1.0 turn out to meet this criterion. A caliper radius ratio (ρ) of 0.8 requires that the difference in a matching criterion variable between loans in two racial groups should not exceed 80% of the pooled standard deviation of that variable.

Using this matching process, we generate a matched sample for African American borrowers (target group) and non-Hispanic white borrowers (control group) and a second sample for Hispanic borrowers (target group) to non-Hispanic white borrowers (control group). These samples are generated separately based on the *color-neutral* probabilities and the *regulation-free* probabilities and reported for two caliper radii (ρ = 0.8 and ρ = 1.0) resulting in 8 samples. In the appendix, Tables 12,13,14,15 report the matching balance diagnosis for the *color-neutral* and the *regulation-free* scenarios for each of the matched samples. All of the matching criterion variables are balanced in those matching samples except for the original loan amount variable. The original loan amount variable is unbalanced in the matched sample for African American borrowers (target group) and non-Hispanic white borrowers (control group) when the caliper is equal to 1. The matched balanced samples allow us to compare the mean contract rate between the control and target groups to measure whether there is a systematic difference in the contract rate between the two groups. We also estimate the loan contract rate determination model (Eq. (6)) over each of the matched balanced samples to account for the potential locational/neighborhood variation in the matched loans.[40]

Table 7 reports the results on nearest 1-to-1 matching between loans of African American borrowers and loans of non-Hispanic white borrowers based on two matching samples (ρ = 0.8 and ρ = 1.0), while Table 8 demonstrates the results on two matching samples between Hispanics and non-Hispanic whites. The samples in both Tables 7 and 8 are generated using the *color-neutral* probabilities. Panel A of the Tables report the mean difference in loan contract rate between the control and target groups and a *t-value* to test if the mean difference is zero. Panel B of these tables contains the regression adjustment results for each matching sample. Based on the matching of African American borrowers to non-Hispanic white borrowers, the results show little evidence of a significant difference in the contract rates between the two

[39] A caliper is applied to impose a tolerance level on the maximum distance on a matching criterion variable between two groups. Specifically, the caliper is defined as$Caliper = \pm\rho\sqrt{\left(s^2_{group1} + s^2_{group2}\right)}\ /2$, where ρ is the caliper radius, s^2_{group1} and s^2_{group2} denote the sample variance of a matching criterion variable in target and control groups.

[40] Applying the same diagnosis technique to other variables included in the loan contract rate determination model (Eq. (6)) reveals that the neighborhood traits variables are typically unbalanced, providing justification for the regression adjustment to be applied.

Table 7 Results on nearest 1-to-1 Matching between African Americans and non-hispanic whites - color-neutral [a]

Panel A Mean Difference in Loan Contract Rate without Regression Adjustment								
	Matching Sample 1				Matching Sample 2			
	Caliper radius ratio (ρ): 0.8 [b]				Caliper radius ratio (ρ): 1.0 [b]			
	Mean (African American)	Mean (non-Hispanic White)	Mean Diff.	Mean Diff. *t* test	Mean (African American)	Mean (non-Hispanic White)	Mean Diff.	Mean Diff. *t* test
Variable								
Contract rate at origination	8.5951	8.2260	0.3691	2.1700	8.5551	8.1477	0.4074	2.4100
Number of matched pairs	86				94			
Panel B Loan Contract Rate Estimates on Matched Samples -2SLS								
	Matching Sample 1				Matching Sample 2			
Variables	Coef.	*p value* [c]			Coef.		*p value* [c]	
Borrower Race (Base group: Non-Hispanic whites)								
African Americans	0.0208	0.8178			0.0877		0.3974	
IV-Predicted termination probabilities [d]	YES				YES			
10-year treasury yield at origination (y_0) [d]	YES				YES			
Loan characteristics [d]	YES				YES			
Borrower characteristics [d]	YES				YES			
Collateral characteristics [d]	YES				YES			
Neighborhood characteristics [d]	YES				YES			
Time Trend [d]	YES				YES			
Number of loans	172				188			
Adjusted R^2	0.7289				0.7226			

[a] Nearest 1-to-1 matching with replacements and with calipers is used to match loans of African American borrowers to loans of non-Hispanic white borrowers. The matching was conducted based on *color-neutral* predicted loan termination probabilities ignoring the effects of race and ethnicity on loan performance. For each loan of an African American borrower, a loan of a non-Hispanic white borrower in the pool with the shortest Mahalanobis distance on the *color-neutral* predicted loan termination probabilities is selected as the sole and best match

[b] A caliper is applied to impose a tolerance level for the maximum distance on a matching criterion variable between borrowers in the two racial (ethnic) groups. Those matching criterion variables include *color-neutral* predicted probability of each event, the 10-year yield at origination, the original LTV ratio, FICO, and original loan amount. The caliper radius (the width of the caliper) on a matching criterion variable is specified as a ratio (ρ) of the pooled standard deviation of that variable

[c] The method outlined in Appendix 6A of Wooldridge (2010) was employed to correct the standard errors and to calculate the *p values* to account for the presence of generated regressors

[d] We only report the estimates on the individual race variable. Estimates on other variables in the loan contract rate determination equation are omitted here, but are available upon request

Table 8 Results on nearest 1-to-1 matching between hispanics and non-hispanic whites - color-neutral [a]

Panel A Mean Difference in Loan Contract Rate without Regression Adjustment								
	Matching Sample 1				Matching Sample 2			
	Caliper radius ratio (ρ): 0.8 [b]				Caliper radius ratio (ρ): 1.0 [b]			
	Mean (Hispanic)	Mean (non-Hispanic White)	Mean Diff.	Mean Diff. *t* test	Mean (Hispanic)	Mean (non-Hispanic White)	Mean Diff.	Mean Diff. *t* test
Variable								
Contract rate at origination	7.8980	7.9126	−0.0146	−0.1700	7.8566	7.9139	−0.0572	−0.6800
Number of matched pairs	283				298			
Panel B Loan Contract Rate Estimates on Matched Samples -2SLS								
	Matching Sample 1				Matching Sample 2			
Variables	Coef.		*p value* [c]		Coef.		*p value* [c]	
Borrower Race (Base group: Non-Hispanic whites)								
Hispanics	−0.0045		0.9101		−0.0176		0.6531	
IV- Predicted termination probabilities [d]	YES				YES			
10-year treasury yield at origination (y_0) [d]	YES				YES			
Loan characteristics [d]	YES				YES			
Borrower characteristics [d]	YES				YES			
Collateral characteristics [d]	YES				YES			
Neighborhood characteristics [d]	YES				YES			
Time Trend [d]	YES				YES			
Number of loans	566				596			
Adjusted R^2	0.8217				0.8085			

[a] Nearest 1-to-1 matching with replacements and with calipers is used to match loans of Hispanic borrowers to loans of non-Hispanic white borrowers. The matching was conducted based on *color-neutral* predicted loan termination probabilities ignoring the effects of race and ethnicity on loan performance. For each loan of a non-Hispanic white borrower, a loan of a Hispanic borrower in the pool with the shortest Mahalanobis distance on the *color-neutral* predicted loan termination probabilities is selected as the sole and best match

[b] A caliper is applied to impose a tolerance level for the maximum distance on a matching criterion variable between borrowers in the two racial (ethnic) groups. Those matching criterion variables include the *color-neutral* predicted probability of each event, the 10-year yield at origination, the original LTV ratio, FICO, and original loan amount. The caliper radius (the width of the caliper) on a matching criterion variable is specified as a ratio (ρ) of the pooled standard deviation of that variable

[c] The method outlined in Appendix 6A of Wooldridge (2010) was employed to correct the standard errors and to calculate the *p values* to account for the presence of generated regressors

[d] We only report the estimates on the individual race variable. Estimates on all other variables in the loan contract rate determination equation are omitted here, but are available upon request

groups. While the mean difference in the $\rho = 1.0$ matched sample is statistically significant, it is not found to be significant once neighborhood characteristics are controlled for in the contract rate equation based on the same matched sample. The results based on the matching of Hispanic borrowers to non-Hispanic white borrowers show no evidence of a significant difference in the contract rates between the two groups.

Tables 9 and 10 contain the matched sample results based on the results of the *regulation-free* scenario. The results based on the matched sample of African American borrowers to non-Hispanic white borrowers indicate African American borrowers pay a significantly higher contract rate. This result is consistent across the mean comparison and the contract rate estimation. In Table 10, the results based on the matching of Hispanic borrowers to non-Hispanic white borrowers indicate a significant difference in the contract rates between the two groups once neighborhood characteristics are controlled for in the contract rate equation.

The matching results provide evidence that minority borrowers are disadvantaged by fair-lending laws and regulations in the subprime mortgage market over this study period. Based on the regression results of the *color-neutral* matching samples, we fail to find any evidence of adverse pricing against minority borrowers. However, the results in Tables 9 and 10 demonstrate both African American borrowers and Hispanic borrowers pay a significantly higher contract rate, given their prepayment habits, than their non-Hispanic white counterparts. In other words, because lenders are not allowed to use race and ethnicity to evaluate risk, minority borrowers are charged a higher interest rate. The magnitudes of the race-based contract rate disparities shown in Tables 9 and 10 are slightly larger than those reported in Table 6.

Discussion

A limitation of the data used in this study is that the number of points paid by a borrower is not observed. As points are usually used by borrowers to buy down a loan contract rate, one might argue that a possible explanation for the higher contract rates paid by African American borrowers and Hispanic borrowers is they tend to choose less points on average at loan closing. However, in this study both African American borrowers and Hispanic borrowers are shown to be less likely to prepay, other things equal. Rational borrowers, anticipating a lower likelihood of prepaying a loan in the near future, would choose to pay more points, ceteris paribus.[41]

The study is also limited by its focus on the Miami area. However, it is important to note that we are not alone in this prepayment finding, several prior studies have found empirical evidence that a borrower's race and ethnicity are associated with loan prepayment patterns. Kelly (1995), using loan-level data from the VA's mortgage program from 1971 to 1989, found that African American borrowers and Hispanic borrowers prepaid less frequently than whites. Based on a pool of fixed-rate residential

[41] Woodward and Hall (2010, 2012) note that borrowers may suffer an informational disadvantage compared to the brokers and fail to recognize that more upfront cash payments (more points) should lead to a lower interest rate. Under such a scenario, they find that minority borrowers pay significantly higher total origination fees including points than similar white borrowers.

Table 9 Results on nearest 1-to-1 matching between African Americans and non-hispanic whites - regulation-free [a]

Panel A Mean Difference in Loan Contract Rate without Regression Adjustment								
	Matching Sample 1				Matching Sample 2			
	Caliper radius ratio (ρ): 0.8 [b]				Caliper radius ratio (ρ): 1.0 [b]			
	Mean (African American)	Mean (non-Hispanic White)	Mean Diff.	Mean Diff. *t* test	Mean (African American)	Mean (non-Hispanic White)	Mean Diff.	Mean Diff. *t* test
Variable								
Contract rate at origination	8.5447	8.0826	0.4621	2.5400	8.6299	8.0763	0.5536	3.2000
Number of matched pairs	77				88			
Panel B Loan Contract Rate Estimates on Matched Samples -2SLS								
		Matching Sample 1			Matching Sample 2			
Variables		Coef.		*p value* [c]	Coef.		*p value* [c]	
Borrower Race (Base group: Non-Hispanic whites)								
African Americans		0.2472		0.0074	0.4619		<0.0001	
IV- Predicted termination probabilities [d]		YES			YES			
10-year treasury yield at origination (y_0) [d]		YES			YES			
Loan characteristics [d]		m			YES			
Borrower characteristics [d]		YES			YES			
Collateral characteristics [d]		YES			YES			
Neighborhood characteristics [d]		YES			YES			
Time Trend [d]		YES			YES			
Number of loans		154			176			
Adjusted R^2		0.7815			0.8554			

[a] Nearest 1-to-1 matching with replacements and with calipers is used to match loans of African American borrowers to loans of non-Hispanic white borrowers. The matching was conducted based on *regulation-free* predicted loan termination probabilities considering the effects of race and ethnicity on loan performance. For each loan of an African American borrower, a loan of a non-Hispanic white borrower in the pool with the shortest Mahalanobis distance on the *regulation-free* predicted loan termination probabilities is selected as the sole and best match

[b] A caliper is applied to impose a tolerance level for the maximum distance on a matching criterion variable between borrowers in the two racial (ethnic) groups. Those matching criterion variables include the *regulation-free* predicted probability of each event, the 10-year yield at origination, the original LTV ratio, FICO, and original loan amount. The caliper radius (the width of the caliper) on a matching criterion variable is specified as a ratio (ρ) of the pooled standard deviation of that variable

[c] The method outlined in Appendix 6A of Wooldridge (2010) was employed to correct the standard errors and to calculate the *p values* to account for the presence of generated regressors

[d] We only report the estimates on the individual race variable. Estimates on all other variables in the loan contract rate determination equation are omitted here, but are available upon request

Table 10 Results on nearest 1-to-1 matching between hispanics and non-hispanic whites - regulation-free [a]

Panel A Mean Difference in Loan Contract Rate without Regression Adjustment								
	Matching Sample 1				Matching Sample 2			
	Caliper radius ratio (ρ): 0.8 [b]				Caliper radius ratio (ρ): 1.0 [b]			
	Mean (Hispanic)	Mean (non-Hispanic White)	Mean Diff.	Mean Diff. *t* test	Mean (Hispanic)	Mean (non-Hispanic White)	Mean Diff.	Mean Diff. *t* test
Variable								
Contract rate at origination	8.1347	7.9198	0.2150	2.4000	8.1581	7.9126	0.2455	2.8200
Number of matched pairs	284				298			
Panel B Loan Contract Rate Estimates on Matched Samples -2SLS								
	Matching Sample 1				Matching Sample 2			
Variables	Coef.		*p value* [c]		Coef.		*p value* [c]	
Borrower Race (Base group: Non-Hispanic whites)								
Hispanics	0.2388		<0.0001		0.2353		<0.0001	
IV- Predicted termination probabilities [d]	YES				YES			
10-year treasury yield at origination (y_0) [d]	YES				YES			
Loan characteristics [d]	YES				YES			
Borrower characteristics [d]	YES				YES			
Collateral characteristics [d]	YES				YES			
Neighborhood characteristics [d]	YES				YES			
Time Trend [d]	YES				YES			
Number of loans	568				596			
Adjusted R^2	0.7975				0.7584			

[a] Nearest 1-to-1 matching with replacements and with calipers is used to match loans of Hispanic borrowers to loans of non-Hispanic white borrowers. The matching was conducted based on *regulation-free* predicted loan termination probabilities considering the effects of race and ethnicity on loan performance. For each loan of a non-Hispanic white borrower, a loan of a Hispanic borrower in the pool with the shortest Mahalanobis distance on the *regulation-free* predicted loan termination probabilities is selected as the sole and best match

[b] A caliper is applied to impose a tolerance level for the maximum distance on a matching criterion variable between borrowers in the two racial (ethnic) groups. Those matching criterion variables include the *regulation-free* predicted probability of each event, the 10-year yield at origination, the original LTV ratio, FICO, and original loan amount. The caliper radius (the width of the caliper) on a matching criterion variable is specified as a ratio (ρ) of the pooled standard deviation of that variable

[c] The method outlined in Appendix 6A of Wooldridge (2010) was employed to correct the standard errors and to calculate the *p values* to account for the presence of generated regressors

[d] We only report the estimates on the individual race variable. Estimates on all other variables in the loan contract rate determination equation are omitted here, but are available upon request

mortgage loans originated from 1993 to 1994, with their performance tracked through 1998, Clapp et al. (2001) concluded that a borrower's minority status is negatively associated with prepayment rate. The same conclusions could also be found in Deng and Gabriel (2006) and Firestone et al. (2007). While some of these studies are also based on city level data, several are based on national data (e.g., Kelly (1995), Deng and Gabriel (2006), and Firestone et al. (2007)). The result of minority borrowers prepaying with less frequency does not seem unique to the Miami metropolitan area.

Conclusion

Overall, this study provides empirical evidence that fair-lending laws and regulations may actually have unintended consequences with respect to mortgage pricing. What should not be overlooked is this consequence is a direct result of prepayment risk, a phenomenon more generally observed than default. Further, given the consideration that a borrower's race and ethnicity may be correlated with loan termination behavior, we explicitly model and control for loan termination risk (default and prepayment) in the contract rate estimation rather than using a reduced-form approach. The results demonstrate that African American and Hispanic borrowers are less likely to prepay than non-Hispanic white borrowers. This loan performance finding provides us with mortgage pricing implications. In a completely competitive mortgage market, the lower prepayment risk would be reflected in the pricing of those loans.

Fair-lending laws and regulations prohibit lenders from using a borrower's race or ethnicity in assessing a loan's risk. In this *color-neutral* world, minority borrowers with a relatively lower prepayment risk would *not* be offered a lower contract rate given their prepayment habits. The empirical results provide little support for a contract rate disparity in this *color-neutral* world; the rates paid by borrowers, charged by lenders, are consistent with the goals of equal contract rates across racial and ethnic groups.[42]

However, in this *color-neutral* world, the termination risk of minority borrowers does not reflect the true termination risk, minority borrower's termination risk is overestimated given they are less likely to prepay. If the prepayment risk of minority borrowers is considered in determining termination risk, what we term as a *regulation-free* scenario, the estimation results on the loan contract rate model indicates African American borrowers pay an additional 15 basis points and Hispanic borrowers pay roughly 10 extra basis points for their contract rate as a result of regulation. Such an increase in contract rate raises the monthly payment by $19 for African Americans and $13 for Hispanics based on the average size and contract rate of the loans in the sample. These results are robust to a matching approach.

Acknowledgements We would like to thank Stephen L. Ross for his helpful and insightful comments, as well as Lingxiao Li and the other participants of the 2018 ARES Meeting for their helpful feedback. We acknowledge and appreciate the constructive comments of the anonymous referees.

[42] In this "*color neutral*" world, although the contract rate estimation results based on the full sample (Table 5) show that African American borrowers pay a significantly higher contract rate than non-Hispanic white borrowers, this racial disparity disappears with the matched samples in which borrowers in the two racial groups are more homogeneous in terms of loan termination risk (Table 7).

Appendix 1

Table 11 Definition of variables

Variables	Variable definition
Loan Characteristics	
Contract rate at origination (C_0)	Contract interest rate at origination, in percentage
10-year treasury yield at origination (y_0)	10-year treasury constant maturity yield at origination, in percentage
Contract rate spread at origination (C_0-y_0)	Contract interest rate at origination minus 10-year treasury constant maturity yield at origination, in percentage
Original LTV	The LTV ratio at loan origination, in percentage
FICO at origination	Credit score at loan origination
Original loan amount (in \$10,000)	Loan size at loan origination (in \$10,000)
Full income documentation (0,1)	=1 for a loan with full documentation of a borrower's income
Prepayment penalty for 1–3 years	=1 for a loan encumbered by a prepayment penalty for up to the first 3 years
Prepayment penalty for 5 years	=1 for a loan encumbered by a prepayment penalty for the first 5 years
Within prepayment penalty period (0,1) [a]	=1 for a loan encumbered by a prepayment penalty in a given mortgage month *t*
Property Characteristics	
Property owner occupied (0,1)	=1 for a loan secured by a property occupied by the owner
Property condo (0,1)	=1 for a loan secured by a condo
Neighborhood-Level Characteristics	
Recent housing price appreciation rate at origination	The ratio of the median housing sale price in a census tract in a three-year period prior to the month of loan origination to the median housing sale price in the same census tract in another three-year period prior to the three-year pre-origination period, then minus 1
Heterogeneity in housing price [b]	The standard deviation of the housing sale price in a census tract over a three-year period prior to a given month (in \$10,000)
Housing occupancy rate (from Census Survey) [b]	Housing occupancy rate in a census tract
Poverty rate (from Census Survey) [b]	Poverty rate in a census tract
Median applicant income [b]	Ratio of the median applicant income in a census tract to the median applicant income in Miami-Dade County, FL.
Proportion of non-Hispanic white applicants (from HMDA) [b]	Ratio of the number of loan applications of non-Hispanic white borrowers in a census tract to the total number of loan applications in that census tract.
Proportion of Hispanic applicants (from HMDA) [b]	Ratio of the number of loan applications of Hispanic borrowers in a census tract to the total number of loan applications in that census tract.
Proportion of African American applicants (from HMDA) [b]	Ratio of the number of loan applications of African American borrowers in a census tract to the total number of loan applications in that census tract.
Market Characteristics	
Relative house price at time *t* (RHP_t) [a]	The ratio of the median house sale price index at time *t* to the median house sale price index at loan origination [c]
Market interest rate change at time *t* (y_0-y_{t-2}) [a]	10-year treasury yield at loan origination minus 10-year treasury yield at time *t*, lagged by 2 months, in percentage
Others	
Time Trend	A variable that equals to 1 for a loan originated in year 1997, 2 for a loan originated in year 1998, and so forth

[a] Denotes time-varying variables in default/prepayment hazard model

[b] Denotes variables that are time-varying variables in the default/prepayment hazard model, but represent values at loan origination in the contract rate determination model

[c] This median house price index is at the census tract level and is generated from the property transaction database in Miami-Dade County, FL. For each month, this index is calculated based on the inflation-adjusted median housing sale price over a three-year window around that month, eighteen months before and eighteen months after that given month. All prices are defined in 2009 dollars

Appendix 2

Matching Algorithm Descriptions

The matching approach is implemented as follows.

Step 1: Examine common support region.

Before matching, the density distribution of each predicted probability ($\hat{P}_k$) in the target group is compared to that in the control group to check if they overlap with each other. Loans in the target group that fall in this overlapping region are defined as falling within the common support pool.

Step 2: Apply calipers restrictions to select potential control group matches for each target group loan.

For each target group loan which falls within the common support pool, loans in the control group that satisfy the calipers restrictions for the set of matching variables are identified as potential control group matches.[43] The matching variables include the predicted probability of each termination event ($\hat{P}_k$), the 10-year yield at origination (y_0), as well as several variables that are believed to be the major loan risk factors including the original LTV ratio, FICO score, and original loan amount.[44]

Step 3: Select the nearest 1–1 match based on the Mahalanobis distance.

To identify a specific control loan from the pool of potential matches for each target loan, the Mahalanobis distance, based on the predicted loan termination probabilities ($\hat{P}_k$), is used. Specifically, the Mahalanobis distance (MD) is defined as $MD = \left(\Delta\hat{P}_1, \Delta\hat{P}_{2,}\right)\mathbf{V}^{-1}\left(\Delta\hat{P}_1, \Delta\hat{P}_{2,}\right)^T$, where $\Delta\hat{P}_k$is the difference in the predicted probability of each event k ($k = 1, 2$) between the target loan and a loan in the control group, and$\mathbf{V}$is the sample covariance matrix of the predicted termination probabilities from the full set of loans in the control racial group. The Mahalanobis distance is commonly used for multi-dimensional matching because it accounts for the variance of each predicted probability as well as the covariance between the two predicted probabilities. The loan with the smallest MD is identified as the control loan.

[43] A caliper is applied to impose a tolerance level on the maximum distance on a matching criterion variable between two groups. Specifically, the caliper is defined as$Caliper = \pm\rho\sqrt{\left(s^2_{group1} + s^2_{group2}\right)}/2$, where ρ is the caliper radius, $s^2_{group1}$and s^2_{group2}denote the sample variance of a matching criterion variable in target and control groups. Austin (2011), Cochran and Rubin (1973), and Rosenbaum and Rubin (1985) all examine the extent to which the caliper radius (ρ) reduces the bias between the two groups.

[44] Although LTV is a continuous variable, the loans in the sample tend to fall on common LTV ratios (e.g., 80%, 85%, 90%, etc.). Thus, we convert the LTV ratios into three categories based on the original LTV ratio: loans with original LTV ratio below 80, loans with original LTV ratio above 80 but below 100, and loans with original LTV ratio above 100. Matched loans must fall in the same original LTV ratio category.

Table 12 Balance diagnosis on matching samples between African Americans and non-hispanic whites (Color neutral) [a]

Variables [b]	Mean		Std. Dev.		Mean diff.	Std. Mean Diff. [b]	Ratio of the Var. [b]	Mean Diff. t-Test [b]
	African American	White	African American	White				
Panel A Matching Sample 1 (Caliper radius ratio, ρ: 0.8)								
10-year treasury yield at origination (y_0)	4.9499	4.8816	0.7560	0.7159	0.0683	0.0927	1.1152	0.6100
Predicted default probability	0.0801	0.0726	0.0733	0.0706	0.0074	0.1033	1.0781	0.6800
Predicted prepayment probability	0.7538	0.7520	0.1008	0.1069	0.0019	0.0179	0.8894	0.1200
LTV	88.1047	87.5698	11.0454	12.3979	0.5349	0.0456	0.7937	0.3000
FICO	684.9651	687.5000	55.9604	55.1995	−2.5349	−0.0456	1.0278	−0.3000
Original loan amount (in $10,000)	12.1518	12.9489	7.8258	6.7319	−0.7971	−0.1092	1.3514	−0.7200
Number of Matched Pairs	86							
Panel B Matching Sample 2 (Caliper radius ratio, ρ: 1.0)								
10-year treasury yield at origination (y_0)	4.9079	4.8207	0.7730	0.7694	0.0871	0.1130	1.0094	0.7700
Predicted default probability	0.0850	0.0759	0.0812	0.0694	0.0091	0.1206	1.3685	0.8300
Predicted prepayment probability	0.7466	0.7511	0.1123	0.1060	−0.0045	−0.0410	1.1235	−0.2800
LTV	88.5532	88.0106	11.2972	12.9760	0.5426	0.0446	0.7580	0.3100
FICO	685.8830	689.0957	56.2410	52.7369	−3.2128	−0.0589	1.1373	−0.4000
Original loan amount (in $10,000)	12.2669	14.7457	7.7455	8.3005	−2.4787	−0.3088	0.8707	−2.1200
Number of Matched Pairs	94							

[a] The matching was conducted based on *color-neutral* predicted loan termination probabilities ignoring the effects of race and ethnicity on loan performance

[b] The variables demonstrated in this table are criterion variables used for matching. The standardized difference of the mean, the ratio of the variance, and mean difference t-test of those variables between the two racial groups are used to diagnose whether each matching sample is balanced or not, following the matching literature (Rubin, 2001; Austin 2011). Following Rubin (2001), a variable is well balanced if and only if the standardized difference of the mean falls in the range of (−0.25, 0.25), and the ratio of the variance falls in the range of (0.5, 2). Based on this rule, all of the matching criterion variables in *color-neutral* matching sample 1 are well balanced; while in matching sample 2, all of those variables are balanced except for original loan amount. This balance diagnosis technique is also applied to other variables included in the loan contract rate determination model (Eq. (6)) that describe the traits of the borrower, the property, and the neighborhood. A few of them, especially those neighborhood traits variables, turned out to be unbalanced. Therefore, a regression adjustment following Eq. (6) specifications is applied to each matching sample to account for any difference in those unbalanced variables between these two racial groups. The matching balance diagnosis results on other variables are omitted here, but are available upon request

Table 13 Balance diagnosis on matching samples between African Americans and non-hispanic whites (Regulation free) [a]

Variables [b]	Mean		Std. Dev.		Mean diff.	Std. Mean Diff. [b]	Ratio of the Var. [b]	Mean Diff. *t*-Test [b]
	African American	White	African American	White				
Panel A Matching Sample 1 (Caliper radius ratio, ρ: 0.8)								
10-year treasury yield at origination (y_0)	4.9544	4.8974	0.7807	0.8021	0.0570	0.0720	0.9471	0.4500
Predicted default probability	0.1058	0.1074	0.0969	0.1050	−0.0016	−0.0154	0.8516	−0.1000
Predicted prepayment probability	0.6901	0.6956	0.1214	0.1165	−0.0055	−0.0461	1.0855	−0.2900
LTV	87.4286	86.5584	11.1136	10.8549	0.8701	0.0792	1.0482	0.4900
FICO	685.0779	686.4156	54.6079	54.8755	−1.3377	−0.0244	0.9903	−0.1500
Original loan amount (in $10,000)	11.1124	12.3965	6.5882	6.0371	−1.2841	−0.2032	1.1909	−1.2600
Number of Matched Pairs	77							
Panel B Matching Sample 2 (Caliper radius ratio, ρ: 1.0)								
10-year treasury yield at origination (y_0)	4.9505	4.9036	0.7627	0.8521	0.0468	0.0579	0.8011	0.3800
Predicted default probability	0.1069	0.0986	0.0966	0.0973	0.0084	0.0861	0.9866	0.5700
Predicted prepayment probability	0.6947	0.7016	0.1254	0.1232	−0.0070	−0.0560	1.0361	−0.3700
LTV	88.0227	86.9545	11.1334	10.7167	1.0682	0.0978	1.0793	0.6500
FICO	684.4091	686.6477	54.2056	52.8646	−2.2386	−0.0418	1.0514	−0.2800
Original loan amount (in $10,000)	11.9180	13.7135	7.7905	7.7556	−1.7955	−0.2310	1.0090	−1.5300
Number of Matched Pairs	88							

[a] The matching was conducted based on *regulation-free* predicted loan termination probabilities considering the effects of race and ethnicity on loan performance

[b] The variables demonstrated in this table are criterion variables used for matching. The standardized difference of the mean, the ratio of the variance, and mean difference *t*-test of those variables between the two racial groups are used to diagnose whether each matching sample is balanced or not, following the matching literature (Rubin, 2001; Austin 2011). Following Rubin (2001), a variable is well balanced if and only if the standardized difference of the mean falls in the range of (−0.25, 0.25), and the ratio of the variance falls in the range of (0.5, 2). Based on this rule, all of the matching criterion variables in the matching samples ($\rho = 0.8$ and $\rho = 1$) under the *regulation-free* scenario are well balanced. This balance diagnosis technique is also applied to other variables included in the loan contract rate determination model (Eq. (6)) that describe the traits of the borrower, the property, and the neighborhood. Several variables, especially neighborhood traits variables, turned out to be unbalanced. Therefore, considering those unbalanced variables, a regression adjustment following Eq. (6) specifications is applied to each matching sample to account for any difference in those unbalanced variables between these two racial groups. The matching balance diagnosis results on other variables are omitted here, but are available upon request

Table 14 Balance diagnosis on matching samples between Hispanics and non-hispanic whites (Color neutral) [a]

Variables [b]	Mean		Std. Dev.		Mean diff.	Std. Mean Diff. [b]	Ratio of the Var. [b]	Mean Diff. *t*-Test [b]
	Hispanic	White	Hispanic	White				
Panel A Matching Sample 1 (Caliper radius ratio, ρ: 0.8)								
10-year treasury yield at origination (y_0)	5.0881	5.1260	0.7327	0.7801	−0.0380	−0.0501	0.8820	−0.6000
Predicted default probability	0.0783	0.0776	0.1177	0.1182	0.0007	0.0059	0.9912	0.0700
Predicted prepayment probability	0.7383	0.7364	0.1351	0.1392	0.0019	0.0137	0.9410	0.1600
LTV	81.7244	81.4064	11.3291	11.1623	0.3180	0.0283	1.0301	0.3400
FICO	709.3039	712.4205	43.4811	48.9410	−3.1166	−0.0673	0.7893	−0.8000
Original loan amount (in $10,000)	19.5431	20.2996	12.8730	12.9828	−0.7565	−0.0585	0.9831	−0.7000
Number of Matched Pairs	283							
Panel B Matching Sample 2 (Caliper radius ratio, ρ: 1.0)								
10-year treasury yield at origination (y_0)	5.0693	5.1272	0.7419	0.7819	−0.0580	−0.0760	0.9002	−0.9300
Predicted default probability	0.0783	0.0803	0.1203	0.1240	−0.0020	−0.0160	0.9411	−0.2000
Predicted prepayment probability	0.7355	0.7329	0.1376	0.1442	0.0025	0.0181	0.9110	0.2200
LTV	81.0604	81.2517	11.4314	11.0475	−0.1913	−0.0170	1.0707	−0.2100
FICO	707.0973	712.2919	43.1939	50.7199	−5.1946	−0.1103	0.7253	−1.3500
Original loan amount (in $10,000)	19.8576	20.9379	12.8972	13.6348	−1.0803	−0.0814	0.8947	−0.9900
Number of Matched Pairs	298							

[a] The matching was conducted based on *color-neutral* predicted loan termination probabilities ignoring the effects of race and ethnicity on loan performance

[b] The variables demonstrated in this table are criterion variables used for matching. The standardized difference of the mean, the ratio of the variance, and mean difference *t*-test of those variables between the two ethnic groups are used to diagnose whether each matching sample is balanced or not, following the matching literature (Rubin, 2001; Austin 2011). Following Rubin (2001), a variable is well balanced if and only if the standardized difference of the mean falls in the range of (−0.25, 0.25), and the ratio of the variance falls in the range of (0.5, 2). Based on this rule, all of the matching criterion variables in the matching samples ($\rho = 0.8$ and $\rho = 1$) under the *color-neutral* scenario are well balanced. This balance diagnosis technique is also applied to other variables included in the loan contract rate determination model (Eq. (6)) that describe the traits of the borrower, the property, and the neighborhood. Only a few neighborhood traits variables are unbalanced. To account for any difference in those unbalanced variables between these two ethnic groups, a regression adjustment following Eq. (6) specifications is applied to each matching sample. The matching balance diagnosis results on other variables are omitted here, but are available upon request

Table 15 Balance diagnosis on matching samples between hispanics and non-hispanic whites (Regulation free) [a]

Variables [b]	Mean		Std. Dev.		Mean diff.	Std. Mean Diff. [b]	Ratio of the Var. [b]	Mean Diff. t-Test [b]
	Hispanic	White	Hispanic	White				
Panel A Matching Sample 1 (Caliper radius ratio, ρ: 0.8)								
10-year treasury yield at origination (y_0)	5.1205	5.1311	0.7495	0.7895	−0.0106	−0.0137	0.9011	−0.1600
Predicted default probability	0.0715	0.0725	0.1115	0.1129	−0.0010	−0.0086	0.9742	−0.1000
Predicted prepayment probability	0.7359	0.7415	0.1332	0.1363	−0.0056	−0.0415	0.9549	−0.4900
LTV	81.9437	81.4824	11.4063	11.1623	0.4613	0.0409	1.0442	0.4900
FICO	709.1162	712.6127	46.6836	50.3837	−3.4965	−0.0720	0.8585	−0.8600
Original loan amount (in $10,000)	19.7110	20.1171	12.1708	13.0097	−0.4061	−0.0322	0.8752	−0.3800
Number of Matched Pairs	284							
Panel B Matching Sample 2 (Caliper radius ratio, ρ: 1.0)								
10-year treasury yield at origination (y_0)	5.1415	5.1278	0.7362	0.7824	0.0138	0.0182	0.8853	0.2200
Predicted default probability	0.0710	0.0716	0.1132	0.1148	−0.0006	−0.0051	0.9722	−0.0600
Predicted prepayment probability	0.7400	0.7428	0.1347	0.1393	−0.0028	−0.0201	0.9345	−0.2500
LTV	81.1174	81.2651	12.2786	11.0411	−0.1477	−0.0126	1.2367	−0.1500
FICO	704.8557	712.1510	44.7148	50.5697	−7.2953	−0.1528	0.7818	−1.8700
Original loan amount (in $10,000)	20.5224	20.9714	12.4156	13.7308	−0.4490	−0.0343	0.8176	−0.4200
Number of Matched Pairs	298							

[a] The matching was conducted based on *regulation-free* predicted loan termination probabilities considering the effects of race and ethnicity on loan performance

[b] The variables demonstrated in this table are criterion variables used for matching. The standardized difference of the mean, the ratio of the variance, and mean difference t-test of those variables between the two ethnic groups are used to diagnose whether each matching sample is balanced or not, following the matching literature (Rubin, 2001; Austin 2011). Following Rubin (2001), a variable is well balanced if and only if the standardized difference of the mean falls in the range of (−0.25, 0.25), and the ratio of the variance falls in the range of (0.5, 2). Based on this rule, all of the matching criterion variables in matching sample 1 and 2 under the *regulation-free* scenario are well balanced. This balance diagnosis technique is also applied to other variables included in the loan contract rate determination model (Eq. (6)) that describe the traits of the borrower, the property, and the neighborhood. A few variables, especially neighborhood traits variables, turned out to be unbalanced. To account for any difference in those unbalanced variables between these two ethnic groups, a regression adjustment following Eq. (6) specifications is applied to each matching sample. The matching balance diagnosis results on other variables are omitted here, but are available upon request

References

Allison, P. D. (2010). Survival analysis using SAS: A practical guide, 2nd edition. SAS Institute.

Austin, P. C. (2008). A Critical Appraisal of Propensity-Score Matching in the Medical Literature between 1996 and 2003. *Statistics in Medicine, 27*(12), 2037–2049

Austin, P. C. (2009). Balance Diagnostics for Comparing the Distribution of Baseline Covariates between Treatment Groups in Propensity-Score Matched Samples. *Statistics in Medicine, 28*(25), 3083–3107

Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics, 10*(2), 150–161.

Avery, R., Brevoort, K., & Canner, G. (2007). Opportunities and issues in using HMDA data. *Journal of Real Estate Research, 29*(4), 351–380.

Bayer, P., Ferreira, F., & Ross, S. L. (2017). What drives racial and ethnic differences in high-cost mortgages? The role of high-risk lenders. *The Review of Financial Studies, 31*(1), 175–205.

Black, H., Schweitzer, R. L., & Mandell, L. (1978). Discrimination in mortgage lending. *The American Economic Review, 68*(2), 186–191.

Black, H. A., Boehm, T. P., & DeGennaro, R. P. (2003). Is there discrimination in mortgage pricing? The case of overages. *Journal of Banking & Finance, 27*(6), 1139–1165.

Boehm, T. P., Thistle, P. D., & Schlottmann, A. (2006). Rates and race: An analysis of racial disparities in mortgage rates. *Housing Policy Debate, 17*(1), 109–149.

Cheng, P., Lin, Z., & Liu, Y. (2015). Racial discrepancy in mortgage interest rates. *The Journal of Real Estate Finance and Economics, 51*(1), 101–120.

Chernov, M., Dunn, B. R., & Longstaff, F. A. (2017). Macroeconomic-driven prepayment risk and the valuation of mortgage-backed securities. *The Review of Financial Studies, 31*(3), 1132–1183.

Clapp, J. M., Goldberg, G. M., Harding, J. P., & LaCour-Little, M. (2001). Movers and Shuckers: Interdependent prepayment decisions. *Real Estate Economics, 29*(3), 411–450.

Cochran, W. G., & Rubin, D. B. (1973). Controlling Bias in Observational Studies: A Review. Sankhyā*: The Indian Journal of Statistics*, Series A, 417–446.

Courchane, M. (2007). The pricing of home mortgage loans to minority borrowers: How much of the APR differential can we explain? *Journal of Real Estate Research, 29*(4), 399–440.

Cox, J. C., Ingersoll Jr, J. E., & Ross, S. A. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica, 53*(2), 385–408

Crawford, G. W., & Rosenblatt, E. (1999). Differences in the cost of mortgage credit implications for discrimination. *The Journal of Real Estate Finance and Economics, 19*(2), 147–159.

Deng, Y., Quigley, J. M., & Van Order, R. (2000). Mortgage terminations, heterogeneity and the exercise of mortgage options. *Econometrica, 68*(2), 275–307.

Deng, Y., & Gabriel, S. (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*(6), 1431–1460.

Duan, J. C., & Simonato, J. G. (1999). Estimating and testing exponential-affine term structure models by Kalman filter. *Review of Quantitative Finance and Accounting, 13*(2), 111–135.

Duca, J. V., & Rosenthal, S. S. (1994). Do mortgage rates vary based on household default characteristics? Evidence on rate sorting and credit rationing. *The Journal of Real Estate Finance and Economics, 8*(2), 99–113.

Dunn, K. B., & McConnell, J. J. (1981a). A comparison of alternative models for pricing GNMA mortgage-backed securities. *The Journal of Finance, 36*(2), 471–484.

Dunn, K. B., & McConnell, J. J. (1981b). Valuation of GNMA mortgage-backed Securiti. *The Journal of Finance, 36*(3), 599–616.

Fang, L., & Munneke, H. J. (2016). Gender equality in mortgage lending. *Real Estate Economics*. https://doi.org/10.1111/1540-6229.12198.

Firestone, S., Van Order, R., & Zorn, P. (2007). The performance of low-income and minority mortgages. *Real Estate Economics, 35*(4), 479–504.

Getter, D. E. (2006). Consumer credit risk and pricing. *Journal of Consumer Affairs, 40*(1), 41–63.

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

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

Holmes, A., & Horvitz, P. (1994). Mortgage Redlining: Race, Risk, and Demand. *The Journal of Finance, 49*(1), 81–99.

Imai, K., King, G., & Stuart, E. A. (2008). Misunderstandings Between Experimentalists and Observationalists about Causal Inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 171*(2), 481–502.
Kau, J. B., Keenan, D. C., & Munneke, H. J. (2012). Racial discrimination and mortgage lending. *The Journal of Real Estate Finance and Economics, 45*(2), 289–304.
Kau, J. B., Keenan, D. C., Muller, W. J., & Epperson, J. F. (1992). A generalized valuation model for fixed-rate residential mortgages. *Journal of Money, Credit and Banking, 24*(3), 279–299.
Kelly, A. (1995). Racial and ethnic disparities in mortgage prepayment. *Journal of Housing Economics, 4*(4), 350–372.
LaCour-Little, M. (1999). Discrimination in mortgage lending: A critical review of the literature. *Journal of Real Estate Literature, 7*(1), 15–50.
Ladd, H. F. (1998). Evidence on discrimination in mortgage lending. *Journal of Economic Perspectives, 12*(2), 41–62.
Mayer, C., Piskorski, T., & Tchistyi, A. (2013). The inefficiency of refinancing: Why prepayment penalties are good for risky borrowers. *Journal of Financial Economics, 107*(3), 694–714.
Munnell, A. H., Tootell, G. M., Browne, L. E., & McEneaney, J. (1996). Mortgage lending in Boston: Interpreting HMDA data. *The American Economic Review*, 25–53.
Nothaft, F. E., & Perry, V. G. (2002). Do mortgage rates vary by neighborhood? Implications for loan pricing and redlining. *Journal of Housing Economics, 11*(3), 244–265.
Petrin, A., & Train, K. (2003). Omitted Product Attributes in Discrete Choice Models. (No. w9452). *National Bureau of Economic Research*.
Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.
Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39*(1), 33–38.
Ross, S. L., & Tootell, G. M. (2004). Redlining, the community reinvestment act, and private mortgage insurance. *Journal of Urban Economics, 55*(2), 278–297.
Ross, S. L., & Yinger, J. (2002). *The color of credit: Mortgage discrimination, research methodology, and fair-lending enforcement*. MIT Press.
Ross, S. L., Turner, M. A., Godfrey, E., & Smith, R. R. (2008). Mortgage lending in Chicago and Los Angeles: A paired testing study of the pre-application process. *Journal of Urban Economics, 63*(3), 902–919.
Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association, 95*(450), 573–585.
Schwartz, E. S., & Torous, W. N. (1989). Prepayment and the valuation of mortgage-backed securities. *The Journal of Finance, 44*(2), 375–392.
Schwartz, E. S., & Torous, W. N. (1992). Prepayment, default, and the valuation of mortgage pass-through securities. *Journal of Business, 65*, 221–239.
Tootell, G. M. (1996). Redlining in Boston: Do mortgage lenders discriminate against neighborhoods? *The Quarterly Journal of Economics, 111*(4), 1049–1079.
Turner, M. A., & Skidmore, F (1999). Mortgage lending discrimination: A review of existing evidence. The Urban Institute.
Woodward, S. E., & Hall, R. E. (2010). Consumer confusion in the mortgage market: Evidence of less than a perfectly transparent and competitive market. *American Economic Review, 100*(2), 511–515.
Woodward, S. E., & Hall, R. E. (2012). Diagnosing consumer confusion and sub-optimal shopping effort: Theory and mortgage-market evidence. *American Economic Review, 102*(7), 3249–3276.
Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.
Zhang, Y. (2013). Fair lending analysis of mortgage pricing: Does underwriting matter? *The Journal of Real Estate Finance and Economics, 46*(1), 131–151.