What Drives Racial and Ethnic Differences in High Cost Mortgages? The Role of High Risk Lenders

Patrick Bayer, Duke University and NBER

Fernando Ferreira, The Wharton School, University of Pennsylvania, and NBER

Stephen L Ross, University of Connecticut

Corresponding Author Address and Author Emails: Stephen Ross, Department of Economics, University of Connecticut, 309 Oak Hall, 365 Fairfield Way U-1063, Storrs, CT 06269-1063; 806-486-3533; stephen.l.ross@uconn.edu; patrick.bayer@duke.edu; fferreir@wharton.upenn.edu.

What Drives Racial and Ethnic Differences in High Cost Mortgages? The Role of High Risk Lenders

Abstract

This paper examines racial and ethnic differences in high cost mortgage lending in seven diverse metropolitan areas from 2004-2007. Even after controlling for credit score and other key risk factors, African-American and Hispanic home buyers are 105 and 78 percent more likely to have high cost mortgages for home purchases. The increased incidence of high cost mortgages is attributable to both sorting across lenders (60-65 percent) and differential treatment of equally qualified borrowers by lenders (35-40 percent). The vast majority of the racial and ethnic differences across lender can be explained by a single measure of the lender's foreclosure risk and most of the within-lender differences are concentrated at high-risk lenders. Thus, differential exposure to high-risk lenders combined with the differential treatment by these lenders explains almost all of the racial and ethnic differences in high cost mortgage borrowing.

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 is motivated by several related issues. First, a large literature has studied racial discrimination in mortgage lending. While historically focused on discrimination in mortgage underwriting and redlining against minority neighborhoods, attention has increasingly turned to differences in the price of mortgage credit, which has also been the focus of several high profile U.S. Department of Justice cases in the wake of the recent financial crisis.[1] Second, higher credit costs create an obvious barrier to African-American and Hispanic homeownership, which has historically lagged that of white households by large margins.[2] Finally, the last housing cycle was characterized by a growing subprime sector (Mian and Sufi 2009). African-American and Hispanic borrowers were more likely to receive subprime loans at higher costs, possibly contributing to especially high foreclosure rates for these borrowers.[3]

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. While some previous studies ask the question of whether similar borrowers receive different prices from the same lender (e.g., disparate treatment discrimination),[4] our use of market-wide data shifts

---

[1] Munnell, Tootell, Browne and McEneaney (1996) and Ross and Yinger (2002) examine mortgage underwriting, Holmes and Horvath (1994) and Tootell (1996) examine redlining, and Ross (2005) and Chan, Haughwout and Tracy (In Press) study racial price differences. 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] See Quercia, McCarthy and Wachter (2003), Belsky, Retsinas and Duda (2005), and Herbert, Haurin, Rosenthal and Duda (2005) for a detailed discussion of racial and ethnic differences in homeownership.

[3] See Gerardi and Willen (2009), Reid and Laderman (2009) and Edminston (2009).

[4] For example, see Courchane and Nickerson (1997), Black, Boehm and DeGennaro (2003), Nelson (2005), and Courchane (2007). These studies have found very small, if any, within-lender differences between white and minority borrowers in the incidence of high cost mortgage credit.

the question to whether unexplained racial and ethnic differences exist in market outcomes, a phenomenon that Heckman (1998) described as market discrimination.[5] Significant market level differences in the price of credit may 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 the Home Mortgage Disclosure Act (HMDA) data on home purchases and refinance mortgages from 2004 and 2007 to public records data on housing transactions and liens in seven distinct metropolitan housing markets.[6] 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. HMDA contains a rate spread variable that allows us to create an indicator for the loan, which is equal to one when the Annual Percentage Rate (APR) exceeds the interest rate on treasury securities of comparable maturity by at least 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 loans.[7] We drew a sample of matched mortgages that were originated between May and August of each year and provided the names and addresses from this sample to one of the major credit reporting agencies. The credit reporting agency used the name and address to

---

[5] Other studies documenting market wide differences in the prevalence of high cost loans include Bhutta and Ringo (2013) using HMDA and credit repository data and Courchane (2007), Haughwout, Mayer and Tracy (2009) and Ghent, Hernandez-Murillo and Owyang (In Press) using proprietary loan data. Studies using proprietary data have typically been 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.

[6] The data also includes a sample of 2008 originations, which we do not include in our analysis in order to focus on lending prior to the financial crisis that began to accelerate with the failure of Bear-Stearns during the fall of 2007 and winter of 2008. See Harding and Ross (2010) for a discussion of the timing of the on-set of the broad financial crisis. All results are robust to including 2008 originations.

[7] The Annual Percentage Rate (APR) estimates cost of credit including interest rate and closing costs. These high cost or rate spread loans are sometimes referred to as subprime loans, but other authors study the subprime market based on a list of top subprime lenders, e.g. Ferreira and Gyourko (2015), based on borrowers who have a low credit score, e.g. Mian and Sufi (2009), or private label securitized loans, e.g. Ghent, Hernandez-Murillo and Owyang (In Press).

match borrowers to archival credit reporting data from March 31 in the year preceding the mortgage origination and from March 31 for every subsequent year through 2009, providing in each year a vantage credit score plus detailed credit line information from each individual's report.

With these linked data on housing transactions, mortgage originations and credit information, we estimate models that relate the presence of a high cost loan to race, ethnicity and other common risk factors. We find significant unexplained racial and ethnic differences in the incidence of high cost mortgage credit. These differences persist after controlling for detailed measures of borrower and loan characteristics including credit score, loan to value ratio, the presence of subordinate liens, and housing and debt expenses relative to income. Relative to a model based on only the control variables available in HMDA, the inclusion of these additional controls erodes about half of the racial and ethnic differences in mortgage pricing. Still, the remaining differences are sizable: African-American and Hispanic borrowers have a 9.1 and 6.8 percentage point higher likelihood of a rate spread or high cost loan, respectively, in the home purchase market relative to an incidence of 8.7 for white borrowers. Significant loan-pricing differences exist across all of the metropolitan housing markets in the study, including not only faster-growing markets in California and Florida that experienced especially sharp housing booms and busts in the 2000s, but also slower-growing Eastern and Midwestern housing markets. Similar, but more modest results, arise for refinance mortgages.

The inclusion of lender fixed effects in the model substantially reduces the unexplained differences for African-American and Hispanic borrowers by 60 and 65 percent, respectively.[8] These findings imply that sorting across (or differential access to) lenders plays a significant role

---

[8] Avery, Canner and Cooke (2005) and Avery, Brevoort and Canner (2007) using HMDA data and Bhutta and Ringo (2014) using HMDA data matched to a 1 percent sample of credit reports also find that lender fixed effects explain a substantial fraction of racial differences in the cost of credit. Also see Ross and Yinger (2002) for evidence of significantly larger market level differences in mortgage underwriting than the differences observed at the lender level.

in creating market wide differences in mortgage pricing, even after controlling for detailed borrower and loan attributes.

To better understand the role of lenders we examine a series of models that replace lender fixed effects with characteristics of the lender. More specifically, we focus on a measure of lender foreclosure risk. The intuition is that certain lenders can be extremely aggressive in their mortgage underwriting, increasing market shares by issuing high cost mortgages to a group of individuals negatively selected on the basis of unobservables (individuals more likely to face future negative income and health shocks, less likely to keep up with mortgage payments, etc.). We provide a novel way of identifying these high risk lenders by directly estimating their underlying risk of foreclosure from actual foreclosure data.[9] This ex-post foreclosure risk measure emerges as the key explanatory variable, accounting for 75-90 percent of the racial and ethnic differences explained by lender fixed effects. Qualitatively similar results hold regardless of whether this measure of foreclosure risk is based on the full sample of a loans or only loans to white borrowers and regardless of whether lender risk is based on foreclosure or 30 day delinquencies. Conditioning the measure of lender foreclosure risk on whether an individual loan is high cost results in only modest reductions in the ability of lender risk to explain racial and ethnic differences in high cost lending.

In additional analyses, we examine which borrowers are more likely to be served by these high-risk lenders and explore heterogeneity in the magnitude of racial and ethnic differences across borrowers related to individual demographic and risk factors, loan attributes, neighborhood attributes, and ex-post lender foreclosure risk. Race, ethnicity, credit score and high loan to value

---

[9] We explain the construction of this measure in detail in section 4. Briefly, we estimate lender fixed effect models of ever receiving a foreclosure notice, and control for the fixed effect estimate in the high cost lending model using a split-sample instrumental variables strategy (Case and Shiller 1989; Angrist and Krueger 1995).

ratios stand out as the most important factors for explaining borrower allocation to high risk lenders. The modest effects of income and neighborhood poverty suggest that minority status plays an especially important role in the sorting of borrowers across lenders, but the large effects for credit score and loan to value ratio suggest that there is substantial sorting across lenders based on risk factors. Finally, most of the within lender racial and ethnic differences in high cost lending arise for borrowers with loans at high risk lenders.

Taken as a whole, the results of our analysis imply that the substantial market-wide racial and ethnic differences in the incidence of high cost mortgages arise because African-American and Hispanic borrowers tend to be more concentrated at high-risk lenders. Strikingly, this pattern holds for all borrowers even those with relatively unblemished credit records and low-risk loans. High-risk lenders are not only more likely to provide high cost loans overall, but are especially likely to do so for African-American and Hispanic borrowers. In fact, these lenders are largely responsible for the differential treatment of equally qualified borrowers; minimal racial and ethnic differences exist among lenders that serve less risky segments of the market.

The paper proceeds as follows: Section 2 presents our data in detail. Section 3 presents the baseline estimates. Section 4 examines the role of high-risk lenders, and Section 5 examines heterogeneity in racial and ethnic differences. Section 6 provides a brief conclusion.

2. **Data**

Our data are based on public Home Mortgage Disclosure Act (HMDA) data between 2004 and 2007 and proprietary housing transaction/lien and assessor's databases purchased from Dataquick Inc. 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, all Maryland Counties excluding Baltimore City, and San Francisco CA CMSA.[10] 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 tract or the zip code was missing. This eliminates very few records due to the high quality of the name and address records in the assessor files. 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.[11]

We use the HMDA rate spread variable to create a dummy variable that is one whenever the APR on the mortgage is at least 3 percentage points above the interest rate on treasury securities. This threshold was chosen by federal regulators based on Government Sponsored Enterprise data in order to capture the high interest rates often seen in the subprime sector in the early 2000's. Naturally, this outcome variable provides somewhat limited information on interest rates due to its discrete nature. Further, the share of rate spread loans is sensitive to the yield curve over bond maturities because APR is compared to treasury rates of comparable maturity to the term of the mortgage and mortgages are often pre-paid, see Avery, Brevoort, and Canner (2007). Our core results are robust to analyses that adjust the high cost loan threshold by year in order to keep the share of high cost loans constant over time, anchored to 2004 which had the lowest share.

---

[10] Also see Bayer, Ferreira and Ross (2016) on the data.
[11] 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 HMDA respondent. Less restrictive match criteria can yield a match rate closer to 80 percent.

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 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. 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,[12] and initialized so that each site receives equal weight in the pooled sample.

This sample is provided to Experian 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

---

[12] 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.

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. Bayer, Ross and Ferreira (2016) show that the final weighted sample composition is quite comparable to the population of HMDA data for these sites, except for a moderate decline in share white and moderate increase in loan amount arising from the difficulty of matching lender names between HMDA and the Dataquick provided assessor files, see Appendix Table 1.

Table 1 shows the weighted means for our final home purchase and refinance subsamples that were successfully merged to pre-mortgage credit report data.[13] 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 and county level price indices or assessed value when a previous sale is unobserved for refinance mortgages, and detailed property attributes including whether a single family home, a condominium, and number of units on the property. Notably, information on subordinate liens is typically not available in other studies because only individual loans are tracked in most mortgage samples, not entire housing-mortgage transactions. 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. The Vantage Score is a proprietary credit score developed by the credit

---

[13] Some small lenders could not be identified based on the reporting restrictions. If the lender was not identified, the observation is dropped from the regression sample. Similar results are observed using the full sample.

reporting agencies as an alternative to the traditional FICO index of credit score. The credit report

observation following the year of the mortgage is used to obtain monthly mortgage payment, which

when combined with HMDA income is used to calculate the mortgage payment to income ratio.

The monthly mortgage payment is combined with debt payments from the pre-mortgage credit

data and HMDA income to calculate the debt payment to income ratio. Finally, age, which has

not typically been available in studies of mortgage lending, is observed for many borrowers and

co-borrowers in the credit history files.

3. Rate Spread Models

Table 2 presents the rate spread regression results for the pooled MSA samples. The first

four specifications reported in Table 2 can be characterized by the following equation:

(1) $\qquad y_{ij} = \alpha X_i + \varepsilon_{ij}$

where *y* indicates the presence of a high cost loan (i.e., a HMDA rate spread loan) and *X* represent

characteristics of the borrower *i* or loan. The table only shows the estimates for the following race

and ethnicity categories: Asian, African-American and Hispanic. The econometric model omits

the dummy for Whites, so all estimates should be interpreted as relative to a White borrower.[14]

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 year-by-site fixed effects. The second

---

[14] We focus on the linear probability model due to concerns about the bias arising from estimating non-linear models with fixed effect estimates based on small cell sizes. We confirm that the magnitude and significance of estimated racial and ethnic differences are robust to trimming the data to eliminate predicted values outside of the unit interval (Horrace and Oaxaca 2006).

column includes additional controls made available by merging the HMDA data with Dataquick housing transaction data including combined loan to value ratio, whether the primary lien is an adjustable rate mortgage, number of subordinate liens, and year-by-week fixed effects. The combined loan to value ratio is included in the model as a series of dummy variables associated with LTV falling below 0.6, 0.6 to 0.8, 0.8 to 0.85, 0.85 to 0.90, 0.90 to 0.95, 0.95 to 1.00, 1.00 to 1.05, and 1.05 and above. The third column adds dummy variables for credit score in 20 point bins, housing expense to income ratio in bins as small as 0.02 around the traditional secondary market criteria of 0.28, and total debt expense to income ratio categories with bins as small as 0.03 around the threshold of 0.36. The fourth column includes additional controls for the potential effect of subprime lending, identifying borrowers with Vantage scores below 701 as subprime borrowers[15] and interacting this subprime dummy variable with variables associated with key thresholds of loan to value ratio, debt to income ratio, mortgage payment to income ratio,[16] a dummy variable whether there are subordinate liens, and whether the primary lien is adjustable rate. The fifth column adds lender fixed effects ω_j:

$$(2) \qquad y_{ij} = \alpha X_i + \omega_j + \varepsilon_{ij}$$

The results shown in the first column reveal that African-American and Hispanic borrowers have an increased likelihood of having a rate spread loan of 20.3 and 14.0 percentage points

---

[15] The credit reporting agencies that developed the Vantage score algorithms describe 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 during our sample period. Subprime borrowers make up about 25 percent of our weighted home purchase sample.

[16] The loan to value thresholds used are 0.80, 0.90, 0.95 and 1.00 with each bin containing 30, 12, 35 and 3 percent of our weighted home purchase sample, the debt to income thresholds used are 0.36 and 0.45 with 13 and 42 percent in the middle and upper bins, and the mortgage payment to income ratio thresholds used are 0.28 and 0.33 with 9 and 43 percent in the middle and upper bins.

relative to white borrowers, respectively, for a home purchase mortgage when conditioning only on the standard controls available in HMDA. The difference between white and Asian borrowers is small in this specification and in all other specifications reported below.

The addition of standard underwriting controls in columns 2 and 3 reduces the estimated differences for African-American and Hispanic borrowers to 9.1 and 6.8 percentage points for the home purchase and 4.6 and 1.4 for the refinance market, reductions on the order of 55-60 percent for racial differences and 65 percent for ethnic differences,[17] while the inclusion of additional subprime controls in Column 4 has little impact on the estimated differences.[18]

We refer to the model presented with the full set of borrower and loan control variables reported in column 4 as our baseline model. In the home purchase market, the remaining racial and ethnic differences represent 104.6 and 78.2 percent, respectively, of the incidence of rate spread loans for white borrowers. Comparing the results of column 4 to those of column 1 reveals both (i) 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 (ii) that economically and statistically significant differences remain even after controlling for these most commonly used measures of credit worthiness and risk, especially in the home purchase market.

The addition of lender fixed effects in column 5 substantially erodes the differential incidence of high cost loans. The point estimates in the home purchase sample decline from 8.7 and 6.9 to 3.6 and 2.4 percentage point differences for African-Americans and Hispanics,

---

[17] 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 in the expected directions, and we find that the likelihood of a rate spread loan is higher for jumbo loans and for transactions that use subordinate liens.

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

respectively, and for the refinance sample differences decline from 4.3 and 1.4 to 1.9 and 0.5. Thus, in all cases, a majority of the racial and ethnic differences that remain after controlling for standard underwriting variables can be explained by differential access to traditional lenders and/or selection into high cost lenders.

The inclusion of lender fixed effects shifts the interpretation of racial and ethnic differences from measures of market-level disparities to differences in the treatment of equally qualified minority and white borrowers by the same lender. As evidence of lender discrimination, the lender fixed effect estimates are comparable to the findings in the Munnell et al. (1996) study of underwriting discrimination in Boston, which also used lender fixed effects in a sample of loan applications from 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 6 and 41 percent.

In terms of the cost of credit, Avery, Canner and Cooke (2005) and Avery, Brevoort and Canner (2007) using 2004 and 2005/06 HMDA data, respectively, and Bhutta and Ringo (2014) using 2006 HMDA data matched to a 1 percent sample of credit reports find that lender fixed effects can explain a substantial portion of the unexplained racial and ethnic differences. Neither Avery, Canner and Cooke (2005) nor Avery, Brevoort and Canner (2007) can determine whether the across lender differences explained by lender fixed effects are due to the sorting of observationally equivalent borrowers or due to key underwriting variables that are unobserved in HMDA. Unlike the earlier studies, Bhutta and Ringo (2014) and this paper show that the unexplained racial and ethnic differences after controlling for detailed credit variables,[19] which are

---

[19] Bhutta and Ringo (2014) do not observe loan to value ratio.

often attributed to discrimination in the market, are primarily the result of the systematic selection

of African American and Hispanic borrowers into lenders who tend to issue high cost mortgages.

Neither Bhutta and Ringo (2014) nor this paper, however, observe some of the key loan attributes

that were sometimes associated with high cost loans or subprime loans during the run up to the

recent crisis, such as no documentation of income or initial teaser interest rates combined with pre-

payment penalties.

An Expositional Note

The baseline results shown in Table 2 imply that racial and ethnic differences are

significantly greater for home purchase versus refinance mortgages in all specifications. For

expositional simplicity, therefore, we focus our presentation of the remaining results for the home

purchase sample. We include a full set of comparable tables for refinance mortgages in the

appendix. Most of the key patterns highlighted in the next section arise in the refinance sample as

well.

4. Understanding the Role of High Risk Lenders

Having shown that the inclusion of lender fixed effects substantially reduces the estimated

racial and ethnic differences in the incidence of rate spread loans, we now demonstrate that most

of the racial differences explained by lenders are associated with lenders that have high ex-post

foreclosure risks. We also include additional lender attributes as controls: (i) the type of lending

institution (agency code),[20] (ii) the share of mortgages securitized, and (iii) the share of securitized

---

[20] The agency code in HMDA identifies the lender's regulator and the regulator identifies whether the lender is a national bank, commercial bank, a state chartered bank, a savings and loan, a credit union, or a non-depository mortgage bank.

mortgages sold to each type of purchaser.[21] The lender characteristics are represented by Z in equation (3).

(3) $y_{ij} = \alpha X_i + \beta Z_j + \varepsilon_{ij}$

Appendix Table 2 presents means and standard deviations for each variable.

In order to create proxies for lender foreclosure risk, we use our sample of home purchase mortgage originations. These proxies are based on estimated lender fixed effects, ϕ_j, from models of whether foreclosure notices, f_{ij}, ever appear in the borrower's credit report between March 31 the year after origination through March 31, 2009:

(4) $f_{ij} = \gamma X_i + \phi_j + u_{ij}$

The ever foreclosure model specification is analogous to the lender fixed effects model shown in equation (2) but with the presence of a foreclosure notice replacing the presence of a high cost loan as the dependent variable.

If we observed a large enough number of loans for each lender, we could simply include the estimated lender fixed effects from the foreclosure model in equation (4) as a measure of lender foreclosure risk when estimating the rate spread model shown in equation (3). Because only a

---

[21] The variables related to securitization are calculated using the full sample of HMDA loans between 2004 and 2007 for the seven sites in our credit history sample. These securitization variables are merged into our sample using the respondent id leading to a slightly reduced sample size. We calculate the share of loans securitized for each lender and the share of securitized loans sold to each type of purchaser including the Government Sponsored Enterprises (GSE), Federal Housing Administration (FHA), private securitizers, commercial or savings banks, insurance/ companies/credit unions/mortgage banks, affiliated lenders, and other buyers.

limited number of loans are observed for each lender, however, our measure of foreclosure risk represents a noisy measure of the actual foreclosure risk faced by each lender. To consistently estimate the rate spread model with lender foreclosure risk, therefore, we use a split sample instrumental variables strategy.[22]

Specifically, we restrict our sample to borrowers at lenders with at least 10 loans in our home purchase sample, and then randomly allocate half of the loans for each lender to a hold-out sample and the other half to the regression sample. We then estimate the foreclosure model shown in equation (4) separately for the regression and hold-out samples. The lender fixed effect estimate from the regression sample is included in the high cost lending model regression shown in equation (3), and the fixed effect estimate from the hold-out sample is used as an instrument.[23] Standard errors are bootstrapped by sampling lenders with replacement.

Columns 1 and 2 of Table 3 replicate the baseline and lender fixed effect results from Columns 4 and 5 in Table 2. The next column reports the split-sample IV estimates for a specification that includes ex-post lender foreclosure risk instead of lender fixed effects. These results indicate that lender foreclosure risk is a strong predictor (both economically and statistically) of the presence of a rate-spread loan and, strikingly, that lender foreclosure risk explains the vast majority (74 and 91 percent) of the racial and ethnic differences explained by

---

[22] This procedure was first used by Case and Shiller (1989) to address measurement error in estimated housing price indices, but was described more recently and more generally by Angrist and Krueger (1995) who named this procedure split-sample IV. In addition to measurement error, this approach also eliminates other forms of small sample bias. For instance, in our setting, if the loans sampled for a specific lender happen to be bad loans by random chance, then those loans are likely to be charged a high interest rate and experience a foreclosure. This random variation will create a correlation between high cost lending and foreclosure rates unless the foreclosure fixed effect estimate is based on a separate sample of loans, i.e. a split sample IV hold-out sample.

[23] Estimates presented are an average of results from 20 separate runs for different draws of the regression and hold-out samples. The standard deviation of estimates across these sample draws is about one thirtieth of the mean estimate for the lender fixed effect and about one tenth of the estimates of racial and ethnic differences conditional on foreclosure risk.

lender fixed effects in the home purchase market.[24] As shown in the fourth column, comparable results hold for a model that includes additional controls for lender type (agency code) and securitization, which have little additional predictive power once foreclosure risk is included in the analysis. Notably, prior to including the control for foreclosure risk, lender type and securitization variables have substantial explanatory power, with mortgage banks (non-depository lenders) and lenders that sell a substantial fraction of loans into private-label securitization having an especially high incidence of rate spread loans, but even then those variables do little to explain racial and ethnic differences.[25]

Table 4 reports results using alternative measures of lender foreclosure risk. First, we estimate a model based only on loans to white borrowers. The goal of this analysis is to examine whether certain lenders specialize in providing high-risk loans to all borrowers regardless of race and ethnicity, or whether the results presented in Table 3 are driven by lenders that specialize in providing high-risk loans primarily to African-American and Hispanic borrowers and then have higher foreclosure risks on average due to the large number of minority loans. To conduct this analysis, we further restrict the data to lenders with at least 10 loans to white borrowers, and re-estimate the "ever foreclosed" model controlling for lender-by-race/ethnicity fixed effects for both the hold-out and regression samples. We then use the lender fixed effects for white borrowers from the two samples to implement split sample IV in the high cost loan model. Column 1 in Table 4 contains the estimates of the baseline model from Table 3 column 2 with the new restricted sample showing that the sample restriction has little impact on the estimates, and column 2 shows the estimates for a model that uses foreclosure risk associated with white loans only. The estimated

---

[24] Not surprisingly given the design, the first stage is very powerful with F-statistics in the thousands so there are no problems associated with weak instruments.
[25] See appendix Table 3.

racial and ethnic differences are remarkably similar whether the estimate of lender foreclosure risk is based on all loans or only those to white borrowers.[26] This suggests that these lenders specialize in providing high-risk loans to the market as a whole, not just to African-American and Hispanic borrowers.

A second possible interpretation of these results is that high cost lenders also happen to be lenders that aggressively enter the foreclosure process as loans become delinquent. In order to rule out this possibility, we estimate a model of ever received a 30-180 day delinquency using the original sample of all lenders with at least 10 loans in our sample. The estimated lender fixed effects from this model are used as alternative measure of lender risk again using split-sample IV. These results are shown in column 3. The racial and ethnic differences are very similar to our benchmark model that controls for foreclosure risk and the coefficient on the proxy for delinquency is large and statistically significant. These findings suggest that across lender racial and ethnic differences arise from borrower delinquency and default rather than lender foreclosure behavior.

Next, we re-estimate the ever foreclosed model including the high cost loan variable as an additional control. The resulting measure of ex-post foreclosure risk is now conditional on whether the loans issued by the lender were high cost loans. To the extent that lenders price observable differences between borrower and loan attributes, the inclusion of this control should erode the ability of foreclosure risk to explain racial differences that arise due to lender observable factors, such as loan terms like whether the loan had low or no income documentation requirements.

The resulting estimates are shown in column 4 of Table 5. For these models, the racial and ethnic differences that remain after controlling for foreclosure risk rise notably, by 45 and 32

---

[26] Additional noise in the white-lender fixed effect estimates (due to a lower incidence of foreclosure) increase the standard errors on the foreclosure risk estimates. Very similar results arise for models that include the additional lender controls based on agency code and securitization patterns.

percent, respectively. However, the vast majority of the racial and ethnic differences explained by lender foreclosure risk are still captured by the new measure of foreclosure risk, 68 and 78 percent, respectively. Accordingly, the majority of the racial and ethnic differences in high cost loans across lenders are associated either with borrowers sorting across lenders on factors unobservable to the lender or observable differences between borrowers that were not aggressively priced into loans.

Sorting across Lenders

As a final exercise designed to better understand the implications of these high risk lenders for mortgage markets, we examine whether these lenders also provide credit to a disproportionate number of borrowers or loans with specific risk attributes. In particular, Table 5 presents estimates of models that relate lender foreclosure risk ϕ_j to borrower and loan attributes X including (i) borrower demographic and financial variables, (ii) neighborhood attributes and (iii) loan attributes.[27]

(5) $\phi_j = \theta X_i + v_{ij}$

For this analysis, lender foreclosure risk is based on the lender fixed effects from the ever foreclosed model using the sample of loans from lenders with 10 or more loans. For ease of interpretation, we standardize the foreclosure risk variable prior to estimating the regression. Given the inclusion of a full set of borrower, neighborhood and loan attributes in the foreclosure model used to create ϕ_j, the results presented in Table 6 cannot be driven by the ability of these

---

[27] The model also includes site by purchase year fixed effects. Standard errors are clustered at the census tract level.

variables to explain foreclosures directly. Rather, the regression effectively tests for whether borrowers are negatively selected into these lenders based on borrower and loan observables.[28]

The race and ethnicity correlations are quite large, approximately 4 percent of a standard deviation for African-Americans and 5 percent for Hispanics. While many estimates are statistically significant, the estimates on the other demographic variables, age, gender, presence of a co-borrower (a proxy for marital status), and the logarithm of income tend to be smaller, ranging between one half and 2 percent of a standard deviation.[29] Similarly, the estimates on the tract composition variables are modest. A standard deviation increase in any variable never has an effect much larger than one percent of a standard deviation in lender foreclosure risk,[30] and the doubling of the present value of rents relative to housing price is associated with an increase of only 2 percent of a standard deviation in lender riskiness.

On the other hand, the borrower's vantage score is strongly associated with borrowing from a high-risk lender. Having a subprime credit score is associated with about 10 percent of a standard deviation increase in lender foreclosure risk, and having an above median credit score within one's corresponding market segment (prime or subprime) is associated with a 3-5 percent of a standard deviation reduction in lender foreclosure risk. If borrowers sort across lenders based on key risk observables, likely credit score, they presumably will sort across borrowers on unobservables, as

---

[28] We do not use a hold-out sample in this analysis. There is no bias from measurement error because the measurement error is on the left hand side of the equation. Similarly, incidental parameter bias is no longer a concern. The fixed effects are estimated conditional on the borrower and loan attributes so any conditionally bad draw of loans in terms of foreclosure that enter ex-post foreclosure risk through the expected residual of the "ever foreclosed" model are independent of the draw on observable loan and borrower attributes because the model is conditional on those attributes.

[29] These differences are even larger when the model just controls for demographics 5.5 and 7 for blacks and Hispanics and only 1.5 percent higher for loans without a co-borrower, the largest demographic effects in this model.

[30] A reasonable concern is that we have included many correlated neighborhood variables that have diluted the effect of any individual neighborhood variable. However, we run a model only including one tract variable, share households in poverty which has one of the largest effects of any of the composition variables, and the estimate on tract poverty barely moves.

well. Similarly, having a loan to value ratio above 0.95 is associated with 7 percent of a standard deviation increased exposure to high-risk lenders. This last finding might be consistent with sorting over product type, but also could arise because borrowers with few assets available for down payment, a key unobservable, sort into these lenders.

Taken as a whole, the results of Table 5 imply that a major reason that African-American and Hispanic borrowers pay more for mortgage credit is that they tend to do business with lenders who specialize in providing high-risk loans in terms of both observable risk factors (e.g., credit score and LTV) and unobserved foreclosure risk, but we cannot fully rule out the possibility that these results are driven by the terms of the mortgage products that these lenders tend to issue.

Variation across Metropolitan Sites

Tables 6 presents the estimated results separately for each metropolitan housing market for three main specifications: (i) a baseline model comparable to column 4 of Table 2, (ii) a model controlling for overall lender foreclosure risk measure comparable to column 3 of Table 3, and (iii) a lender fixed effects model comparable to column 5 of Table 2. Note that the foreclosure risk model is estimated as a reduced form model simply including the estimated fixed effects for each lender across all sites as a regressor because the much smaller site specific sample sizes raised concerned about precision and small sample bias if we were to either use the split sample IV estimator or use site specific lender fixed effects.[31] The columns represent in order Chicago, Cleveland, Denver, Los Angeles, Maryland Counties, Miami-Palm Beach Corridor, and San Francisco Bay Area.

---

[31] Further, bias from measurement error leads to conservative estimates, and concerns about incidental parameters bias are substantially mitigated because the lender fixed effect is based on the entire sample, not just the loans in a given site.

While there is some variation, racial and ethnic differences in the home purchase sample exist for all seven sites for African-Americans and six sites for Hispanics in models both with and without lender FE's. In the home purchase market without lender FE's, differences range between 5.9 and 11.7 for African-Americans, and 6.2 and 7.6 for Hispanics. The inclusion of lender FE's lowers these differences to ranges of 2.1 to 5.1, and 1.8 to 2.5, respectively. Further, adding the controls for lender foreclosure risk to the baseline model again result in racial and ethnic differences that are closer to the differences in the lender fixed effect model with ex-post foreclosure risk explaining between 49 and 77 percent of the across lender racial and ethnic differences. Notably, the estimates on the foreclosure risk range between 2.10 and 2.91, which is comparable to the reduced form estimate of 2.56 for the entire sample.

Taken together, we conclude that the market wide differences in the incidence of high cost loans are present in all of our market areas in the home purchase sample. Further, as with the overall sample, lender fixed effects significantly erode the estimates primarily because some lenders have both a disproportionate share of minority borrowers and those lenders tend to have unusually high ex-post foreclosure risks.

5. Heterogeneity in Racial and Ethnic Differences

In order to further assess how widespread is the incidence of racial and ethnic differences in high cost loans, we estimate models in which race and ethnicity are interacted with three key classes of variables: borrower risk factors, census tract attributes associated with the property, and the ex-post foreclosure risk of the lender. These results are presented in Table 7 for home purchase mortgages. The first two columns of the table present estimates based on adding interaction terms to the baseline model shown in column 4 of Table 2, and column 3 presents estimates built on the

foreclosure risk regression in column 3 of Table 3. The last three columns present estimates based on the lender fixed effect model shown in column 5 of Table 2.

Columns 1 and 4 present the interactions with three key risk variables: subprime credit score or Vantage score below 701, non-conforming loan to value ratio above 0.95, and a debt to income ratio above 0.45.[32] The results imply large differences in the likelihood of a having a high cost loan even for low-risk African-American borrowers, (i.e. those with prime credit scores, conforming loan to value ratios and reasonably low debt to income ratios) relative to their white counterparts. In particular, for our baseline model, low-risk African-American borrowers have an 8.5 percentage point higher likelihood of receiving a rate spread loan compared to low-risk white borrowers, very close to the 8.7 estimate for the entire sample reported in column 4 of Table 2. African-American borrowers with subprime credit scores are more likely to have high cost loans, but the effect is small and does not persist with the inclusion of lender fixed effects. In the lender fixed effects model, the racial difference for low risk African-American borrowers is 2.6 percentage points, a modest reduction from the 3.6 difference estimated for the full sample in column 5 of Table 3.

Low-risk Hispanic borrowers, on the other hand, have a substantially lower likelihood of high cost loans relative to the likelihood for the full sample, 2.4 versus 6.9 for the baseline model and 0.1 versus 2.4 for the lender fixed effects model. The changes for low-risk Hispanic borrowers are driven by the fact that Hispanics with high LTV loans are much more likely to have high cost loans. While smaller, the LTV effect for Hispanics persists in the lender fixed effect model, and no ethnic differences exist for low LTV borrowers.

---

[32] In our weighted sample, African-American and Hispanic loans make up 54 and 40 percent of their racial and ethnic subsamples where only 18 percent of white borrowers in our sample have subprime credit scores. Similarly, the share of African-American and Hispanic loans with an LTV above 0.95 is 62, 54, and 31 percent, respectively, and the shares for DTI above 0.45 are 49, 47, and 40 percent.

24

In columns 2 and 5, we estimate models that interact geographic controls for borrower location with race and ethnicity. We include geographic controls for the percent of households in poverty, the percent of residences that are owner occupied, racial and ethnic composition and the mean rent to value ratio, all within the census tract where the borrower will reside upon closing. Percent poverty is included as a general proxy for a disadvantaged neighborhood, while rent to value ratio is used as a measure of perceived equity risk and is scaled to capture the ratio of the present value of all future rents at the current rental rates to current value so that higher values are associated with lower expected rates of price appreciation.[33] We only interact share African-American and share Hispanic with the African-American and Hispanic identifiers, respectively.

Focusing on the baseline model for the home purchase sample in column 2, the results imply that racial and ethnic differences are somewhat concentrated in neighborhoods characterized by variables typically associated with disadvantage or risk. A higher neighborhood percent poverty implies significantly higher rates of high cost lending for African-Americans. For Hispanic borrowers ethnic differences are higher in neighborhoods with more Hispanics, lower home ownership rates and higher rent to price ratios. In order to examine unexplained racial differences for low-risk borrowers, we evaluate the race and ethnicity coefficients for the 25th percentile on all tract attributes, except for percent owner-occupied, which we evaluate at the 75th percentile. Low risk African-American borrowers in relatively advantaged neighborhoods have only a 1.3 percentage point difference compared to the 8.5 difference for low risk borrowers in column 1, and therefore most of the racial differences are concentrated in high poverty neighborhoods, even if the individual minority borrowers and loans are observably low risk. On

---

[33] Specifically, we assume an annual discount rate of 0.06 or a monthly discount rate of 0.005, and multiply monthly rents by 200 prior to dividing by housing prices.

the other hand, the Hispanic interactions lead to only a modest reduction to 2.1 for low risk

borrowers in advantaged neighborhoods from 2.4 for low risk borrowers overall.

The finding that racial and ethnic differences vary by neighborhood is consistent with the

literature that documents the concentration of subprime lending in poor and minority

neighborhoods (Calem, Gillen, Wachter 2004; Mayer, Pence 2008; Reid, Laderman 2009; Fisher,

Lambie-Hanson, 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 concentration of racial and ethnic differences in the incidence of

high cost lending in high poverty rate neighborhoods.

In columns 3 and 6, we present estimates of a model where we also interact race and

ethnicity with the lender foreclosure risk variable. The coefficients on the interactions of race and

ethnicity with the foreclosure risk variable are both positive, and the coefficient on the interaction

with race is sizable and statistically significant. Strikingly, the coefficients reported in the first

two rows of the table fall towards zero with the exception of the coefficient on the interaction of

race with tract rent to price ratio. Now evaluating lender risk and neighborhood disadvantage

variables at the 25th percentile, we find that the racial and ethnic differences for low risk borrowers

at low foreclosure risk lenders have both fallen below zero to -0.4 and -2.4 percentage points,

respectively. Therefore, African-American and Hispanic borrowers that do business with low-risk

lenders are no more likely to have high cost loans than comparable white borrowers at these

institutions, and these differences explain all of the neighborhood effects on racial and ethnic

differences. Looking at Columns 5 and 6, the inclusion of lender fixed effects has basically the

same effect as the inclusion of lender foreclosure risk capturing all remaining racial differences,

eliminating the interaction between race and subprime credit score, cutting in half the effect for

26

Hispanic borrowers with high LTVs, and eliminating or weakening all of the originally significant interactions between race and ethnicity and the neighborhood attributes.

Taken as a whole, the results presented in Table 7 provide a picture of the main drivers of racial and ethnic differences in the incidence of high cost loans. Regardless of individual risk factors, African-American borrowers are substantially more likely to have high cost mortgages due primarily to the increased likelihood of doing business with high-risk lenders, which are especially likely to provide high-cost loans to African-American borrowers. A similar pattern emerges for Hispanic borrowers.

6. Summary and Conclusion

In this paper, we identify large racial and ethnic differences 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. Differential sorting across lenders and the differential treatment of equally qualified borrowers by the same lender both emerge as important drivers of market-wide differences. These findings are consistent with evidence that African American and Hispanic borrowers are less likely to compare prices across lenders (Courchane, Surette, Zorn 2004; Alexandrov and Koulayev 2015). Of course, this sorting may also be influenced by lender behavior. In field studies, Ross, Turner, Godfrey and Smith (2008) and Hanson, Hawley, Martin and Liu (In Press) find that lenders offer more information and assistance in response to borrower inquiries and in the case of Hanson et al. are more likely to follow-up on initial contact from white borrowers.

The majority of the racial and ethnic differences due to sorting across lenders can be explained by lender ex-post foreclosure risk, which is strongly predictive of the likelihood of

receiving a high cost loan. Substantial market-wide racial and ethnic differences in the incidence of high cost lending arise because African-American and Hispanic borrowers tend to be concentrated at these high-risk lenders, even when their own credit scores are relatively unblemished.

Interestingly, high-risk lenders are also largely responsible for the differential treatment of equally qualified borrowers within lender. African-American and Hispanic borrowers are especially likely to receive high cost loans from these lenders, while minimal differences exist for lenders that specialize in serving less risky segments of the market. Further, the higher racial and ethnic differences observed among high risk borrowers and in disadvantaged neighborhoods are entirely explained by higher minority concentration and higher racial and ethnic differences at high foreclosure risk lenders.

The results of our analysis may 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 is also associated with a higher risk of future delinquency and default, with serious long-term consequences for long-term credit scores and home ownership rates. These effects can be expected to exacerbate existing wealth gaps (Charles and Hurst 2002; Gittleman and Wolff 2004).

In fact, using the same sample of loans, Bayer, Ross and Ferreira (2016) 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 0.60 and 0.50 percentage point increase in foreclosure rates for African-

28

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.

The results of our analysis also have implications for policymakers when developing strategies to reduce 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 borrowers applying for credit through the same credit market channel.

While we cannot directly test the above hypothesis with our data, the fact that controlling for ex-post foreclosure risk eliminates any correlation between private label securitization or non-depository lenders and high cost lending is again suggestive that our measure of high foreclosure risk lender captures important information about the non-conforming mortgage market and the impact of this market on minority borrowers. 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 African American and Hispanic borrowers (Savage 2011, 2012). At the same time, the substantial differences across lenders in terms of borrower credit score, loan to value ratio and ex-post foreclosure risk suggest that these lenders may be operating in submarkets with different unobserved risk profiles.

References

Alexandrov, Alexi and Sergei Koulayev. 2015. No Shopping in the U.S. Mortgage Market: Direct and Strategic Effects of Providing Information. Working Paper.

Angrist, Joshua D. and Alan B. Krueger. 1995. Split-Sample Instrumental Variables Estimates of the Return to Schooling. *Journal of Business & Economic Statistics*, 13 (2), 225-235.

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.

Avery, Robert B., Kenneth P. Brevoort, and Glenn B. Canner. 2007. The 2006 HMDA Data. Federal Reserve Bulletin, December.

Bayer, Patrick, Fernando Ferreira and Stephen L. Ross. 2016. The Vulnerability of Minority Homeowners in the Housing Boom and Bust. *American Economic Journal: Economic Policy*, 8 (1), 1-30.

Belsky, Eric S., Nicolas P. Retsinas and Mark Duda. 2007. The Financial Returns to Low-Income Homeownership, In: *Chasing the American Dream: new perspectives on affordable homeownership,* eds. W. Rohe and H. Watson, 191-212. Ithaca and London: Cornell University Press.

Bhutta, Neil and Daniel R. Ringo. 2014. The 2013 Home Mortgage Disclosure Act Data. Federal Reserve Bulletin, 100 (6).

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), 1139-65.

Calem, Paul S., Kevin Gillen, and Susan Wachter. 2004. The Neighborhood Distribution of Subprime Mortgage Lending *Journal of Real Estate Finance and Economics*, 29 (4), 393-410.

Case, Karl E. and Robert J. Shiller. 1989. The Efficiency of the Market for Single Family Homes. *American Economic Review*, 79 (1), 125–137.

Chan, Sewin, Andrew Haughwout and Joseph Tracy. 2015. *How Mortgage Finance Affects the Urban Landscape*, In the Handbook of Urban and Regional Economics, Eds. G. Duranton, W. Strange and V. Henderson, 987-1046. Amsterdam: North Holland.

Charles, Kerwin Kofi and Erik Hurst. 2002. The Transition to Home Ownership and the Black-White Wealth Gap. *Review of Economics and Statistics*, 84 (2), 281-297.

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 (4), 399-440.

Courchane, Marsha J. and David Nickerson. 1997. Discrimination Resulting from Overage Practices. *Journal of Financial Services Research*, 11 (1), 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.

Ferreira, Fernando and Joseph Gyourko. 2015. A New Look at the U.S. Foreclosure Crisis: Panel Data Evidence of Prime and Subprime Borrowers from 1997 to 2012. NBER Working Paper #21261.

Fisher, Lynn, Lauren Lambie-Hanson, and Paul S. Willen. 2010. A Profile of the Mortgage Crisis in a Low and Moderate Income Community. Boston Federal Reserve Public Policy Discussion Papers 10-6.

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

Ghent, Andra C., Rubén Hernández-Murillo, and Michael T. Owyang. In Press. Differences in Subprime Loan Pricing Across Races and Neighborhoods. *Regional Science and Urban Economics*.

Gittleman, Maury and Edward N. Wolff. 2004. Racial Differences in Patterns of Wealth Accumulation. *Journal of Human Resources*, 39 (1), 193-227.

Hanson, Andrew, Zackary Hawley, Hal Martin and Bo Liu. In Press. Discrimination in Mortgage Lending: Evidence from a Correspondence Experiment. *Journal of Urban Economics*.

Harding, John and Stephen L. Ross. 2010. Regulation of Large Financial Institutions: Lessons from Corporate Finance Theory. *Connecticut Insurance Law Journal*, 16 (1), 243-260

Haughwout, Andrew, Christopher Mayer, and Joseph Tracy. 2009. Subprime Mortgage Pricing: The Impact of Race, Ethnicity and Gender on the Cost of Borrowing. Brookings-Wharton Papers on Urban Affairs, Eds. G. Burtless and J. Pack, 33-63. Washington D.C.: Brookings Institution Press.

Heckman, James J. 1998. Detecting Discrimination. *Journal of Economic Perspectives*, 12 (2), 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. https://www.huduser.gov/Publications/pdf/HomeownershipGapsAmongLow-IncomeAndMinority.pdf

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

Horrace, William C. and Ronald L. Oaxaca. 2006. Results on the Bias and Inconsistency of Ordinary Least Squares for the Linear Probability Model. *Economics Letters*, 90 (3), 321-327.

Mayer, Christopher and Karen Pence. 2008. Subprime Mortgages: What, Where and To Whom? Finance and Economics Discussion Series, Federal Reserve Board, 2008-29.

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), 1449–1496.

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 (1), 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), 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 L. 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 (3): 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).

Tootell, Geoffrey M.B. 1996. Redlining in Boston: Do Lenders Discriminate Against

    Neighborhoods? *The Quarterly Journal of Economics*, 111 (4), 1049-1079.

Table 1: Descriptive Statistics

	Purchase Sample		Refinance Sample	
	Mean	Std. Dev.	Mean	Std. Dev.
HMDA Data				
Rate Spread	0.161	0.367	0.185	0.388
American Indian	0.003	0.053	0.003	0.057
Asian	0.094	0.292	0.082	0.274
African American	0.093	0.290	0.133	0.340
Hispanic	0.614	0.487	0.593	0.491
White	0.197	0.398	0.189	0.392
Male	0.649	0.477	0.650	0.477
Female	0.348	0.476	0.347	0.476
Loan Amount (in 1000s)	292.9	217.2	267.1	206.0
Applicant Income (in 1000s)	109.3	113.1	102.7	111.4
Coborrower Present	0.385	0.487	0.503	0.500
Jumbo Loan	0.283	0.450	0.190	0.393
Tract Median Income (in 1000s)	60158	23489	60123	23380
Tract Share African American	10.0	18.5	12.7	22.1
Tract Share Hispanic	15.9	20.1	17.2	20.8
Tract Share Asian	6.3	10.1	7.0	10.6
Tract Share Owner Occupant	68.5	23.9	69.4	22.7
Tract Share in Poverty	7.7	7.3	8.0	7.5
Tract Rent/Price	0.004	0.002	0.004	0.002
Dataquick Data				
Loan to Value Ratio >=0.80	0.800	0.400	0.518	0.500
Loan to Value Ratio >= 0.95	0.377	0.485	0.362	0.481
Subordinate Lien	0.003	0.055	0.066	0.248
First Lien Adjustable Rate	0.499	0.500	0.470	0.499
Condo	0.216	0.411	0.135	0.341
Mobile	0.001	0.037	0.001	0.036
Single Family	0.776	0.417	0.842	0.365
Lot Size (sf in 1000s)	12384	92140	14345	152661
Unit square feet (in 1000s)	1804	26329	1768	20173
Number of Bathrooms	1.98	5.64	2.01	1.14
Number of Bedrooms	2.15	5.88	2.02	1.76
Number of Stories	1.18	1.64	1.22	1.33
Units in Building	1.41	17.25	1.45	18.25
Credit Data				
Vantage Score	782	105	773	109
Mortgage Payment to Income Ratio	0.257	0.313	0.249	1.131
Debt Payment to Income Ratio	0.323	0.382	0.352	1.275
Borrower Age	43.7	11.3	49.6	11.0
Sample Size	94,699		91,242	

Notes: The table presents means and standard errors of the key variables in the home purchase and refinance regression samples used in Table 2.

Table 2: Rate Spread Models (04-07 Originations)

Variable Names	HMDA	Dataquick	Credit Data	Baseline	Lender FE
	Home Purchase Sample				
Asian	0.008	0.013**	0.010**	0.010**	0.006*
	(0.006)	(0.005)	(0.005)	(0.004)	(0.003)
African American	0.203***	0.152***	0.090***	0.087***	0.036***
	(0.021)	(0.015)	(0.011)	(0.009)	(0.007)
Hispanic	0.140***	0.090***	0.068***	0.069***	0.024***
	(0.017)	(0.011)	(0.010)	(0.009)	(0.004)
Observations	94,699	94,699	94,699	94,699	94,699
R-squared	0.186	0.297	0.368	0.417	0.609
	Refinance Sample				
Asian	0.001	0.001	0.008	0.007	0.005
	(0.007)	(0.007)	(0.006)	(0.006)	(0.004)
African American	0.118***	0.104***	0.046***	0.043***	0.019***
	(0.015)	(0.014)	(0.009)	(0.009)	(0.004)
Hispanic	0.046***	0.027***	0.014**	0.014*	0.005
	(0.010)	(0.008)	(0.007)	(0.007)	(0.003)
Observations	91,242	91,242	91,242	91,242	91,242
R-squared	0.116	0.196	0.336	0.354	0.551

Notes: These models are estimated using OLS with standard errors clustered at the lender level. The first column controls for just the variables available in the Home Mortgage Disclosure Act data, the second column includes additional controls made available from assessor's data including Combined Loan to Value Ratio and the presence of subordinate liens, column three includes credit score and income ratios based on data from the credit repository, column 4 adds interactions between key loan terms and whether the borrower has a subprime credit score (Vantage<701), and column 5 presents the model from column 4 after including lender fixed effects.
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Rate Spread Loans and Lender Risk

| Variable Names | Home Purchase Sample | | | |
	Baseline Model	Lender Fixed Effects	Baseline Model + Lender Foreclosure Risk	Baseline Model + Lender Attributes
African American	0.087***	0.036***	0.049***	0.046***
	(0.009)	(0.007)	(0.009)	(0.010)
Hispanic	0.069***	0.024***	0.028***	0.026***
	(0.009)	(0.004)	(0.007)	(0.006)
Foreclosure risk by lender			3.181***	3.438***
			(0.303)	(0.432)
Lender Type (National Banks omitted)				
Commercial Banks				0.004
				(0.046)
State Chartered Banks				-0.036
				(0.103)
Savings and Loans				-0.039
				(0.028)
Credit Unions				0.027
				(0.075)
Mortgage Banks				-0.025
				(0.033)
Share of Mortgages Re-Sold				-0.042
				(0.038)
Loan Buyer (GSE omitted)				
FHA				-0.119
				(0.168)
Private Securitization				-0.105
				(0.102)
Commercial/Savings Bank				-0.059
				(0.058)
Ins, CU, Mort Bank				0.018
				(0.065)
Affiated Lender				-0.067*
				(0.039)
Other				-0.042
				(0.038)
Observations	94,699	94,699	46,933	44,284
R-squared	0.417	0.609	0.490	0.489

Notes: This table presents estimates for the home purchase sample. The table presents estimates from the subprime model for a subsample restricted to only lenders with 10 or more loans. The estimates from the subprime model are shown in column 1, and the estimates for the lender fixed effects model are shown in column 2. These models are estimated using OLS. Column 3 includes a control for the lender's ex-post foreclosure risk based on the lender fixed effect from a model of whether the borrower ever faces foreclosure, and column 4 adds controls for the type of lender based on the lender's regulator, the share of loans sold by the lender on the secondary market, and the share of sold loans by type of purchaser. Columns 3 and 4 are estimated using split sample IV estimates. The split sample IV estimates are based on a subsample with at least 10 loans per lender, and the estimates presented are the average of 20 hold-out sample draws. Standard errors are clustered at the lender level except for the last two columns where they are bootstrapped by sampling lenders with replacement.

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 4: Alternative Lender Risk Measures

| Variable Names | Home Purchase Sample | | | |
| | Lenders with >=10 white loans | | Lenders with >= 10 loans | |
	Baseline	White Foreclosure Risk	30 Day Delinquency	Conditional on High Cost
African American	0.046***	0.047***	0.052***	0.061***
	(0.009)	(0.010)	(0.011)	(0.010)
Hispanic	0.025***	0.020***	0.028***	0.037***
	(0.006)	(0.007)	(0.006)	(0.009)
Lender Foreclosure Risk (based on all loans)	3.331***			
	(0.261)			
Lender Foreclosure Risk (based on white loans)		4.295***		
		(0.851)		
Lender 30 Day Delinquency			2.142***	
			(0.277)	
Foreclosure net of rate spread loans				3.652
				(0.570)
Observations	46,763	46,752	46,933	46,933
R-squared	0.493	0.349	0.460	0.419

Notes: Column 1 presents the results from Column 3 in Table 3 for the home purchase sample restricted only to lenders with at least 10 white loans. Column 2 contains estimates from a model controlling for lender foreclosure fixed effects based only on the lender's white loans. Column 3 contains estimates using ex-post lender risk of 30 to 180 day delinquency, and Column 4 contains estimates using ex-post foreclosure risk based on an ever foreclose model that controls for having a high cost loan. Columns 3 and 4 use the original home purchase sample of lenders with at least 10 loans overall. All models are estimated using split sample IV, and standard errors are bootstrapped.
Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5: Sorting over Ex-post Foreclosure Risk

Variable Names	Coefficient Estimates
Borrower Characteristics	
African American	0.038***
	(0.003)
Hispanic	0.049***
	(0.002)
Asian	0.004*
	(0.002)
Female	0.007***
	(0.001)
Co-borrower present	-0.013***
	(0.001)
Age second quintile	0.005***
	(0.002)
Age third quintile	0.012***
	(0.002)
Age fourth quintile	0.020***
	(0.002)
Age fifth quintile	0.020***
	(0.002)
Log Income	0.015***
	(0.001)
Subprime Credit Score	0.106***
	(0.003)
Subprime Score above Median	-0.055***
	(0.003)
Prime Score above Median	-0.028***
	(0.001)
Neighborhood Characteristics	
Tract Percent African-American	0.017***
	(0.005)
Tract Percent Hispanic	0.044***
	(0.006)
Tract Percent Owner-Occupied	0.030***
	(0.004)
Tract Percent Poverty	0.044***
	(0.015)
Tract Rent to Price Ratio	0.013***
	(0.002)

Table 5: Sorting over Ex-post Foreclosure Risk

Variable Names	Coefficient Estimates
Loan Risk Factors	
Loan to Value Ratio above 0.80	0.014***
	(0.001)
Loan to Value Ratio above 0.90	0.018***
	(0.002)
Loan to Value Ratio above 0.95	0.071***
	(0.002)
Loan to Value Ratio above 1.00	-0.004
	(0.004)
Housing Expense to Income Ratio above 0.26	-0.015***
	(0.002)
Housing Expense to Income Ratio above 0.33	-0.012***
	(0.002)
Debt to Income Ratio above 0.36	-0.004**
	(0.002)
Debt to Income Ratio above 0.42	0.010***
	(0.002)
Observations	94,481
R-squared	0.232

Notes: This model regresses the standardized measure of ex-post foreclosure risk on borrower, neighborhood and loan attributes in the home purchase sample. Standard errors are clustered at the census tract level.

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Metropolitan Specific Estimates Model - Home Purchase Sample

				Baseline Model			
Variable Names	Chicago	Cleveland	Denver	Los Angeles	Maryland	Miami	San Francisco
African American	0.094***	0.066***	0.082***	0.059***	0.085***	0.117***	0.062***
	(0.016)	(0.020)	(0.014)	(0.011)	(0.011)	(0.014)	(0.014)
Hispanic	0.062***	0.011	0.063***	0.063***	0.076***	0.071***	0.075***
	(0.015)	(0.018)	(0.010)	(0.010)	(0.021)	(0.013)	(0.014)
Observations	12,752	11,976	13,182	13,515	13,693	13,395	14,293
R-squared	0.429	0.445	0.414	0.435	0.412	0.431	0.407

				Lender Foreclosure Risk (Reduced Form)			
Variable Names	Chicago	Cleveland	Denver	Los Angeles	Maryland	Miami	San Francisco
African American	0.068***	0.034*	0.050***	0.036***	0.057***	0.075***	0.039***
	(0.011)	(0.017)	(0.014)	(0.008)	(0.008)	(0.011)	(0.011)
Hispanic	0.034***	0.014	0.046***	0.033***	0.036***	0.046***	0.034***
	(0.012)	(0.018)	(0.010)	(0.007)	(0.013)	(0.010)	(0.011)
Foreclosure Risk	2.743***	2.877***	2.256***	2.515***	2.915***	2.435***	2.105***
	(0.228)	(0.238)	(0.207)	(0.207)	(0.200)	(0.235)	(0.277)
Observations	12,752	11,976	13,182	13,515	13,693	13,395	14,293
R-squared	0.531	0.549	0.502	0.545	0.546	0.521	0.513

				Lender Fixed Effect Model			
Variable Names	Chicago	Cleveland	Denver	Los Angeles	Maryland	Miami	San Francisco
African American	0.039***	0.030*	0.032**	0.021**	0.039***	0.051***	0.018*
	(0.010)	(0.016)	(0.013)	(0.010)	(0.009)	(0.011)	(0.010)
Hispanic	0.018**	0.011	0.021***	0.021***	0.024**	0.032***	0.022***
	(0.008)	(0.015)	(0.008)	(0.006)	(0.010)	(0.008)	(0.007)
Observations	12,752	11,976	13,182	13,515	13,693	13,395	14,293
R-squared	0.600	0.620	0.601	0.641	0.645	0.603	0.654

Notes: Table 6 presents estimates by individual sites in each column for the home purchase sample. Panel 1 presents the subprime model, panel 2 presents reduced form models that control for estimated lender fixed effects from the ever foreclose regression, and the third panel presents the lender fixed effects estimates. Standard errors are clustered at the lender level.

Robust standard errors shown in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 7: Racial and Ethnic Differences for Home Purchase Market

Variable Names	Baseline Model			Lender Fixed Effects Model		
African American	0.085***	0.042*	0.001	0.025***	0.004	0.010
	(0.013)	(0.025)	(0.023)	(0.008)	(0.014)	(0.037)
Hispanic	0.024**	0.036	-0.005	0.001	0.008	0.013
	(0.010)	(0.022)	(0.021)	(0.005)	(0.014)	(0.008)
<u>Interactions with Common Loan Risk Factors</u>						
African American*Subprime	0.032*	0.027*	-0.010	0.015	0.013	0.007
	(0.017)	(0.016)	(0.014)	(0.011)	(0.011)	(0.006)
Hispanic*Subprime	-0.029*	-0.030*	-0.012	-0.009	-0.010	-0.017
	(0.017)	(0.017)	(0.013)	(0.010)	(0.010)	(0.004)
African American*High DTI	0.005	0.005	0.010	0.007	0.007	0.011
	(0.009)	(0.008)	(0.012)	(0.007)	(0.007)	(0.010)
Hispanic*High DTI	-0.006	-0.005	-0.001	0.000	0.000	0.001
	(0.010)	(0.010)	(0.011)	(0.006)	(0.006)	(0.014)
African American*High LTV	-0.015	-0.018	0.006	0.009	0.007	0.003
	(0.013)	(0.013)	(0.011)	(0.008)	(0.008)	(0.004)
Hispanic*High LTV	0.113***	0.110***	0.061***	0.056***	0.054***	0.049***
	(0.013)	(0.013)	(0.016)	(0.011)	(0.011)	(0.014)
<u>Interactions with Neighborhood Characteristics</u>						
African American*Pct Poverty		0.145**	0.058		0.057	0.059
		(0.063)	(0.072)		(0.067)	(0.179)
Hispanic*Pct Poverty		0.001	0.040		0.051	0.057
		(0.066)	(0.074)		(0.050)	(0.069)
African American*Pct Black		-0.021	-0.008		-0.005	-0.004
		(0.018)	(0.019)		(0.013)	(0.003)
Hispanic*Pct Hispanic		0.042***	0.018		0.001	0.001
		(0.016)	(0.025)		(0.012)	(0.009)
African American*Pct Owner-Occupied		-0.022	-0.022		-0.018	-0.020**
		(0.021)	(0.022)		(0.015)	(0.009)
Hispanic*Pct Owner-Occupied		-0.050**	-0.027		-0.035**	-0.035***
		(0.020)	(0.021)		(0.014)	(0.013)
African American*Rent to Price ratio		-0.021	0.048***		0.031***	0.028***
		(0.018)	(0.015)		(0.010)	(0.008)
Hispanic*Rent to Price Ratio		0.042***	0.018		0.013	0.011
		(0.016)	(0.011)		(0.009)	(0.007)

Table 7: Racial and Ethnic Differences for Home Purchase Market

Variable Names	Baseline Model			Lender Fixed Effects Model		
Interactions with Lender Foreclosure Risk						
African American*Lender Foreclosure Risk			0.524**			0.187
			(0.225)			(0.208)
Hispanic*Lender Foreclosure Risk			0.185			0.197**
			(0.217)			(0.080)
Observations	94,481	94,481	46,424	94,481	94,481	46,424
R-squared	0.421	0.436	0.492	0.610	0.610	0.198

Notes: This table presents the estimates for the home purchase sample of models that interact race and ethnicity with key borrower, location and lender controls to the baseline subprime and lender fixed effect models. The first and fourth rows present the estimates interacting dummies for subprime credit score, debt to income ratio above 0.45 and loan to value ratio above 0.95 with race and ethnicity for the subprime and lender fixed effects models, respectively. Columns 2 and 5 also add interactions between race and ethnicity and census tract attributes for the tract in which the housing unit is located. Columns 3 and 6 also interact race and ethnicity with lender ex-post foreclosure risk. Columns 1, 2, 4 and 5 are estimated using OLS with standard errors clustered at the lender level, and columns 3 and 6 are estimated using split-sample IV and standard errors are bootstrapped.
Standard errors shown in paretheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix Table 1: Sample Selection of HMDA Variables

Variable	HMDA full sample Mean	HMDA full sample Stand. Dev.	DQ High quality match Mean	DQ High quality match Stand. Dev.	DQ High quality match May-Aug Mean	DQ High quality match May-Aug Stand. Dev.	Sample, weighted Mean	Sample, weighted Stand. Dev.	Credit Data Matched Sample, weighted[2] Mean	Credit Data Matched Sample, weighted[2] 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: This table presents means and standard errors for the pooled sample from HMDA for all seven sites. The first column is the full sample, the second column is just loans that match our assessors data on all key fields including lender name, the third column restricts the sample to transactions between May and Aug, the fourth column is our random sample using the sampling weights, and the final column is for the final sample of observations with a pre-origination Vantage score.

Appendix Table 2: Sample Means for Lender Variables

	Purchase Sample		Refinance Sample	
	Mean	Std. Dev.	Mean	Std. Dev.
Agency Code				
National Banks	0.287	0.452	0.260	0.439
Commercial Banks	0.129	0.335	0.128	0.334
State Chartered Banks	0.042	0.201	0.045	0.207
Savings and Loans	0.188	0.391	0.223	0.417
Credit Unions	0.009	0.092	0.018	0.132
Mortgage Banks	0.344	0.475	0.326	0.469
Sample Size	94,699		91,242	
Securitization				
Share sold	0.256	0.262	0.287	0.287
Share to GSE	0.315	0.343	0.322	0.351
Share to FHA	0.024	0.052	0.012	0.036
Private Securitization	0.070	0.161	0.073	0.164
Commercial/Savings Bank	0.074	0.136	0.065	0.128
Ins, CU, Mort Bank	0.109	0.230	0.128	0.246
Affiated Lender	0.149	0.277	0.178	0.304
Other	0.259	0.343	0.223	0.320
Sample Size	89885		84748	
Foreclosure Risk Variables				
Foreclosure risk by lender	0.000	0.058	0.000	0.038
Sample Size	92,806		89,078	
Foreclosure risk using white loans	0.000	0.283	0.000	0.260
Sample Size	91,675		87,770	

Notes: The table presents means and standard errors of key lender variables for the home purchase and refinance regression samples used in Table 3 and appendix Table 3, which are restricted to lenders with 10 or more loans.

Appendix Table 3: Rate Spread Loans and Lender Attributes

Variable Names	Baseline Model	Baseline Model + Agency Code	Baseline Model + Securitization Shares	Baseline Model + Lender Attributes
	Home Purchase Sample			
African American	0.087***	0.084***	0.075***	0.046***
	(0.009)	(0.008)	(0.007)	(0.010)
Hispanic	0.069***	0.064***	0.058***	0.026***
	(0.009)	(0.008)	(0.008)	(0.006)
Foreclosure risk by lender				3.438***
				(0.432)
Commercial Banks		0.056**		0.004
(omit national banks)		(0.022)		(0.046)
State Chartered Banks		0.145		-0.036
		(0.112)		(0.103)
Savings and Loans		0.053***		-0.039
		(0.020)		(0.028)
Credit Unions		0.007		0.027
		(0.021)		(0.075)
Mortgage Banks		0.113***		-0.025
		(0.022)		(0.033)
Share sold			0.039	-0.042
			(0.038)	(0.038)
FHA			0.166	-0.119
(Omit sold to GSE)			(0.104)	(0.168)
Private Securitization			0.495***	-0.105
			(0.091)	(0.102)
Commercial/Savings Bank			0.008	-0.059
			(0.047)	(0.058)
Ins, CU, Mort Bank			0.214***	0.018
			(0.064)	(0.065)
Affiated Lender			0.071**	-0.067*
			(0.031)	(0.039)
Other			0.185***	-0.042
			(0.033)	(0.038)
Observations	94,699	94,699	89,885	44,284
R-squared	0.417	0.432	0.459	0.489

Notes: This table presents estimates for the home purchase sample. The first column of the table presents estimates for the subprime model for a subsample restricted to only lenders with 10 or more loans. The next two columns present estimates after including either agency code or securitization controls, respectively. The first three models are estimated using OLS with standard errors clustered at the lender level. The final column presents estimates for a model that contains all of our lender controls, and is estimated using split-sample IV with bootstrapped standard errors.

Standard errors shown in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Appendix Table 4: Rate Spread Loans and Lender Risk

Variable Names	Baseline Model	Lender Fixed Effects	Baseline Model + Lender Foreclosure Risk	Baseline Model + Lender Attributes
		Refinance Sample		
African American	0.042***	0.018***	0.031***	0.023***
	(0.009)	(0.004)	(0.009)	(0.005)
Hispanic	0.013*	0.005	0.011	-0.004
	(0.007)	(0.003)	(0.011)	(0.008)
Foreclosure risk by lender			3.884***	3.921***
			(0.561)	(0.854)
Commercial Banks				0.082
(omit national banks)				(0.076)
State Chartered Banks				-0.057
				(0.052)
Savings and Loans				-0.038
				(0.043)
Credit Unions				-0.045
				(0.062)
Mortgage Banks				-0.020
				(0.040)
Share sold				0.075
				(0.056)
FHA				-0.136
(Omit sold to GSE)				(0.185)
Private Securitization				0.167**
				(0.075)
Commercial/Savings Bank				0.038
				(0.093)
Ins, CU, Mort Bank				0.149**
				(0.063)
Affiated Lender				-0.044
				(0.059)
Other				0.047
				(0.047)
Observations	91,242	91,242	45,101	41,649
R-squared	0.354	0.551	0.351	0.389

Notes: This table presents estimates for the refinance sample. The table presents estimates from the subprime model for a subsample restricted to only lenders with 10 or more loans. The estimates from the subprime model are shown in column 1, and the estimates for the lender fixed effects model are shown in column 2. These models are estimated using OLS. Column 3 includes a control for the lender's ex-post foreclosure risk based on the lender fixed effect from a model of whether the borrower ever faces foreclosure, and column 4 adds controls for the type of lender based on the lender's regulator, the share of loans sold by the lender on the secondary market, and the share of sold loans by type of purchaser. Columns 3 and 4 are estimated using split sample IV estimates. The split sample IV estimates are based on a subsample with at least 10 loans per lender, and the estimates presented are the average of 20 hold-out sample draws. Standard errors are clustered at the lender level except for the last two columns where they are bootstrapped by sampling lenders with replacement.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix Table 5: Lender Foreclosure Risk and Racial Composition of Loans

Variable Names	Refinance Sample			
	Lenders with >=10 white loans		Lenders with >= 10 loans	
	Baseline	White Foreclosure Risk	30 Day Delinquency	Conditional on High Cost
African American	0.029***	0.031***	0.026***	0.036***
	(0.009)	(0.007)	(0.008)	(0.011)
Hispanic	0.001	0.011	-0.012	0.09
	(0.012)	(0.009)	(0.015)	(0.017)
Lender Foreclosure Risk (based on all loans)	4.225***			
	(0.622)			
Lender Foreclosure Risk (based on white loans)		4.359***		
		(1.531)		
Lender 30 Day Delinquency			1.992***	
			(0.273)	
Foreclosure net of rate spread loans				2.695***
				(0.884)
Observations	44,856	44,851	45,101	45,101
R-squared	0.345	0.270	0.369	0.343

Notes: Column 1 presents the results from Column 3 in Appendix Table 4 for the refinance sample restricted only to lenders with at least 10 white loans. Column 2 contains estimates from a model controlling for lender foreclosure fixed effects based only on the lender's white loans. Column 3 contains estimates using ex-post lender risk of 30 to 180 day delinquency, and Column 4 contains estimates using ex-post foreclosure risk based on an ever foreclose model that controls for having a high cost loan. Columns 3 and 4 use the original refinance sample of lenders with at least 10 loans overall. All models are estimated using split sample IV, and standard errors are bootstrapped.restricted to lenders that have at least 10 white loans. For comparison purposes, columns 1 and 2 present split sample IV estimates for the simple lender foreclosure fixed effect model using the restricted sample. Standard errors are bootstrapping sampling lenders with replacement.

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix Table 6: Sorting over Ex-post Foreclosure Risk (Complete Estimates)

Variable Names	Home Purchase Sample			Refinance Sample		
	Borrower Attributes	+ Census Tract Attributes	+ Loan Attributes	Borrower Attributes	+ Census Tract Attributes	+ Loan Attributes
African American	0.055***	0.045***	0.038***	0.022***	0.014***	0.014***
	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Hispanic	0.071***	0.058***	0.049***	0.022***	0.016***	0.015***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Asian	0.004*	0.002	0.004*	-0.003	-0.004**	-0.004*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Female	0.006***	0.007***	0.006***	0.002	0.002*	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Co-borrower	-0.015***	-0.015***	-0.012***	-0.004***	-0.004***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age second quintile	0.001	0.001	0.005***	0.004	0.004	0.004*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Age third quintile	0.004**	0.004**	0.012***	-0.001	-0.001	0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Age fourth quintile	0.011***	0.010***	0.019***	0.0020	0.002	0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Age fifth quintile	0.007***	0.008***	0.019***	0.003	0.003	0.005**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Logarithm of Income	0.009***	0.014***	0.015***	0.002**	0.005***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Subprime Credit Score	0.114***	0.111***	0.105***	0.099***	0.097***	0.095***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Subprime Score above Median	-0.055***	-0.053***	-0.054***	-0.051***	-0.050***	-0.049***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Prime Score above Median	-0.042***	-0.040***	-0.028***	-0.031***	-0.030***	-0.028***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tract Percent African-American		0.023***	0.016***		0.016***	0.016***
		(0.005)	(0.005)		(0.003)	(0.003)
Tract Percent Hispanic		0.056***	0.043***		0.016***	0.014***
		(0.006)	(0.006)		(0.004)	(0.004)
Tract Percent Owner-Occupied		0.029***	0.030***		0.018***	0.017***
		(0.004)	(0.004)		(0.003)	(0.003)
Tract Percent Poverty		0.049***	0.044***		0.065***	0.057***
		(0.015)	(0.015)		(0.010)	(0.011)
Tract Rent to Price Ratio		0.018***	0.013***		0.006***	0.007***
		(0.003)	(0.002)		(0.002)	(0.002)

Appendix Table 6: Sorting over Ex-post Foreclosure Risk (Complete Estimates)

Variable Names	Home Purchase Sample			Refinance Sample		
	Borrower Attributes	+ Census Tract Attributes	+ Loan Attributes	Borrower Attributes	+ Census Tract Attributes	+ Loan Attributes
Loan to Value Ratio above 0.80			0.013***			0.011***
			(0.001)			(0.002)
Loan to Value Ratio above 0.90			0.018***			0.016***
			(0.002)			(0.002)
Loan to Value Ratio above 0.95			0.071***			0.019***
			(0.002)			(0.003)
Loan to Value Ratio above 1.00			-0.004			0.022***
			(0.004)			(0.001)
Housing Expense to Income Ratio above 0.26			-0.014***			-0.005***
			(0.002)			(0.001)
Housing Expense to Income Ratio above 0.33			-0.011***			-0.014***
			(0.002)			(0.002)
Debt to Income Ratio above 0.36			-0.004**			0.007***
			(0.002)			(0.002)
Debt to Income Ratio above 0.42			0.010***			0.017***
			(0.002)			(0.002)
Observations	94,481	94,481	94,481	91,070	91,070	91,070
R-squared	0.204	0.209	0.232	0.136	0.138	0.144

Notes: This tables present the estimates from models that regress the standardized measure of ex-post foreclosure risk on borrower, neighborhood and loan attributes in the home purchase and refinance samples of lenders with at least 10 loans. Columns 1-3 present the home purchase sample estimates beginning with a model that adds borrower attributes, followed by the addition of neighborhood attributes, and finally by loan terms. Columns 4-6 contain estimates of the same models for the refinance sample. Standard errors are clustered at the census tract level.

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Appendix Table 7: Metropolitan Specific Estimates Model Refinance Sample

				Subprime Model			
Variable Names	Chicago	Cleveland	Denver	Los Angeles	Maryland	Miami	San Francisco
African American	0.051***	0.029	0.027**	0.034**	0.040***	0.052***	0.028*
	(0.015)	(0.018)	(0.014)	(0.014)	(0.012)	(0.019)	(0.015)
Hispanic	0.026**	-0.008	0.031***	0.013360*	0.007	0.01	0.006
	(0.013)	(0.020)	(0.012)	(0.007)	(0.016)	(0.013)	(0.007)
Observations	12,385	11,365	12,793	13,103	12,920	12,974	13,538
R-squared	0.396	0.362	0.306	0.325	0.423	0.388	0.231

			Lender Foreclosure Risk (Reduced Form)				
Variable Names	Chicago	Cleveland	Denver	Los Angeles	Maryland	Miami	San Francisco
African American	0.043***	0.025	0.024*	0.027**	0.035***	0.032*	0.023
	(0.016)	(0.017)	(0.013)	(0.013)	(0.011)	(0.017)	(0.014)
Hispanic	0.022100*	-0.001	0.019*	0.005	0.003	-0.001	0
	(0.012)	(0.019)	(0.011)	(0.008)	(0.015)	(0.012)	(0.007)
Foreclosure Risk	2.763***	2.319***	2.012***	2.110***	3.053***	2.682***	2.138***
	(0.836)	(0.806)	(0.727)	(0.311)	(0.297)	(0.377)	(0.361)
Observations	12,385	11,365	12,793	13,103	12,920	12,974	13,538
R-squared	0.438	0.391	0.339	0.368	0.479	0.440	0.293

				Lender Fixed Effect Model			
Variable Names	Chicago	Cleveland	Denver	Los Angeles	Maryland	Miami	San Francisco
African American	0.018**	0.012	0.001	0.020**	0.015622**	0.032***	0.01
	(0.009)	(0.014)	(0.010)	(0.009)	(0.007)	(0.010)	(0.008)
Hispanic	0.014*	0.004	0.016*	0.006	0.003424	0.002	0.003
	(0.007)	(0.015)	(0.008)	(0.006)	(0.010)	(0.009)	(0.005)
Observations	12,385	11,365	12,793	13,103	12,920	12,974	13,538
R-squared	0.613	0.577	0.548	0.492	0.619	0.572	0.483

Notes: This table presents estimates by individual sites in each column for the refinance sample. Panel 1 presents the subprime model, panel 2 presents reduced form model that controls for lender fixed effects from the ever foreclose regression, and the third panel presents the lender fixed effects estimates. Standard errors are clustered at the lender level.

Robust standard errors shown in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix Table 8: Racial and Ethnic Differences for Refinance Subprime model

Variable Names	Subprime			Lender FE		
	Credit Quality	Neighborhood	Foreclosure	Credit Quality	Neighborhood	Foreclosure
African American	0.023**	0.026	0.021	0.007*	-0.007	0.000
	(0.009)	(0.026)	(0.034)	(0.004)	(0.020)	(0.028)
Hispanic	0.011	0.0026	0.005	0.008*	0.010	0.020
	(0.007)	(0.017)	(0.026)	(0.004)	(0.017)	(0.024)
African American*Subpri	0.017	0.016	-0.003	0.009	0.007	-0.001
	(0.013)	(0.013)	(0.018)	(0.009)	(0.009)	(0.015)
Hispanic*Subprime	0.009	0.007	-0.001	0.000	-0.001	-0.013
	(0.009)	(0.009)	(0.016)	(0.007)	(0.007)	(0.011)
African American*High D	0.008	0.008	-0.003	-0.002	-0.002	-0.007
	(0.007)	(0.007)	(0.011)	(0.006)	(0.006)	(0.009)
Hispanic*High DTI	0.004	0.004	0.004	-0.009	-0.009	-0.010
	(0.008)	(0.008)	(0.010)	(0.007)	(0.007)	(0.008)
African American*High L	0.030***	0.029***	0.011	0.022***	0.021***	0.015**
	(0.008)	(0.008)	(0.009)	(0.006)	(0.006)	(0.007)
Hispanic*High LTV	0.005	0.004	-0.007	0.000	-0.001	-0.0047
	(0.006)	(0.006)	(0.009)	(0.005)	(0.005)	(0.007)
African American*Pct Poverty		0.023	-0.029		0.037	0.036
		(0.067)	(0.095)		(0.052)	(0.075)
Hispanic*Pct Poverty		-0.005	-0.045		0.0132	-0.004
		(0.076)	(0.098)		(0.076)	(0.089)
African American*Pct Black		-0.030	-0.015		-0.012	-0.011
		(0.020)	(0.024)		(0.016)	(0.0189)
Hispanic*Pct Hispanic		0.009	0.017		0.000	0.000
		(0.017)	(0.024)		(0.013)	(0.016)
African American*Pct Owner-Occupied		-0.022	-0.032		-0.008	-0.007
		(0.023)	(0.030)		(0.019)	(0.025)
Hispanic*Pct Owner-Occupied		-0.014	-0.037		-0.020	-0.025
		(0.016)	(0.024)		(0.016)	(0.021)
African American*rent to price ratio		0.018	0.030**		0.019*	-0.031**
		(0.014)	(0.015)		(0.011)	(0.013)
Hispanic*Rent to Price Ratio		0.017*	0.024**		0.011	0.016
		(0.010)	(0.012)		(0.009)	(0.013)

Appendix Table 8: Racial and Ethnic Differences for Refinance Subprime model

Variable Names	Subprime Credit Quality	Neighborhood	Foreclosure	Lender FE Credit Quality	Neighborhood	Foreclosure
African American*Lender Foreclosure			0.516			0.408
			(0.486)			(0.267)
Hispanic*Lender Foreclosure			0.293			0.389
			(0.593)			(0.245)
Observations	91,070	91,070	44,600	91,070	91,070	44,600
R-squared	0.371	0.371	0.370	0.551	0.551	0.201

Notes: This table presents the estimates for the refinance sample of models that interact race and ethnicity with key borrower, location and lender controls to the baseline subprime and lender fixed effect models. The first and fourth rows present the estimates interacting dummies for subprime credit score, debt to income ratio above 0.45 and loan to value ratio above 0.95 with race and ethnicity for the subprime and lender fixed effects models, respectively. Columns 2 and 5 also add interactions between race and ethnicity and census tract attributes for the tract in which the housing unit is located. Columns 3 and 6 also interact race and ethnicity with lender ex-post foreclosure risk. Columns 1, 2, 4 and 5 are estimated using OLS with standard errors clustered at the lender level, and columns 3 and 6 are estimated using split-sample IV and standard errors bootstrapped.
Standard errors shown in paretheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$