RACE, ETHNICITY AND HIGH-COST MORTGAGE LENDING

Patrick Bayer
Fernando Ferreira
Stephen L. Ross

Race, Ethnicity and High-Cost Mortgage Lending
Patrick Bayer, Fernando Ferreira, and Stephen L. Ross
NBER Working Paper No. 20762
December 2014
JEL No. G0,G21,R21,R3

ABSTRACT

This paper examines how high cost mortgage lending varies by race and ethnicity. It uses a unique panel data that matches a representative sample of mortgages in seven large metropolitan markets between 2004 and 2008 to public records of housing transactions and proprietary credit reporting data. The results reveal a significantly higher incidence of high costs loans for African-American and Hispanic borrowers even after controlling for key mortgage risk factors: they have a 7.7 and 6.2 percentage point higher likelihood of a high cost loan, respectively, in the home purchase market relative to an overall incidence of 14.8 percent among all home purchase mortgages. Significant racial and ethnic differences are widespread throughout the market – they are present (i) in each metro area, (ii) across high and low risk borrowers, and (iii) regardless of the age of the borrower. These differences are reduced by 60 percent with the inclusion of lender fixed effects, implying that a significant portion of the estimated market-wide racial differences can be attributed to differential access to (or sorting across) mortgage lenders.

Patrick Bayer
Department of Economics
Duke University
213 Social Sciences
Durham, NC 27708
and NBER
patrick.bayer@duke.edu

Fernando Ferreira
The Wharton School
University of Pennsylvania
1461 Steinberg - Dietrich Hall
3620 Locust Walk
Philadelphia, PA 19104-6302
and NBER
fferreir@wharton.upenn.edu

Stephen L. Ross
University of Connecticut
Department of Economics
341 Mansfield Road, Unit 1063
Storrs, CT 06269-1063
stephen.l.ross@uconn.edu

1. Introduction

Whether African-American and Hispanic mortgage borrowers face a higher cost of credit than comparable white borrowers has been a long-standing question in academic and policy debates about inequities in financial markets. Interest in this question has been motivated by several related issues. Most directly, a large legal history and scholarly literature has considered claims of racial discrimination in mortgage lending. While historically focusing on discrimination in mortgage underwriting (Ross and Yinger 2002) and redlining against minority neighborhoods (Holmes and Horvath 1994; Tootell 1996), attention has increasingly turned to differences in the price of mortgage credit (Ross 2005; Haughwout, Tracy, and Chen In Press), including several high profile U.S. Department of Justice cases in the wake of the recent financial crisis.[1] A second motivation for estimating racial and ethnic differences in mortgage lending has been the concern that these differences create barriers to homeownership, contributing to low minority homeownership rates and growing wealth disparities (Belsky, Retsinas and Duda 2005; Herbert, Haurin, Rosenthal and Duda 2005; and Quercia, McCarthy and Wachter 2003). And, finally, risk based pricing and high cost mortgage lending has been a defining feature of the subprime mortgage market.[2] In the recent housing boom, lending in the subprime market was heavily concentrated in minority neighborhoods, potentially contributing to especially high foreclosure rates in these neighborhoods in the subsequent financial crisis

---

[1] Recent cases have been filed or settled against National City Bank, Wells Fargo, GFI Mortgage Bankers and Bank of America based on the past actions of Countrywide Mortgage.

[2] Demyanyk and Van Hemert (2011) find that the quality of loans deteriorated for six consecutive years before the crisis and that securitizers were aware of this quality decline. They argue that the subprime mortgage market followed a classic lending boom-bust scenario with unsustainable growth leading to the collapse of the market.

(Gerardi and Willen 2009; Reid and Laderman 2009; Edminston 2009; Calem, Gillen, and Wachter 2004).[3]

The Home Mortgage Disclosure Act of 1975 was amended in 1989, expanding regulatory coverage to include non-depository lenders and mandating the creation of a loan-level database that eventually captured virtually the entire population of mortgage applications nationwide. Historically, the HMDA database has been used to document large racial and ethnic differences in the likelihood of having a mortgage approved (Avery, Beeson and Sniderman 1996). While a substantial fraction of the racial differences are certainly attributable to differences in borrower and loan risk factors, Munnell, Tootell, Browne, McEneany (1996) in a seminal, but controversial,[4] study showed that substantial racial differences remained in a sample of HMDA loans in Boston even after controlling for detailed mortgage risk factors. With the growing size of the subprime mortgage market and the increased use of risk based pricing, the HMDA database was expanded in 2004 to contain information on whether the APR[5] on a loan exceeded the interest rate on 10 year treasury notes by 3 percentage points. Loans that exceed this threshold are often described as rate spread loans, and this threshold is typically used to identify high cost or subprime loans. Avery, Canner and Cooke (2005) documented large racial and ethnic differences in the incidence of rate spread loans in HMDA, but were unable to control for common mortgage risk factors (e.g., borrower credit score), which are not included in the HMDA database.[6]

---

[3] Bhutta and Canner (2013) and Bayer, Ferreira and Ross (In Press) show that substantial racial and ethnic differences in foreclosure exist during the crisis even after controlling for traditional mortgage risk factors.
[4] See Ross and Yinger (2002) for a detailed review of the debate surrounding the "Boston Fed" study.
[5] The Annual Percentage Rate or APR includes both the interest or note rate on the loan and the effect of closing costs on the cost of credit.
[6] The sometimes very high rates of interest charged in the subprime sector has led to a significant debate about whether these loans are predatory. Bond, Musto and Yilmaz (2009) develop a model of predatory lending that implies highly collateralized loans, inefficient refinancing of subprime loans, lending without due regard to ability to

In light of this limitation of the HMDA database, most of the studies that have documented differences in the prevalence of high cost loans have used one of two sources: proprietary data aggregated from individual lenders or data obtained directly from individual lenders. Ghent, Hernandez-Murillo, and Owyang (2013), Haughwout, Mayer and Tracy (2009), Bocian, Ernst and Lee (2006) used proprietary data aggregated from the reports of many lenders and merged with the Home Mortgage Disclosure Act database.[7] While these studies have documented substantial unexplained differences by race, they are often restricted to samples that represent a subset of the market, usually emphasizing loans that are securitized privately or lenders that operate primarily in the subprime sector. Several other studies have examined the mortgage pricing behavior of individual lenders (Black, Boehm, DeGennaro 2003; Nelson 2005; Courshane 2007; Courshane and Nickerson 1997). These studies have found very small, if any, within-lender differences between white and minority borrowers in the incidence of high cost mortgage credit.

In this study, we examine racial and ethnic differences in the incidence of high cost mortgage loans in a market-wide sample covering several large U.S. metropolitan areas or regions. The shift to market-wide data changes the question being asked from whether similar borrowers receive different prices from the same lender (e.g., disparate treatment discrimination) to whether unexplained racial differences exist in market outcomes, a phenomenon that Heckman (1998) described as market discrimination.[8] Significant market level differences in the

---

pay, prepayment penalties, balloon payments, and poorly informed borrowers. Agarwal, Amromin, Ben-David, Chomsisengphet, and Evanoff (2014) examine the effects of an anti-predatory loan program in Chicago finding that the program cut lending activity in half mostly through the exit of lenders specializing in especially risky loans.

[7] A larger set of studies examine racial and ethnic patterns of high cost lending at the neighborhood level solely using proprietary data (Mayer and Pence 2007; Mayer, Pence and Sherlund 2009; Reid and Laderman 2009; Fisher, Lambie-Hanson and Willen 2010).

[8] See Ross and Yinger (2002) and Blackburn and Vermilyea (2006) for evidence of significantly larger market level differences in mortgage underwriting than the differences observed at the lender level.

price of credit have important consequences for the dynamics of racial and ethnic inequality in homeownership, wealth, and credit worthiness, even if only small differences exist at the lender level.

To build the database for our analysis, we first linked HMDA data on home purchase and refinance mortgages between 2004 and 2008 to public records data on housing transactions and liens in seven distinct metropolitan housing markets. The public records data contain information on all liens as well as the name and address of the individual purchasing the housing unit or refinancing their mortgage and in many cases the name of the individual's spouse.[9] We drew a sample of matched mortgages that were originated between May and August of each year and provide the names and addresses resulting from this sample to one of the major credit reporting agencies. The credit reporting agency then used the name and address to match borrowers to archival credit reporting data from March 31 in the year preceding the mortgage origination and from March 31 for every following year through 2009 providing in each year a vantage credit score plus detailed credit line information from each individual's report. The credit reporting agency also provided information on borrower age, which has not typically been available in studies of mortgage lending or pricing.

Our empirical analysis reveals significant unexplained racial and ethnic differences in the incidence of high cost or subprime mortgage credit. These differences persist after controlling for detailed measures of borrower and loan characteristics including credit score, ratio of the loan amount to housing price, presence of subordinate liens, and housing and debt expenses relative to individual income. Relative to a model based on control variables available in HMDA, the

---

[9] Information on subordinate liens is typically not available except in lender provided samples of mortgages because only individual loans are tracked in most mortgage samples, not entire mortgage transactions. See Foote, Gerardi, Goette, and Willen (2010) for another example where HMDA is matched to housing transaction data in order to obtain information on subordinate liens.

inclusion of these additional controls erodes about half of the racial and ethnic differences in loan pricing. Still, the remaining differentials are sizable with African American and Hispanic borrowers having a 7.7 and 6.2 percentage point higher likelihood of a rate spread or high cost loan, respectively, in the home purchase market relative to an overall incidence of 14.8 percent among home purchase mortgages. In the refinance sample, the estimated differences are smaller at 4.2 and 1.7 percentage points compared to a base rate of 17.1 percent. These loan-pricing differences exist across a variety of large metropolitan housing markets including not only faster-growing markets in California and Florida that experienced especially severe housing market downturns, but also slower-growing Eastern and Midwestern housing markets.

The further inclusion of lender fixed effects substantially erodes the unexplained differences for all groups by 60 to 70 percent.[10] These findings suggest that sorting across or differential access to lenders plays a significant role in creating market wide differences in mortgage pricing, a finding that is consistent with the high concentration of subprime loans in minority neighborhoods.

In the home purchase market, further analysis indicates that differences in the likelihood of a high cost loan are distributed widely across the distribution for African-American borrowers. In particular, substantial differences in the incidence of high cost loans remain for the subsample of borrowers who have prime credit scores, conventional loan to value ratios, and reasonable debt expense to income ratios. On the other hand, the differences for Hispanic borrowers are concentrated primarily among borrowers with high LTV ratios. For refinance mortgages, the

---

[10] These lender fixed effect estimates are comparable to the findings in Munnell et al. (1996) of underwriting discrimination in Boston, except that the racial differences arising from their within lender comparisons were 80 percent of sample denial rates, significantly larger than the 2.9 to 22.3 percent within lender racial and ethnic differences found in this study.

pattern for African-Americans is similar to that of Hispanic borrowers with the greatest differences arising for borrowers with subprime credit scores and high LTV mortgages.

Finally, we examine the geographic dimension of racial and ethnic differences in high cost lending by interacting two measures of neighborhood and market demographics with borrower race and ethnicity. The first control variable is percentage of households in poverty within the borrower's census tract, which is intended to proxy for neighborhood level disadvantage. The second variable is intended to capture information on the broader geographic housing and mortgage submarket in which current homebuyers participate. Specifically, we use data at the county level from the American Community Survey in order to measure the fraction of recent movers into owner-occupied housing in a county who have less than two years of college by race and ethnicity in order to capture the demographic composition of recent homebuyers in each county. We find that the observed racial and ethnic differences in the incidence of high cost loans are concentrated in high poverty census tracts for both African-Americans and Hispanics. Further, for African-Americans, racial differences tend to be concentrated in the counties where recent African-American homebuyers have lower levels of education.

Taken as a whole, the results of our analysis imply that African-American and Hispanic borrowers have a higher incidence of high cost loans, even after controlling for a detailed set of standard underwriting variables designed to measure credit risk, and that these racial and ethnic differences are not confined to a particular segment of the market. In this way, even among borrowers with comparable credit scores, loan terms, homes and locations, African-Americans or Hispanics are much more likely to have a high cost loan. Further, even African-Americans with favorable credit scores and loan terms experience a significantly higher incidence of high cost

loans than equivalent white borrowers, and these differences are most prevalent in specific neighborhoods and submarkets where disadvantaged borrowers are concentrated. While these market-level differences do not necessarily imply discrimination on the part of individual lenders, the differential exposure to high cost loans can impact a wide array of subsequent outcomes including wealth accumulation, rates of delinquency and default, credit scores, and long-term home ownership.

The remainder of this paper proceeds as follows. The next section presents the data used in the analysis and our econometric model. Section 3 then presents the main results, and heterogeneity estimates by several measures. Finally, section 4 concludes the paper.

2. **Data and Model Specification**

Our data are based on public Home Mortgage Disclosure Act (HMDA) data from between 2004 and 2008 and proprietary housing transaction/lien and assessor's databases purchased from Dataquick Inc.[11] We begin with a convenience sample of seven major housing markets where Dataquick has information on refinance mortgages going back to at least 2004: Chicago IL CMSA, Cleveland OH MSA, Denver CO MSA, Los-Angeles CA CMSA, Miami-Palm Beach Corridor, Maryland Counties excluding Baltimore City, and San Francisco CA CMSA. We restrict our HMDA data to home purchase or refinance mortgages on owner-occupied, 1-4 family properties. In the Dataquick sample, we eliminate non-arm's length transactions, transactions where the name field contains the name of a church, trust, or where the first name is missing, and transactions where the address could not be matched to a 2000 Census

---

[11] The property transaction data is collected by Dataquick or by intermediaries from county assessor's offices and contains a population of all sales and liens of all types including refinance mortgages, home improvement loans, and home equity lines of credit from the present back to 1995 to 1997 for most states and back to 1988 for California.

tract or the zip code was missing.[12] The HMDA and Dataquick data are then merged based on year, loan amount, name of lender, state, county and census tract. We obtain high quality matches for approximately 50% of our HMDA sample.[13]

Next, we draw a sample of mortgages to provide to a credit reporting agency. These mortgages were sampled from May through August so that the March 31st archival credit report for the year of the mortgage provides appropriate information on the borrowers' credit quality prior to obtaining the mortgage. We oversample mortgages to minority borrowers, mortgages to white borrowers in minority or low-income neighborhoods, and high cost mortgages as designated in HMDA as high rate spread loans. In order to maximize the number of minority loans given the likelihood of sample saturation, we first draw the following oversamples based on race and ethnicity: 500 in each site, year and group (400 for 2004) selected randomly from mortgages to African-American borrowers, mortgages to Hispanic borrowers, and mortgages to white borrowers in minority or low-income neighborhoods.[14] We then split the remaining sample into rate spread and non-rate spread loans drawing 1000 borrowers associated with rate spread loans in each year and site (800 for 2004) and 2714 borrowers (2286 for 2004) from the non-rate spread sample in each year and site. Weights are developed based on the probability of selection,[15] and initialized so that each site receives equal weight in the pooled sample.[16]

---

[12] This eliminates very few records due to the high quality of the name and address records in the assessor files.

[13] The key factor limiting the match rate is the lender name because the lender of record in the local assessor's data often differs from the respondent in HMDA. Less restrictive match criteria can yield a match rate closer to 80%, but in order to be conservative we restricted ourselves only to instances where we successfully match on lender name.

[14] Our budget at the credit reporting agency is based on number of borrowers so whenever a mortgage is sampled which contains the name of the co-borrower, typically the borrower's spouse, this was counted as having sampled two borrowers.

[15] The sampling is explicitly based on 8 strata for each site: African-American borrowers, Hispanic borrowers, white borrowers in minority or low-income neighborhoods, and all other borrowers divided into rate spread and non-rate spread loans. All loans from the same strata and year receive equal weight.

[16] We have a convenience sample of housing markets so it would be inappropriate to weight based on the number of mortgages. In any stratified sampling scheme, Los Angeles, which dominates our sample in terms of total number of HMDA mortgages, would be selected with certainty while housing markets like Denver and Cleveland would be

This sample is provided to a major credit repository who matches the name and address of each borrower and co-borrower to archival credit report data from the March 31st preceding the mortgage transaction and March 31st for every year that follows this transaction through 2009. Our match rate for the pre-mortgage archive is 81.4 and 84.5 percent in the home purchase and refinance samples, respectively. For years following the mortgage, the match rate rises by 4 to 5 percentage points. In many cases, these individuals also may not have had sufficient information on record when the lender requested a report for the credit reporting agency to provide a credit score, in which case lack of a score matches the information that the lender would have had when approving and pricing the loan, but lenders can enter by hand additional information that is not available to us such as social security number or previous addresses.[17]

Table 1 illustrates the impact of our match process on the sample mean of whether the loan is a high cost (rate spread) loan,[18] race and ethnicity of the borrower, gender, loan amount, family income of the borrower, whether there is a co-borrower, whether the loan is with a non-depository lender,[19] relative lender size,[20] whether the loan is a jumbo,[21] and census tract

---

assigned to a stratum with other similarly sized and located metropolitan areas and if chosen would receive a higher weight (offsetting the smaller number of mortgages) based on the probability of being selected from the stratum.

[17] For home purchase mortgages, we only observe the address of the new housing unit, but in practice this does not present a major problem for the credit data match because the archival data can be matched based on current and several past addresses and in practice we observe only a small difference between the home purchase and refinance match rate.

[18] The rate spread variable is based on whether the Annual Percentage Rate or APR, which includes both the interest rate and the effect of closing costs on the cost of credit, is 2 or more percentage points above the yield on the 10 year treasury bond. HMDA only reports the actual APR if it exceeds this rate spread, and so our dependent variable is defined as a binary variable capturing whether the APR exceeds the reporting threshold or the rate spread.

[19] This variable is based on the lenders regulator or the agency code variable in HMDA.

[20] The relative lender size is based on number of loans in market divided by the maximum number of loans for a single lender in that market so that it always falls between zero and one. The mean is relatively high because the very largest lenders dominate the sample of loans.

[21] The jumbo variable is set to one if the loan amount exceeds the jumbo threshold for loans on single family homes which was $333,700 in 2004, 359,650 in 2005, and 417,000 in 2006 through 2008. In the second half of 2008, higher thresholds were temporarily approved for high cost markets. However, loan amounts above $417,000 have

variables including median income, percent African-American, Hispanic and Asian residents, percent of households in poverty, percent of properties owner-occupied, percentage of households in poverty and the ratio of mean rents to mean house values.[22] The first column shows the mean for the entire HMDA sample for our seven sites where each site receives equal weight in the mean. The second column shows the mean for our HMDA-Dataquick match, and the third column restricts our sample to mortgages between May and August. The fourth column shows the weighted mean for the sample of mortgages that was provided to the credit reporting agency.   The last column in Table 1 shows the weighted means on these common variables for just the subsample where the name and addressed was matched to the minimum amount of credit line data in order to generate a record.[23] The sample composition is quite stable except for a moderate decline in share white and moderate increase in loan amount between columns 1 and 2 associated with the difficulty of matching lender names between HMDA and the Dataquick provided assessor files. While our HMDA-Dataquick match algorithm loses 50 percent of the HMDA mortgages, the composition of the match sample is quite similar to the composition of the population of mortgages, and the other aspects of our sample construction have virtually no impact on the composition of mortgages over key attributes.

---

never been viewed as fully conforming by the GSE's, and early concerns about the temporary nature of these higher limits limited the impact of these limits especially during the second half of 2008.

[22] The last variable is a common measure of neighborhood equity risk because current rents can only be high relative to values if investors expect rents to fall in the future.

[23] Some borrowers are not matched with a credit score because insufficient credit information was available for that borrower or co-borrower. If a credit score is not observed for both the borrower and co-borrower, the observation will be dropped in our regression analysis. Similar results are observed using the full sample with dummies for missing credit score data, but the resulting racial and ethnic differences are slightly larger in those models. In order to be conservative, we present results using a regression sample that is restricted to observations where a credit score is matched.

Table 2 shows the weighted means for our final home purchase and refinance subsamples that were successfully merged to pre-mortgage credit report data.[24] The first two columns show the mean and standard deviation for our sample of home purchase mortgages, and the last two columns show these values for refinance mortgages. The first set of rows present the full set of demographic, loan and census tract variables that are available in HMDA and that we use in our regressions. From the match with transaction data, we observe the presence and size of subordinate liens, whether the liens are fixed or variable rate mortgages, the loan to value ratio based on sales price for home purchase mortgages and on an estimated value based on either previous sales price[25] or assessed value when a previous sale is unobserved for refinance mortgages,[26] and detailed property attributes including whether a single family home, a condominium, and number of units on the property. The borrowers' (or if unavailable co-borrower's) Vantage score is drawn from the credit report data from the March 31st prior to the mortgage origination.[27] The credit report observation following the mortgage is used to obtain monthly mortgage payment, which when combined with HMDA income is used to calculate the

---

[24] This sample is somewhat smaller than the last column in Table 1 because some small lenders could not be identified based on the reporting restrictions of the credit reporting agency. If the lender was not identified, the observation is dropped from the regression sample. As with credit score, similar results are observed using the full sample with dummies for missing lender identity, but the resulting racial and ethnic differences are larger in those models (primarily in the model with lender FE's). Again, in order to be conservative, we present results using the smaller sample.

[25] We use our extensive housing transaction data to develop both a hedonic and repeat sales quarterly price index for each county. When we observe a previous sale of the property, we simply adjust that earlier sales price to estimate current value based on the hedonic index. However, the repeat sales index yields quite similar estimates.

[26] When a previous sale is not observed, we use the county assessment and adjust that value by the average ratio of sales price to assessed value for that county and quarter, see Clapp, Nanda and Ross (2008). In California, our refinance sample is restricted to mortgages where a previous purchase is observed because property assessments are uninformative as to the value of the underlying property. This restriction is feasible because the Dataquick data in California contains transactions back to the late 1980s.

[27] The Vantage Score is a proprietary credit score developed by the credit reporting agencies as an alternative to the traditional FICO index of credit score. The two scores are very highly correlated.

mortgage payment to income ratio.[28]   The monthly mortgage payment is combined with debt

payments from the pre-mortgage credit data and HMDA income to calculate debt payment to

income ratio. Finally, age is observed for many borrowers and co-borrowers in the credit history

files.

In terms of our model specification, we regress whether the loan was high cost (i.e., a

HMDA rate spread loan) on the detailed borrower, tract, and loan attributes from Table 2 plus

year by week fixed effects and site by origination year fixed effects.[29]   Separate models are

estimated for home purchase and refinance mortgages, as well as by site in a later analysis. The

loan to value ratio is included as bins below 0.6, 0.6 to 0.8, 0.8 to 0.84, 0.85 to 0.89, 0.90 to 0.94,

0.95 to 1.00, 1.00 to 1.04, and 1.05 and above. In addition to controlling for the combined loan to

value ratio, we control for the number of subordinate liens and whether each is a fixed or

variable rate loan. The vantage score is included as a series of dummy variables based on 20-

point intervals. The mortgage payment and debt to income ratios are also divided into bins. The

bins vary in size because the data is thin for unusual income ratios. For mortgage payment to

income ratios, the smallest bins are 0.02 near the pre-crisis secondary market criteria of 0.33, and

for total debt payment to income ratios the smallest bins are 0.03 near the pre-crisis threshold of

---

[28] The mortgage payment for the current mortgage is only observed in the credit line data from the year following
the mortgage. However, in most instances, borrowers who are matched by the credit reporting agency prior to the
mortgage are also matched in the following year.

[29] The fraction of loans classified as rate spread loans in HMDA is affected by the spread between treasury and
market mortgage rates. These spreads changed substantially in late 2004 so that the fraction of rate spread loans is
much lower in 2004 than other years. Using the information available on APR above the rate spread threshold, we
defined an alternative rate spread variable holding constant the fraction of rate spread loans in a housing market,
year and sector (home purchase or refinance) at the fraction observed in 2004. While magnitudes may vary, results
presented are robust to alternative definitions of the rate spread variable that define rate spread loans as a constant
fraction of the market.

0.45.[30]   Finally, we include either controls for lender size based on loan volume and dummy

variables for the agency that regulates the lender or in some models lender fixed effects.

3.   Rate Spread Models

Table 3 presents the rate spread regression results for the pooled samples with the

estimates for home purchase mortgages in panel 1 and for refinance mortgages in panel 2. For

comparison with results in the previous literature, the first column presents the model with just

the standard HMDA controls including the demographic variables, family income, a jumbo loan

dummy amount, the census tract attributes, and the year by site fixed effects.   The second

column adds additional controls made available by merging the HMDA data with the Dataquick

housing transaction data including loan to value ratio, whether the loan is an adjustable rate

mortgage, information on subordinate liens, and year by week fixed effects. The third column

adds the dummies for credit score and housing and debt expense to income ratio categories. The

fourth column adds additional controls for the potential effect of subprime lending, identifying

borrowers with Vantage scores below 701 as subprime borrowers[31] and then interacting the

subprime dummy with dummy variables associated with key thresholds of loan to value ratio,

debt to income ratio, mortgage payment to income ratio,[32] the presence of subordinate debt and

whether the primary mortgage is adjustable rate. The fifth column includes lender fixed effects.

The rows present the coefficient estimates for race and ethnicity categories.

---

[30] The formal GSE income ratio guidelines were 0.28 and 0.36, respectively, but these guidelines were relaxed substantially during the period leading up to the crisis due in part to increased reliance of the GSE's on automated underwriting programs and the data indicates that the GSE's purchased many loans with income ratios above these formal guidelines.

[31] The credit reporting agencies that developed the Vantage score algorithms describes scores below 701 as non-prime. Further, a Vantage score of 701 is comparable to a FICO score of 660, a common FICO threshold for subprime, in that in both cases approximately 30% of individuals have credit scores below these thresholds.

[32] The loan to value thresholds of 0.80, 0.90, 0.95 and 1.00, the debt to income thresholds used are 0.36 and 0.45, and the mortgage payment to income ratio thresholds used are 0.28 and 0.33

Based on the standard HMDA controls, we find 17.1 and 11.6 percentage point differences in the likelihood of receiving a rate spread loan for African-American and Hispanic borrowers relative to whites for home purchase mortgages, while differences are small for Asians. For refinance mortgages, the estimated differences for African-Americans and Hispanics are smaller, 10.6 and 4.3 percentage points, respectively.

The addition of standard underwriting controls in columns 2 and 3 reduces the estimated differences for African-American and Hispanic borrowers to 8.0 and 6.1 for the home purchase and 4.5 and 1.7 for the refinance market, reductions on the order of 50 percent for all four coefficients,[33] while the inclusion of additional subprime controls in Column 4 has little impact on the estimated differences.[34] In the home purchase market, these differences are consistent with 54.0 and 41.2 percent differentials measured as a share of the overall incidence of rate spread loans, and in the refinance market the differentials are 26.3 and 9.9 percent for African-Americans and Hispanics, respectively. These results imply both that a significant portion of the observed racial and ethnic differences of the receipt of high cost loans by race and ethnicity can be explained by differences in standard underwriting variables *and* that economically and statistically significant differences remain even after controlling for these most commonly used measures of credit worthiness and risk.

The addition of lender fixed effects model in column 5 leads to substantially eroded, but still statistically significant differences in the incidence of high cost loans. The point estimates in the home purchase sample decline from 8.0 and 6.1 to 3.3 and 2.5 percentage point differences

[33] The coefficients on the additional controls suggest that the model is well specified. For example, we find that the likelihood of rate spread loans changes monotonically with the vantage score, loan to value ratio, housing expense to income ratio and debt expense to income ratio bins in the expected directions, and we find that the likelihood of a rate spread loan is higher for jumbo loans and for primary loans with subordinate liens.

[34] The addition of LTV in column 2 and credit score and income ratios in column 3 both explain a significant fraction of the racial and ethnic differences, especially in the home purchase market.

for African-Americans and Hispanics, respectively, and for the refinance sample differences decline from 4.5 and 1.7 to 1.7 and 0.5.[35] The inclusion of lender fixed effects shifts the focus from understanding market disparities between equally qualified minority and white borrowers[36] to racial and ethnic differences observed within lenders. As evidence of discrimination, the lender fixed effect estimates are comparable to the findings in the Munnell et al. study of underwriting discrimination in Boston, which also used lender fixed effects in a sample combining loan applications for many lenders in a common market. However, the racial differences arising from their within lender comparisons were significantly larger, 80% or 8 percentage point difference over a 10 percent rejection rate, than the within lender racial and ethnic differences in the incidence of rate spread loans, which fell between 2.9 and 22.3 percent.

These results imply that a sizeable majority of the unexplained racial and ethnic differences in market outcomes can be explained by the differential access to traditional lenders and/or selection into high cost or subprime lenders. Many users of the subprime market are qualified for financing in the primary market based on assessment using automated underwriting tools (FreddieMac, 2000), and Lax, Manti, Raca, and Zorn (2004) find that only half of the two percentage point difference between prime and subprime interest rates can be explained by differential credit risk and servicing costs. As noted earlier, subprime lending tends to be concentrated in predominantly minority neighborhoods (Geradi and Willen 2009; Reid and Laderman 2009; Edminston 2009; Calem, Gillen, and Wachter 2004). Further, a National Community Reinvestment Coalition paired testing study found that minority testers were never counseled on up referral to an affiliated prime lender, while 7 percent of white testers were

---

[35] These findings are comparable to Avery, Canner, and Cooke (2005) who estimate racial and ethnic differences in the incidence of rate spread loans controlling for information available in HMDA and find that lender fixed effects explain a majority of the unexplained racial and ethnic differences.
[36] This phenomenon has been described by Heckman (1998) as market discrimination.

counseled (Harney, 2006), and in Chicago Ross, Turner, Godfrey and Smith (2008) found that minority testers received less time and attention than white testers during pre-application inquires. Stein and Libby (2001) found that three-quarters of all subprime borrowers in California that they surveyed did not approach a bank prior to applying for a mortgage, one-third experienced aggressive marketing as the lender attempted to initiate a loan, and almost three-quarters claimed that loan terms changed for the worse at closing; and Courchane, Surette and Zorn (2004) find that subprime borrowers are less likely to search for a better interest rate and are less likely to be offered a choice of mortgage products.

Heterogeneity in Racial and Ethnic Differences

Having presented our baseline findings in Table 3, the remainder of our analysis aims to provide further insight in the nature of the observed racial and ethnic differences by exploring how these effects vary along a number of dimensions including (i) metropolitan area, (ii) borrower and loan characteristics, and (iii) residential location.

Table 4 presents the estimated results for each metropolitan housing market. The structure of the table follows Table 3 except that the columns represent in order Chicago, Cleveland, Denver, Los Angeles, Maryland Counties, Miami-Palm Beach Corridor, and San Francisco Bay Area. The first panel presents the estimates for the subprime model in column 4 of Table 3 for the home purchase sample, the second panel presents the lender fixed model for the same sample, and the third and fourth panels present the results for the corresponding models using the refinance sample. While there is some variation, racial and ethnic differences exist for all seven sites in the home purchase sample in models both with and without lender FE's. In the home purchase market without lender FE's, differences range between 4.5 and 10.3 for African-

Americans, and 5.4 and 10.9 for Hispanics. The inclusion of lender FE's lowers these differences to ranges of 1.7 to 4.1, and 1.1 to 3.3, respectively. Significant differences exist for all groups in most sites for the refinance sample in models without lender FE's, but differences are confined to a small number of areas for refinance mortgages once lender FE's are included in the models. For refinance mortgages in the model without lender FE's, the significant estimated differences range between 2.1 and 5.4 for African-Americans and 1.1 and 3.0 for Hispanics. Taken together, we conclude that the market wide differences in the incidence of high cost loans are present in all of our market areas especially in the home purchase sample. Estimated differences for Asians are small in all sites except for Denver in the home purchase sample.

Next in order to assess how widespread racial and ethnic differences are across the mortgage market, we estimate models in which group membership is interacted with three key risk variables: subprime credit score or Vantage score below 701, non-conforming loan to value ratio or a ratio above 0.95, and a high debt to income ratio, i.e. above 0.45. Panel 1 replicates the results from Table 3 for comparison purposes, and Panel 2 presents the estimated effects by race, ethnicity and loan terms. Starting with the subprime model for the home purchase sample shown in Table 5 Column 1, we continue to find large racial differences in the likelihood of a high cost loan for low-risk African-American borrowers, i.e. those with prime credit scores, conforming loan to value ratios and reasonable debt to income ratios. In particular, low-risk African-American borrowers have a 7.6 percentage point higher likelihood of receiving a rate spread compared to low-risk white borrowers (an identical gap to the estimate reported for the full sample in Panel 1). This estimate falls to 2.6 percentage points (column 2) when lender fixed effects are included in the model, suggesting that about two-thirds of the racial differences for low-risk borrowers can be attributed to differential sorting across (or access to) lenders. For the

18

refinance mortgage market as show in columns 3 and 4, the estimates imply a greater concentration of the racial difference in the incidence of the high cost loans among higher-risk borrowers, although significant differences remain for low-risk African-American borrowers relative to low-risk whites.

There are much smaller differences between low-risk Hispanic and white borrowers in the incidence of high cost loans (2.2 percentage points without lender fixed effects and basically zero with fixed effects). Instead, having a non-conforming LTV increases the Hispanic-white difference in the incidence of high cost loans by 10.5 and 5.1 percentage points in models without and with lender fixed effects, respectively. In this way, the overall differences estimated for Hispanic borrowers in Table 3 appear to be driven by especially large differences for a group of borrowers with a particular mortgage risk factor, rather than by widespread differences throughout the entire market.

Table 6 presents the results for models where we interact race and ethnicity with age dummies. Lax, Manti, Raca and Zorn (2004) observe that older, potentially more vulnerable borrowers are more likely to have subprime loans, and so we examine whether older African-American borrowers have especially likely to have a rate spread loan. While racial and ethnic differences are slightly lower for the youngest African-American and Hispanic borrowers, we do not observe any systematic relationship between age and estimated racial and ethnic differences for other borrowers.

Finally, we estimate additional models that interact geographic controls for borrower location with race and ethnicity. Our first control is simply the percent of households in poverty within the census tract where the borrower resides, which is included as a general proxy for a disadvantaged neighborhood. Second, we use the American Community Survey (ACS) to

measure the education level of borrowers in particular housing and mortgage submarkets. Specifically, we pool data from the 2005-2009 ACSs selecting heads of households who reside in owner-occupied housing and for which everyone in the household has resided in the house for less than one year as a proxy for individuals who recently purchased a home with a mortgage. Then, by county and by race/ethnicity category we measure the fraction of these recent movers into owner-occupied housing who do not have at least two years of college.[37] In this way, we identify counties for which specific groups have below average levels of education. These models are only estimated for the home purchase models because the ACS does not contain information on whether households refinanced their mortgage.

Table 8 presents the results for this model using the home purchase sample. The first and the third column present the baseline results from Columns 4 and 5 of Table 3 after the direct inclusion of these geographic variables.[38] The second and four columns present the results after the inclusion of the interactions of these two variables with the race and ethnicity dummy variables. A higher neighborhood percent poverty implies significantly higher rates of high cost lending for both African-Americans and Hispanics, and a higher share of same race borrowers with less than two years of college in a county is associated with more rate spread loans for African-American borrowers. In fact, racial and ethnic differences are near zero for borrowers in low poverty rate neighborhoods and in counties where borrowers typically have at least two years of college. These results hold for both the models without and with lender fixed effects.

---

[37] Similar results are obtained using fractions by county by group by year matching the 2005 ACS with the 2004 originations etc, but the sample sizes within many cells are quite small leading to substantial measurement error, which attenuates the estimated effects.

[38] Percent poverty was in the original Table 3 specifications as part of the broad vector of neighborhood controls and is positively associated with a higher incidence of rate spread loans. The share homebuyers in the county without two years of college is added to this specification, but is statistically insignificant and its inclusion has virtually no effect on the estimates of racial and ethnic differences.

The finding that racial and ethnic differences are concentrated in neighborhoods with elevated poverty rates is consistent with the literature that documents the concentration of subprime lending in poor and minority neighborhoods (Mayer and Pence 2007; Mayer, Pence and Sherlund 2009; Reid and Laderman 2009; Fisher, Lambie-Hanson and Willen 2010). To our knowledge, however, ours is one of the few studies to document such neighborhood effects after the inclusion of detailed underwriting controls, and ours is the first study to document the increased level of racial and ethnic differences in the incidence of high cost lending in disadvantaged neighborhoods.

Several studies also document the correlation between high cost lending and either borrower or neighborhood education levels (Courchane, Surette and Zorn 2004; Calem, Hershaff and Wachter 2004; Smith, Fink and Huston 2011).[39] As with many of those studies, we cannot distinguish between differences that arise between borrowers with different levels of education and differences that arise across locations or submarkets with different average or typical education levels because education is not observed in the vast majority of mortgage data sets. Regardless, the finding that the county average education level of African-American borrowers is associated with larger racial differences in the incidence of high costs loans, with no similar finding for Hispanics, is suggestive that less educated African-American borrowers are at a significant disadvantage in the mortgage market, especially given the earlier findings that racial differences in high cost lending are distributed throughout the quality distribution of African-American mortgages.

---

[39] Also see Germais (In Press) who finds a link between neighborhood education levels and mortgage borrower mistakes on loan terms, and Campbell (2006) who documents the role of borrower education in explain the rate of poor decisions in financial investment.

4. Summary and Conclusion

In this paper, we identify robust differences between whites and minority borrowers in the likelihood of receiving a rate spread mortgage in both the home purchase and refinance market after controlling for detailed borrower and loan attributes. A substantial fraction of these differences are attributable to sorting across (or differential access to) lenders as opposed to differential treatment of equally qualified applications by lenders. Racial and ethnic differences are observed in all metropolitan or regional markets in most samples and specifications with the exception of the refinance sample when lender fixed effects are included in the pricing model.

In the home purchase sample where estimated differences are the largest, racial differences are widespread with large differences arising even for borrowers with prime credit scores, conforming loan to value ratios, and reasonable debt to income ratios. On the other hand, ethnic differences are concentrated among borrowers with either subprime credit scores or non-conforming loan to value ratios. For both groups, differences are concentrated among minority borrowers residing in higher poverty rate neighborhoods, and for Hispanics differences are very small in low poverty rate neighborhoods. For African-American borrowers, substantial differences remain, but these differences are entirely explained by a second location control, i.e. the education composition of African-American borrowers at the county level. Therefore, while racial differences may persist among credit worthy borrowers in low poverty neighborhoods, the remaining racial differences are concentrated in submarkets where African-American borrowers have lower levels of education.

The results of our analysis have important implications for the dynamics of racial and ethnic differences along a number of dimensions related to wealth, credit-worthiness and home ownership. In particular, the greater financial burden associated with high cost loans not only

leads directly to slower wealth accumulation due to the higher mortgage payments, but also to a higher risk of future delinquency and default, with serious long-term consequences for long-term credit scores and home ownership rates. In fact, using the same sample of loans, Bayer, Ross and Ferreira (2012) show that having a rate spread loan is associated with an 8.0 percentage point higher incidence of foreclosure notices on an individual's credit report among home purchase loans. This implies that the unexplained racial and ethnic differences estimated in this paper (in models without lender fixed effects) are associated with a 0.62 and 0.50 percentage point increase in foreclosure rates for African-Americans and Hispanics, respectively, relative to a population average foreclosure rate of 5.3 percent, i.e. 12 and 9 percent increases in foreclosure rates. Importantly, these implications hold regardless of the underlying explanation for the large observed market-wide differences in the incidence of high cost loans.

The results of our analysis also have several implications for strategies that are likely to be effective at reducing racial and ethnic differences in the incidence of high cost loans. First, across our entire analysis, the inclusion of lender fixed effects substantially reduced the estimated coefficients, by over 50 percent in every specification. The strong explanatory power of lender fixed effects suggests that the structure of the mortgage market involving separate prime and subprime subsidiaries for most major lenders may play a much larger role in creating mortgage market disparities than differential treatment of whites and minorities applying for credit through the same credit market channel. In fact, recent Justice Department settlements with Countrywide and Wells Fargo specifically focused on the impact of cost differentials between the prime and subprime subsidiaries and the potential steering of minority borrowers (Savage, 2011, 2012).

References

Apgar, William, Alegra Calder, and Gary Fauth. 2004. *Credit, Capital and Communities: The Implications of the Changing Mortgage Banking Industry for Community-Based Organizations*. Cambridge, MA: Joint Center for Housing Studies, Harvard University.

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, and Douglas D. Evanoff. 2014. Predatory lending and the subprime crisis. *Journal of Financial Economics*, 113(1), 29-52.

Avery, Robert B., Patricia E. Beeson, and Mark S. Sniderman. 1996. Accounting for Racial Differences in Housing Credit Markets. In *Mortgage Lending, Racial Discrimination, and Federal Policy* (J. Goering and R. Wienk, Eds.). Washington, DC: Urban Institute Press.

Avery, Robert B., Glen B. Canner, and Robert E. Cooke. 2005. New Information Reported under HMDA and Its Implications in Fair Lending Enforcement. *Federal Reserve Bulletin* (Summer 2005).

Bayer, Patrick, Fernando Ferriera and Stephen L. Ross. In Press. The Vulnerability of Minority Homeowners in the Housing Boom and Bust. *American Economic Journal: Economic Policy*.

Bayer, Patrick, Fernando Ferriera and Stephen L. Ross. 2012. The Vulnerability of Minority Homeowners in the Housing Boom and Bust. NBER Working Paper #19020.

Belsky, Eric S., Nicolas P. Retsinas and Mark Duda. 2007. Chapter 9, The Financial Returns to Low-Income Homeownership, In: *Chasing the American Dream: new perspectives on affordable homeownership,* ed. William M. Rohe and Harry L. Watson, p. 191-212. Ithaca and London: Cornell University Press.

Bhutta, Neil and Glenn B. Canner. 2013. Mortgage Market Conditions and Borrower Outcomes: Evidence from the 2012 HMDA Data and Matched HMDA--Credit Record Data. Federal Reserve Bulletin, 99.

Black, Harold A., Thomas P. Boehm, and Ramon P. DeGennaro, 2003, "Is There Discrimination in Mortgage Pricing? The Case of Overages." *Journal of Banking and Finance*, 27(6), pp. 1139-65.

Blackburn, McKinley and Todd Vermilyea. 2006. *Journal of Financial Services Research*, 29, 125-47

Bocian, Debbie G., Keith S. Ernst, and Wei Li. 2006. *Unfair Lending: The Effect of Race and Ethnicity on the Price of Subprime Mortgages*. Durham, NC: Center for Responsible Lending.

Bond, Philip, David K. Musto, and Bilge Yilmaz. 2009. Predatory mortgage lending. *Journal of Financial Economics*, 94(3), 412-427.

Calem, Paul S., Kevin Gillen, and Susan Wachter. 2004. The Neighborhood Distribution of Subprime Mortgage Lending, *Journal of Real Estate Finance and Economics*, 29, 393-410.

Calem, Paul S., Jonathan E. Hershaff, and Susan M. Wachter. 2004. Neighborhood Patterns of Subprime Lending: Evidence from Disparate Cities. *Housing Policy Debate*, 15, 603-622.

Campbell, John Y. 2006. Household Finance. *The Journal of Finance*, 61(4), 1553–1604.

Chan, Sewin, Andrew Haughwout, and Joseph Tracy. In Press. How Mortgage Finance Affects the Urban Landscape. In the Handbook of Urban and Regional Economics (Eds. G. Duranton, W. Strange and V. Henderson).

Courchane, Marsha J. 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, 399-439.

Courchane, Marsha and David Nickerson. 1997. Discrimination Resulting from Overage Practices. *Journal of Financial Services Research*, 11, 133-151.

Courchane, Marsha J., Brian J. Surette, and Peter M. Zorn. 2004. Subprime Borrowers: Mortgage Transitions and Outcomes, *Journal of Real Estate Finance and Economics*, 29(4): 365-392.

Demyanyk, Yuliya and Otto Van Hemert. 2011. Understanding the Subprime Mortgage Crisis. *Review of Financial Studies*, 24(6), 1848-1880.

Edminston, Kelly D. 2009. Characteristics of High-Foreclosure Neighborhoods in the Tenth District. *Economic Review (2nd Quarter)*, 51-75.

Fisher, Lynn, Lauren Lambie-Hanson, and Paul S. Willen. 2010. A Profile of the Mortgage Crisis in Low and Moderate Income Neighborhoods. Boston Federal Reserve Working Paper.

Foote, Christopher, Kristopher Gerardi, Lorenz Goette, and Paul S. Willen. 2010. "Reducing Foreclosures: No Easy Answers." In NBER Macroeconomics Annual Volume 24, ed. Daron Acemoglu, Kenneth Rogoff, and Michael Woodford, 89–138. Chicago: University of Chicago Press.

Germaise, Mark J. In press. Borrower Misreporting and Loan Performance. *The Journal of Finance*.

Gerardi, Kristopher S. and Paul Willen. 2009. Subprime Mortgages, Foreclosures, and Urban Neighborhoods. *BE Journal of Economic Analysis and Policy*, 9 (3), Symposium Article 12.

Ghent, Andra C., Rubén Hernández-Murillo, and Michael T. Owyang. 2013. Race, Redlining and Subprime Mortgage Pricing. St Louis Federal Reserve Working Paper.

Harney, Kenneth. 2006. Study Finds Bias in Mortgage Process. *Washington Post* (June 17, 2006, page F01).

Haughwout, Andrew , Christopher Mayer, and Joseph Tracy. 2009. Subprime Mortgage Pricing: The Impact of Race, Ethnicity and Gender on the Cost of Borrowing. New York Federal Reserve Working Paper.

Heckman, James J. 1998. Detecting Discrimination. *Journal of Economic Perspectives*, 12, 101-116.

Herbert, Christopher E., Donald R. Haurin, Stuart S. Rosenthal, and Mark Duda. 2005. "Homeownership Gaps Among Low Income and Minority Borrowers and Neighborhoods." Washington, DC: U.S. Department of Housing and Urban Development. http://www.huduser.org/publications/HOMEOWN/HGapsAmongLInMBnN.html.

Holmes, Andrew and Paul Horvitz. 1994. Mortgage Redlining: Race, Risk, and Demand. *The Journal of Finance*, 49(1) 81-99.

Lax, Howard, Michael Manti, Paul Raca, and Peter Zorn. 2000. Subprime Lending: An Investigation of Economic Efficiency. *Housing Policy Debate*, 15, 533-571.

Mayer, Chris and Karen Pence. Subprime Mortgages: What, Where and To Whom? Columbia University Working Paper.

Mayer, Chris, Karen Pence and Shane M. Sherlund. 2009. The Rise in Mortgage Defaults. *Journal of Economic Perspectives*, 1.

Mian, Atif and Amir Sufi . 2009. "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis." *Quarterly Journal of Economics*, 124(4), 1449–96.

Munnell, Alicia H., Geoffrey M. B. Tootell, Lynn E. Browne, and James McEneaney. 1996. Mortgage Lending in Boston: Interpreting HMDA Data. *American Economic Review*, 86, 25-53.

Nelson, Ashlyn Aiko. 2005. *Better Credit, Better Schools: Credit Quality, Race, and Discrete Choice Residential Sorting*. Doctoral Dissertation. Stanford University.

Quercia, Roberto G., George W. McCarthy, and Susan M. Wacther. 2003. "The Impacts of Affordable Lending Efforts on Homeownership rates." *Journal of Housing Economics* 12(1), p. 29-59.

Reid, Carolina and Elizabeth Laderman. 2009. The Untold Costs of Subprime Lending: Examining the Links among Higher-Priced Lending, Foreclosures and Race in California, Federal Reserve Bank of San Francisco Working Paper.

Ross, Stephen. 2005. The Continuing Practice and Impact of Discrimination. University of Connecticut Working Paper 2005-19.

Ross, Stephen L., Margery Austin Turner, Erin Godfrey, and Robin R. Smith. "Mortgage Lending in Chicago and Los Angeles: A Paired Testing Study of the Pre-application Process." *Journal of Urban Economics* 63 (May 2008): 902-919.

Ross, Stephen L. and John Yinger. 2002. *The Color of Credit: Mortgage Lending Discrimination, Research Methodology, and Fair-Lending Enforcement.* Cambridge, MA: MIT Press.

Savage, Charlie. 2012. Wells Fargo Will Settle Mortgage Bias Charges. The New York Times (July 12, 2012).

Savage, Charlie. 2011. Countrywide Will Settle a Bias Suit. The New York Times (July 12, 2012).

Smith, Hyrum, Michael S. Finke, and Sandra J. Huston. 2011. The Impact of Financial Sophistication on Adjustable Rate Mortgage Ownership. *Journal of Financial Counseling and Planning* 22: 3-15.

Stein, Kevin and Margaret Libby. 2001. *Stolen Wealth: Inequities in California's Subprime Mortgage Market*. Durham, NC: Center for Responsible Lending.

Tootell, Geoffrey. 1996. Redlining in Boston: Do Lenders Discriminate Against Neighborhoods. *The Quarterly Journal of Economics*, 111, 1049-1079.

Table 1: Sample Selection of HMDA Variables

Variable	HMDA full sample		DQ High quality match		DQ High quality match May-Aug		Sample, weighted		Credit Data Matched Sample, weighted[2]	
	Mean	Stand. Dev.	Mean	Stand. Dev.	Mean	Stand. Dev.	Mean	Stand. Dev.	Mean	Stand. Dev.
African American	0.111	0.339	0.116	0.345	0.114	0.343	0.114	0.343	0.112	0.315
Hispanic	0.174	0.410	0.194	0.427	0.193	0.426	0.192	0.426	0.185	0.388
Asian	0.075	0.284	0.086	0.303	0.085	0.301	0.086	0.303	0.089	0.284
White	0.678	0.505	0.601	0.529	0.605	0.528	0.605	0.528	0.611	0.487
Loan Amount (in 1000s)	247	243	271	221	274	224	274	227	278	211
Applicant Income (in 1000s)	107	142	105	128	105	127	106	132	106	115
Tract Median Income (in 1000s)	59.1	25.6	59.6	25.2	59.7	25.2	59.7	25.2	60.4	23.5
Tract Pct African American	0.126	0.238	0.116	0.225	0.115	0.224	0.115	0.224	0.113	0.204
Tract Pct Hispanic	0.169	0.227	0.165	0.221	0.164	0.220	0.165	0.221	0.163	0.202
Tract Pct Asian	0.063	0.109	0.065	0.112	0.065	0.111	0.065	0.111	0.066	0.104
Number of Observations	9,345,709		4,002,996		1,459,468		273,589		238,785	

Notes: The first column presents the means and standard deviations for all Home Mortgage Disclosure Act (HMDA) home purchase and refinance mortgages in our seven market areas between 2004 and 2008 where each market is given equal weight in the means. The second column presents the means for the subsample where we have a high quality match between HMDA and the housing transaction file based on loan amount, type of loan, census tract and lender name, and the third column is based on further restricting the sample to mortgages originated between May and August. The fourth column presents means for the stratified sample that was merged to the credit history data weighted by the inverse of the sampling probabilities again with equal weights for each market, and the fifth column presents weighted means for the subsample that was successfully merged to an established credit history prior to mortgage origination.

Table 2: Descriptive Statistics

	Purchase Sample		Refinance Sample	
HMDA Data	Mean	Std. Dev.	Mean	Std. Dev.
Rate Spread	0.148	0.110	0.171	0.122
American Indian	0.003	0.020	0.003	0.023
Asian	0.098	0.110	0.086	0.118
African American	0.090	0.078	0.129	0.105
Hispanic	0.188	0.128	0.183	0.132
White	0.621	0.164	0.599	0.176
Male	0.642	0.165	0.639	0.174
Female	0.356	0.165	0.358	0.173
Loan Amount (in 1000s)	289.3	74.4	262.7	69.1
Applicant Income (in 1000s)	107.9	37.1	100.8	34.8
Coborrower Present	0.347	0.169	0.467	0.184
Jumbo Loan	0.263	0.155	0.180	0.151
Tract Median Income (in 1000s)	60059	8272	60263	8746
Tract Share African American	9.995	5.428	12.484	7.112
Tract Share Hispanic	15.799	6.786	16.994	7.232
Tract Share Asian	6.402	3.606	7.067	4.271
Tract Share Owner Occupant	68.27	8.17	69.17	8.14
Tract Share in Poverty	0.004	0.001	0.004	0.001
Tract Rent/Price	7.699	2.333	7.936	2.558
Dataquick Data				
Loan to Value Ratio	0.884	0.086	1.310	5.964
Subordinate Lien	0.003	0.018	0.063	0.081
First Lien Adjustable Rate	0.453	0.174	0.445	0.183
Condo	0.218	0.143	0.142	0.128
Mobile	0.001	0.011	0.001	0.014
Single Family	0.771	0.145	0.835	0.135
Lot Size (sf in 1000s)	12902	42999	12789	35464
Unit square feet (in 1000s)	1788	10552	1678	5144
Number of Bathrooms	1.975	2.511	1.983	0.396
Number of Bedrooms	2.150	2.618	1.990	0.621
Number of Stories	1.180	0.564	1.214	0.460
Units in Building	1.387	4.703	1.510	7.830
Credit Data				
Vantage Score	784.9293	0.3571101	779.8701	0.4003712
Mortgage Payment to Income Ratio	0.2626551	0.0010006	0.2498142	0.0024357
Debt Payment to Income Ratio	0.3274709	0.0012041	0.3505712	0.0028699
Borrower Age	27.77746	0.0801383	35.20718	0.0903688
Sample Size	120,732		115,763	

Notes: The first two columns present means and standard deviations for our sample of post-mortgage credit reports for each home purchase mortgage borrower for which a credit score was observed prior to mortgage origination. The last two columns present the same information for the post-mortgage reports of mortgage refinancers.

Table 3. Rate Spread Models

			Home Purchase Sample		
Variable Names	HMDA	Dataquick	Credit Data	Subprime	Lender FE
Asian	0.0084***	0.01263***	0.0098***	0.0102***	0.0051**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
African American	0.1711***	0.1329***	0.0800***	0.0768***	0.0327***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.003)
Hispanic	0.1169***	0.0792***	0.0613***	0.0620***	0.0249***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	120,732	120,732	120,732	120,732	120,732
R-squared	0.224	0.311	0.369	0.418	0.594

			Refinance Sample		
Variable Names	HMDA	Dataquick	Credit Data	Subprime	Lender FE
Asian	0.0069*	0.0055	0.0099***	0.0097***	0.0043
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
African American	0.1064***	0.0967***	0.0455***	0.0419***	0.0172***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Hispanic	0.0432***	0.0290***	0.0170***	0.0173***	0.0050*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	115,763	115,763	115,763	115,763	115,763
R-squared	0.169	0.234	0.347	0.384	0.555

Notes: The table presents the estimates on racial and ethnic dummy variables for a linear probability model of whether the borrower received a rate spread loan as defined in the Home Mortgage Disclosure Act (HMDA) data. The first panel presents estimates for a sample of home purchase mortgages, and the second panel presents estimates for a sample of refinance mortgages. The first column presents estimates for a model that includes the standard controls available in HMDA, the second column is based on adding controls made available by the merge with the housing transaction data, and the third column is based on adding controls made available by the merge with credit reporting data prior to the mortgage origination. The fourth column specification includes the interaction of key loan terms with whether the borrower had a subprime credit score, and the final column specification includes lender fixed effects.

Table 4A. Metropolitan Area Estimates: Home Purchase Sample

	Subprime Model						
Variable Names	Chicago	Cleveland	Denver	Los Angeles	Maryland	Miami	San Francisco
Asian	-0.016601**	-0.000598	0.041067***	0.008678	-0.002892	0.038082**	0.019873***
	(0.008)	(0.009)	(0.011)	(0.007)	(0.007)	(0.018)	(0.004)
Black	0.087678***	0.057813***	0.070384***	0.046933***	0.074821***	0.103266***	0.055526***
	(0.012)	(0.009)	(0.011)	(0.010)	(0.007)	(0.010)	(0.009)
Hispanic	0.055575***	0.010872	0.058828***	0.054424***	0.068167***	0.067046***	0.068724***
	(0.008)	(0.011)	(0.007)	(0.006)	(0.008)	(0.008)	(0.006)
Observations	16,673	15,449	17,282	17,670	17,867	17,416	18,375
R-squared	0.427	0.407	0.398	0.434	0.418	0.457	0.409

	Lender Fixed Effect Model						
Variable Names	Chicago	Cleveland	Denver	Los Angeles	Maryland	Miami	San Francisco
Asian	-0.012413*	-0.001078	0.034481***	0.003173	-0.012308**	0.020421	0.004621
	(0.007)	(0.007)	(0.009)	(0.007)	(0.006)	(0.015)	(0.004)
Black	0.039574***	0.027588***	0.027102***	0.016852**	0.032805***	0.047284***	0.015321**
	(0.010)	(0.008)	(0.010)	(0.009)	(0.006)	(0.009)	(0.008)
Hispanic	0.019452***	0.011615	0.023155***	0.019768***	0.020666***	0.033113***	0.022658***
	(0.007)	(0.010)	(0.006)	(0.005)	(0.007)	(0.007)	(0.005)
Observations	16,673	15,449	17,282	17,670	17,867	17,416	18,375
R-squared	0.586	0.583	0.585	0.627	0.636	0.601	0.637

Table 4B. Metropolitan Area Estimates: Refinance Sample

Variable Names	Subprime Model						
	Chicago	Cleveland	Denver	Los Angeles	Maryland	Miami	San Francisco
Asian	-0.002592	0.017576	-0.008664	0.010181	0.011034	0.010618	0.007224*
	(0.012)	(0.021)	(0.013)	(0.009)	(0.012)	(0.022)	(0.004)
Black	0.048830***	0.024649*	0.021422*	0.035434***	0.043667***	0.054432***	0.028042***
	(0.012)	(0.013)	(0.011)	(0.009)	(0.009)	(0.010)	(0.008)
Hispanic	0.029117***	-0.005653	0.030217***	0.021825***	0.015993*	0.020390**	0.010716**
	(0.008)	(0.017)	(0.009)	(0.006)	(0.009)	(0.008)	(0.005)
Observations	16,358	14,030	16,673	17,031	17,171	16,902	17,598
R-squared	0.420	0.371	0.333	0.360	0.443	0.429	0.283

Variable Names	Lender Fixed Effect Model						
	Chicago	Cleveland	Denver	Los Angeles	Maryland	Miami	San Francisco
Asian	-0.006155	0.010619	-0.021589*	0.009468	0.003507	-0.000479	0.002306
	(0.009)	(0.018)	(0.011)	(0.008)	(0.011)	(0.019)	(0.003)
Black	0.015606	0.010079	-0.000242	0.018679**	0.015747**	0.027021***	0.007116
	(0.011)	(0.012)	(0.011)	(0.009)	(0.007)	(0.009)	(0.007)
Hispanic	0.012992*	0.009339	0.017030**	0.009017	0.005642	0.001030	0.002403
	(0.007)	(0.015)	(0.007)	(0.006)	(0.008)	(0.007)	(0.004)
Observations	16,358	14,030	16,673	17,031	17,171	16,902	17,598
R-squared	0.615	0.574	0.554	0.500	0.621	0.576	0.492

Notes: The table presents the estimates on racial and ethnic dummy variables for a linear probability model of whether the borrower received a rate spread loan. The first and second panels present estimates for a sample of home purchase mortgages, and the third and fourth panels present estimates for a sample of refinance mortgages. The first and third panels present estimates for models that include all of the standard control variables plus the interaction of key loan terms with whether the borrower had a subprime credit score, and the second and fourth panel specifications are based on the same controls plus lender fixed effects. Each column presents estimates for the subsample of mortgages in a particular metropolitan market.

Table 5. Rate Spread Racial and Ethnic Differences by Credit Quality Indicators

	Baseline Racial and Ethnic Differences			
	Home Purchase Sample		Refinance Sample	
Variable Names	Subprime	Lender FE	Subprime	Lender FE
Black	0.076816***	0.032651***	0.041889***	0.017197***
	(0.004)	(0.003)	(0.004)	(0.004)
Hispanic	0.061963***	0.024886***	0.017307***	0.004977*
	(0.003)	(0.003)	(0.003)	(0.003)
Observations	120,732	120,732	115,763	115,763
R-squared	0.418	0.594	0.384	0.555

	Differences by Credit Quality			
	Home Purchase Sample		Refinance Sample	
Variable Names	Subprime	Lender FE	Subprime	Lender FE
Black	0.076060***	0.025619***	0.027124***	0.008291*
	(0.006)	(0.005)	(0.006)	(0.005)
Hispanic	0.021514***	0.003590	0.014641***	0.008470**
	(0.004)	(0.003)	(0.004)	(0.003)
Black*Subprime	0.025590***	0.014424**	0.018174**	0.011244*
	(0.008)	(0.006)	(0.007)	(0.006)
Hispanic*Subprime	-0.028763***	-0.008953	0.003810	-0.000408
	(0.007)	(0.006)	(0.007)	(0.006)
Black*High DTI	0.004354	0.005799	-0.001503	-0.008372
	(0.007)	(0.006)	(0.007)	(0.006)
Hispanic*High DTI	-0.003893	0.000611	-0.000139	-0.011027**
	(0.006)	(0.005)	(0.006)	(0.005)
Black*High LTV	-0.010247	0.003475	0.019656***	0.018107***
	(0.007)	(0.006)	(0.007)	(0.006)
Hispanic*High LTV	0.105151***	0.050799***	0.004803	0.001866
	(0.006)	(0.005)	(0.005)	(0.005)
Observations	120,732	120,732	115,763	115,763
R-squared	0.423	0.595	0.384	0.555

Notes: The table presents the estimates on racial and ethnic dummy variables for a linear probability model of whether the borrower received a rate spread loan. The first panel repeats the baseline estimates from Table 3 for the model that includes the standard control variables plus the interaction of key loan terms with whether the borrower had a subprime credit score (columns one and three) and for the model that also includes lender fixed effects (columns two and four). The second panel presents the estimates after interacting the race and ethnicity with dummy variables for whether the borrower has a subprime vantage credit score below 701, a debt expense to income ratio above 0.45 and a loan to value ratio above 0.95.

Table 6. Rate Spread Racial and Ethnic Differences by Age

Variable Names	Home Purchase Sample		Refinance Sample	
	Subprime	Lender FE	Subprime	Lender FE
Black	0.069855***	0.034293***	0.043922***	0.028605**
	(0.008)	(0.007)	(0.015)	(0.012)
Hispanic	0.060058***	0.026722***	0.037297***	0.017732**
	(0.007)	(0.005)	(0.009)	(0.008)
Age 36-41	0.025561***	0.017057***	0.024362***	0.017438***
	(0.003)	(0.002)	(0.006)	(0.005)
Age 42-48	0.038654***	0.021316***	0.029717***	0.017570***
	(0.003)	(0.003)	(0.005)	(0.005)
Age 49-55	0.041781***	0.025544***	0.032271***	0.015073***
	(0.003)	(0.003)	(0.005)	(0.004)
Age 56+	0.047956***	0.027959***	0.039665***	0.021322***
	(0.003)	(0.003)	(0.005)	(0.004)
Age 36-41*Black	-0.010492	-0.011064	0.004582	-0.007823
	(0.011)	(0.009)	(0.018)	(0.015)
Age 42-48*Black	-0.033818***	-0.015694*	-0.025124	-0.030570**
	(0.011)	(0.009)	(0.016)	(0.013)
Age 49-55*Black	-0.004663	-0.017080*	0.004696	-0.004491
	(0.012)	(0.010)	(0.016)	(0.014)
Age 56+*Black	-0.015515	-0.011349	-0.005326	-0.018321
	(0.012)	(0.010)	(0.016)	(0.013)
Age 36-41*Hispanic	-0.022467**	-0.014061*	-0.027264**	-0.021000**
	(0.009)	(0.008)	(0.011)	(0.010)
Age 42-48*Hispanic	-0.029809***	-0.010636	-0.021917**	-0.010864
	(0.010)	(0.008)	(0.011)	(0.009)
Age 49-55*Hispanic	-0.022291**	-0.016259	-0.029702***	-0.023095**
	(0.011)	(0.010)	(0.011)	(0.010)
Age 56+*Hispanic	-0.021837*	-0.014994	-0.028763***	-0.011105
	(0.012)	(0.010)	(0.011)	(0.010)
Observations	120,732	120,732	115,763	115,763
R-squared	0.419	0.594	0.384	0.555

Notes: The table presents the estimates on racial and ethnic dummy variables for a linear probability model of whether the borrower received a rate spread loan. All model specification interact the race and ethnicity variables with dummy variables for age categories where the omitted category is for borrowers younger than age 36. The first and third columns present estimates for the model that includes the standard control variables plus the interaction of key loan terms with whether the borrower had a subprime credit score, and the second and fourth columns present estimates for the model that also includes lender fixed effects.

Table 7. Rate Spread Racial and Ethnic Differences by Neighborhood Poverty and County Average Homebuyer Educational Level

Variable Names	Subprime Model		Lender FE Model	
	Baseline	Location Controls	Baseline	Location Controls
African-American	0.0754***	0.0141	0.0345***	-0.012
	(0.004)	(0.016)	(0.004)	(0.013)
Hispanic	0.0595***	0.0257	0.0289***	0.006
	(0.005)	(0.016)	(0.004)	(0.014)
African American*Share < 2 Years		0.1008***		0.0849***
		(0.034)		(0.028)
Hispanic*Share < 2 Years		0.0425		0.0269
		(0.028)		(0.024)
African American*Pct Poverty		0.0019***		0.0010***
		(0.000)		(0.000)
Hispanic*Pct Poverty		0.0019***		0.0015***
		(0.000)		(0.000)
Observations	120,422	120,422	120,422	120,422
R-squared	0.418	0.419	0.594	0.594

Notes: The table presents the estimates on racial and ethnic dummy variables for a linear probability model of whether the borrower received a rate spread loan using the home purchase sample. The first and third columns present estimates for the models that include the standard control variables plus the interaction of key loan terms with whether the borrower had a subprime credit score and for the model that also includes lender fixed effects, respectively, with the addition of a control for share borrower by county and race with less than two years of education (the standard controls already include percent poverty). The second and third columns present estimates for models that also include the interaction of race and ethnicity with the census tract percent poverty and the county average education level among homebuyers.