ELSEVIER

Journal of Economics and Business 60 (2008) 110–124

Journal of Economics & Business

Race, ethnicity and subprime home loan pricing

Debbie Gruenstein Bocian [a,*], Keith S. Ernst [b], Wei Li [b]

[a] *Center for Responsible Lending, 910 17th Street NW, Suite 500, Washington, DC 20006, United States*
[b] *Center for Responsible Lending, 301 W. Main Street, Durham, NC 27701, United States*

Received 30 March 2007

Abstract

This study examines whether borrowers' race and ethnicity affect subprime loan pricing after accounting for objective determinants, including credit scores and loan-to-value ratios. The results show that African-American and Latino borrowers are more likely to receive higher-rate subprime home loans than non-Latino white borrowers.

Keywords: Subprime; Mortgage; Pricing; Race; Ethnicity; Discrimination; Fair lending

1. Introduction

For the first time in 2004, lenders were required to report details on the costs of subprime home loans—mortgages intended to serve borrowers with blemished credit or other high-risk characteristics. Lenders disclosed pricing information related to the most expensive subprime loans (referred to here as "higher-rate" loans), while lower-rate subprime loans and virtually all prime loans were exempt from this reporting requirement. When the data was first released to the public in the fall of 2005, several analyses showed that African-American and Latino borrowers received a disproportionate share of higher-rate home loans, even when controlling for factors such as borrower income and property location (Avery, Canner, & Cook, 2005).

A number of concerned groups pointed to these disparities as evidence of discrimination against communities who already lag far behind in homeownership and wealth. Others contended, however, that the pricing disparities were not meaningful since they did not fully account for differences in legitimate risk factors. In this article, we provide a more detailed examination of pricing patterns in the subprime home loan market. More specifically, to our knowledge, this is the first full research

* Corresponding author.
E-mail address: Debbie.Bocian@responsiblelending.org (D.G. Bocian).

0148-6195/$ – see front matter
doi:10.1016/j.jeconbus.2007.10.001

effort that examines 2004 HMDA data to assess the effects of race and ethnicity on the cost of subprime home loans while controlling for the major risk factors typically used by lenders to set prices.

Our findings show that, for most types of subprime home loans, African-American and Latino borrowers are at greater risk of receiving higher-rate loans than non-Latino white borrowers, even after controlling for objective risk factors. Specifically, we find that African-Americans were more likely to receive higher-rate home purchase and refinance loans than similarly situated non-Latino white borrowers in the set of subprime loans that carried prepayment penalties. Latino borrowers were more likely to receive higher-rate loans than similarly situated non-Latino white borrowers in the set of subprime loans used to purchase homes.

Our analysis does not allow us to estimate precisely how much race and ethnicity increase the prices charged to borrowers. It is also beyond the scope of this paper to determine definitively why these disparities exist. However, we do explore some possible causes, including the role of discretionary pricing and the potential contribution of market segmentation.

2. Background

2.1. HMDA

In 1975, Congress enacted the Home Mortgage Disclosure Act (HMDA). HMDA was a legislative response to the widespread practice of mortgage "redlining," that is, the systematic exclusion of neighborhoods of color when marketing or originating home loans. Initially, HMDA required regulated institutions to disclose summaries of their mortgage lending by census tracts. Disclosure requirements under HMDA have evolved over time to reflect the changing nature of mortgage lending and potential discriminatory practices, broadening both the range of lenders under its purview and the information those lenders are required to disclose. Specifically, HMDA's scope has expanded to encompass non-depository institutions, such as mortgage companies, which have significantly increased their share of the mortgage market over the past three decades. In addition, the disclosure requirements of lenders have evolved to include a range of application information, loan characteristics, borrower demographics, and property characteristics.

One of the most important changes to HMDA is the recent inclusion of limited pricing information related to the annual percentage rate (APR) of certain "higher-rate" loans. For many loans originated in 2004, lenders were required to report the spread between the APR of certain loans and the yield on a U.S. Treasury security of comparable maturity. More specifically, lenders submitted this information on first-lien loans if the spread was equal to or greater than three percentage points, and they submitted this information on loans secured by subordinate liens if the spread was at least five percentage points.

This information on higher-rate loans makes it possible for the first time to use HMDA directly not just to detect disparities in loan dispositions (i.e., the proportion of loans that were approved or denied) between demographic groups, but also differences in loan pricing. Since borrowers can be vulnerable to discrimination at both the underwriting and pricing stages of the loan process, the ability to detect discriminatory patterns in both areas is critical for ensuring that all racial and ethnic communities have equal opportunities.

2.2. Select literature review

Though direct analyses of mortgage pricing based on HMDA data have not been possible prior to the release of the 2004 data, many studies have examined HMDA data to evaluate other issues related to possible inequities in the mortgage market. Such studies have focused on differences in loan disposition (i.e., whether loan applications were approved or denied) by race and ethnicity or on whether certain groups are disproportionately served by subprime lenders.

Perhaps the most well-known research along these lines was published in 1996 by the Federal Reserve Bank of Boston (Munnell, Tootell, Browne, & McEneaney, 1996), often referred to simply as the "Boston Fed Study." This study combined publicly available HMDA data from the Boston area with a number of additional variables, including information on credit history collected from area lenders. The study found that the risk of loan denial for African-Americans and Latinos was about 80% greater than that of white applicants, controlling for a host of applicant, loan, property, and neighborhood characteristics.

Several studies have used HMDA data to analyze whether specific racial and ethnic groups receive a disproportionate share of subprime loans. Because HMDA data does not specifically identify subprime loans, most of these studies have approximated which loans were subprime by using annual lists of predominately subprime lenders published by the U.S. Department of Housing and Urban Development (HUD). The HUD-Treasury Joint Task Force (2000) explored the relationship between subprime lending and neighborhood racial composition. Relying on HMDA data, the study reported that subprime lending accounted for 51% of all refinance loans in predominately African-American communities in 1998, compared to only 9% in predominately white neighborhoods. The study also found that these disparities persisted even when controlling for neighborhood income.

A national study by the Center for Community Change (Bradford, 2002) analyzed the proportion of borrowers receiving subprime refinance loans by race and ethnicity and found pervasive disparities among African-American, Latino and white borrowers. In addition, the report found that disparities persisted within income categories and actually increased as income went up. Specifically, while lower-income African-American borrowers were 2.4 times as likely to receive a loan from a subprime lender as lower-income white borrowers, upper-income African-American borrowers were 3.0 times as likely to receive such loans as upper-income white borrowers. At the same time, lower-income Latino borrowers were 1.4 times as likely to receive a subprime loan as lower-income white borrowers, and upper-income Latinos were 2.2 times as likely to receive such loans as upper-income whites.

Another study (Calem, Gillen, & Wachter, 2004) evaluated the effect of borrower race and neighborhood racial composition on receiving subprime loans in the Chicago and Philadelphia areas. The authors found that both the race of individual borrowers and neighborhood racial composition have statistically significant impacts on the likelihood of receiving a loan from a subprime lender, even after controlling for the borrower information available in HMDA and tract-level factors. For complete literature reviews, see White (2004) and Ross and Yinger (2006).

2.3. The 2004 HMDA data: pricing disparities surface

In September 2005, the Federal Financial Institutions Examination Council (FFIEC) released the 2004 data for all HMDA reporters in electronic form, making it possible to analyze the entire U.S. mortgage market. An analysis of the 2004 HMDA conducted by Avery et al. (2005) found that pricing disparities persisted even after controlling for borrower-specific information such as income, origination amount, gender, property location and presence of a co-applicant.

Avery et al. (2005) first confirmed the existence of substantial disparities between the proportion of African-American and Hispanic white borrowers receiving higher-rate loans to that of non-Hispanic white borrowers in both the home purchase and refinance markets. The authors next made a series of adjustments to account for differences between white borrowers and borrowers of color by controlling for the following demographic information contained in HMDA data: borrower income, loan amount, location (MSA) of the property, presence of a co-applicant, and gender. These adjustments marginally lowered observed disparity ratios between borrowers of color and whites, but disparities persisted. They then adjusted for differences in lender composition between the groups. Interestingly, these "lender adjustments" reduced the disparity ratios considerably, though significant differences remained. However, the analysis did not control for several important risk factors, such as credit scores or loan-tovalue ratios (LTVs), since these variables are not part of HMDA's disclosure requirements.

Avery et al. (2005) did, however, present results from an analysis conducted by the Credit Research Center (CRC). CRC's analysis relied on 2004 HMDA data from eight unidentified subprime lenders, supplemented by proprietary information provided by those lenders. No information is provided on whether these lenders were representative of the subprime market as a whole. In fact, there is evidence that the composition of loans from these lenders is fundamentally different from that of the overall subprime market. For example, over 80% of both the purchase loans and refinance loans analyzed by CRC were higher-rate while less than 50% of subprime loans overall in 2004 appear to have been identified as higher-rate in HMDA (Ernst & Goldstein, 2005). Moreover, the predominance of higher-rate loans results in an unbalanced sample that complicates efforts to identify pricing disparities. In addition, the paper does not contain important information on CRC's methodology nor does it provide results for the control variables used in the analysis. This lack of clarity makes it difficult to understand the context for CRC's results and, consequently, what conclusions one can reasonably draw. For example, while CRC controlled for whether the loan was originated by a broker, this variable might confound results by being positively correlated with both pricing and race or ethnicity.

Our analysis adds significantly to this body of research by combining HMDA data with information from a proprietary database. By combining information from each of these two datasets, we are able to incorporate important risk factors into a multivariate analysis of mortgage pricing.

3. Data and methods

3.1. Data

To include additional information on risk factors that might account for higher prices charged to African-American and Latino borrowers, we combined the 2004 HMDA data with a proprietary database of securitized subprime loans. Like HMDA data, the proprietary database contains specific information on individual loans, including borrower and property characteristics. While several types of information can be found in both datasets, including data on the location of the property, the originating lender, lien status, loan purpose, property type, and loan amount, each dataset contains some information that the other does not. For example, the proprietary database includes critical pieces of information on loan risk at origination that are not included in HMDA, such as the LTV, credit score (FICO), and whether the loan was covered by private mortgage insurance. On the other hand, HMDA contains information on the race and ethnicity of borrowers, which the proprietary database lacks. In addition, while HMDA contains information on APR spreads (which incorporates information on certain fees), the proprietary database has information on the mortgage note rate and whether the loan includes a prepayment penalty, but no information on APRs or up-front fees. Finally, while the proprietary database is among the largest subprime home loan datasets available, accounting for an estimated 85% of U.S. subprime originations in 2004 when compared to estimates carried in trade publications (Inside Mortgage Finance, 2006), it only contains securitized subprime loans. For its part, the HMDA dataset is the single largest publicly available dataset on U.S. mortgage originations, and it includes both prime and subprime loans for covered lenders.

Using information common to both HMDA and the proprietary database, we were able to match loans from the two databases, creating a new dataset of 177,487 subprime loans originated in 2004. The procedures used to merge the databases were designed to minimize false positive matches (detailed information on our matching methodology is available upon request). The merged dataset includes individual loan information on borrower characteristics (race, ethnicity, income, FICO credit score, income documentation level); loan characteristics (LTV, loan amount, purpose, existence and duration of prepayment penalties); property characteristics (location, property type); and pricing (APR spread for higher-rate loans). An examination of the distribution of variables in the merged dataset indicates it is representative of the subprime market (comparisons of average values of variables between the proprietary dataset and the merged dataset are available upon request). To complement this loan-level data, we added publicly available information on prevailing interest rates and state-specific information on housing prices, demographics and state judicial foreclosure and deficiency judgment laws as defined by Ambrose, LaCour-Little, and Sanders (2004).

The HMDA data allows borrowers to report both an ethnicity designation (either "Hispanic or Latino" or "Not Hispanic or Latino") and up to five racial designations (including both "white" and "African-American or Black"). We coded and refer to any borrower who was identified as "Hispanic or Latino" as "Latino," and any borrower who was identified as "African-American or Black" in any of the race fields as "African-American." We coded borrowers and refer to them as "white" if they were associated with "Not Hispanic or Latino" and only identified as "white" in the race fields. The remaining loans were not coded into racial or ethnic categories and were excluded from the analysis. In practice, the Latino and African-American categories are not mutually exclusive, but the overlap in our merged dataset is small (about 2%), and using this method ensures maximum inclusion for members for each group. For the list of variables used in our analysis, see Table 1 .

3.2. Methods

Our statistical analysis adapts a mortgage pricing model created by Ambrose et al. (2004). In the referenced study, the authors examined whether conforming to the conventional loan guidelines set by Fannie Mae and Freddie Mac had an impact on mortgage prices. Although our purpose is different and, consequently, the specific variables that we analyze are not identical to those included in their study, we adapted their general analytical approach.

Like Ambrose et al. (2004), we used multiple regression analysis to estimate the impact of different borrower, property, loan and geographic factors on the APR spread of a loan. However, whereas Ambrose et al. (2004) had the actual APR spread for all of the loans in their database, we only have the spread for those loans that exceeded HMDA's APR spread-reporting threshold. As a result, while the Ambrose study was able to use regression analysis to estimate the actual APR spread, our analysis with logistic regressions allowed us to compare the odds of different racial and ethnic groups receiving higher rate loans, but did not allow us to estimate the magnitude of differences in APR spreads themselves.

Logistic regression procedures assume that there is no endogeneity between the dependent and independent variables. We recognize that this assumption may not be valid in our case since a

Table 1
Variables and descriptions

Variable name	Description
HCOST	Dummy variable = 1 if APR spread is reported in HMDA, else = 0
Borrower characteristics (BOR)	
AA	Dummy variable = 1 if the borrower is African-American, else = 0
LAT	Dummy variable = 1 if the borrower is Latino, else = 0
INC	Monthly income, in US$ 1000
INC^2	Monthly income squared
FICO	FICO credit score
DOC	Dummy variable = 1 if the borrower provided full documentation of income, else = 0
Loan/property characteristics (LOAN)	
LTV	Loan-to-value ratio at origination
LAMT	Loan origination amount, in US$ 10,000
PPP	Dummy variable = 1 if the loan carries a prepayment penalty, else = 0
MULTI	Dummy variable = 1 if the loan is secured by a property with 2–4 units, else = 0
CONDO	Dummy variable = 1 if the loan is secured by a condominium, else = 0
AG1–AG4	Categorical dummy variables representing the regulatory agency of the originating lender
Economic variables (ECO)	
CSPREAD	Monthly difference between AAA and Baa bond yields
YCURVE	Monthly difference between 10-year and 1-year treasury yields
HPIV	8 quarter standard deviation in the OFHEO state Housing Price Index
RATE	Interest rate for subprime loans with 30-year term, B-credit, 80% LTV from Inside B&C Lending
RATEV	15 month standard deviation in 1-year treasury yield
Q2–Q4	Categorical dummy variables for the second, third and fourth quarters of 2004
Geographic variables (GEO)	
CDIV2–CDIV9	Categorical dummy variables for the Census division in which the property is located
LAW2–LAW4	Categorical dummy variables for state laws created by Ambrose et al. (2004). Based on rules pertaining to judicial foreclosure and deficiency judgment
NCCITY	Dummy variable = 1 if the property is located in an MSA but outside of a central city, else = 0
RURAL	Dummy variable = 1 if the property is located outside an MSA, else = 0
AAST	Proportion of state population that is African-American
LATST	Proportion of state population that is Latino

loan's APR may affect some of the variables we hold constant, namely LTV, loan amount, and whether the loan carried a prepayment penalty. Therefore, like Ambrose et al. (2004), our model includes statistical adjustments to account for the possible interdependence of these variables, providing a more reliable estimate of the effect of race and ethnicity on the risk of receiving a higher-rate loan. Like Ambrose et al. (2004), we specify LTV and loan amount as a simultaneous equation system and utilize the predicted values of these two variables in the subsequent logistic regression. The following system of equations are estimated via three-stage least square regression (Zellner & Theil, 1962):

$$\begin{aligned}\text{LAMT} = {}& \beta_0 + \beta_1 \times \text{AA} + \beta_2 \times \text{LAT} + \beta_3 \times \text{LTV} + \beta_4 \times \text{RATE} + \beta_5 \times \text{INC} \\ & + \beta_6 \times \text{INC}^2 + \beta_7 \times \text{HPI} + \beta_8 \times \text{DOC} + \text{Error}\end{aligned} \tag{1}$$

$$\begin{aligned}\text{LTV} = {}& \beta_0 + \beta_1 \times \text{AA} + \beta_2 \times \text{LAT} + \beta_3 \times \text{LAMT} + \beta_4 \times \text{RATE} + \beta_5 \times \text{INC} \\ & + \beta_6 \times \text{INC}^2 + \beta_7 \times \text{HPI} + \beta_8 \times \text{DOC} + \beta_9 \times \text{FICO} + \text{Error}\end{aligned} \tag{2}$$

Based on the model coefficients estimated from the third stage of the three-stage least squares regression, we calculated the predicted LTV and loan amount for each loan, which is then used as two of the independent variables in the following logistic regression model[1]:

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \sum_{i=1}^{6}\beta_i \times \mathrm{BOR}_i + \sum_{i=1}^{9}\beta_{i+6} + \mathrm{LOAN}_i + \sum_{i=1}^{8}\beta_{i+15} \times \mathrm{ECO}_i + \sum_{i=1}^{15}\beta_{i+23} \times \mathrm{GEO}_i + \mathrm{Error} \quad (3)$$

where π is the probability of receiving a higher-rate loan. To select explanatory variables, we proceed from the theoretical observations and constructs presented by Ambrose et al. (2004). This leads us to include four types of explanatory variables: borrower characteristics (BOR), loan or property characteristics (LOAN), economic variables (ECO) and geographic variables (GEO).

In addition to the African-American dummy variable (AA) and the Latino dummy variable (LAT), we utilize borrower's credit score (FICO) to take into account key borrower traits (BOR) used by lenders to set pricing at origination. We expect higher FICO scores to be associated with lower probability of receiving a higher-rate subprime loan. In addition, we included borrower's monthly income (INC) and whether a borrower's income is supported by full documentation (DOC). We expect that unobserved factors correlated with lower borrower income and loans where borrowers furnish less than full documentation of their income to be associated with a higher probability of receiving a higher-rate subprime loan.

To account for key loan or property traits (LOAN) used by lenders to set pricing at origination, we included predicted loan-to-value (LTVE) ratio and predicted loan amount (LAMTE). We expect that larger loan amounts and lower LTV ratios should be associated with a lower probability of receiving a higher-rate subprime loan. In addition, we included dummy variables representing the type of the property and the regulatory agency of the originating lender of the loan (AG1–AG4).

Also, consistent with Ambrose et al. (2004), we included macroeconomic variables (ECO) that measure yield curve slope (YCURVE), interest rate volatility (RATEV), spread between yields on bonds with higher and lower credit ratings (CSPREAD), housing price index volatility (HPIV), and quarter dummy for the effect of seasonal economic changes (Q2–Q4).

Finally, we included a series of geographically based variables (GEO): three dummy variables that assess the effects related to differing levels of judicial involvement under foreclosure laws in states (LAW2–LAW4), eight census division dummy variables (CDIV2–CDIV9), two dummy variables for the urban status (NCCITY and RURAL), and two variables of the proportion of African-American or Latino of the population at state level (AAST and LATST).

To achieve a homogenous dataset, all analyses were restricted to loans secured by first-liens on owner-occupied properties. Further, we excluded loans secured by manufactured housing units, those backed by private mortgage insurance, those with non-standard amortization schedules, and those with origination amounts exceeding the maximum limit for purchase by Fannie Mae and Freddie Mac, which was US$ 333,700 in most states.

To address potential endogeneity concerns between APR spread and the existence of prepayment penalties (which is a binary variable), we segment our models by prepayment penalty status

[1] Results obtained from the three-stage least squares models to predict LTV ratios (LTVE) and loan amounts (LAMTE) at origination are available from the authors upon request.

(with and without). Recognizing that loan pricing may depend in part on the type of interest rate (adjustable or fixed) and/or the loan purpose (purchase or refinance), we further segment our models by the following four loan categories: (1) purchase fixed-rate mortgages (FRMs); (2) purchase adjustable-rate mortgages (ARMs); (3) refinance FRMs; and (4) refinance ARMs. In addition, to limiting variations in loan products within each category, we included only the dominant types of loans. For ARMs, we limit the data to loans with a fixed interest rate for 2 years followed by a 28-year term with semi-annual interest rate adjustments calculated by adding a margin to an index based on 6-month LIBOR rates, with either no prepayment penalty or a prepayment penalty of 2 years. For FRM, we limit the data to loans with 30-year terms, with either no prepayment penalty or a prepayment penalty with a term of 3 years.

Table 2 provides descriptive statistics (mean and standard deviation) for variables included in the models by loan products. Across eight loan products, the proportion of higher-rate loans range from 26 to 62%, with 44% of all loans coded as higher-rate, consistent with previous estimates that less than half of subprime loans are reported as higher-cost in 2004 HMDA data (Ernst & Goldstein, 2005). Among the samples, 15–25% of loans went to African-American borrowers, and 15–25% went to Latino borrowers. On average, FICO scores range from 580 to 688, LTV ratios are 74–85, monthly incomes average US$ 4900–5800, and 53–75% of borrowers provided full income documentation.

Logistic regression models were estimated using SAS 9.1.3 STAT module using PROC LOGISTIC. Three-stage least squares models were estimated using SAS 9.1.3 ETS module using PROC SYSLIN.

4. Findings

The bottom panel of Table 3 provides two types of goodness-of-fit statistics for the logistic regression models. Across the eight models, the Nagelkerke *R*-squared valued from 42 to 68%, showing that all models did a good job predicting the likelihood of receiving a higher-cost loan. To provide a non-parametric goodness-of-fit measurement of our logistic regression models, we calculated Kolmogorov–Smirnov (KS) statistics. As noted by Lax, Manti, Raca, and Zorn (2004) good models tend to have a KS statistic greater than 50. To show that the model fit equally well for all three race/ethnicity groups, we calculated KS statistics for African-American, Latino, and non-Latino white borrowers separately, in addition to the KS statistics for all groups together. Table 3 shows that all the non-parametric KS statistics are greater than 50. In other words, the models fit the data well for all borrower groups.

Table 3 reports the estimated impact of an independent variable on the odds of receiving a higher-rate loan. For continuous variables, the odds ratio is a change in the estimated odds of receiving a higher-rate loan when the variable increases by the number of units shown within the parenthesis. For categorical variables, the odds ratio is between one category and the control category, which is shown within the parenthesis. The results associated with objective factors are consistent with rational risk-based pricing practices in the subprime mortgage industry. For example, all else being equal, an increase of 10 percentage points in an LTV ratio increases the odds of receiving a higher-rate loan by a factor that ranges from 1.73 for fixed rate loans for refinancing with prepayment penalties to 5.53 for adjustable-rate loans for purchasing without prepayment penalties. Similarly, for an increase of 60 units in FICO, the odds of receiving a higher-rate loan decreased by a factor that ranges from 0.20 for fixed rate loans for refinancing with prepayment penalties to 0.36 for adjustable rate loans for purchasing without prepayment penalties. An increase of US$ 10,000 in loan amount decreased the odds of receiving a higher-rate loan by a factor that

Table 2
Descriptive statistics of the variables

Variable	FRM				ARM			
	Purchase		Refinance		Purchase		Refinance	
	Penalty ($N=2235$)	No penalty ($N=1444$)	Penalty ($N=5918$)	No penalty ($N=2881$)	Penalty ($N=4657$)	No penalty ($N=13{,}321$)	Penalty ($N=11{,}950$)	No penalty ($N=6520$)
HCOST	0.419 (0.493)	0.262 (0.440)	0.267 (0.443)	0.403 (0.491)	0.521 (0.500)	0.532 (0.499)	0.558 (0.497)	0.620 (0.485)
AA	0.162 (0.369)	0.155 (0.362)	0.147 (0.354)	0.198 (0.398)	0.187 (0.390)	0.245 (0.430)	0.157 (0.364)	0.199 (0.400)
LAT	0.213 (0.409)	0.124 (0.330)	0.174 (0.379)	0.152 (0.359)	0.247 (0.431)	0.206 (0.405)	0.161 (0.368)	0.146 (0.353)
FICO	641 (60)	688 (71)	624 (60)	618 (65)	625 (55)	630 (59)	583 (52)	580 (52)
LTV	84 (12)	82 (11)	74 (18)	75 (15)	85 (9)	84 (9)	78 (14)	77 (14)
AG1	0.142 (0.349)	0.604 (0.489)	0.060 (0.237)	0.126 (0.332)	0.083 (0.275)	0.046 (0.209)	0.045 (0.208)	0.023 (0.148)
AG2	0.031 (0.174)	0.022 (0.147)	0.061 (0.239)	0.069 (0.254)	0.012 (0.107)	0.019 (0.138)	0.022 (0.146)	0.025 (0.156)
AG3	0.054 (0.225)	0.033 (0.179)	0.052 (0.222)	0.040 (0.195)	0.137 (0.344)	0.084 (0.277)	0.103 (0.304)	0.071 (0.257)
AG4	0.012 (0.109)	0.006 (0.079)	0.009 (0.093)	0.012 (0.111)	0.009 (0.095)	0.007 (0.085)	0.019 (0.138)	0.013 (0.113)
AAST	11.0 (7.3)	13.4 (8.5)	10.1 (7.7)	13.5 (7.4)	10.3 (6.8)	14.0 (8.1)	10.2 (7.6)	13.7 (7.6)
CDIV2	0.085 (0.278)	0.174 (0.379)	0.074 (0.262)	0.205 (0.404)	0.047 (0.212)	0.323 (0.468)	0.063 (0.243)	0.348 (0.476)
CDIV3	0.116 (0.320)	0.168 (0.374)	0.107 (0.309)	0.128 (0.334)	0.146 (0.354)	0.237 (0.425)	0.165 (0.371)	0.182 (0.386)
CDIV4	0.028 (0.164)	0.098 (0.298)	0.038 (0.192)	0.053 (0.224)	0.056 (0.230)	0.061 (0.239)	0.067 (0.251)	0.056 (0.230)
CDIV5	0.144 (0.351)	0.259 (0.438)	0.167 (0.373)	0.182 (0.386)	0.165 (0.371)	0.200 (0.400)	0.168 (0.374)	0.162 (0.369)
CDIV6	0.056 (0.231)	0.033 (0.178)	0.048 (0.213)	0.016 (0.124)	0.045 (0.206)	0.014 (0.118)	0.041 (0.199)	0.011 (0.102)
CDIV7	0.255 (0.436)	0.086 (0.280)	0.087 (0.281)	0.288 (0.453)	0.144 (0.351)	0.009 (0.096)	0.053 (0.225)	0.138 (0.345)
CDIV8	0.065 (0.247)	0.053 (0.223)	0.074 (0.262)	0.043 (0.203)	0.094 (0.292)	0.038 (0.191)	0.099 (0.299)	0.029 (0.167)
CDIV9	0.183 (0.387)	0.080 (0.271)	0.276 (0.447)	0.046 (0.209)	0.183 (0.386)	0.043 (0.203)	0.210 (0.408)	0.021 (0.144)
CONDO	0.089 (0.286)	0.117 (0.322)	0.068 (0.252)	0.054 (0.226)	0.112 (0.316)	0.105 (0.306)	0.088 (0.284)	0.063 (0.244)
CSPREAD	0.763 (0.046)	0.758 (0.048)	0.760 (0.047)	0.759 (0.047)	0.765 (0.047)	0.761 (0.048)	0.765 (0.047)	0.764 (0.048)
DOC	0.718 (0.450)	0.695 (0.460)	0.749 (0.434)	0.713 (0.453)	0.615 (0.487)	0.527 (0.499)	0.673 (0.469)	0.671 (0.470)
LATST	15.0 (12.5)	10.5 (10.4)	14.0 (11.9)	16.7 (12.6)	13.8 (11.6)	10.7 (8.2)	11.9 (10.9)	13.2 (10.1)
HPIV	13.2 (10.4)	14.0 (10.0)	18.1 (11.3)	12.6 (10.4)	15.6 (11.4)	18.0 (10.8)	17.5 (11.5)	16.7 (11.2)
INC	5.1 (4.1)	5.7 (3.7)	4.9 (3.6)	5.2 (4.0)	5.2 (3.2)	5.8 (4.2)	5.2 (5.8)	5.6 (9.6)
MULTI	0.046 (0.209)	0.063 (0.243)	0.048 (0.214)	0.056 (0.230)	0.058 (0.233)	0.118 (0.323)	0.044 (0.204)	0.085 (0.279)
NCCITY	0.437 (0.496)	0.474 (0.499)	0.542 (0.498)	0.489 (0.500)	0.501 (0.500)	0.578 (0.494)	0.549 (0.498)	0.586 (0.493)
LAMT	12.8 (7.2)	13.9 (7.6)	15.0 (7.5)	13.5 (7.5)	14.4 (7.3)	15.9 (8.1)	15.8 (7.5)	15.8 (8.0)
Q2	0.379 (0.485)	0.325 (0.468)	0.388 (0.487)	0.367 (0.482)	0.304 (0.460)	0.288 (0.453)	0.298 (0.458)	0.276 (0.447)
Q3	0.299 (0.458)	0.292 (0.455)	0.257 (0.437)	0.278 (0.448)	0.349 (0.477)	0.346 (0.476)	0.339 (0.473)	0.341 (0.474)
Q4	0.149 (0.356)	0.226 (0.418)	0.147 (0.354)	0.179 (0.384)	0.199 (0.399)	0.233 (0.423)	0.195 (0.396)	0.212 (0.409)
RATEV	0.220 (0.121)	0.248 (0.130)	0.213 (0.123)	0.228 (0.129)	0.248 (0.124)	0.258 (0.127)	0.246 (0.125)	0.252 (0.126)
RURAL	0.140 (0.347)	0.173 (0.378)	0.126 (0.331)	0.138 (0.345)	0.112 (0.316)	0.119 (0.323)	0.115 (0.319)	0.126 (0.332)
LAW2	0.429 (0.495)	0.241 (0.428)	0.371 (0.483)	0.402 (0.490)	0.356 (0.479)	0.148 (0.355)	0.305 (0.460)	0.227 (0.419)
LAW3	0.334 (0.472)	0.387 (0.487)	0.328 (0.470)	0.370 (0.483)	0.310 (0.463)	0.458 (0.498)	0.309 (0.462)	0.411 (0.492)
LAW4	0.038 (0.190)	0.031 (0.174)	0.036 (0.187)	0.011 (0.106)	0.023 (0.150)	0.054 (0.227)	0.027 (0.163)	0.061 (0.239)
YCURVE	2.482 (0.412)	2.376 (0.451)	2.503 (0.422)	2.447 (0.442)	2.391 (0.426)	2.353 (0.443)	2.397 (0.426)	2.377 (0.434)

Mean (standard deviation) are reported in each cell.

Table 3
Estimated odds ratio

Variables	FRM				ARM			
	Purchase		Refinance		Purchase		Refinance	
	Penalty	No penalty	Penalty	No penalty	Penalty	No penalty	Penalty	No penalty
AA (non-Hispanic White)	1.837**	1.636	1.616***	1.236	1.405***	1.397*	1.174*	1.040
LAT (non-Hispanic White)	1.709*	2.886*	1.071	1.289	1.660***	1.517**	0.941	0.845
FICO (+60 units)	0.237***	0.252***	0.198***	0.223***	0.301***	0.361***	0.223***	0.252***
LTVE (+10 units)	2.746***	2.460*	1.733***	1.954***	3.490***	5.529***	1.896***	1.822***
LAMTE (+US$ 10,000)	0.807***	0.503***	0.875***	0.864***	0.698***	0.708***	0.953***	0.943***
INC (+US$ 1000)	1.050*	2.042***	0.998	0.998	1.406***	1.288***	0.988	1.002
DOC (low/no doc)	0.234***	0.229***	0.320***	0.299***	0.204***	0.157***	0.463***	0.505***
MULTI (single family)	1.275	1.610	1.461	1.119	1.763***	1.432*	1.073	1.250
CONDO (single family)	0.988	0.728	1.038	1.298	0.978	0.876	0.820*	0.957
AG1 (AG5)	1.937**	0.412**	1.223	0.660*	1.435***	1.240	0.642***	0.556*
AG2 (AG5)	0.892	0.766	0.877	0.879	4.055***	6.719***	2.983***	2.575***
AG3 (AG5)	1.513	2.056	1.362	1.010	52.300***	60.703***	21.911***	31.031***
AG4 (AG5)	1.865	1.235	1.387	2.335	0.031***	0.032***	0.191***	0.245***
CSPREAD (+0.1 units)	1.258	1.010	1.332	1.010	0.599***	0.705*	0.546***	0.616**
YCURVE (+1 unit)	1.657	3.047	0.224*	1.085	0.712	0.506	0.516*	1.010
HPIV (+1 unit)	1.035*	1.104***	0.998	0.991	1.059***	1.060***	1.014*	1.009
RATEV (+0.1 units)	1.232	2.353	0.745	1.290	2.112***	1.769*	1.941***	2.290***
Q2 (Q1)	0.341***	0.410*	0.467***	0.397***	0.341***	0.378***	0.323***	0.322***
Q3 (Q1)	0.763	0.438	0.872	0.554	0.316***	0.284***	0.305***	0.266***
Q4 (Q1)	1.212	0.322	0.808	0.705	0.218***	0.210**	0.248***	0.257**
CDIV2 (CDIV1)	0.970	1.221	1.106	2.030	0.851	0.748	1.513*	1.446
CDIV3 (CDIV1)	1.439	0.470	1.682*	0.646	0.946	0.671	2.109***	1.242
CDIV4 (CDIV1)	0.781	0.961	0.873	2.094	0.578**	0.972	1.543*	1.642
CDIV5 (CDIV1)	0.906	0.220	1.740*	1.408	0.512***	0.930	1.402*	1.872
CDIV6 (CDIV1)	0.799	0.183	1.397	1.099	0.488***	0.491	2.206***	0.976
CDIV7 (CDIV1)	1.770	0.052*	1.351	0.929	0.621*	0.760	2.181**	3.543*
CDIV8 (CDIV1)	0.745	0.326	1.012	0.916	0.420***	0.497	1.480*	2.014
CDIV9 (CDIV1)	0.544	0.392	0.784	0.834	0.329***	0.827	1.327	2.951**
LAW2 (LAW1)	0.599	1.702	1.065	0.860	1.087	0.473***	1.000	0.829
LAW3 (LAW1)	0.566*	0.565	0.834	0.723	0.639***	0.868	0.933	0.869

Table 3 (*Continued*)

Variables	FRM				ARM			
	Purchase		Refinance		Purchase		Refinance	
	Penalty	No penalty	Penalty	No penalty	Penalty	No penalty	Penalty	No penalty
LAW4 (LAW1)	0.427*	0.407	1.120	1.254	0.547	0.640	1.273	0.964
RURAL (Central City)	1.314	1.586	1.142	1.685*	1.178*	1.112	1.015	1.239
NCCITY (Central City)	1.059	0.572*	0.859	0.971	0.868*	0.919	1.022	0.986
AAST (+1 unit)	1.019	1.080*	0.997	1.022	1.010	0.985	1.001	0.997
LATST (+1 unit)	0.980*	1.006	0.983*	1.012	0.979***	1.006	0.988*	0.976*
Summary statistics								
Number of lenders	29	31	28	35	33	32	30	32
Number of observations	2235	1444	5918	2881	4657	13,321	11,950	6520
Nagelkerke *R*-squared (%)	57	68	54	55	45	54	46	42
KS statistics (All)	62	77	63	61	54	60	53	53
KS statistic (Black = 1)	63	59	65	55	58	61	56	51
KS statistic (Hispanic = 1)	64	73	67	62	52	59	53	61
KS statistic (Black = 0, Hispanic = 0)	61	80	61	62	54	59	53	51

For categorical variables, the odds ratio is between one category and the control category, which is shown within the parenthesis. For continuous variables, the odds ratio is a change in the estimated odds of receiving a higher-rate loan when the variable increases by the number of units shown within the parenthesis.

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

ranges from 0.50 for fixed rate loans for purchasing without prepayment penalties to 0.95 for adjustable-rate loans for refinancing with prepayment penalties. The odds of a borrower who provided less than full documentation receiving a higher-rate loan are 1.98–6.37 times greater than for a borrower who provided full documentation. In addition, while not reported here, findings were also robust across models with alternative specifications of independent variables (results available on request).

In general, our analyses show that race and ethnicity were significant factors in determining whether borrowers received higher-rate home loans. That is, African-American and Latino borrowers are more likely to receive higher-rate loans than non-Latino white borrowers with similar risk factors for many categories of subprime loans. The significance of race was particularly consistent for loans with prepayment penalties, while the impact of ethnicity was concentrated in loans for home purchases. Specifically, for loans with prepayment penalties, all else being equal, the odds of an African-American borrower receiving a higher-rate loan range from 1.17 to 1.84 times greater than for a non-Latino white borrower (depending on loan product). With respect to ethnicity, all else being equal, for purchase loans the odds of a Latino borrower receiving a higher-rate loan range from 1.52 to 2.89 times greater than for a non-Latino white borrower (again, depending on loan product). All of the above odds ratios are significantly different from 1.0 at a 95% confidence-level. For most classes of loans without prepayment penalties, the odds of receiving a higher-rate loan are not significantly different between African-American and non-Latino white borrowers at a 95% confidence level. Similarly, in refinance products, the odds of receiving a higher-rate loan are not significantly different between Latino and non-Latino white borrowers at that level.

5. Discussion

Efficient financial markets should provide similarly situated borrowers with equally competitive prices on subprime home loans. In fact, subprime lenders construct complicated pricing matrices in the form of "rate sheets" in an effort to meet this challenge. These rate sheets describe how to calculate applicable interest rates from a borrower's credit score, the amount of equity held by the borrower in the home, and several other factors that measure risk. Lenders' internal fair lending compliance operations aim to ensure that these criteria are valid and not based on impermissible discriminatory factors. This investment is prudent, since lenders face serious legal and reputation risks if they violate fair lending standards.

Yet, in multiple analyses that control for the major risk factors lenders explicitly use to set prices, we find that borrowers' race and ethnicity continue to exert a statistically significant influence on the cost of their subprime mortgages. Disparities tended to be larger for fixed-rate loans than for their adjustable-rate counterparts. In addition, African-American and Latino borrowers face the highest risks for pricing disparities under different circumstances. Relative to similarly situated non-Latino white borrowers, African-American borrowers were at greatest risk of receiving a higher-rate loan when their subprime mortgage included a prepayment penalty while Latino borrowers were at greatest risk when they used their mortgage to purchase a home rather than to refinance an existing home loan. What then could explain these results?

While rate sheets do present objective pricing schedules for calculating a loan's interest rate, they are not definitive statements of a loan's price for a given borrower. In addition to interest rate, yield-spread premiums, points and fees are significant elements of a loan's pricing. Variations on these discretionary charges could account for part or all of the differences in subprime loan pricing among non-Latino white, African-American, and Latino borrowers.

The difference between the price paid for a loan with an inflated interest rate and the price that would have been paid for the loan had the borrower received the lowest rate for which he or she qualified is called a yield-spread premium (YSP). We note that subprime lenders' rate sheets routinely stipulate that brokers can only maximize the amount of a YSP if the loan carries a prepayment penalty, which ensures that a subprime lender will receive either extra-interest or penalty income sufficient to offset the up-front cash payment to a broker. This lending practice is consistent with our finding that African-American borrowers were at greatest risk of receiving a higher-rate loan when their subprime mortgage included a prepayment penalty. Jackson and Burlingame (2007) also found that African-American and Latino borrowers paid mortgage brokers more for their services than other borrowers.

Borrowers of color also would be more likely to receive higher-rate subprime loans if they tended, on average, to receive their loans from lenders that generally charge more than the lenders predominately serving white borrowers. Apgar and Calder (2005) have noted that "even though mortgage loans are now readily available in low-income minority communities, by employing high-pressure sales practices and deceptive tactics, some mortgage brokers push minority borrowers into higher-cost subprime mortgages that are not well suited to their needs and can lead to financial problems down the road." This sort of targeting might help explain the disparities observed among Latino borrowers in our dataset. It might be the case that Latinos who take a subprime mortgage to purchase a home are more likely to be recent immigrants. If so, the higher disparities we observe in the purchase market for Latinos may arise from the targeting of recent immigrants by higher-cost lenders. While we generally expect that efficient markets will result in borrowers selecting for themselves loan options with the lowest costs, substantial evidence, apart from the findings presented in this paper, exists to support the notion that borrowers are not finding their way to the best-priced home loan (Calem, Hersaff, & Wachter, 2004; Lax et al., 2004).

If brokers or other more expensive loan originators are disproportionately providing loans to borrowers of color, it is fair to ask whether lower-cost lenders are under-serving such customers. In this explanation, white borrowers receive disproportionately fewer higher-rate loans not because borrowers of color are targeted for such loans, but because the latter are excluded from lower-cost subprime loans. For example, in 2004, the U.S. Department of Justice filed two cases against lenders for failing to lend in communities of color.

Like all social science studies, the study presented here has limitations. First, APR spread is an imperfect measure for examining pricing data, since it essentially blends interest rates with points and fees in a way that assumes that borrowers will keep the loan for its entire term and, consequently, it tends to underemphasize costs arising from fees. However, in the context of this study, it is unlikely this limitation would undercut our basic findings, since it is unlikely that preferences for fee-rate tradeoffs systematically vary by race or ethnicity in ways that are uncorrelated with credit score, income, LTV, or other factors already included in our analysis. Moreover, to the extent that borrowers of color are targeted for high-fee predatory lending (Engel & McCoy, 2003; Renuart, 2003), such patterns would tend to lead to underestimated pricing disparities between these borrowers and white borrowers due to APR's tendency to minimize the cost of fees.

Second, because HMDA only provides APR-spread information for higher-rate loans, our analysis is limited to comparing the relative odds and likelihoods of receiving these higher-rate loans. Unlike Ambrose et al. (2004), we did not estimate the magnitude of the differences in APRs between loans. Moreover, since the proprietary data only contains subprime loans, this analysis neither allows for an evaluation of pricing disparities that includes the prime market, nor provides any insight into how different borrowers end up with subprime rather than prime loans. In addition, the dataset only includes securitized loans. However, since the proprietary dataset used in this analysis covers about 85% of the subprime market originated in 2004, the results can reasonably be generalized to the subprime market.

Third, unlike Avery et al. (2005), our database was not large enough to control for specific metropolitan statistical areas or lenders. Therefore, our findings do not prove that any one lender or group of lenders is explicitly discriminating based on race or ethnicity of the borrower. However, our analysis does account for potential geographical disparities on loan pricing arising from location (central city, non-central city or rural) where the property was located and state housing prices, census regions, status of state laws regarding judicial foreclosure and deficiency judgment, as well as state racial and ethnic compositions. In addition, we included regulating agencies as independent variables, which essentially controlled for lender type. Moreover, if market segmentation caused the disparities, inclusion of lender as a control variable will be inappropriate.

Finally, our models may omit information that is correlated with both APR and the race and ethnicity of borrowers (e.g., recent mortgage delinquency history might exhibit such a pattern). Though we were able to control for the majority of risk-based characteristics that lenders generally use to price loans, at least according to rate sheets, it is nevertheless possible that omitted variables could influence our results.

6. Conclusion and recommendations

In this article, we find evidence that African-American and Latino borrowers are more likely to receive higher-priced subprime credit than similarly situated white borrowers. In other words, we conclude that the market is not providing similarly priced subprime credit to similarly situated borrowers. Given the importance of homeownership to wealth building and the current wealth gap between white Americans and communities of color, these findings suggest a critical need for increased scrutiny of the mortgage market by regulators, policymakers and researchers. Along these lines, regulatory enforcers of fair lending laws should evaluate whether the higher-cost lenders "reverse red-line" by targeting communities of color for high-priced products unrelated to individual borrower risk and whether lower-cost lenders are neglecting such communities. Further, while additional research on the precise mechanisms through which the disparate outcomes observed in this article result would be helpful, we note with interest that recent policy debates have already raised a variety of relevant issues ranging from discretionary pricing structures that give mortgage brokers the incentive to steer borrowers to more expensive loan options (referred to as "yield spread premiums") to the application of a "suitability" requirement on loan originators similar to the rules currently applicable to investment advisors.

Finally, in the course of this article, we have noted limitations of publicly available data collected pursuant to the Home Mortgage Disclosure Act. Certainly, the collection of APR information described in this report is a positive first step in assessing pricing information, but it is of only limited value without a full disclosure of points and fees on subprime mortgages, including up-front fees, yield spread premiums, and prepayment penalties. In addition, as this study shows, HMDA data currently lacks information that would be helpful in evaluating how lenders serve their markets. HMDA would be more helpful if it contained information on factors such as loan-to-value ratios and credit scores of borrowers. Moreover, HMDA data could be markedly improved simply by requiring the reporting of origination channel for each application, so that researchers could better assess pricing differences or similarities among broker, correspondent and retail originations.

Acknowledgements

We would like to thank Alan White for his comments on a version of this paper as it was presented at the 2007 Federal Reserve System's Fifth Community Affairs Research Conference in Washington D.C. Thanks also to David Rodda, Eric Belsky and Raphael Bostic for their thoughtful insights and suggestions on a previous draft of this paper and to Loretta Mester and one anonymous referee for their helpful suggestions. In addition, we would like to thank current and former colleagues for their assistance: Peter Smith, Jamie Goodson, and Mary Moore.

References

Ambrose, B., LaCour-Little, M., & Sanders, A. (2004). The effect of conforming loan status on mortgage yield spreads: A loan level analysis. *Real Estate Economics*, *32*(4), 541–569.

Apgar, W. C., & Calder, A. (2005). *The dual mortgage market: The persistence of discrimination in mortgage lending*. Joint Center for Housing Studies, Harvard University.

Avery, R. B., Canner, G. B., & Cook, R. E. (2005). New information reported under HMDA and its implication in fair lending enforcement. *Federal Reserve Bulletin*, *91*, 344–394.

Bradford, C. (2002). *Risk or race? Racial disparities and the subprime refinance market*. Center for Community Change.

Calem, P., Gillen, K., & Wachter, S. (2004). The neighborhood distribution of subprime mortgage lending. *Journal of Real Estate Finance and Economics*, *29*(4), 393–410.

Calem, P. S., Hersaff, J. E., & Wachter, S. M. (2004). Neighborhood patterns of subprime lending: Evidence from disparate cities. *Housing Policy Debate*, *15*(3), 603–622.

Engel, K. C., & McCoy, P. A. (2003). Revisiting a tale of three markets: The law and economics of predatory lending. *Texas Law Review*, *80*, 439–444.

Ernst, K., & Goldstein, D. (2005). *Comment on federal reserve analysis of home mortgage disclosure act data*. Center for Responsible Lending.

HUD-Treasure Joint Task Force. (2000). *Curbing predatory home mortgage lending*. Washington, DC: U.S. Department of Housing and Urban Development.

Inside Mortgage Finance Publications. (2006). *The 2006 mortgage market statistical annual*. Bethesda, MD: Inside Mortgage Finance Publications.

Jackson, H. E., & Burlingame, L. (2007). Kickbacks or compensation: The case of yield spread premiums. *Stanford Journal of Law, Business and Finance*, *12*, 289–358.

Lax, H., Manti, M., Raca, P., & Zorn, P. (2004). Subprime lending: An investigation of economic efficiency. *Housing Policy Debate*, *15*(3), 533–571.

Munnell, A. H., Tootell, G. M. B., Browne, L. E., & McEneaney, J. (1996). Mortgage lending in Boston: Interpreting HMDA data. *The American Economic Review*, *86*(1), 25–53.

Renuart, E. (2003). Toward one fair and competitive mortgage market: Suggested reforms in a tale of three markets point in the right direction. *Texas Law Review*, *82*, 421–438.

Ross, S. L., & Yinger, J. (2006). Uncovering discrimination: A comparison of the methods used by scholars and civil rights enforcement officials. *American Law and Economics Review*, *8*(3), 562–614.

White, A. M. (2004). Risk-based mortgage pricing: Present and future research. *Housing Policy Debate*, *15*(3), 503–531.

Zellner, A., & Theil, H. (1962). Three-stage least squares: Simultaneous estimation of simultaneous equations. *Econometrica*, *30*, 54–78.