J Financ Serv Res (2018) 54:323–343
DOI 10.1007/s10693-017-0274-5

Unintended Consequences of Risk Based Pricing: Racial Differences in Mortgage Costs

Brent C Smith[1] **· Kenneth N. Daniels**[1]

Received: 27 June 2014 / Revised: 20 February 2017 / Accepted: 22 February 2017 /
Published online: 9 March 2017

Abstract The following analysis focuses on the role that risk pricing has had in the allocation and access to mortgage funds, specifically how it results in cost differences by race. Using a sample of fixed-rate first lien mortgages, we control for the risk characteristics of borrowers and assets. We find that borrowers with comparable credit quality experience significantly higher costs for mortgages in neighborhoods with a high density of minority households. Further, when the pricing differential is controlled for in a model of mortgage default, there is no support for neighborhood price differences. This finding illustrates a potential inequity that results from efficient/risk pricing in mortgage underwriting.

Keywords Residential mortgage default · Mortgage underwriting · Consumer credit

1 Introduction

In the aftermath of the financial crisis, the literature on the fallout from the mortgage market and how the system generated such high default rates has expanded dramatically. Many papers assert that the expansion of the market and a general reduction in the underwriting standards are to blame. This expansion resulted in the highest level of homeownership in the last forty years and was in part facilitated by the employment of risk pricing in the underwriting process. In this analysis we answer two questions. Do borrowers face higher costs for fixed-rate mortgages in neighborhoods where a significant number of minorities (black and Hispanic) reside? And does the cost differential in mortgage pricing indicate efficient pricing where banks charge higher interest rates for riskier loans with higher probabilities of default? Our

✉ Brent C Smith
bcsmith@vcu.edu

Kenneth N. Daniels
kndaniel@vcu.edu

1 Snead Hall, 301 W. Main Street, Box 844000, Richmond, VA 23284-4000, USA

sample comprises first lien mortgages originated in Florida over the period from 2001 to 2008. In the sample we control for borrower, contract, and default risk characteristics. The results suggest that the cost of mortgages increases for borrowers as the proportion of minority residents increases in the neighborhood. We also examine the probability of default after matching loans by the degree of over- and underpricing relative to the term structure at origination and find that there is no relationship between the probability of default and the mortgage's fixed-rate interest costs across neighborhoods does not exist. Our analysis implies that two borrowers with similar credit criteria are charged different interest rates based on the racial composition of the neighborhood. This suggests pricing differences do not necessarily reflect accurate or efficient pricing.

By recognizing that the underwriting process for mortgage lending is intertwined with the borrower and the asset, our method allows for the investigation of the impact that the racial composition of the locale (not the borrower) has on the lending process (Hunter and Walker 1996; Bostic 2003; Bostic and Lee 2008).

A number of studies explore the relation between a neighborhood's racial composition and access to a mortgage (see, e.g., Berkovec et al. 1994; Courchane et al. 2004). However, few studies examine the impact of the neighborhood's racial composition on the cost of mortgage debt. A study by Haughwout et al. (2009) indicates no such relationship exists between the neighborhood's demographics and the cost of mortgage funds. This result is inconsistent with our findings, but not surprising. Their data set comprises a different pool of borrowers, and they do not take the second step in relating contract costs to default probability. A study by Kau et al. (2012) represents a recent effort to tie together the neighborhood's racial composition and mortgage pricing. They obtain similar results to our analysis but there are a number of differences between the two studies including data, and analytical approach.

Using a more current time period and a rich set of variables, our analysis finds a significant statistical and economic impact on the cost of mortgages due to the racial composition of the neighborhood. In addition, taking a quantile regression approach, we show that at all levels the spread of the mortgage increases as the black proportion in the ZIP code increases. This approach provides more granularities in the results. As a control we also consider the disposition of the loan post origination and how the investor's type is connected to different risk profiles in the mortgages. We find that private investors are more likely to purchase riskier mortgage portfolios relative to those retained in house or sold to government sponsored enterprises (GSEs). Further a second analysis using propensity score matching allows us to compare borrowers with similar credit characteristics across neighborhoods.

The literature offers a number of possible explanations for our results. Kau et al. (2012) conclude that their findings, that are similar to ours, are the result of pricing discrimination. Another view is restrictive lending, which results in lenders offering fewer marginal quality loans and as a result there are fewer defaults. There could be an issue of differential information obtained in underwriting. Lacour-Little and Yang (2013) report that the degree of income exaggeration is positively related to default probability, suggesting there exists a potential for underwriters to receive partial or misleading information in the application. This theme is expanded in a recent paper by Darolia et al. (2016) where they explore the potential for temporary credit score manipulation on the part of borrowers. Their findings indicate that borrowers with lower credit scores can benefit from even small alterations in their credit score in the form of reduced borrowing costs. Griffin and Maturana (2016) observe that the frequency of silent second loans varies inversely with credit score suggesting. This research and the work by Carillo (2011) illustrate the potential for rational borrowers to withhold or provide fraudulent information. Thus, the source of our observed differential could be either the lenders or the borrowers. Both are potentially culpable.

The remainder of this paper is organized as follows: Section 2 presents the background and empirical framework for the study. Section 3 presents the data, multivariate analysis, and a discussion of the results. In Section 4, we support our findings of differential pricing with a test of default probability. Section 5 concludes the paper.

2 Hypotheses development via the literature

The early research on racial differentials in the access to residential mortgages offers inconsistent conclusions. This is due largely to a lack of controls for many risk factors that impact the underwriting process. The 1992 study by the Boston Federal Reserve (Munnell et al. 1996) addressed a number of these concerns by observing significant lending disparities between black and white applicants. Jackson (1994) also finds that lending officers are more likely to deny black applicants than white applicants when both have comparable financial characteristics. As a counter to these findings, Horne (1997) offers caution in rigidly interpreting denial rate studies and the significance of the race variable. Horne's analyses illustrate the potential for bias inherent in underwriting data that can offer spurious results.

The literature on mortgage access and pricing such as Benston (1981), Barth, et al. (1979), Berkovec et al. (1994), and Karikari (2009) show the dependence on quality information in the mortgage lending process. Another view asserts varying lending standards as presented in Musulis (1982), Gruben et al. (1990), Calem (1996), Calem et al. (2004). The authors imply that pricing differentials are the result of variations in the information that emanates from neighborhood characteristics rather than the individual borrower. Gruben et al. (1990) and Musulis (1982) support the idea that different search costs lead to different patterns of lending based on the racial composition of the neighborhood. Crews-Cutts and Van Order (2005) build on this theme by indicating that information access is a significant barrier to the efficient pricing of mortgages.

For our purposes the model presented by Longhofer and Peters (2005) and Blanchard et al. (2008) provide a foundation for the efficient pricing rational. Because institutions are profit maximizers, they seek to optimize the underwriting process subject to traditional concerns of cost, risk (default or prepayment), and return. Likewise, exogenous variations in lender costs for a borrower have the potential to instill bias in the underwriting process (Mayer and Pence 2009). Credit worthiness is assumed to be distributed throughout the population according to the probability density function $f(\theta)$ with the cumulative distribution function $F(\theta) = \int_0^{\phi} f(t)dt$. Although each individual's creditworthiness, θ, is given exogenously, an individual's application decision is endogenous. Thus, borrowers decide whether to apply for a loan within the constraints of their creditworthiness.[1] Given a set of costs and available funds, financial institutions identify a minimum acceptable θ^* they are willing to approve. In a full

[1] In Yezer et al. (1994), θ is variable throughout the application process.

information world, lenders know every applicant's θ and approve only those applicants where $\theta \geq \theta^*$ (Longhofer and Peters 2005). Borrowers know in advance if their credit meets the threshold and use that information when deciding to apply.

Regardless of the extent of underwriting diligence, lenders still have limited ability to control a borrower's default (Ambrose and Buttimer 2000), or to monitor events, beyond payment history, that provide additional insight into the prospects for default.[2] For example, trigger events and the price volatility in housing are virtually impossible to foresee or observe. Consistent with this idea, most mortgage pricing models rely on the interaction of interest rates and property values after origination to determine the default probability (Kau and Kim 1994; Kau et al. 1993; Foster and Van Order 1984).[3] Lacking perfect knowledge lenders must rely on limited information obtained via signals defined as, $s = \theta + \varepsilon$, where ε represents the errors in assessing the borrower's risk. Further, there is a threshold value of the signal s^* such that the bank approves every application where $s \geq s^*$. Imbedded in the ε are the costs to the lender in foregone business and the potential costs of default. The objective of the loan pricing process is to maximize the yield subject to the information constraints regarding the borrower, the localized market, and future events. It is assumed that applying for a loan is costly; therefore individuals will only do so if their chance of being approved is sufficiently high.

For our purposes we define discrimination as occurring when lenders require a subset of potential borrowers to meet a higher cutoff signal s^* than for members of another group. The signaling relation presented above allows for borrowers to self-select, but in our case this is neither relevant nor plausible. Borrowers are fixed in location by virtue of the asset, and the location variation (across neighborhoods) exposes all borrowers within a neighborhood to similar levels of discrimination.

For example, consider pools of applicants residing in two neighborhoods that the banks identify as 1) low cost for default and 2) high cost for default, then borrowers apply for loans with a single financial institution that services both neighborhoods. The neighborhood with a lower estimated cost of default motivates the lender to institute more liberal underwriting practices than for loans in the high cost area. Given the borrower's location constraint, they face different requirements expressed as $s_h^* > s_l^*$ for all θ^*. Thus, lenders require applicants in high cost neighborhoods to meet higher underwriting standards. This is particularly relevant to the period of observation in this paper given the cycle in the housing market.

There are a number of reasons why lenders perceive minority neighborhoods as comparably more risky. Historically, minority neighborhoods have experienced greater price volatility (Peng and Thibodeau 2013) that translates into higher risk profiles. Furthermore, minority dominated neighborhoods that are stable typically have lower overall average rates of appreciation. Lenders often characterize minority neighborhoods that are in some state of flux (e.g., rehabilitation or gentrification) as more heterogeneous housing markets and consequently as more risky. The potential outcome of this characterization is a variation based on location regardless the borrower's ability to pay. Therefore, our first hypothesis is as follows:

H1: The neighborhood's racial composition matters in the financing of fixed-rate mortgages regardless of the race of the individual borrower. Individuals face a higher cost for

[2] The threat of prepayment is typically not under the lender's control in residential mortgages.

[3] As Ambrose and Buttimer (2000) report, numerous studies that examine the time-to-default indicate that borrowers' characteristics have a limited impact on predicting default three years after origination (von Furstenberg and Green 1974; Williams et al. 1974).

fixed-rate mortgages if the neighborhood comprises a significant number of minorities (black and Hispanic).

However, we do not expect this relation to hold equally for black and Hispanic residents. Racial minorities cannot be classified as broad categories, and our analysis examines the extent of price variation by the neighborhood's racial composition. We expect significant links between the mortgage interest costs and the typical measures of credit capacity (e.g., FICO score). However, we also expect the neighborhood's racial composition to be significantly related to interest costs (higher for minority neighborhoods). If this is a form of efficient pricing then the subsequent loan performance should be reflected uniformly across neighborhoods by the interest costs to the obligor. This statement leads to the second countervailing hypothesis.

H2: We anticipate that a uniform relationship between the probability of default and interest costs across neighborhoods does not exist. We expect that a significant disconnect will be observed in the cost of the mortgage and the observed extent of the risk of default. This disconnect can be tied back to the racial profile of the neighborhood.

If the probability of default is not directly related to higher interest cost, this disconnect suggests that pricing differences do not reflect accurate or efficient pricing. Munnell et al. (1996) and Black et al. (1997), for example, find that lenders reject minority applicants for mortgages more frequently than comparable white applicants. On the other hand, the implication that default rates should be lower among black applicants due to reduced access is not borne out in a study of the Federal Housing Administration (FHA) mortgage market by Berkovec et al. (1994). As a caveat we note that our results could be biased to the time period observed as there was extensive price appreciation observed in the Florida housing market. However, studies such as Musulis (1982) indirectly infer support for this hypothesis, and Archer and Smith (2013) observe the upward trend in house prices varied dramatically across the state of Florida with across the board reductions initially observed beginning in 2006.

3 Data and methodology

3.1 Descriptive analysis

The loan data utilized in this analysis are from LPS Analytics Inc. and represent the servicing reports on individual loans spatially identified by the five-digit ZIP code over the period from 2001 to 2008. This time frame marks a period of increasing prices for all years except 2008.[4] Table 1 contains the descriptive statistics for the sample and the definition of each variable. The data set includes over 1.1 million 30-year fixed-rate mortgages obtained for refinance or purchase across the state of Florida. We apply a number of filters to the loan data to ensure a robust data set. The dependent variable, SPREAD, is the difference between the contract's interest rate and the rate of the 20-year Treasury rate at the time of origination. The variable

[4] The data were made available via the author's research affiliation with the Federal Reserve Bank of Richmond and an access agreement between the author and LPS Analytics Inc.

Table 1 Variable descriptions and summary statistics, figures reported here are in actual dollar values ($)

Variable	Mean	Standard error	Minimum	Maximum	Description
SPREAD	1.552	1.044	-4.670	10.965	Actual v/ 20-year T-Bonds at origination
FICO	-6.84E-09	1.000	-6.075	2.281	Standardized FICO score
LTV	78.006	11.720	50.000	149.380	Loan-to-value ratio at origination
DTI	36.528	17.552	1.000	99.000	Debt-to-income at origination
VALUE	$ 263,862	$ 254,330	$ 7000	$18,500,000	Value at origination
CONDO	0.254	0.562	0.000	4.000	Coded 1 if condominium
FHA	0.061	0.239	0.000	1.000	Coded 1 if FHA loan
CONV	0.698	0.459	0.000	1.000	Coded 1 if conventional loan
PURCH	0.571	0.495	0.000	1.000	Purpose purchase
RURAL	0.028	0.166	0.000	1.000	ZIP code in rural county
URBAN	0.900	0.206	0.000	1.000	Percent of population in urban areas
% VAC	0.119	0.093	0.000	0.839	Census reported residential vacancy
% BLACK	0.116	0.148	0.000	0.980	Census % of population black
% HISP	0.161	0.192	0.005	0.929	Census % of population Hispanic
SPANISH	0.041	0.075	0.000	0.669	Census % of population Spanish isolated
MEDINC	48,777.130	5077.896	28,186.000	63,728.000	Median household income by ZIP
MEDYEAR	1981	12.56	1942	1998	Census median year built
TERM	1.503	1.323	-0.520	3.680	Term structure proxy for 30 day to 20 year US Security
PRIVATE	0.251	0.434	0.000	1.000	Investor private
PORTFOLIO	0.121	0.327	0.000	1.000	Investor retained in portfolio

SPREAD ranges from −4.67 to nearly 11%. The average for the data set is 155 basis points.[5] We include a number of loan risk measurements such as the FICO score at origination, the loan-to-value (LTV), and the debt-to-income (DTI). The reported FICO score is an unscaled number; to address this deficiency we standardize the FICO score at origination. The resulting variable FICO has a mean of approximately zero but ranges from −6.075 to 2.281 suggesting a sample with diverse credit risk. The LTV has a mean of 78 with a range of 50 to 149. The mean DTI of roughly 37% is close to the industry standard cutoff for prime rate loans of 38%.

The mean VALUE for the residences in the data set over the entire observation period is $264,000 with a wide range of $7000 to $18.5 million. Twenty-five percent of the observed loans are condominiums. We also include the variables PRIVATE and PORTFOLIO representing the distribution of the loan after origination (sold on the private market, retained in portfolio, or sold to GSEs). The lender's decision to sell (and to who) or to retain the loan is directly related to both the conditions in the market and the price that can be attained relative to the risk of the loan. GSEs monitor loan performance and restrict or prohibit purchases from lenders with poor performance. Additionally, lenders that sell loans to GSEs are frequently contracted to service the loans after origination. Such a relation creates a brand and operational connection between the lender and the loan. We also have a set of variables that capture the characteristics or types of mortgages used in the sample. Fifty-seven percent of the observed mortgages were obtained for the purpose of purchasing the residence and 43% for refinancing.[6]

[5] We experimented with estimating the spread using the cost of funds (mortgage interest rate) and the prevailing risk free Treasury rates of various maturities. The results are insensitive to these permutations.

[6] A number of socioeconomic controls were considered and/or tested with no significant alteration in the model outcomes or coefficient estimates.

Our analysis is based at the neighborhood level. The neighborhood is identified by ZIP code boundaries, and the Census Bureau provides socioeconomic variables, including racial composition, at the ZIP code. In our sample, %BLACK and %HISP represent the racial composition of the neighborhood; with an additional control variable SPANISH.[7] The variable SPANISH is distinct from Hispanic and represents the percentage of the population in a ZIP code that is classified by the Census Bureau as linguistically isolated; in this case isolation specifically refers to those Spanish speaking families with no person in the house that speaks English. For our sample, the mean %BLACK is 11.6%, the mean %HISP is 16.1%, and the mean SPANISH is 4.1%. We also have additional controls for the term structure of the interest rates with a mean of 1.50%.[8] As previously mentioned, the disposition of the individual loans at origination is allocated into three categories: private, portfolio, and GSEs. Twelve percent are held in portfolio, 25% are sold on the secondary market to private investors, and the remaining 63% are sold to GSEs.[9]

Because the interest in the analysis is the relation between racial composition and loan pricing (SPREAD), we divide the ZIP codes into quartiles based on the percent of the population identified as black (Panel A) or Hispanic (Panel B) in Table 2. The first quartile represents those ZIP codes with the smallest percentage of the race of interest and the fourth is the highest. The mean of each discrete variable indicates the percentage of the sample. The standard deviation is in italics and all variables are defined in Table 1. As expected the SPREAD increases from 1.44% to 1.73% for black and 1.47% to 1.66% for Hispanics from the 1st quartile to the 4th quartile. The average value of the mortgage at origination is $264,000 compared to a range of $329,000 to $219,000 for black quartiles and $260,000 to $280,000 for the Hispanic quartiles. We note the inverse trend between the black and Hispanic mean house values. As the percentage of the population identified as black increases, the average house value decreases. The inverse is true for the Hispanic quartiles. The reason is likely the distribution of the two minority populations between urban and rural areas. As the Hispanic proportion increases, the percentage of observations in rural areas decreases from 24 to 11%. The black proportions vary, but only slightly, around 17%. The house values in Florida, as with most states in the United States, are on average higher in urban areas when compared to rural areas, thus driving this inverse relation between black and Hispanic quartiles and values. The values for the variable SPANISH are relatively small and do not exceed 13% in any ZIP code.

[7] The U.S. Office of Management and Budget's definitions for race are used in the 2010 Census. The definition of Hispanic or Latino used in the 2010 Census refers to a person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin regardless of race. Black or African American refers to a person having origins in any of the black racial groups of Africa. It includes people who indicated their race(s) as black, African Am., or Negro or reported entries such as African American, Kenyan, Nigerian, or Haitian. The Census Bureau defines a linguistically isolated household as one in which no one in the house, 14 years old and over, speaks English or speaks English "very well" as a second language. In other words, all members of the household 14 years old and over have at least some difficulty with English. This variable serves as additional evidence of the concentration of the Hispanic population in the observed zip codes.

[8] The variable TERM denoted by the 20-year T-bond compared with the 30-day T-bill serves as a measure of the term structure comparing short- and long-term interest rates at the time of origination and exhibits a significant range in the short- and long-term spread over the observation period. We use the TERM spread as an external control where a positive TERM spread indicates an upward sloping yield curve and indicates a positive likelihood that economic activity will be expanding.

[9] This is a static entry provided by LPS. It should be noted that this entry only applies to reported action or intent of the lender at origination. For example, if the loan is held in portfolio and in a year sold on the market, the eventual sale to the secondary market is not observed.

Table 2 Descriptive statistics by African American quartiles and Hispanic quartiles

	1st Quartile	2nd Quartile	3rd Quartile	4th Quartile
Panel A: African American quartiles				
SPREAD	1.435	1.484	1.563	1.729
	1.017	*1.012*	*1.009*	*1.111*
FICO	0.156	0.062	-0.036	-0.184
	0.956	*0.974*	*1.003*	*1.034*
LTV	76.125	77.419	79.102	79.396
	11.450	*11.559*	*11.751*	*11.818*
DTI	35.289	36.348	37.338	37.149
	17.620	*17.470*	*17.578*	*17.462*
VALUE	320,372.900	280,563.300	234,952.900	218,951.400
	330,374.400	*273,190.700*	*196,393.600*	*171,877.500*
CONDO	0.299	0.236	0.198	0.283
	0.554	*0.524*	*0.507*	*0.648*
FHA	0.043	0.053	0.071	0.076
	0.204	*0.225*	*0.257*	*0.264*
CONV	0.771	0.725	0.662	0.635
	0.420	*0.447*	*0.473*	*0.481*
PURCH	0.573	0.573	0.580	0.557
	0.495	*0.495*	*0.494*	*0.497*
RURAL	0.027	0.015	0.039	0.033
	0.162	*0.120*	*0.194*	*0.178*
URBAN	0.880	0.880	0.909	0.933
	0.235	*0.041*	*0.177*	*0.172*
% VAC	0.173	0.113	0.088	0.101
	0.138	*0.075*	*0.047*	*0.054*
% BLACK	0.013	0.045	0.094	0.313
	0.007	*0.013*	*0.018*	*0.181*
% HISP	0.147	0.155	0.167	0.176
	0.249	*0.187*	*0.161*	*0.153*
SPANISH	0.045	0.041	0.034	0.045
	0.104	*0.086*	*0.045*	*0.048*
MEDINC	48,365.520	49,216.470	49,300.810	48,229.970
	5, 391.325	*5, 185.749*	*4, 717.106*	*4, 895.687*
MEDYEAR	1, 980.957	1, 983.640	1, 983.205	1, 975.131
	12.988	*16.300*	*7.801*	*9.498*
TERM	1.597	1.543	1.470	1.398
	1.316	*1.320*	*1.322*	*1.323*
PRIVATE	0.254	0.243	0.236	0.270
	0.435	*0.429*	*0.425*	*0.444*
PORTFOLIO	0.137	0.123	0.107	0.118
	0.344	*0.329*	*0.309*	*0.323*
Panel B: Hispanic quartiles				
SPREAD	1.470	1.496	1.578	1.665
	0.999	*1.019*	*1.063*	*1.083*
FICO	0.094	0.047	-0.022	-0.120
	1.019	*1.003*	*0.996*	*0.968*
LTV	78.101	78.004	77.885	78.032
	12.167	*11.820*	*11.409*	*11.463*
DTI	35.988	36.016	36.631	37.479
	18.429	*17.917*	*16.957*	*16.804*
VALUE	260,148.000	255,016.000	259,575.100	280,703.400
	273,018.300	*253,848.100*	*226,110.700*	*261,054.900*
CONDO	0.189	0.198	0.269	0.361
	0.510	*0.519*	*0.577*	*0.620*
FHA	0.067	0.064	0.055	0.057
	0.251	*0.245*	*0.228*	*0.232*

Table 2 (continued)

	1st Quartile	2nd Quartile	3rd Quartile	4th Quartile
CONV	0.701	0.702	0.705	0.685
	0.458	*0.457*	*0.456*	*0.465*
PURCH	0.588	0.576	0.556	0.563
	0.492	*0.494*	*0.497*	*0.496*
RURAL	0.062	0.016	0.021	0.014
	0.242	*0.124*	*0.142*	*0.118*
URBAN	0.821	0.900	0.932	0.950
	0.270	*0.212*	*0.157*	*0.135*
% VAC	0.159	0.114	0.105	0.099
	0.124	*0.068*	*0.074*	*0.080*
% BLACK	0.096	0.086	0.142	0.139
	0.166	*0.112*	*0.151*	*0.149*
% HISP	0.028	0.062	0.130	0.426
	0.008	*0.013*	*0.030*	*0.217*
SPANISH	0.004	0.010	0.025	0.127
	0.003	*0.005*	*0.013*	*0.112*
MEDINC	48,140.790	48,714.790	50,867.730	47,385.790
	5, 913.397	*5, 051.428*	*3, 829.336*	*4, 600.585*
MEDYEAR	1, 980.640	1, 981.125	1, 981.904	1, 979.275
	15.755	*8.966*	*8.458*	*15.011*
TERM	1.541	1.538	1.475	1.456
	1.321	*1.323*	*1.319*	*1.324*
PRIVATE	0.219	0.226	0.259	0.301
	0.413	*0.418*	*0.438*	*0.459*
PORTFOLIO	0.117	0.126	0.121	0.121
	0.322	*0.332*	*0.326*	*0.327*

Note: standard deviations in italics

The average LTV for the entire sample, 78%, is consistent throughout the Hispanic quartiles and relatively stable throughout the black quartiles, ranging from 76 to 79%. However, both the DTI and FICO indicate deterioration in the capacity of the average borrower in both black and Hispanic populations as the minority proportion grows. The percentage of Hispanics in condos increases significantly from approximately 19% to 36% as the neighborhood becomes increasingly Hispanic.

Federal Housing Administration loans increase in Panel A for the black quartiles. The conventional loans decrease in the black quartiles from 77% of the total loans in quartile 1 (Panel A) to 64% in quartile 4. For the same variables in Panel B both loan types are generally consistent across all quartiles. The PRIVATE and PORTFOLIO variables for the black population suggest slight variations across the quartiles with an increase in loans marketed to PRIVATE investors and a reduction in the portion of loans retained in the lender's portfolio. In Panel B the Hispanic quartiles suggest a significant increase in PRIVATE investors with a static PORTFOLIO value for all four quartile levels. The increase indicates that the growth in private investors reduces the dominance of GSE loans in the sample and is consistent with the growth in the secondary mortgage market over the observation period.

As illustrated in Panels A and B, the SPREAD between the prevailing mortgage rate and the rate on an individual mortgage increases as the proportion of black and Hispanic residents increases in a ZIP code. To allow for further comparison, Figs. 1 and 2 present the SPREAD of observed mortgages by year over the observation period with each line representing one of the four quartiles for the black and Hispanic populations. Throughout the observation period the

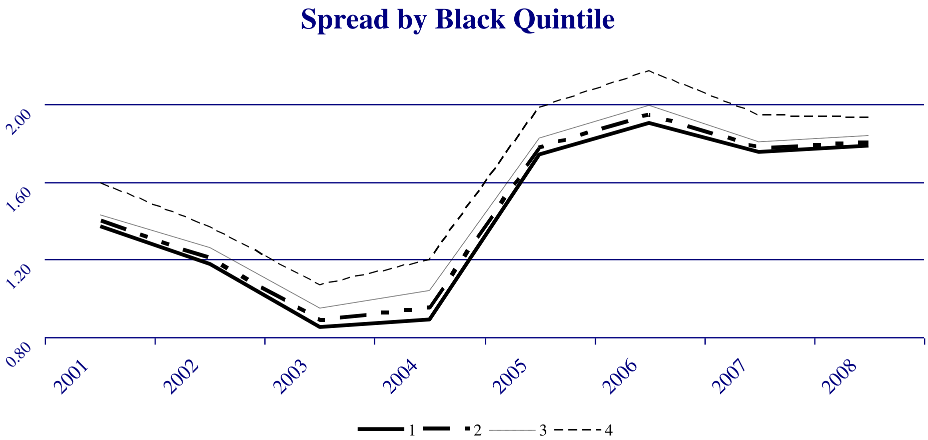

Fig. 1 African-American spread by ZIP code quintile. Data for this Figure were derived from Freddie Mac's monthly average fixed-rate mortgage data, loan-level mortgage rate data from LPS, and ZIP code-level population distribution by race from the 2000 Census Bureau

relation illustrated in Panels A and B in Table 2 holds; the higher the proportion of minorities in a ZIP code the higher the average price of mortgages.

The roughly 90 basis point variation in the SPREAD from 2003 to 2006 is highly correlated with the indices on which many mortgage interest rates are pegged, the London inter-bank offer rate (LIBOR) and the T-bill rate. Fig. 3 shows that while the data are restricted to fixed-rate loans, the movement of the SPREAD between the actual and the prevailing trend during the period from 2003 to 2006 is nearly identical to the two index rates.[10] However, Table 2 also confirms that individual borrowers' risk increases across the quartiles (e.g., FICO decreases and LTV increases).

3.2 Multivariate analysis

For the first test we rely on the quantile regression approach developed by Koenker and Bassett (1978). This approach is an extension of the linear model for estimating rates of change at various levels of the distribution of the response variable SPREAD. The examples of the quantile regression abound in the fields of finance (Santa-Clara and Valkanov 2003), economics (Kuester et al. 2006), and regional science (McMillen and Thorsnes 2006; Coulson and McMillen 2007); but the approach has received limited attention in studies on mortgage markets. The estimates are semiparametric in the sense that we do not assume a distributional form for the random error portion of the model, although we assume a parametric form for the deterministic portion (Cade and Noon 2003). In an ordinary least squares, the objective is to estimate the conditional mean of the random dependent variable, Y, given a set of explanatory variables, x_i, such that $E[Y|x_i]$. A quantile regression generalizes the concept of a univariate quantile to a conditional quantile

[10] As most readers are aware, this was a period of initially low overall interest rates (2004) and then a series of events triggered inflation concerns and overheated markets. That marked a period when the U.S. Federal Reserve (Fed) and counterparts to the Fed in other nations began a systematic increase in interest rates through announcements and open market actions.

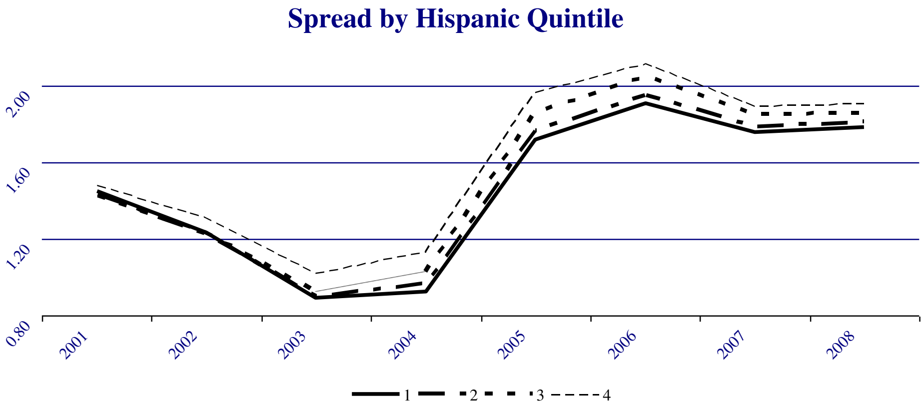

Fig. 2 Hispanic spread by ZIP code quintile. Data for this Figure were derived from Freddie Mac's monthly average fixed-rate mortgage data, loan-level mortgage rate data from LPS, and ZIP code-level population distribution by race from the 2000 Census Bureau

given one or more covariates (Chen 2005). With the random variable Y, the probability distribution function has the following form:

$$F(y) = \Pr(Y \leq y). \tag{1}$$

The τ^{th} quantile of Y^* is defined as the inverse function,

$$Q(\tau) = \inf\{y : F(y) \geq \tau\} \tag{2}$$

where $0 < \tau < 1$. The τ^{th} sample quantile, with analogue $Q(\tau)$ expressed as $\zeta(\tau)$, is an optimization problem of the form,

$$\min_{\varsigma \upsilon R} \sum_{i=1}^{n} \rho_\tau(y_{ii} - \xi) \tag{3}$$

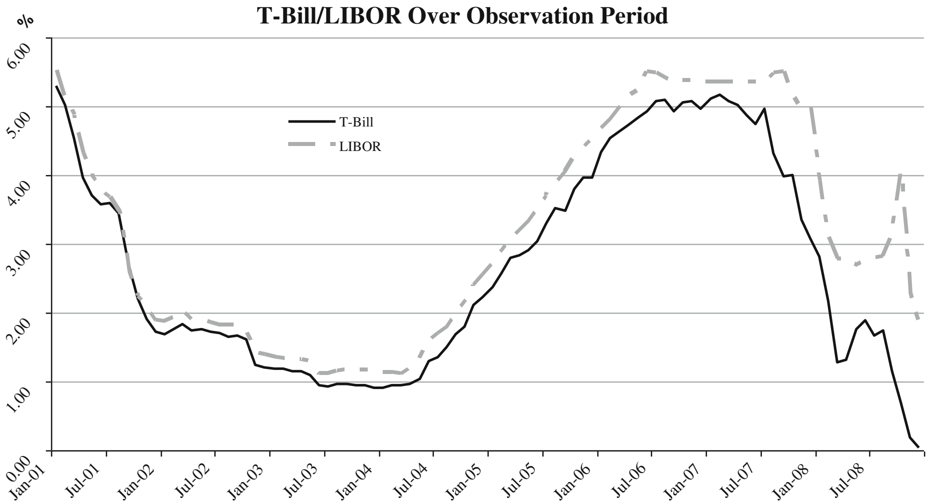

Fig. 3 Presents the prevailing 3-month T-bill/LIBOR rate data were derived from Freddie Mac's monthly average fixed-rate mortgage data and loan-level mortgage rate data from LPS

where $\rho_\tau(\kappa) = \kappa(\tau - I(\kappa < 0))$, $0 < \tau < 1$. Here $I(.)$ denotes the indicator function, and the linear conditional quantile function is,

$$Q(\tau|X = x) = x^{'}\beta(\tau) \tag{4}$$

for any quantile $\tau \in (0, 1)$. The quantity $\tilde{\beta}{\sim}(\tau)$ is the τ^{th} regression quantile. For example, the case where τ = ½, which minimizes the sum of absolute residuals, corresponds to the median regression.

3.3 Quantile regression analysis

To examine the association between mortgage pricing and the racial composition of the neighborhood, we estimate the following model with the quantile regression approach:

$$\begin{aligned} &SPREAD = \alpha_1 + \alpha_2 Std(\mathrm{FICO}) + \alpha_3 LTV + \alpha_4 DTI + \alpha_5 VALUE + \alpha_6 CONDO + \alpha_7 FHA + \\ &\alpha_8 CONV + \alpha_9 PURCH + \alpha_{10} RURAL + \alpha_{11}\% VAC + \alpha_{12}\% BLACK + \alpha_{13}\% HISP + \\ &\alpha_{14} SPANISH + \alpha_{15} MEDINC + \alpha_{16} MEDYEAR + \alpha_{17} TERM + \alpha_{18} PRIVATE + \\ &\alpha_{19} PORTFOLIO + \alpha_{20\text{–}26}(Y02\text{–}Y08) + \varepsilon_{it}. \end{aligned} \tag{5}$$

Table 3 contains the results of the multivariate analysis. For the analysis, ten quantiles are used to allow for more context and texture in observing the variations in the spread over the incremental changes in the coefficient estimates.[11] The output in Table 3 represents the same model run nine times with the median fixed at the specified quantile. The pseudo R^2 is modest and increases on the tails of the spread distribution, and the tails are most likely to respond to our hypotheses.

The borrower and asset risk variables are all significant with the sign in the expected direction, except for the home value (VALUE). The FICO score is negative and decreases in the quantiles. Thus, as the FICO score decreases, the spread increases, and the incremental change increases as the spread increases (moving from quantile 0.10 to 0.9). Both the DTI and the LTV are generally constant across the spread quantiles. As expected both are positive and increase the level of spread at each quantile. The variable that represents condominiums (CONDO) is positive and increases over the quantiles. This result is reasonable because the mortgages on condominiums in the sample, on average, are charged slightly higher interest rates compared to the overall sample.[12] In the case of the appraised value of the property, the spread decreases as the value increases at all quantiles, although at a decreasing rate. The likely explanation is that buyers with a higher capacity to repay the loan purchase higher priced homes.

The loan characteristics are also significant, but with relatively few surprises. The spread for FHA and conventional loans is lower than the remaining loans. The coefficient estimates for conventional purchase loans (versus refinance) are generally flat across the quantiles, while the coefficients for the FHA decrease (becoming more negative) as the spread increases. This is consistent given the stated goals of the FHA that include providing access to home ownership via insurance against mortgage default (Smith 2011). The dummy variables for the annual fixed effects are generally significant across the quantiles, rotating around 2001. There are a few interesting trends in the coefficient estimates that warrant further consideration in a future

[11] A number of different quantile breaks were tested with all yielding similar conclusions.

[12] Because many of the condos are second homes, it is important to note that the mortgages on second homes average 25 to 50 basis points higher than the first mortgages.

Table 3 Quantile regression estimates

Variable	0.1 Quantile coefficient	0.2 Quantile coefficient	0.3 Quantile coefficient	0.4 Quantile Coefficient	0.5 Quantile coefficient	0.6 Quantile coefficient	0.7 Quantile coefficient	0.8 Quantile coefficient	0.9 Quantile coefficient
FICO	−0.110 ***	−0.129 ***	−0.146 ***	−0.165 ***	−0.186 ***	−0.211 ***	−0237 ***	−0261 ***	−0.308 ***
	−97.92	*−153.95*	*−189.66*	*−229.11*	*−240.68*	*−254.36*	*−246.07*	*−243.72*	*−198.17*
LTV	0.001 ***	0.001 ***	0.001 ***	0.002 ***	0.002 ***	0.003 ***	0.003 ***	0.004 ***	0.005 ***
	5.89	*11.09*	*17.77*	*24.81*	*27.97*	*32.83*	*34.98*	*40.18*	*35.54*
DTI	0.002 ***	0.002 ***	0.001 ***	0.001 ***	0.001 ***	0.001 ***	0.002 ***	0.002 ***	0.002 ***
	33.85	*35.35*	*32.97*	*33.10*	*33.31*	*34.76*	*33.99*	*34.03*	*24.91*
VALUE	−4.23E-07 ***	−3.68E-07 ***	−3.38E-07 ***	−320E-07 ***	−3.16E-07 ***	−3.14E-07 ***	−3.09E-07 ***	−2.98E-07 ***	−2.70E-07 ***
	−127.86	*−140.89*	*−132.22*	*−126.27*	*−108.81*	*−94.82*	*−75.75*	*−62.17*	*−35.25*
CONDO	0.023 ***	0.033 ***	0.041 ***	0.050 ***	0.058 ***	0.067 ***	0.080 ***	0.089 ***	0.104 ***
	13.20	*24.46*	*32.80*	*42.06*	*45.61*	*49.87*	*52.02*	*54.30*	*46.75*
FHA	−0.154 ***	−0.191 ***	−0.223 ***	−0.258 ***	−0.305 ***	−0.354 ***	−0398 ***	−0.449 ***	−0.561 ***
	−31.11	*−51.96*	*−66.50*	*−83.43*	*−93.43*	*−103.10*	*−102.12*	*−107.17*	*−97.75*
CONV	−0.149 ***	−0.151 ***	−0.151 ***	−0.149 ***	−0.152 ***	−0.152 ***	−0.146 ***	−0.145 ***	−0.173 ***
	−50.63	*−69.65*	*−75.24*	*−80.12*	*−76.13*	*−70.95*	*−59.20*	*−54.12*	*−45.97*
PURCH	0.016 ***	0.024 ***	0.023 ***	0.023 ***	0.022 ***	0.020 ***	0.021 ***	0.027 ***	0.045 ***
	7.33	*14.20*	*15.20*	*16.40*	*14.22*	*12.62*	*11.59*	*13.48*	*16.49*
RURAL	0.031 ***	0.027***	0.024 ***	0.018 ***	0.011 ***	0.005 ***	−0.002 ***	−0.006	0.003
	4.96	*5.72*	*5.51*	*4.64*	*2.60*	*1.13*	*−0.32*	*−1.03*	*0.47*
%VAC	−0.111 ***	−0.048 ***	−0.006	0.018 ***	0.051 ***	0.087 ***	0.127 ***	0.171 ***	0.185 ***
	−9.26	*−5.34*	*−0.68*	*2.38*	*6.36*	*10.23*	*13.27*	*16.70*	*13.23*
%BLACK	0.105 ***	0.149 ***	0.180 ***	0.204 ***	0.235 ***	0.266 ***	0.299 ***	0331 ***	0386 ***
	13.72	*26.21*	*34.52*	*42.22*	*45.78*	*49.05*	*48.55*	*50.06*	*42.61*
%HISP	0.030 **	0.059 ***	0.076 ***	0.078 ***	0.077 ***	0.082 ***	0.078 ***	0.052 ***	−0.010
	2.11	*5.54*	*7.85*	*8.66*	*8.10*	*8.17*	*6.80*	*4.28*	*−0.62*
SPANISH	0.132 ***	0.113 ***	0.093 ***	0.107 ***	0.137 ***	0.129 ***	0.130 ***	0.166 ***	0.252 ***
	3.62	*4.14*	*3.68*	*4.59*	*5.55*	*4.95*	*4.39*	*5.23*	*5.86*
MEDINC	−3.85E-06 ***	−3.06E-06 ***	−2.67E-06 ***	−2.48E-06 ***	−220E-06 ***	−1.96E-06 ***	0.000 ***	0.000***	0.000 ***
	−18.09	*−19.31*	*−18.32*	*−18.38*	*−15.37*	*−12.97*	*−10.79*	*−9.71*	*−8.5.1*
MEDYEAR	−8.02E-04 ***	−5.55E-04 ***	−5.24E-04 ***	−5.85E-04	−6.25E-04 ***	−6.65E-04 ***	−8.88E-04 ***	−9.92E-04 ***	−1.05E-03 ***
	-6.18	*-5.79*	*-5.98*	*-7.26*	*-7.32*	*-7.40*	*-8.72*	*-9.09*	*-7.02*
TERM	−0.262 ***	−0.223 ***	−0.203 ***	−0.182 ***	−0.164 ***	−0.151 ***	−0.145 ***	−0.145 ***	−0.154 ***
	-111.58	*-130.38*	*-131.18*	*-128.67*	*-110.03*	*-95.98*	*-80.87*	*-75.24*	*-58.63*

Table 3 (continued)

Variable	0.1 Quantile coefficient	0.2 Quantile coefficient	0.3 Quantile coefficient	0.4 Quantile Coefficient	0.5 Quantile coefficient	0.6 Quantile coefficient	0.7 Quantile coefficient	0.8 Quantile coefficient	0.9 Quantile coefficient
PRIVATE	0.217 ***	0.276 ***	0.327 ***	0.398 ***	0.495 ***	0.636 ***	0.856 ***	1.128 ***	1.382 ***
	80.11	*137.76*	*179.65*	*239.63*	*285.67*	*355.43*	*430.91*	*538.52*	*485.70*
PORTFOLIO	−0.223 ***	−0.104 ***	−0.030 ***	0.052 ***	0.154 ***	0.279 ***	0.450 ***	0.711 ***	1.027 ***
	-67.10	*-41.95*	*-13.44*	*24.76*	*69.79*	*121.00*	*173.63*	*257.94*	*274.79*
YEAR 2002	0.106 ***	0.064 ***	0.038 ***	0.013 ***	−0.007	−0.017 ***	-0.022 ***	−0.026 ***	−0.022 ***
	13.77	*10.95*	*703*	*262*	*−1.39*	*−2.96*	*−3.39*	*−3.77*	*−2.28*
YEAR 2003	−0.424 ***	−0.369 ***	−0.318 ***	−0.297 ***	−0.290 ***	−0.285 ***	−0.278 ***	−0.272 ***	−0.265 ***
	−59.79	*−67.96*	*−62.69*	*−62.80*	*−57.23*	*−52.97*	*−45.32*	*−41.18*	*−29.31*
YEAR 2004	−0.573 ***	−0.468 ***	−0.397 ***	−0.372 ***	−0.363 ***	−0.358 ***	−0.350 ***	−0.346 ***	−0.341 ***
	−80.60	*−86.39*	*−78.66*	*−79.17*	*−72.35*	*−67.29*	*−57.89*	*−53.20*	*−38.33*
YEAR 2005	−0.131 ***	−0.131 ***	−0.112 ***	−0.086 ***	−0.055 ***	−0.018 ***	0.016 ***	0.053 ***	0.103 ***
	−17.08	*−23.39*	*−22.07*	*−18.54*	*−11.28*	*−3.55*	*2.76*	*8.57*	*12 20*
YEAR 2006	−0.190 ***	−0.148 ***	−0.110 ***	−0.065 ***	−0.019 ***	0.023 ***	0.049 ***	0.055 ***	0.051 ***
	−21.15	*−22.88*	*−18.92*	*−12.29*	*−3.45*	*4.02*	*7.45*	*7.74*	*5.23*
YEAR 2007	−0.249 ***	−0.196 ***	-0.150 ***	−0.091 ***	−0.033 ***	0.016 ***	0.044 ***	0.060 ***	0.073 ***
	−28.75	*−31.21*	*−26.44*	*−17.44*	*−6.06*	*2.69*	*6.57*	*8.35*	*7.37*
YEAR 2008	0.409 ***	0.398 ***	0.426 ***	0.457 ***	0.492 ***	0.524 ***	0.557 ***	0.600 ***	0.676 ***
	53.13	*69.32*	*80.64*	*93.55*	*94.98*	*95.90*	*89.78*	*90.31*	*74.69*
CONSTANT	3.122 ***	2.720 ***	2.699 ***	2.843 ***	2.942 ***	3.037 ***	3.514 ***	3.794 ***	4.113 ***
	12.18	*14.37*	*15.59*	*17.85*	*17.45*	*17.11*	*17.48*	*17.61*	*13.97*
Prenda R7	21.64%	18.90%	18.16%	18.37%	19.10%	20.57%	23.16%	27.32%	32.52%
n=	1,120,178								

***99% level of significance

**95% level of significance

*90% level of significance

Note: standard deviations in italics

analysis, including the relation between the temporal controls and the coefficient estimates for the variable TERM (Term Structure), which is negatively related to the SPREAD. For this analysis, however, it is sufficient that the estimates are significant and reasonably stable across the quantiles.

The neighborhood fixed effects include the designation as rural versus urban (RURAL), the level of vacancy in 2000 (%VAC), the median household income (MEDINC), the median year built (MEDYEAR), and the race variables. We acknowledge that a valid argument can be made that the actual neighborhood and ZIP code boundaries are not consistent. The ZIP code is the finest detail provided in the loan data. Further, according to work by Cannon et al. (2006), postal routes and the associated ZIP codes are organized around natural terrain and obstacles. This observation suggests that the fit of the ZIP code areas is superior to the U.S. Census or the political boundaries.

The rural coefficients appear to work well in the lower quantiles, which suggests that the spreads are higher in rural areas compared to urban. This relation does not hold up for the quantiles above 0.5. The vacancy rate (%VAC) increases across quantiles, with coefficients that are negative and significant through quantile 0.2 and positive and increasing from quantile 0.4. At the upper quantiles a 1% increase in the vacancy rate increases the median of the spread by as much as 19 basis points holding all else constant. The spread is negatively related to both the MEDINC in the ZIP code and the MEDYEAR for the stock of housing in the ZIP code. Both variables have reasonably consistent coefficient estimates across the quantiles.

Consistent with Hypothesis 1, we anticipate higher comparable borrowing costs in neighborhoods with higher concentrations of minorities. The results from the quantile regression iterations support this assertion on one level. However, when we consider the variable for the percent of the population that is black (% BLACK) in Table 3 and reproduced in Fig. 4, the coefficient is always positive and statistically significant at the 1% level. This result is consistent with the research on the flow of mortgage credit and minority populations (Holmes 2000). As the quantiles advance, the estimated value of the coefficient presents a nearly linear upward progression. At all quantile levels the spread increases as the black proportion in the ZIP code increases. The increase in the spread also increases at higher quantiles. Similarly, the coefficient that identifies the racial composition of Hispanics is always positive and significant at the 1% level. The spread for Hispanics increases in quantiles 0.10 through 0.60. But, beginning at quantile 0.70 the coefficient decreases and creates a convex relation between the coefficient estimate for the variable HISP and the SPREAD. The variable SPANISH is significant and positive across all quantiles with the largest coefficient estimates at the two highest points (0.8 and 0.9), which serves as a counter-balance to the HISP results, especially in the higher quantiles. The data in Table 3 also provide evidence that private investors are likely to purchase riskier portfolios relative to those retained in portfolios or sold to GSEs. The negative signs on the lower quantiles for PORTFOLIO are, in part, a result of the lender retaining jumbo loans and selling prime loans.

Table 4 presents the results for the inter-quantile tests of coefficient equality across quantiles. All of the inter-quantile tests significantly reject the null hypothesis of coefficient equality for both minority coefficients, which further supports Hypothesis 1 that mortgage rates vary as a function of the racial composition of the neighborhood.

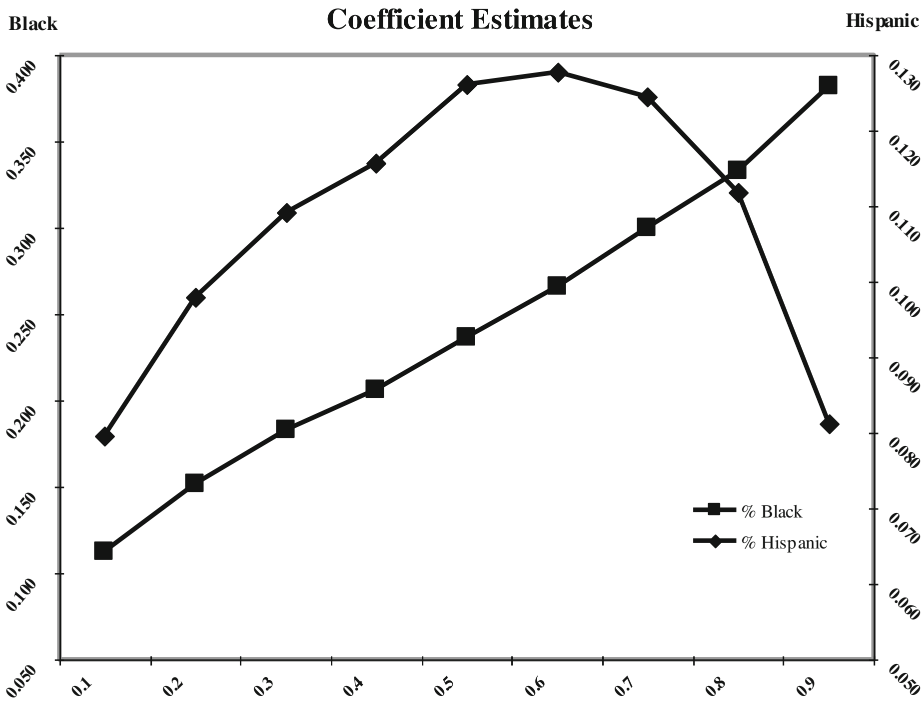

Fig. 4 Quantile regression minority coefficient estimates. This graph represents the coefficients from the quantile regression transformed into graph form. The trends are consistent with the results. As the %BLACK advances the spread increases. As the %HISPANIC advances the spread increases until quantile 0.7 where it begins a steep drop

4 Test of the efficient pricing argument

Over the last two decades financial innovations in risk priced mortgages have enabled lenders to access an expanding pool of potential borrowers. Deng and Gabriel (2006) illustrate that the increased default risks of loans originated among lower credit quality and minority borrowers are offset by reduced prepayments. We examine Hypothesis 2 as an efficient pricing mechanism of default risk. Our assertion is that if we control for the borrower's risk factors and then match borrowers by the SPREAD, a model of default probability will indicate that loans in minority neighborhoods will have a similar likelihood of default when compared to predominantly white neighborhoods. If this is correct, then by extension minority dominated neighborhoods will pay differentially higher interest charges relative to the potential increased location risk inherent in the asset. To use this model, we rely on the same data set used for the SPREAD models and create a propensity score logit model of default probability with the observations coded one if the borrower has defaulted (DEFAULT) over the observation period and zero otherwise. Propensity score matching has a long history in treatment and control studies on subjects in the medical and labor industries (Barnow et al. 1980).[13]

The observations are matched by the SPREAD allowing for a comparison of the variable effects between observations with similar interest costs. By matching borrowers to the

[13] For illustrations see, Dehejia and Wahba (1999), Heckman and Hotz (1989), LaLonde (1986), Rosenbaum and Rubin (1983), and Smith and Todd (2001).

Table 4 Significance tests of coefficient estimates. This table represents a t-test of the significance of the difference between the two adjoining coefficients

t-test of race coefficients

Coefficient comparison	%BLACK	%HISPANIC
0.1 & 0.2	4430.621 ***	2728.330 ***
0.2 & 0.3	4477.714 ***	2066.945 ***
0.3 & 0.4	3597.933 ***	1302.544 ***
0.4 & 0.5	4633.547 ***	2058.141 ***
0.5 & 0.6	4486.380 ***	316.137 ***
0.6 & 0.7	4578.000 ***	568.844 ***
0.7 & 0.8	4113.561 ***	1989.169 ***
0.8 & 0.9	4830.315 ***	3851.576 ***

SPREAD, we are controlling for some portion of the unobserved bias in the data that occurs when borrowers of certain characteristics face differential borrowing costs. Thus, we compare the performance of borrowers of high cost loans with similar borrowers and vice versa. We use a Mahalanobis matching algorithm with a prespecified caliper of 0.20 to match one to many. The use of this algorithm addresses the constraint that there are relatively few observed defaults in our data set. Under this matching constraint and a prespecified $\delta > 0$, a defaulted unit i is matched to those non-defaulted units i such that:

$$\delta > \left|p_i - p_j\right| = \min_{k \in \{D=0\}} \{|p_i - p_k|\}. \tag{6}$$

If none of the non-defaulted units is within the caliper of the treated unit i, it is left unmatched. The model is in the following form:

$$\begin{aligned} DEFAULT = {} & \alpha_{0+}\alpha_1 \mathrm{SPREAD} + \alpha_2 Std(\mathrm{FICO}) + \alpha_3 LTV + \alpha_4 DTI + \alpha_5 VALUE + \alpha_6 \mathrm{CONDO} + \alpha_7 FHA + \\ & \alpha_8 \mathrm{CONV} + \alpha_9 PURCH + \alpha_{10} URBAN + \alpha_{11} + \alpha_{12}\%BLACK + \alpha_{13}\%HISP + \alpha_{14} SPANISH + \alpha_{15} MEDINC + \\ & \alpha_{16} \mathrm{MEDYEAR} + \alpha_{17} \mathrm{TERM} + \alpha_{18} \mathrm{PRIVATE} + \alpha_{19} \mathrm{PORTFOLIO} + \alpha_{20\text{–}26}(Y02\text{–}Y08) + \varepsilon_{it}. \end{aligned} \tag{7}$$

Table 5 presents the results from the logit equation with the matched observations. The control variables for the borrower, property type, and origination year are all significant and generally of the expected sign except for the dichotomous variable YEAR 2008. The lack of significance in this variable is not surprising given the relatively short seasoning of the loans over the period of observation. For example, in a pool of borrowers with similar SPREADs, higher FICO scores lower the probability of default while higher LTVs and DTIs increase the probability of default. Although the observations are matched by the SPREAD, we retain the variable to control for differences in the probability of default across the entire observation set. As before the variables of interest are those representing the racial composition of the ZIP code.

For example, when we control for the SPREAD in the loans in neighborhoods with higher proportions of black residents and compare the results against other loans with similar SPREAD values, we see that the default probability is negatively related to the percent of the population that is identified as black. Given this evidence we fail to reject Hypothesis 2. The results call into question the additional interest premium on fixed-rate mortgages that is charged to residents in neighborhoods with a significant concentration of black residents. More importantly, the results imply that the race of

Table 5 Foreclosure test of pricing efficiency. The propensity score matching by spread is accomplished with a caliper range of 0.2. The dependent variable is binary and coded 1 if the mortgage entered into default during the observation period otherwise 0

Variable	Coefficient	z-score
SPREAD	0.370	74.81 ***
FICO	-0.551	-97.74 ***
LTV	0.035	58.10 ***
DTI	0.003	10.14 ***
VALUE	5.91E-07	41.67 ***
CONDO	0.162	19.14 ***
FHA	-0.089	-3.46 ***
CONV	0.260	19.98 ***
PURCH	0.531	44.80 ***
URBAN	-0.056	-2.03 ***
%BLACK	-0.121	-3.35 ***
%HISP	0.507	7.40 ***
SPANISH	-0.423	-2.47 ***
MEDINC	1.79E-05	15.92 ***
MEDYEAR	0.002	3.52 ***
TERM	-0.399	-29.97 ***
PRIVATE	0.451	33.53 ***
PORTFOLIO	0.129	7.26 ***
YEAR 2002	0.353	4.78 ***
YEAR 2003	0.423	6.14 ***
YEAR 2004	0.740	11.43 ***
YEAR 2005	0.799	13.25 ***
YEAR 2006	0.953	15.16 ***
YEAR 2007	0.624	9.90 ***
YEAR 2008	-0.098	-1.40
CONSTANT	-13.932	-10.28 ***
n=	1,120,158	
Pseudo R2	18.63%	

the borrower is not necessarily driving the decision on the mortgage's pricing. Our analysis implies that two borrowers with a similar level of credit are charged a different SPREAD based on differentials in the racial composition of the neighborhood. If the risk based pricing is the norm, then we should expect to see an increase in default probability with increases in black residents. We observe the opposite.

In the case of Hispanics the results are less certain but potentially enlightening from another dimension. In the analysis of the SPREAD, there is only tacit evidence of higher borrowing costs in neighborhoods with Hispanic concentrations. However, in both the SPREAD and the default models there is support for the argument that a significant and direct relation exists between higher borrowing costs and Spanish isolation. Further, and similar to the conclusions for the neighborhoods with black residents, Spanish isolation decreases the probability of default. A valuable extension of this work would be to examine the distinctions and characteristics of borrowers in urban enclaves where Spanish isolation is prominent. The results might be a reflection of the predatory lending that figured prominently during the financial crisis.

Table 5 also provides evidence of a relation between the probability of default and the variables PRIVATE and PORTFOLIO for the secondary market. It is interesting to note that the loans retained in portfolio and the loans sold to private investors both have a higher probability of default when compared to GSE loans. This result presents opportunities for further analysis as the mortgage market's restructuring of GSEs is likely to continue in response to events over the last decade.

5 Conclusion

Prior to the U.S. Fair Housing Act, mortgage discrimination took the form of outright denial of credit on the basis of the borrower's race. Further, prior to 1965, when the FHA became part of the Department of Housing, its mandate comprised restrictions on minority lending and spatial redlining. It appears that racial bias in the mortgage market persists. This analysis indicates that differential pricing does not necessarily reflect the inherent risk of the mortgage. The results suggest that mortgage borrowers in neighborhoods characterized as having high concentrations of minority borrowers face higher borrowing costs. The results from the analysis indicate that borrowers in proportionately white neighborhoods receive preferential loan pricing relative to borrowers with the similar financial characteristics in proportionately minority neighborhoods. Contrary to Musulis (1982) we find that higher default rates are not associated with neighborhoods that have high minority concentrations. This finding suggests that the additional risk premium might not be warranted after controlling for the necessary risk characteristics, and has a significant impact on the pricing characteristics of the market for fixed-rate mortgages. We observe the differential cost, or perceived risk, in the interest rate that is charged in individual mortgages and is evident when the loan risk factors are controlled for in the model.

When risk factors are controlled for and subjects are matched by the spread of the loan, a model of the propensity for default indicates there is no support for the argument that default is higher across the board in minority neighborhoods. It is important to note that the results from the analysis are reflective of an increasing market. The dataset does not span the period marked by the financial crisis and as a result does not include an entire cycle. The evaluation by underwriters of the higher asset risk associated with high minority neighborhoods (i.e., higher default probability or a lower rate of appreciation) is a self-fulfilling prophecy. As with all components of the housing asset, interest rates are capitalized into the value of the house. If, through differential pricing, lending results in one neighborhood facing higher borrowing costs for mortgage funds, then that differential exerts downward pressure on the values in those neighborhoods. This pressure could result in lowering the ceiling on potential appreciation and increasing the value of the default option for the borrower. The economic costs and benefits of risk based pricing in the mortgage market still need to be fully explored and represent fertile ground for additional research and legislative policy on underwriting. Our findings suggest that federal regulatory agencies should incorporate expanded analyses and appropriate constraints on the potential for risk-based mortgage pricing that create inequities in the mortgage market.

References

Ambrose BW, Buttimer RJ (2000) Embedded options in the mortgage contract. J Real Est Finance Econ 21(2): 95–111

Archer WR, Smith BC (2013) Residential mortgage default: the roles of house price volatility, euphoria and the borrower's put option. J Real Est Finance Econ 46(2):355–378

Barnow B, Cain GG, Goldberger AS (1980) Issues in the analysis of selectivity bias. In: Stromsdorfer E, Farkas G (eds) Evaluation studies, vol 5. Sage, San Francisco, pp 42–59

Barth JR, Cordes JJ, Yezer AMj (1979) Financial institution regulations, redlining and mortgage markets. The Regulation of Financial Institutions 21:101–143

Benston GJ (1981) Mortgage redlining research: a review and critical analysis. J Bank Res 11:8–23

Berkovec JA, Canner GB, Gabriel SA, Hannan TH (1994) Race, redlining, and residential mortgage loan performance. J Real Est Finance Econ 9(3):263–294

Black HA, Collins MC, Cyree KB (1997) Do black-owned banks discriminate against black borrowers? J Financ Serv Res 11(1–2):189–204
Blanchard L, Zhao B, Yinger J (2008) Do lenders discriminate against minority and woman entrepreneurs? J Urban Econ 63(2):467–497
Bostic RW (2003) A test of cultural affinity in home mortgage lending. J Financ Serv Res 23(2):89–112
Bostic RW, Lee KO (2008) Mortgage, risk, and homeownership among low-and moderate-income families. Amer Econ Rev 98(2):310–314
Cade BS, Noon BR (2003) A gentle introduction to quantile regression for ecologists. Frontiers Eco Environ 1(8):412–420
Calem PS (1996) Mortgage credit availability in low and moderate-income minority neighborhoods: are information externalities critical? J Real Est Finance Econ 13(1):71–89
Calem PS, Gillen K, Wachter S (2004) The neighborhood distribution of subprime mortgage lending. J Real Est Finance Econ 29(4):393–410
Cannon S, Miller NG, Pandher GS (2006) Risk and return in the US housing market: a cross-sectional asset-pricing approach. Real Est Econ 34(4):519–552
Carillo P (2011) Testing for fraud in the residential mortgage market: how much did early-payment-defaults overpay for housing? J of Real Est Finance and Econ 47:36–64
Chen C (2005) An introduction to quantile regression and the Quantreg procedure. SUGI 30 Proceedings paper # 213–30. Cary: SAS Institute Inc
Coulson NE, McMillen DP (2007) The dynamics of intraurban quantile house price indexes. Urban Stud 44(8):1517–1537
Courchane MJ, Surette BJ, Zorn PM (2004) Subprime borrowers: mortgage transitions and outcomes. J Real Est Finance Econ 29(4):365–392
Crews-Cutts A, Van Order RA (2005) On the economics of subprime lending. J Real Est Finance Econ 30(2):167–196
Darolia R, Smith BC, Yezer AM (2016) Mortgage markets with rational applicants: theory, evidence, and research implications, working paper
Dehejia R, Wahba S (1999) Causal effects in nonexperimental studies: reevaluating the evaluation of training programs. J Amer Statist Assoc 94(448):1053–1062
Deng Y, Gabriel S (2006) Are underserved borrowers lower risk? New evidence on the performance and pricing of FHA-insured mortgages. J of Money, Credit and Banking 38(6):1431–1460
Foster C, Van Order R (1984) An option-based model of mortgage default. Housing Finance Rev 3(4):351–372
Griffin JM, Maturana G (2016) Who facilitated misreporting in securitized loans? Rev of Fin Studies 29(2):384–419
Gruben WC, Neuberger JA, Schmidt RH (1990) Imperfect information and the community reinvestment act. Federal Reserve Bank of San Francisco Economic Review, Summer:27–46
Haughwout A, Mayer C, Tracy J (2009) Subprime mortgage pricing: the impact of race, ethnicity and gender on the cost of borrowing. In: Burtless G, Rothenberg Pack J (eds) Brookings papers on urban affairs. Brookings Institute, Washington, pp 33–64
Heckman J, Hotz J (1989) Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training. J Amer Statist Assoc 84(408):862–874
Holmes MD (2000) Minority threat and police brutality: determinants of civil rights criminal complaints in US municipalities. Criminology 38(2):343–368
Horne DK (1997) Mortgage lending, race, and model specification. J Finance Serv Res 11(1):43–68
Hunter WC, Walker M-B (1996) The cultural affinity hypothesis and mortgage lending decisions. J Real Est Finance Econ 13(1):57–70
Jackson WE III (1994) Discrimination in mortgage lending markets as rational economic behavior: theory, evidence, and public policy. In: Lashley ME, Jackson MN (eds) African Americans and the new policy consensus. Greenwood Press, New York, pp 157–178
Karikari JA (2009) Neighborhood patterns of racial steering of subprime mortgage lending. Available at SSRN: http://ssrncom/abstract=1439854
Kau JB, Kim T (1994) Waiting to default: the value of delay. Real Est Econ 22(3):539–551
Kau JB, Keenan DC, Kim T (1993) Transaction costs, suboptimal termination and default probabilities. Real Est Econ 21(3):247–263
Kau JB, Keenan DC, Munneke H (2012) Racial discrimination and mortgage lending. J Real Est Finance Econ 45(2):289–304
Koenker R, Bassett G Jr (1978) Regression quantiles. Econometrica 46(1):33–50
Kuester K, Mittnik S, Paolella MS (2006) Value-at-risk prediction: a comparison of alternative strategies. J Financ Economet 4(1):53–89

Lacour-Little M, Yang J (2013) Taking the lie out of liar loans: the effect of reduced documentation on the performance and pricing of alt-a and subprime mortgages. J of Real Est Research 35(4):507–553

LaLonde R (1986) Evaluating the econometric evaluations of training programs with experimental data. Amer Econ Rev 76(4):604–620

Longhofer SD, Peters SR (2005) Self-selection and discrimination in credit markets. Real Est Econ 33(2):237–268

Musulis RW (1982) Government intervention in the mortgage market: a study of anti-redlining regulations. J Monet Econ 10(2):193–213

Mayer C, Pence K (2009) The rise in mortgage defaults. J Econ Persp 23(1):27–50

McMillen DP, Thorsnes P (2006) Housing renovations and the quantile repeat-sales price index. Real Est Econ 34(4):567–584

Munnell AH, Tootell GMB, Browne LE, McEneaney J (1996) Mortgage lending in Boston: interpreting HMDA data. Amer Econ Rev 86(1):25–53

Peng L, Thibodeau T (2013) Risk segmentation of American homes: evidence from Denver. Real Est Econ 41(3):569–599

Rosenbaum P, Rubin D (1983) The central role of the propensity score in observational studies for causal effects. Biometrika 70(1):41–55

Santa-Clara P, Valkanov R (2003) The presidential puzzle: political cycles and the stock market. J Finance 58(5):1841–1872

Smith BC (2011) Competition and risk factors in mortgage lending: the Federal Housing Administration's evolving role in the primary market. Econ Quar 97(1):95–110

Smith JA, Todd PE (2001) Reconciling conflicting evidence on the performance of propensity-score matching methods. Amer Econ Rev 91(2):112–118

von Furstenberg GM, Green RJ (1974) Home mortgage delinquencies: a cohort analysis. J Finance 29(5):1545–1548

Williams AO, Beranek W, Kenkel J (1974) Default risk in urban mortgages: a Pittsburgh prototype analysis. Real Est Econ 2(2):101–112

Yezer AMJ, Phillips RF, Trost RP (1994) Bias in estimates of discrimination and default in mortgage lending: the effects of simultaneity and self-selection. J Real Est Finance Econ 9(3):197–215