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

This paper examines racial and ethnic differences in high-cost mortgage lending in seven diverse metropolitan areas from 2004 to 2007. Controlling for credit score and other risk factors, African American and Hispanic borrowers are 103% and 78% more likely to receive high-cost mortgages for home purchases. A large part of the increase is attributable to sorting across lenders (55%-65%), and this, in turn, can be largely accounted for by the lender's ex post foreclosure risk. The remaining within-lender differences are also concentrated in high-risk lenders, revealing the central role of these institutions in explaining market-wide racial and ethnic differences. (*JEL* G21, I28, J15, J71, R21)

Received December 17, 2014; editorial decision January 20, 2017 by Editor Philip Strahan.

Whether African American and Hispanic mortgage borrowers face a higher cost of credit compared with white borrowers has been a long-standing question in academic and policy debates about inequities in financial markets. Studies of racial discrimination in mortgage lending historically have focused on

---

The estimates in Table 2, in panels 1 and 3 of Table 6, and in Columns 1 and 4 of Table 7 are based on models similar to those presented in NBER Working Paper #20762. Gordon MacDonald, Kyle Mangum, Yuan Wang, Ailing Zhang, and Samantha Minieri provided outstanding research assistance. The analyses presented in this paper uses information provided by one of the major credit reporting agencies. The credit agency has the legal right to view the research prior to its public release. However, the substantive content of the paper is the responsibility of the authors and does not reflect the specific views of any credit reporting agencies. This work was supported by the Ford Foundation, the Research Sponsors Program of the Zell/Lurie Real Estate Center at Wharton, and the Center for Real Estate and Urban Economic Studies at the University of Connecticut. Professor Ross has worked in the past as a consultant on fair housing and fair lending issues. None of his consulting activities in the last three years have totaled more than $10,000. Supplementary data can be found on *The Review of Financial Studies* Web site. Send correspondence to Stephen L. Ross, Department of Economics, University of Connecticut, 309 Oak Hall, 365 Fairfield Way U-1063, Storrs, CT 06269-1063; telephone: 806-486-3533. E-mail: stephen.l.ross@uconn.edu.

doi:10.1093/rfs/hhx035 Advance Access publication April 29, 2017

discrimination in mortgage underwriting (Munnell et al. 1996; Ross and Yinger 2002) and redlining against minority neighborhoods (Holmes and Horvitz 1994; Tootell 1996; Ross and Tootell 2004). But attention has increasingly turned to differences in the price of mortgage credit (Ross 2005; Chan, Haughwout, and Tracy 2015), which also has been the focus of several high profile U.S. Department of Justice cases in the wake of the recent financial crisis.[1]

In this study, we estimate 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, and we examine a series of factors that explain those disparities. In particular, we investigate how much of the cost of credit differences by race and ethnicity is due to across lender differences and whether specific lender characteristics can explain that effect.[2]

A critical contribution of our work is to build a representative micro database including all types of mortgages (prime and subprime) and containing information about the race and ethnicity of each borrower, their risk factors and whether the mortgage is high cost. We do so by linking the Home Mortgage Disclosure Act (HMDA) data on home purchase and refinance mortgages from 2004 and 2007, which has information on race and ethnicity of each borrower, to public records data on housing transactions and liens in seven distinct metropolitan housing markets.[3] HMDA contains a rate spread variable that allows us to create an indicator for high-cost loans, which is equal to one when the Annual Percentage Rate (APR) exceeds the interest rate on treasury securities of comparable maturity by at least three percentage points.[4] We then provide a sample of matched mortgages that were originated between May and August to one of the major credit reporting agencies. The credit reporting agency uses the name and address (available in the public records) to match borrowers to archival credit repository data, providing a vantage credit score plus detailed credit line information for each borrower.

---

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

2 Higher credit costs create an obvious barrier to African American and Hispanic homeownership, which has historically lagged that of white households by large margins (Quercia, McCarthy, and Wachter 2003; Haurin, Herbert, and Rosenthal 2007; Belsky, Retsinas, and Duda 2007). Moreover, high-cost lending appears to have been concentrated in minority neighborhoods (Mayer and Pence 2008; Chan, Haughwout, and Tracy 2015) in the run up to the housing crisis, and during the crisis especially high foreclosure rates were observed in those same neighborhoods (Edmiston 2009; Gerardi and Willen 2009; Reid and Laderman 2009).

3 The data also include 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 (2011) for a discussion of the timing of the onset of the broad financial crisis. All results are robust to including the 2008 originations.

4 The annual percentage rate (APR) estimates the cost of credit, including interest rate and closing costs. These high-cost or rate spread loans are sometimes referred to as subprime loans (Mayer and Pence 2008; Chan, Haughwout, and Tracy 2015), but other authors study the subprime market based on a list of top subprime lenders, for example, Ferreira and Gyourko (2015), based on borrowers who have a low credit score, for example, Mian and Sufi (2009), or private label securitized loans, for example, Ghent, Hernández-Murillo, and Owyang (2014).

With these data in hand, we begin the empirical analysis by showing that most of the racial and ethnic differences both across and within lender arise in the home purchase segment of the market (as opposed to the mortgage refinance market), similar to the findings of Ghent, Hernández-Murillo, and Owyang (2014) in the subprime market. After controlling for credit score, type of mortgage and other key risk factors, African American and Hispanic borrowers have a 9.0 and 6.8 percentage point higher likelihood of a rate spread or high-cost loan, respectively, in the home purchase market relative to an average incidence of 8.7 for all-white borrowers.[5] As shown earlier by Avery, Canner, and Cook (2005), Avery, Brevoort, and Canner (2007), and Bhutta and Ringo (2014), the inclusion of lender fixed effects in the model substantially reduces the unexplained differences for African American and Hispanic borrowers by 59% and 65% in our home purchase sample, respectively.[6] These overall patterns hold for each individual MSA in the sample.

To understand what drives this variation across lenders, we develop a new measure of ex post lender foreclosure risk, which captures the inherent risk in the lender's portfolio of loans after controlling for each loans' observable risk factors.[7] Indeed, ex post foreclosure risk emerges as the key explanatory variable, accounting for 73% and 89% of the racial and ethnic differences explained by lender fixed effects in the home purchase sample, respectively. Other lender features, such as type of lender based on regulatory agency (e.g., commercial or state chartered bank), the lender's securitization patterns and the lender's share of adjustable rate loans have more limited explanatory power. In fact, after controlling for ex post foreclosure risk, there is no significant relationship between high-cost lending and either regulatory agency or securitization patterns, and the importance of the share of adjustable rate loans provided by each lender is substantially reduced.

The explanatory power of our new measure of lender foreclosure risk holds in several alternative specifications. For example, creating the foreclosure risk measure from subsamples composed only of mortgages issued to white borrowers or only mortgages securitized through the government sponsored enterprises yields similar results. These findings imply that the unobserved riskiness of individual lenders is broad based, that is, that the unobserved riskiness of a lender's portfolio of loans is not a solely a function of the unobserved attributes or actions of its minority borrowers or confined to the

[5] Significant racial and ethnic differences in loan-pricing 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.

[6] See also Bayer, Ferreira, and Ross (2014).

[7] 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 between origination and March 2009 for our sample of mortgages. Then we control for the resultant lender fixed effect estimate in the high-cost lending model using a split-sample instrumental variables strategy (Case and Shiller 1989; Angrist and Krueger 1995).

types of risky loan products that were traditionally financed by nonagency securitization.[8]

Which borrowers are more likely to be served by these high foreclosure risk lenders? Race, ethnicity, credit score and high loan-to-value ratios stand out as the most important factors for explaining borrower allocation to high-risk lenders. The strong sorting of borrowers over credit score and LTV, both of which are predictive of foreclosure, implies that high-risk lenders serve a segment of the mortgage market that is riskier on the basis of both observed and unobserved factors. Interestingly, the relatively modest effects of neighborhood demographic composition and poverty rates imply that minority status is significantly more important than geography in explaining the sorting of borrowers into high-risk lenders.

We close our analysis by examining heterogeneity in the estimated racial and ethnic differences in the likelihood of receiving a high-cost loan across individuals and lenders, yielding two important findings. First, for African American borrowers, significant across- and within-lender differences in the likelihood of receiving a high-cost loan exist even for borrowers with relatively unblemished credit records and low-risk loans. Second, the remaining within-lender racial and ethnic differences are concentrated at the highest foreclosure risk lenders.

In this way, a lender's ex post foreclosure risk emerges as the leading factor in accounting for both the across and within lender components of the observed racial and ethnic differences in the incidence of high-cost mortgages. An important limitation of our analysis is that we are unable to draw firm conclusions about why the market is segmented on the basis of race, ethnicity and unobserved risk in this way. The observed sorting might be driven, for example, by the differential demand of risky borrowers for mortgage products only available at certain lenders, historical market concentration of minority borrowers combined with a lack of price based shopping by those borrowers or by the active steering of minority borrowers to high-risk lenders or to lenders with high-cost products. However, we can rule out the neighborhood of borrowers as the major reason behind this concentration. We are also unable to determine whether the increased incidence of high-cost lending is completely justified in equilibrium by the corresponding increase in unobserved risk that high foreclosure risk lenders anticipate. While additional research is necessary to address these open issues, our analysis clearly establishes the dominant role of high foreclosure risk lenders in explaining racial and ethnic differences in high-cost mortgage lending, making clear where such future research or policy oversight should be focused.

Our analysis complements and extends a long line of prior research that has studied across and within lender racial price differentials in the mortgage

[8] In Section 4 we also rule out heterogeneous foreclosure policies of lenders as a potential explanation by using delinquencies instead of foreclosures to predict risk and find similar effects for the ex post delinquency risk.

market. Our estimates of the substantial role of sorting across lenders in explaining racial differences in the likelihood of receiving a high-cost loan are in line with the results using primarily HMDA data by Avery, Canner, and Cook (2005), Avery, Brevoort, and Canner (2007), and Bhutta and Ringo (2014).[9] Similarly, Nelson (2010) and Courchane (2007) estimate small, if any, within-lender racial differences after conditioning on standard underwriting variables in a convenience samples of lenders.[10] To our knowledge, our paper is the first to examine the relative role of a range of underlying lender attributes, in particular ex post foreclosure risk, in accounting for the strong across-lender racial and ethnic price differentials.[11] Our ability to calculate and study the impact of this measure of foreclosure risk is a key contribution of our study, made possible by the new linked data set that we have assembled.

1. 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 for which 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.[12] 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

[9] Bhutta and Ringo (2014) also include credit profile information merged into HMDA.

[10] In addition, Courchane and Nickerson (1997) and Black, Boehm, and DeGennaro (2003) examine data on overages using specialized samples of lenders and found very small, if any, within-lender differences between white and minority borrowers in the incidence of overages.

[11] Our results are also related, but not strictly comparable, to recent work by Bocian, Ernst, and Li (2008), Haughwout, Mayer, and Tracy (2009), and Ghent, Hernández-Murillo, and Owyang (2014), who examine racial and/or ethnic differences using proprietary data on subprime mortgages only. Bocian, Ernst, and Li (2008) found large racial differences in the incidence of rate spread or high-cost loans for a broad sample of subprime loans. Haughwout, Mayer, and Tracy (2009) find statistically insignificant differences for 2/28 adjustable rate mortgages, while Ghent, Hernández-Murillo, and Owyang (2014) estimate significant differences in the home purchase sector of the market and for loans issued by nondepository institutions. Our analysis is based on a broader sample of all mortgages, but unfortunately, we are unable to observe loan product types beyond whether the loan has a fixed or adjustable rate. Prior to controlling for foreclosure risk, we find a higher incidence of high-cost loans at both nondepository lenders and at lenders with a high share of adjustable rate loans, but only the lender control for share adjustable rate loans explains the observed racial differences.

[12] See Bayer, Ferreira, and Ross (2016) for more details regarding the data set.

and Dataquick data are then merged based on year, loan amount, name of lender, state, county, and census tract. We obtain high-quality matches for approximately 50% of our HMDA sample.[13]

We use the HMDA rate spread variable to create a dummy variable that is one whenever the APR on the mortgage is at least three percentage points above the interest rate on treasury securities. This threshold was chosen by federal regulators to capture the high interest rates often seen in the subprime sector in the late 1990s and early 2000s. 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 prepaid (Avery, Brevoort, and Canner 2007). Our core results are robust to analyses that adjust the high-cost loan threshold by year to keep the share of high-cost loans constant over time anchored to 2004, which had the lowest share.

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. 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 1,000 borrowers associated with rate spread loans in each year and site (800 for 2004) and 2,714 borrowers (2,286 for 2004) from the non-rate-spread sample in each year and site. Weights are developed based on the probability of selection, and they are initialized so that each site receives equal weight in the pooled sample.[14]

This sample is provided to a major credit reporting agency who matches the name and address of each borrower and coborrower 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 premortgage archive is 81.4% and 84.5% in the home purchase and refinance samples, respectively. For years following the mortgage, the match rate rises

[13] The key factor limiting the match rate is the lender name because the lender of record in the local assessor's data often differs from HMDA respondent. Less restrictive match criteria can yield a match rate closer to 80%.

[14] The sampling is explicitly based on eight 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, site, and year receive equal weight.

by 4 to 5 percentage points. In many cases, these individuals may not have had sufficient information on record to provide a credit score when the lender requested a report from the credit reporting agency, in which case lack of a score matches the information that the lender would have had when approving and pricing the loan, but lenders can enter by hand additional information that is not available to us such as social security number or previous addresses. Bayer, Ferreira, and Ross (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 Table A1.

Table 1 shows the weighted means for our final home purchase and refinance subsamples that were successfully merged to premortgage credit report data.[15] 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, whether 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 borrower's (or coborrower's if the borrower is unavailable) Vantage or credit score is drawn from the credit report data from the March 31st prior to the mortgage origination.[16] 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 coborrowers in the credit history files.

15 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.

16 The Vantage Score is a proprietary credit score developed by the credit reporting agencies as an alternative to the traditional FICO index of credit score.

Table 1
Descriptive statistics for control variables from HMDA, Dataquick, and credit data sets

	Purchase sample		Refinance sample	
	Mean	SD	Mean	SD
HMDA data				
Rate spread	0.16	0.36	0.18	0.38
American Indian	0.00	0.05	0.00	0.05
Asian	0.09	0.29	0.08	0.27
African American	0.09	0.29	0.13	0.34
Hispanic	0.19	0.39	0.18	0.39
White	0.61	0.48	0.59	0.49
Male	0.64	0.47	0.65	0.47
Female	0.34	0.47	0.34	0.47
Loan amount (in 1,000s)	292.9	217.2	267.1	206.0
Applicant income (in 1,000s)	109.3	113.1	102.7	111.4
Coborrower present	0.38	0.48	0.50	0.50
Jumbo loan	0.28	0.45	0.19	0.39
Tract median income (in 1000s)	60.1	23.4	60.1	23.3
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 times 200	0.88	0.38	0.89	0.34
Dataquick data				
Loan-to-value ratio >=0.80	0.80	0.40	0.51	0.50
Loan-to-value ratio >= 0.95	0.37	0.48	0.36	0.48
Subordinate lien	0.43	0.49	0.01	0.13
First lien adjustable rate	0.50	0.50	0.47	0.49
Condo	0.21	0.41	0.13	0.34
Mobile	0.00	0.03	0.00	0.03
Single family	0.77	0.41	0.84	0.36
Lot size (sq. ft. in 1,000s)	15.5	102.9	16.7	165.0
Unit square feet (in 1,000s)	2.01	27.78	1.90	20.91
Number of bathrooms	2.27	5.99	2.22	0.98
Number of bedrooms	3.08	6.83	3.14	1.14
Number of stories	1.61	1.73	1.55	1.32
Units in building	1.60	18.4	1.52	18.6
Credit data				
Vantage score	782	105	773	109
Subprime credit score, Vantage < 701	0.24	0.43	0.27	0.44
Mortgage payment-to-income ratio >0.28	0.38	0.48	0.39	0.48
Debt payment-to-income ratio >0.36	0.36	0.48	0.45	0.49
Borrower Age	43.7	11.4	49.6	11.0
Sample size	94,699		91,242	

This table presents means and standard errors for the key control variables that are available in the HMDA, Dataquick and Credit Repository data sets in panels 1, 2 and 3, respectively. Statistics are reported separately for the home purchase and refinance regression samples. Means and standard deviations are weighted based on the inverse of the probability of selection into the sample provided to the credit repository. Observations are only included in the sample if a credit score is observed and the identity of the lender was identified. Note that the lender respondent identification number is suppressed for lenders with less than 10 loans in our sample.

2. 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:

$$y_i = \alpha X_i + \varepsilon_i \tag{1}$$

Table 2
Alternative models of the likelihood of a rate spread or high cost loan

Dependent variable: Indicator for a rate spread loan

Variable names	HMDA	Dataquick	Credit data	Subprime	Lender FE
	(1)	(2)	(3)	(4)	(5)
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)	(0.010)	(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-square	0.18	0.29	0.36	0.41	0.60
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-square	0.11	0.19	0.33	0.35	0.55

This table presents OLS estimates for five models of the likelihood of a rate spread loan, sequentially adding five distinct groups of control variables to the model. Descriptive statistics for the set of variables included in each group are reported in Table 1. Column (1) includes controls for just the variables available in the Home Mortgage Disclosure Act data; Column (2) includes additional controls made available from assessor's data including Combined Loan-to-Value Ratio and the presence of subordinate lien; Column (3) adds 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) adds lender fixed effects. All specifications are based on the sample of loans originated from 2004-2007, and standard errors are clustered at the lender level.
Standard are errors in parentheses.
***p<0.01, **p<0.05, *p<0.1.

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 their 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.[17] All regressions are estimated using within metropolitan area weights based on the inverse of the selection probability, as discussed earlier.[18]

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 amount dummy, the

[17] We focus on the linear probability model due to concerns about the bias arising from estimating nonlinear 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).

[18] Since our sample is based on choice based sampling and oversamples high-cost loans as recommended by Clarke and Courchane (2005), unweighted regressions will have biased intercepts associated with each strata, that is, the coefficients on the minority group dummy variables. The use of sampling weights eliminates any bias caused by the sampling procedure, see Clarke and Courchane (2005) and Manski and Lerman (1977).

census tract attributes and year-by-site fixed effects. The second 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 lending to subprime borrowers, identifying borrowers with Vantage scores below 701 as subprime borrowers.[19] This subprime borrower dummy variable is then interacted with variables based on key thresholds of loan-to-value ratio, debt-to-income ratio and mortgage payment to income and with dummy variables for whether there are subordinate liens and whether the primary lien is adjustable rate.[20] The fifth column adds lender fixed effects ω_j:

$$y_{ij} = \alpha X_i + \omega_j + \varepsilon_{ij} \tag{2}$$

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 relative to white borrowers, respectively, for a home purchase mortgage when conditioning only on the standard controls available in HMDA. The racial and ethnic differences are 11.8 and 4.6 in the refinance market. 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 9.0 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% for racial differences and 50%-70% for ethnic differences.[21] The inclusion of additional subprime borrower

[19] The credit reporting agencies that developed the Vantage score algorithms describe scores below 701 as nonprime. Further, a Vantage score of 701 is comparable to a FICO score of 660, a common FICO threshold for defining a subprime borrower, in that in both cases approximately 30% of individuals had credit scores below these thresholds during our sample period. Subprime borrowers make up about 25% of our weighted home purchase sample.

[20] 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% of our weighted home purchase sample. The debt to income thresholds used are 0.36 and 0.45 with 13% and 42% 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% in the middle and upper bins.

[21] 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, the loan-to-value ratio and the housing-expense-to-income ratio in the expected directions.

controls in Column 4 has little impact on the estimated differences.[22] 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 103.4% and 78.2%, respectively, of the incidence of rate spread loans for white borrowers. Comparing the results of Column 3 or 4 to those of Column 1 reveals both (1) 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 (2) 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 in the subprime or baseline model to 3.6- and 2.4-percentage-point differences for African Americans and Hispanics, 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 borrower access to traditional lenders and/or systematic borrower 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. While not directly comparable, our estimates of within lender racial and ethnic differences are in line with the findings on loan rejection rates in the Munnell et al. (1996) study of underwriting discrimination in Boston, which included lender fixed effects in a sample of loan applications from many lenders in a common market.

In papers that have studied the cost of credit directly, Avery, Canner, and Cook (2005) and Avery, Brevoort, and Canner (2007) using 2004 and 2005/2006 HMDA data, respectively, and Bhutta and Ringo (2014) using 2013 HMDA data matched to a 1% 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 Cook (2005) nor Avery, Brevoort, and Canner (2007) determined 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 are primarily the result of the systematic selection of African American

22 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.

and Hispanic borrowers into lenders who tend to issue high-cost mortgages.[23] Unfortunately, neither we nor Bhutta and Ringo (2014) observe some of the key loan attributes that were sometimes associated with high-cost loans during the run up to the recent crisis, such as no documentation of income or initial teaser interest rates combined with aggressive rate resets and prepayment penalties. Thus, it is fundamentally impossible for us to directly determine whether the estimated across or within lender racial and ethnic differences are associated with demand for these products.

2.1 The role of specific lender attributes

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 replace those lender fixed effects with several lender attributes: (1) the type of lending institution, (2) the share of mortgages securitized, (3) the share of securitized mortgages sold to each type of purchaser, (4) the share of loans that are adjustable rate, and (5) a measure of lender ex post foreclosure risk. Given the substantially smaller differences observed in the refinance sample, we limit this investigation to the home purchase sample.[24] Lenders are classified as national banks, commercial banks, state chartered banks, savings and loans, credit unions, or nondepository mortgage banks.[25] The types of purchasers considered include 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.[26]

To create the proxy for lender foreclosure risk, we use our sample of home purchase mortgage originations. The proxy is based on estimated lender fixed effects, $\hat{\phi}_j$, 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:

$$f_{ij} = \gamma X_i + \phi_j + u_{ij}. \tag{3}$$

The ever-foreclosed 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. The resultant high-cost loan model specification is

$$y_{ij} = \alpha X_i + \beta Z_j + \hat{\phi}_j + \varepsilon_{ij}, \tag{4}$$

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

[24] Comparable estimates for the refinance sample are presented in the Online Appendix (Tables A4–A9).

[25] The type of lender is identified using the regulator provided by the agency code variable in HMDA.

[26] The variables related to securitization shares are calculated using the full sample of HMDA loans between 2004 and 2007 for the seven sites in our credit history sample. These variables are merged into our sample using the respondent ID leading to a slightly reduced sample size. The share adjustable rate is calculated within our matched sample because whether a loan is adjustable rate is not observed in HMDA.

where the lender characteristics are represented by Z in Equation (4) and $\hat{\phi}_j$ is the predicted fixed effect from Equation (3). Table A2 in the on-line appendix presents means and standard deviations for each lender variable in the weighted samples of home purchase and refinance loans.

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 (3) as a measure of lender foreclosure risk when estimating the rate spread model shown in Equation (4). Because only a 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. Therefore, to consistently estimate the rate spread model with lender foreclosure risk, we use a split sample instrumental variables strategy.[27]

Specifically, we restrict our sample to borrowers at lenders with at least 10 loans in our home purchase sample, and then we 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 (3) 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 (4), and the fixed effect estimate from the hold-out sample is used as an instrument. Standard errors are bootstrapped by sampling lenders with replacement.[28]

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 includes the split-sample IV estimates for the specification that includes ex post lender foreclosure risk instead of lender fixed effects.[29] The next three columns report the estimates for models that replace lender fixed effects with either dummy variables for the type of lender, variables on securitization and purchaser shares and the share of loans that are adjustable rate, respectively. The last column includes the split-sample IV estimates for the specification that includes all lender controls above. The lender foreclosure risk estimates shown in Column 3 imply that foreclosure risk explains the vast majority (73% and 89%) of the racial and ethnic

[27] 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 unusually risky loans due to sampling error, 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 estimated foreclosure risk unless the foreclosure fixed effect estimate is based on a separate sample of loans, that is, a split sample IV hold-out sample.

[28] The 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 mean estimates of racial and ethnic differences conditional on foreclosure risk. Standard errors presented are the average of bootstrapped standard errors for five different hold-out samples.

[29] 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.

Table 3
Lender attributes and the likelihood of a rate spread loan

Dependent variable: Indicator for a rate spread loan

	Home Purchase Sample						
Variable names	Baseline (subprime) model	Lender fixed effects	Baseline model + Lender foreclosure risk	Baseline model + Agency code	Baseline model + Securitization shares	Baseline model + Share adjustable rate loans	Baseline model + Lender attributes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Asian	0.010**	0.006*	0.011***	0.012***	0.014***	0.006	0.012
	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.008)
African American	0.087***	0.036***	0.050***	0.085***	0.074***	0.065***	0.045***
	(0.010)	(0.007)	(0.008)	(0.008)	(0.007)	(0.006)	(0.008)
Hispanic	0.069***	0.024***	0.029***	0.064***	0.058***	0.047***	0.027***
	(0.009)	(0.004)	(0.007)	(0.009)	(0.009)	(0.006)	(0.006)
Foreclosure risk by lender			3.225***				3.039***
			(0.288)				(0.436)
Lender share adjustable rate loans						0.462***	0.124**
						(0.065)	(0.060)
Lender type (national banks omitted)							
Commercial banks				0.055**			−0.011
				(0.022)			(0.051)
State chartered bank				0.147			−0.033
				(0.113)			(0.059)
Savings and loan				0.053***			−0.045*
				(0.020)			(0.023)
Credit union				0.008			0.042
				(0.021)			(0.051)
Mortgage banks				0.114***			−0.021
				(0.022)			(0.031)
Share of mortgages resold					0.048		−0.047
					(0.043)		(0.038)
Loan buyer share (GSE omitted)							
FHA					0.260**		−0.174
					(0.115)		(0.144)
Private securitization					0.473***		−0.089
					(0.109)		(0.072)
Commercial/Savings bank					−0.025		−0.060
					(0.056)		(0.068)
Ins., CU, Mort. bank					0.184***		−0.018
					(0.060)		(0.047)
Affiated lender					0.094**		−0.075
					(0.038)		(0.031)
Other					0.211***		−0.041
					(0.035)		(0.035)
Observations	94,699	94,699	46,532	94,699	89,885	94,699	44,279
R-square	0.41	0.60	0.48	0.43	0.46	0.47	0.50

This table presents estimates based on the baseline (subprime) model for the home purchase sample. OLS estimates from the baseline model (Table 2, Column 4) are shown in Column (1), and estimates for the lender fixed effects model (Table 2, Column 5) are shown in Column (2) . 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 Columns (4)-(6) add controls for the type of lender based on the lender's regulator, the share of loans sold and the share for each purchaser type on the secondary market and the share of loans that are adjustable rate, respectvely. Column (7) includes all controls. Columns (3) and (7) are estimated using split sample IV. The split sample IV estimates are based on a subsample with at least ten 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 Columns (3) and (7), where they are bootstrapped by sampling lenders with replacement with standard errors based on the averages of estimates for five different draws of the hold-out sample. For the split-sample IV, R-squares are based on averages across the 20 hold-out samples.
Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

differences explained by lender fixed effects in the home purchase market. Both lender type and the securitization variables have substantial explanatory power, with mortgage banks (nondepository lenders) and lenders that sell a substantial fraction of loans to private securitizers having an especially high incidence of rate spread loans, but those variables do little to explain racial and ethnic differences. The race and ethnicity estimates are essentially unchanged in the lender type model and reduced by only about 15% in the securitization share model. In the next column, lender share of adjustable rate loans both explains the likelihood of a high-cost loan and erodes racial and ethnic differences, but the share of differences explained is smaller than the share explained by the foreclosure risk variable. Finally, comparable lender foreclosure risk effects arise for a model that includes all of the lender controls. Notably, the lender type and securitization controls have little additional predictive power once foreclosure risk is included in the analysis, and the effect of lender share adjustable declines by 73%.

2.2 Variation across metropolitan sites

Table 4 presents the estimated results separately for each metropolitan housing market for three main specifications: (1) a baseline model comparable to Column 4 of Table 2, (2) a model controlling for an overall lender foreclosure risk measure comparable to Column 3 of Table 3, and (3) 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 foreclosure fixed effects for each lender across all sites as a regressor because the much smaller site specific sample sizes raise concerns about precision and small sample bias if we were to either use the split sample IV estimator or use site specific lender foreclosure fixed effects.[30] The columns represent in order Chicago, Cleveland, Denver, Los Angeles, Maryland counties, Miami-Palm Beach Corridor, and San Francisco Bay Area.

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 FEs. In the home purchase market without lender FE's, the significant differences range between 5.9 and 11.8 for African Americans and 6.2 and 7.5 for Hispanics. The inclusion of lender FEs lowers these differences to ranges of 1.8 to 5.1 and 1.9 to 3.3, respectively. Further, adding the controls for lender foreclosure risk to the baseline model again results in racial and ethnic differences that are closer to the differences in the lender fixed effect model with ex post foreclosure risk explaining between 40% and 89% of the across lender racial and ethnic

[30] Further, bias from measurement error leads to conservative estimates, and concerns about bias from sampling error affecting both high-cost lending and foreclosure are substantially mitigated because the lender fixed effect estimates are based on the entire sample, not just the loans in a given site.

Table 4
Estimates for each metropolitan area, Home purchase sample

Dependent variable: Indicator for a rate spread loan

Variable names	Chicago	Cleveland	Denver	Los Angeles	Maryland	Miami	San Francisco
Baseline (subprime) mode							
Asian	−0.016	0.003	0.048***	0.003	−0.004	0.043*	0.019***
	(0.009)	(0.010)	(0.014)	(0.009)	(0.011)	(0.022)	(0.006)
African American	0.095***	0.066***	0.083***	0.059***	0.085***	0.118***	0.062***
	(0.016)	(0.020)	(0.014)	(0.011)	(0.011)	(0.015)	(0.015)
Hispanic	0.063***	0.01	0.062***	0.063***	0.075***	0.072***	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-square	0.42	0.44	0.41	0.43	0.40	0.42	0.40
Lender foreclosure risk (reduced form)							
Asian	−0.014	−0.006	0.036***	0.005	−0.005	0.032	0.011*
	(0.009)	(0.011)	(0.011)	(0.010)	(0.009)	(0.020)	(0.005)
African American	0.068***	0.034**	0.051***	0.035***	0.057***	0.075***	0.039***
	(0.011)	(0.017)	(0.013)	(0.008)	(0.008)	(0.012)	(0.011)
Hispanic	0.034***	0.011	0.046***	0.033***	0.035***	0.047***	0.034***
	(0.012)	(0.018)	(0.010)	(0.007)	(0.013)	(0.010)	(0.011)
Foreclosure risk	2.764***	2.879***	2.273***	2.527***	2.936***	2.464***	2.111***
	(0.230)	(0.241)	(0.205)	(0.208)	(0.202)	(0.238)	(0.276)
Observations	12,752	11,976	13,182	13,515	13,693	13,395	14,293
R-square	0.52	0.54	0.50	0.54	0.54	0.52	0.51
Lender fixed effects model							
Asian	−0.014*	0.001	0.038***	0.002	−0.011	0.024	0.006
	(0.008)	(0.008)	(0.009)	(0.010)	(0.007)	(0.018)	(0.005)
African American	0.039***	0.030*	0.033**	0.020**	0.039***	0.051***	0.018*
	(0.010)	(0.017)	(0.013)	(0.010)	(0.009)	(0.011)	(0.010)
Hispanic	0.019**	0.011	0.022***	0.021***	0.024*	0.033***	0.022***
	(0.008)	(0.016)	(0.008)	(0.006)	(0.011)	(0.008)	(0.007)
Observations	12,752	11,976	13,182	13,515	13,693	13,395	14,293
R-square	0.59	0.62	0.60	0.64	0.64	0.60	0.65

This table presents estimates for individual sites in each column for the home purchase sample. Panel 1 presents the baseline (subprime) model; Panel 2 presents reduced form models that control for estimated lender foreclosore fixed effects from the ever-foreclosed regression; and Panel 3 presents the lender fixed effects estimates. All models in this table are estimated using OLS and standard errors are clustered at the lender level.
Standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

differences.[31] Notably, the estimates on the foreclosure risk range between 2.1 and 2.9, which is comparable to the reduced-form estimate of 2.6 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 tend to have both a disproportionate share of minority borrowers and unusually high ex post foreclosure rates.

31 The minimum portion explained arises in Denver for Hispanics, and the maximum portion explained arises in Cleveland for African Americans.

3. Understanding the Role of High-Risk Lenders

In this section, we present a series of results based on alternative measures of lender foreclosure risk. First, based on the significant heterogeneity in racial differences across loan types identified by Ghent, Hernández-Murillo, and Owyang (2014), we identify two proxies for unobserved product types, such as low documentation loans or loans involving rate resets and prepayment penalties. The two proxies are the type of securitization (agency, nonagency, and held in portfolio) and whether the loan is a fixed or adjustable rate mortgage because a large majority of loans with subprime or Alt-A attributes are adjustable rate mortgages and are securitized outside of the traditional government sponsored enterprises (nonagency). We estimate the foreclosure models controlling for lender by loan type t (either fixed/adjustable or securitization channel) fixed effects (ϕ_{tj}) and restricting our samples to lender by type combinations where at least ten home purchase loans are present.[32]

$$f_{itj} = \gamma X_{it} + \phi_{tj} + u_{itj} \tag{5}$$

We then create unique measures of foreclosure risk for each loan type at each lender. These controls are then included in the high-cost loan model.

$$y_{itj} = \alpha X_{it} + \beta Z_j + \hat{\phi}_{tj} + \varepsilon_{itj} \tag{6}$$

The estimates from these models are shown in Table 5. The first column presents the overall lender foreclosure risk model estimates from Table 3, Column 3. The next two columns present the estimates controlling for lender by securitization channel and lender by adjustable/fixed rate foreclosure risk. If the role of the lender specific foreclosure rate in explaining high-cost lending varied significantly across these categories of loans, then the coefficient in Column 1 should have been biased by measurement error, and the estimates on foreclosure risk in Columns 2 and 3 should be larger and explain more of the racial and ethnic differences.[33] However, the estimates in Columns 2 and 3 are practically identical to the estimates in Column 1. Similarly, in Columns 5 and 6, we present models that control for lender by securitization channel and lender by adjustable/fixed rate fixed effects with the original fixed effect estimates being presented in Column 4, and the results across the three columns are very similar.

Next, we estimate models of lender specific foreclosure risk based only on subsamples of loans. To conduct this analysis, we further restrict the data to lenders with at least ten home purchase loans of this specific type T, and we re-estimate the "ever-foreclosed" model controlling for lender-by-type fixed

[32] These sample restrictions and similar restrictions below never have any impact on the core estimates in our paper.

[33] This measurement error would not be addressed by the split sample IV estimation strategy because the measurement error would be present in both the variable and the instrument, that is, both the main and hold-out sample estimates of foreclosure risk.

Table 5
Lender risk measures and fixed effects by type of loan

Dependent variable: Indicator for a rate spread loan

	Home purchase sample					
	Lender foreclosure risk by			Fixed effect models by		
Variable names	All loans	Securitization	Fixed vs. adjustable	Lender	Lender by securitization	Lender by fix/adjust
	(1)	(2)	(3)	(4)	(5)	(6)
Asian	0.011***	0.011	0.013*	0.006*	0.006*	0.007**
	(0.004)	(0.014)	(0.007)	(0.003)	(0.004)	(0.003)
African American	0.050***	0.049***	0.051***	0.036***	0.032***	0.034***
	(0.008)	(0.014)	(0.013)	(0.007)	(0.006)	(0.006)
Hispanic	0.029***	0.026***	0.026***	0.024***	0.021***	0.023***
	(0.007)	(0.011)	(0.010)	(0.004)	(0.004)	(0.004)
Lender foreclosure risk (based on all loans)	3.225***					
	(0.288)					
Lender foreclosure risk by securitization type		3.123***				
		(0.547)				
Lender foreclosure risk by rate type			3.031***			
			(0.422)			
Observations	46,532	46,083	45,728	94,699	94,699	94,699
R-square	0.48	0.48	0.48	0.60	0.62	0.62

This table presents estimates of the likelihood of a rate spread loan for alternative models. Column (1) repeats the estimates for the lender foreclosure risk model where foreclosure risk is based on all loans (Column 3, Table 3). The next two columns present new lender foreclosure risk models where a distinct risk measure is estimated for each lender and market segment. In Column (2) the market segments are based on government/agency securitization, nonagency securitization and unsecuritized, and in Column (3) separate risk measures are estimated for the fixed and the adjustable rate loans at each lender. For these three columns, the samples are restricted to assure a minimum of ten loans per cell. The final three columns present fixed effect models by lender (Column (4)), by lender-by-securitization type (Column (5)), and by lender-by-fixed versus adjustable rate loan (Column (6)). Standard errors are bootstrapped for the first three columns and clustered by lender for the last three columns. Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

effects for both the hold-out and regression samples just as in Equation (5). We then use the lender fixed effects for that type of borrower for all loans at the lender.

$$y_{ij} = \alpha X_i + \beta Z_j + \hat{\phi}_{Tj} + \varepsilon_{ij}, \tag{7}$$

where $T \in t$.

First, we base lender foreclosure risk on only loans to white borrowers. The goal of this analysis is to examine whether these effects are associated with specific lenders specializing in providing high-risk loans or products to all borrowers regardless of race and ethnicity, or whether the results presented in Table 3 are driven primarily by African American and Hispanic loans at these lenders. Table 6, Column 1, again contains the estimates of the baseline model from Table 3, Column 3, and Column 2 shows the estimates for a model that uses foreclosure risk associated with white loans only. The estimated 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, and if anything white foreclosure risk does a better job than overall risk of

Table 6
Alternative lender risk measures

Dependent variable: Indicator for a rate spread loan

	Home purchase sample				
		Lender foreclosure risk only for			
Variable names	Baseline, all loans	White loans	Agency securitized loans	Conditional on rate spread loan	30-day delinquency
	(1)	(2)	(3)	(4)	(5)
Asian	0.011***	0.012	0.009	0.011***	0.010***
	(0.004)	(0.008)	(0.009)	(0.004)	(0.004)
African American	0.050***	0.044***	0.039**	0.060***	0.052***
	(0.008)	(0.011)	(0.019)	(0.010)	(0.009)
Hispanic	0.029***	0.017*	0.021***	0.038***	0.028***
	(0.007)	(0.009)	(0.008)	(0.009)	(0.006)
Lender foreclosure risk based on all loans	3.225***				
	(0.288)				
Lender foreclosure risk based on white loans		4.514***			
		(1.082)			
Lender foreclosure risk based on agency loans			2.420***		
			(0.661)		
Lender foreclosure risk conditional on rate spread				3.645***	
				(0.609)	
Lender 30-day delinquency risk					2.204***
					(0.264)
Observations	46,532	45,353	24,876	46,532	46,532
R-square	0.48	0.41	0.21	0.42	0.46

This table presents estimates of the likelihood of a rate spread loan for alternative models. Column (1) presents the lender foreclosure risk model (Column 3, Table 3). Columns (2) and (3) report estimates from models that control for alternative measures of lender foreclosure fixed effects based only on the lender's white loans or government securitized loans, respectively. These models use samples restricted to lenders with ten or more white or government securitized loans, respectively. Column (4) reports estimates using an alternative model of ex post foreclosure risk based on an ever-foreclosed model that controls directly for whether the borrower has a high cost or rate spread loan. Column (5) presents estimates using an alternative dependent variable to estimate the first stage lender fixed effects: ex post lender risk of 30- to 180-day delinquency. Columns (4) and (5) use the original home purchase sample of lenders with at least ten 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.

explaining racial differences in high-cost lending.[34] This suggests that these lenders specialize in providing high-risk loans or loans to high-risk borrowers for the market as a whole, not just to African American and Hispanic borrowers.

In Column 3 of Table 6, we present results basing lender foreclosure risk on only agency-securitized loans. As described in the discussion of Table 5 above, we believe that these loans are less likely to be associated with mortgage products, like low doc and rate reset loans, which are not identified in our data. Therefore, we use this model to assess whether foreclosure risk can still explain racial and ethnic differences in high-cost lending even when the measure of foreclosure risk is based on loans that are least likely to have particularly risky attributes. The Column 3 estimates based on agency loan foreclosure risk are

[34] Additional noise in the white-lender fixed effect estimates (due to a lower incidence of foreclosure) increases the standard errors on the foreclosure risk estimates. Similar results arise for models that include the additional lender controls based on agency code, securitization patterns, and share adjustable rate loans.

noisier because many of the large lenders with nonagency securitized loans have no agency securitized loans, but the resultant racial and ethnic differences are basically the same as the white borrower foreclosure risk model in Column 2.[35]

Next, in Table 6, Column 4, we present estimates based on lender fixed effects taken from an ever-foreclosed model that includes the high-cost loan variable as an additional control. The resultant measure of ex post foreclosure risk is now conditional on whether an individual loan issued by the lender was high cost. If our results are driven by factors like loan product that are observed by the lender, but are not observed in our sample, the inclusion of this control should erode the ability of foreclosure risk to explain racial and ethnic differences in high-cost lending, to the extent that lenders price these factors into the cost of credit. The estimated racial and ethnic differences increase somewhat compared to the model using overall lender foreclosure risk. These estimates provide evidence that a modest portion of the across lender racial and ethnic differences may be due to loan product differences across lenders.

A final possible explanation for the results regarding lender foreclosure risk is that high-cost lenders also happen to be lenders that aggressively enter the foreclosure process as loans become seriously delinquent. To rule out this possibility, we estimate a model of ever received a 30- to 180-day delinquency using the original sample of all lenders with at least ten loans in our sample. The estimated lender fixed effects from this model are used as an alternative measure of lender risk again using split-sample IV. These results are shown in Table 6, Column 5. The racial and ethnic differences are similar to our benchmark model that controls for foreclosure risk, and the coefficient on the proxy for delinquency risk is large and statistically significant. These findings suggest that across lender racial and ethnic differences arise from borrower delinquency and default patterns at these lenders rather than lender foreclosure behavior.

3.1 Sorting across lenders

As a final exercise designed to better understand the implications of these high-risk lenders for mortgage markets, we examine which types of borrowers/mortgages tend to receive mortgage credit from these lenders. In particular, Table 7 presents estimates of a model that relates the lender foreclosure risk to which a borrower is exposed $(\hat{\phi}_j)_i$ to borrower and loan attributes X_i including (1) borrower socio-demographic variables and credit score, (2) neighborhood attributes, and (3) loan attributes.[36]

$$\left(\hat{\phi}_j\right)_i = \theta X_i + v_i \tag{8}$$

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

---

[35] We also investigate models that use foreclosure risk associated with fixed rate loans, but the estimates on foreclosure risk and the race and ethnicity dummies were far too noisy to be informative.

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

Table 7
Sorting over lender foreclosure risk

Dependent variable: Overall ex post foreclosure risk of the borrower's lender

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)
Coborrower 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)
Logarithm of 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)
Loan risk factor	
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	−0.004**
	(0.002)
Housing expense-to-income ratio above 0.33	0.010***
	(0.002)
Debt-to-income ratio above 0.36	−0.015***
	(0.002)
Debt-to-income ratio above 0.45	−0.012***
	(0.002)
Observations	94,481
R-square	0.23

This table reports estimates of an OLS regression of the standardized measure of the ex post foreclosure risk of a borrower's lender on the borrower, neighborhood, and loan attributes in the home purchase sample. Standard errors are clustered at the census tract level.
Standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

ten 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 $\hat{\phi}_j$, the results presented in Table 7 cannot be driven by the ability of these 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.[37]

The model includes controls for the following socio-economic variables: race, ethnicity, age, gender, presence of a coborrower (a proxy for marital status), and the logarithm of income. The race and ethnicity correlations are quite large, approximately 4% of a standard deviation for African Americans and 5% for Hispanics. While many estimates are statistically significant, the estimates on the other demographic variables tend to be smaller, ranging between 1.5% and 2% of a standard deviation.[38] Similarly, the estimates on the tract composition variables, which include racial and ethnic composition, percent households in poverty, percent housing units owner-occupied and rent-to-price ratio as a proxy for equity risk, are quite modest in magnitude.[39] A standard deviation increase in any neighborhood composition variable never has an effect larger than 1% of a standard deviation in lender foreclosure risk, and the doubling of the present value of rents relative to housing price is associated with an increase of only a 1.3%t of a standard deviation in lender riskiness.[40]

On the other hand, mortgage risk factors like the borrower's vantage score and the loan-to-value ratio are strongly associated with borrowing from a high-risk lender. Having a subprime credit score is associated with about 10% of a standard deviation increase in lender foreclosure risk, and having an above median credit score within one's corresponding credit quality segment (prime or subprime credit score) is associated with a 3%-5% of a standard deviation reduction in lender foreclosure risk. Similarly, having a loan-to-value ratio above 0.95 is associated with 7% of a standard deviation increased exposure to

[37] 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, the sampling error bias discussed above 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.

[38] These differences are even larger when the model just controls for demographics: 5.5% and 7.1% for blacks and Hispanics compared to only 1.5% for loans without a coborrower, the next largest demographic estimate in this model.

[39] The price-to-rent ratio 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. Specifically, we assume an annual discount rate of 0.06 or a monthly discount rate of 0.005, and we multiply monthly rents by 200 prior to dividing by housing prices.

[40] 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, and the estimate on tract poverty barely moves.

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. If borrowers sort across lenders based on key risk observables, like credit score and downpayment, they presumably will sort across lenders on their unobservables, as well.[41]

Taken as a whole, the results of Tables 2–7 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.

4. Heterogeneity in Racial and Ethnic Differences

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. The coefficients on the black and Hispanic interaction terms are presented in Table 8 for home purchase mortgages.[42] 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 so that this model can interact lender foreclosure risk with race and ethnicity identifiers.[43]

Column 1 presents the interactions with three key risk variables: subprime credit score or Vantage score below 701, nonconforming loan-to-value ratio above 0.95 and a debt to income ratio above 0.45.[44] 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

41 The coefficients on the housing expense and debt-to-income ratio variables are often statistically significant, but their magnitude is relatively small.

42 The models also include interactions with the Asian dummy variable, but given the small Asian differences in high-cost lending throughout our analyses we do not present these interactions.

43 The baseline model specification includes risk variables that subsume the subprime credit score, high DTI and High LTV indicators and also includes the complete list of census tract variables included in the Column 2 interactions. As noted above, the model in Column 3 includes a direct control for lender foreclosure risk.

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

Table 8
Heterogeneity in the likelihood of a rate spread loan by race and ethnicity

Dependent variable: Indicator for a rate spread loan

Variable names	Baseline model plus interactions		
	(1)	(2)	(3)
African American	0.085***	0.077***	0.027***
	(0.013)	(0.014)	(0.010)
Hispanic	0.024**	0.016	−0.010
	(0.010)	(0.010)	(0.008)
Interactions with common loan risk factors			
African American*Subprime	0.032*	0.028*	−0.006
	(0.017)	(0.016)	(0.016)
Hispanic*Subprime	−0.028*	−0.029*	−0.007
	(0.017)	(0.017)	(0.014)
African American*High DTI ratio	0.005	0.006	0.010
	(0.009)	(0.008)	(0.011)
Hispanic*High DTI ratio	−0.007	−0.006	0.005
	(0.010)	(0.010)	(0.009)
African American*High LTV ratio	−0.015	−0.018	0.003
	(0.013)	(0.013)	(0.011)
Hispanic*High LTV ratio	0.113***	0.110***	0.062***
	(0.013)	(0.013)	(0.013)
Interactions with neighborhood characteristics			
African American*Pct poverty		0.141**	0.063
		(0.063)	(0.072)
Hispanic*Pct poverty		−0.001	0.029
		(0.067)	(0.070)
African American*Pct Black		−0.024	−0.009
		(0.019)	(0.021)
Hispanic*Pct Hispanic		0.044***	0.025
		(0.016)	(0.021)
African American*Pct owner-occupied		−0.022	−0.021
		(0.021)	(0.024)
Hispanic*Pct owner-occupied		−0.050**	−0.030
		(0.020)	(0.022)
African American*Rent-to-price ratio		0.051***	0.046***
		(0.012)	(0.013)
Hispanic*Rent-to-price ratio		0.014	0.023*
		(0.010)	(0.013)
Interactions with lender foreclosure risk			
African American*Lender foreclosure risk			0.543*
			(0.220)
Hispanic*Lender foreclosure risk			0.185
			(0.218)
Observations	94,699	94,699	46,532
R-square	0.41	0.41	0.49

This table presents estimates of models that add interactions of race and ethnicity (including the Asian identifier; estimates not shown) with key borrower, location and lender controls to the baseline (subprime) model for the home purchase sample. The first column presents the estimates interacting indicators 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 model. Column (2) also adds interactions between race and ethnicity and census tract attributes for the tract in which the housing unit is located. Census tract attributes are adjusted so that they equal zero at the weighted mean of the white borrower sample. Column (3) also interacts race and ethnicity with lender ex post foreclosure risk. The baseline model specification includes risk variables that subsume the subprime credit score, high DTI and High LTV indicators and also includes the complete list of census tract variables included in the Column (2) interactions. The model in Column (3) also includes a direct control for lender foreclosure risk. Columns (1) and (2) are estimated using OLS with standard errors clustered at the lender level, and Column (3) is estimated using split-sample IV with bootstrapped standard errors.
Standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

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 a control for lender foreclosure risk. 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 percentage points. 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.

In Column 2, we estimate a model that interacts 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. Note that we only interact share African American and share Hispanic with the African American and Hispanic identifiers, respectively.

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. To examine unexplained racial differences for low-risk borrowers, we evaluate and present in the table the race and ethnicity coefficients for a minority borrower residing in a location with the weighted average value on all tract attributes for the white subsample of borrowers. Low-risk African American borrowers in relatively advantaged neighborhoods (those occupied by typical white homebuyers) have a 7.7-percentage-point difference compared to the 8.5 difference for low-risk borrowers in Column 1, and therefore most of the racial differences persist across different neighborhood types. For Hispanics, the inclusion neighborhood interactions lead to a larger reduction in relative terms to 1.6 for low-risk borrowers in advantaged neighborhoods from 2.4 for low-risk borrowers overall. While many studies document the concentration of subprime lending in poor and minority neighborhoods (Calem, Gillen, and Wachter 2004; Mayer and Pence 2008; Reid and Laderman 2009; Fisher, Lambie-Hanson, and Willen 2010), ours is one of the few studies to document neighborhood effects after the inclusion of detailed underwriting controls. However, our analysis suggests that neighborhood at most plays only a modest role in explaining the high incidence of high-cost loans among minority borrowers.

In Column 3, 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 large, and the coefficient on the interaction with race is statistically significant. Strikingly, the significant race and ethnicity interaction coefficients reported in the first two columns of the table fall towards zero in this model with the exception of the coefficient on the interaction of race with tract rent-to-price ratio. Heterogeneity by lender foreclosure risk captures much of the heterogeneity on observables documented in Columns 1 and 2. Now examining the race and ethnicity coefficients in Column 3, we find that the racial and ethnic differences for low-risk borrowers have both fallen substantially when evaluated for average foreclosure risk lenders, racial differences to 2.7 percentage points and ethnic differences to −1.0 percentage points. Further, evaluating racial differences at the 25th percentile of lender foreclosure risk (estimated fixed effect of −0.036) yields racial differences of only 0.7 percentage points.

Therefore, based on this simple linear extrapolation, African American and Hispanic borrowers that do business with low-risk lenders appear no more likely to have high-cost loans than comparable white borrowers at these institutions. Further, these differences explain most of the concentration of racial and ethnic differences in disadvantaged neighborhoods. Taken as a whole, the results presented in Table 8 provide a picture suggesting that high foreclosure risk lenders are 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 foreclosure risk lenders, and these lenders are especially likely to provide high-cost loans to African American borrowers. A similar, but somewhat weaker, pattern emerges for Hispanic borrowers.

5. Summary and Conclusion

In this paper, we have identified large racial and ethnic differences in the likelihood of receiving a rate spread mortgage in the home purchase 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. The importance of differences across lenders is consistent with evidence that African American and Hispanic borrowers are less likely to compare prices across lenders (Courchane, Surette, and Zorn 2004; Alexandrov and Koulayev 2015). Of course, this sorting also may be influenced by lender behavior. In field studies, Ross et al. (2008) and Hanson et al. (2016) find that lenders offer more information and assistance in response to borrower inquiries from whites and in the case of Hanson et al. (2016) 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 a measure of 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. These differences cannot be explained by the lender's regulator, including whether or not the lender is a depository, the securitization patterns of the lender, the differential foreclosure patterns of minority borrowers or nonagency securitized loans, or across lender differences in the treatment of heavily delinquent loans. At the end of the day, we cannot make additional direct statements about why the market is segmented on the basis of race, ethnicity, and unobserved risk in this way. The observed sorting might be driven by a variety of reasons including, for example, product demand, differential shopping behavior, borrower sorting, lender marketing practices, or lender steering.

Nonetheless, while we cannot observe detailed information on product type, like low documentation loans or rate resets, additional analyses suggest that only a modest fraction of these explained differences can be attributed to factors observable to lenders like detailed product attributes. Rather, we find evidence of systematic sorting of higher risk individuals into these lenders based on credit score, a key observable risk factor, which suggests that these borrowers may have sorted into these lenders based on unobservable risk factors as well. Perhaps lender marketing strategies serve to sort higher risk borrowers into higher cost lenders, or borrowers have imperfect information concerning how much lenders can observe and consider unobservable risk factors when selecting lenders.

Interestingly, our analysis of heterogeneity in racial and ethnic differences demonstrates that high-risk lenders are also strongly associated with 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 primarily explained by the fact that African American and Hispanic borrowers in these neighborhoods are much more likely to get loans from high foreclosure risk lenders, that is, the estimated effects weaken or disappear after including the controls for high-risk lender.

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, creditworthiness 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 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, in using the same sample of loans, Bayer, Ferreira, and Ross (2016) show that having a rate spread loan is associated with a six-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 0.5- and 0.4-percentage-point increase in foreclosure rates for African Americans and Hispanics, respectively, relative to a population average foreclosure rate of 5.3%; that is, 9% and 7% 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. Across our entire analysis, the inclusion of lender fixed effects substantially reduced the sample estimates of racial and ethnic differences, by over 50% 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 nondepository lenders and high-cost lending is again suggestive that our measure of high foreclosure risk lender captures important information about the nonconforming mortgage market and the impact of this market on minority borrowers. In fact, recent Justice Department settlements with Bank of America (Countrywide loans) 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 and loan-to-value ratio suggest that these lenders may be operating in submarkets with different unobserved risk profiles, and therefore the segmentation of the market may serve legitimate business purposes.

Appendix

Table A1
Sample selection of HMDA variables

	HMDA full sample		/DQ high-quality match		/DQ high-quality match May-Aug		Sample, weighted		/Credit data matched sample, weighted	
Variable	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std 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 (1000s)	247	243	271	221	274	224	274	227	278	211
Applicant income (1000s)	107	142	105	128	105	127	106	132	106	115
Tract median income (1,000s)	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	

This table was previously published in P. Bayer, F. Ferreira and S. Ross. The Vulnerability of Minority Homeowners in the Housing Boom and Bust. American Economic Journal: Economic Policy 8, 1–27 (2016).

This table presents means and standard errors for the pooled sample from HMDA for all seven sites for 2004 to 2008 originations. 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 stratified random sample with means based on the sampling weights, and the final column is for the final sample of observations with a preorigination Vantage score also using the sample weights.

References

Alexandrov, A., and S. Koulayev. 2015. No shopping in the U.S. mortgage market: Direct and strategic effects of providing more information. Working Paper.

Angrist, J., and A. Krueger. 1995. Split-sample instrumental variables estimates of the return to schooling. *Journal of Business & Economic Statistics* 13:225–35.

Avery, R., G. Canner, and R. Cook. 2005. New information reported under HMDA and its application in fair lending enforcement. *Federal Reserve Bulletin* Summer:344–94.

Avery, R., K. Brevoort, and G. Canner. 2007. The 2006 HMDA data. *Federal Reserve Bulletin* December:A73–A109.

Bayer, P., F. Ferreira, and S. Ross. 2016. The vulnerability of minority homeowners in the housing boom and bust. *American Economic Journal: Economic Policy* 8:1–27.

———. 2014. Race, ethnicity and high-cost mortgage lending. NBER Working Paper No. 20762.

Belsky, E., N. Retsinas, and M. 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, NY: Cornell University Press.

Bhutta, N., and D. Ringo. 2014. The 2013 home mortgage disclosure act data. *Federal Reserve Bulletin* 100:1–37.

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

Bocian, D., K. Ernst, and W. Li. 2008. Race, ethnicity and subprime home loan pricing. *Journal of Economics and Business* 60:110–24.

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

Case, K., and R. Shiller. 1989. The efficiency of the market for single-family homes. *American Economic Review* 79:125–37.

Chan, S., A. Haughwout, and J. Tracy. 2015. How mortgage finance affects the urban landscape. In the *Handbook of urban and regional economics*, vol. 5A, Eds. G. Duranton, V. Henderson, and W. Strange, 987–1045, Amsterdam: North Holland.

Charles, K., and E. Hurst. 2002. The transition to home ownership and the black-white wealth gap. *Review of Economics and Statistics* 84:281–97.

Clarke, J., and M. Courchane. 2005. Implications of stratified sampling for fair lending binary logit models. *Journal of Real Estate Finance and Economics* 30:5–31.

Courchane, M. 2007. The pricing of home mortgage loans to minority borrowers: How much of the APR differential can we explain? *Journal of Real Estate Research* 29:399–439.

Courchane, M., and D. Nickerson. 1997. Discrimination resulting from overage practices. *Journal of Financial Services Research* 11:133–51.

Courchane, M., B. Surette, and P. Zorn. 2004. Subprime borrowers: Mortgage transitions and outcomes. *Journal of Real Estate Finance and Economics* 29:365–92.

Edmiston, K. 2009. Characteristics of high foreclosure neighborhoods in the tenth district. *Economic Review* 94:51–75.

Ferreira, F., and J. 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 No. 21261.

Fisher, L., L. Lambie-Hanson, and P. Willen. 2010. A profile of the mortgage crisis in a low-and-moderate-income community. Boston Federal Reserve Public Policy Discussion Papers No. 10-6.

Gerardi, K., and P. Willen. 2009. Subprime mortgages, foreclosures, and urban neighborhoods. *Berkeley Electronic Journal of Economic Analysis and Policy* 9:89–138.

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

Gittleman, M., and E. Wolff. 2004. Racial differences in patterns of wealth accumulation. *Journal of Human Resources* 39:193–227.

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

Harding, J., and S. Ross. 2011. Regulation of large financial institutions: Lessons from corporate finance theory. *Connecticut Insurance Law Journal* 16:243–60.

Haughwout, A., C. Mayer, and J. Tracy. 2009. Subprime mortgage pricing: The impact of race, ethnicity, and gender on the cost of borrowing. *Brookings-Wharton Papers on Urban Affairs* 10:33–63.

Haurin, D., C. Herbert, and S. Rosenthal. 2007. Homeownership gaps among low-income and minority households. *Cityscape* 9:5–51.

Holmes, A., and P. Horvitz. 1994. Mortgage redlining: Race, risk, and demand. *Journal of Finance* 49:81–99.

Horrace, W., and R. Oaxaca. 2006. Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters* 90:321–7.

Manski, C., and S. Lerman. 1977. The estimation of choice probabilities from choice based samples. *Econometrica* 45:1977–88.

Mayer, C., and K. Pence. 2008. Subprime mortgages: What, where and to whom? Finance and Economics Discussion Series. Federal Reserve Board Working Paper 2008-29.

Mian, A., and A. Sufi. 2009. The consequences of mortgage credit expansion: Evidence from the U.S. mortgage default crisis. *Quarterly Journal of Economics* 124:1449–96.

Munnell, A., G. Tootell, L. Browne, and J. McEneaney. 1996. Mortgage lending in Boston: Interpreting HMDA data. *American Economic Review* 86:25–53.

Nelson, A. 2010. Credit scores, race, and residential sorting. *Journal of Policy Analysis and Management* 29:39–68.

Quercia, R., G. McCarthy, and S. Wachter. 2003. The impacts of affordable lending efforts on homeownership rates. *Journal of Housing Economics* 12:29–59.

Reid, C., and E. 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 2009-09.

Ross, S. 2005. The continuing practice and impact of discrimination. University of Connecticut Working Paper 2005-19.

Ross, S., and G. Tootell. 2004. Redlining, the Community Reinvestment Act, and private mortgage insurance. *Journal of Urban Economics* 55:278–97.

Ross, S., M. Turner, E. Godfrey, and R. Smith. 2008. Mortgage lending in Chicago and Los Angeles: A paired testing study of the pre-application process. *Journal of Urban Economics* 63:902–19.

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

Savage, C. 2012. Wells Fargo will settle mortgage bias charges. *New York Times*, July 12.

———. 2011. Countrywide will settle a bias suit. *New York Times*, Dec 21.

Tootell, G. 1996. Redlining in Boston: Do mortgage lenders discriminate against neighborhoods? *Quarterly Journal of Economics* 111:1049–79.