J Real Estate Finan Econ (2012) 45:289–304
DOI 10.1007/s11146-011-9330-3

Racial Discrimination and Mortgage Lending

James B. Kau · Donald C. Keenan · Henry J. Munneke

Published online: 28 June 2011

Abstract Looking at a sample of conventional fixed-rate mortgages, this paper examines whether lending practices are consistent with the competitive hypothesis that the racial and ethnic composition of the borrower's neighborhood affects the contract rate charged only to the extent that these characteristics objectively influence the probability of the loan defaulting or prepaying. Our results, however, reject this hypothesis, showing instead that borrowers in predominantly black neighborhoods pay a significantly higher contract rate than is consistent with evidence of their behavior.

Keywords Racial discrimination · Mortgage lending

Introduction

This paper examines whether contract rates charged borrowers indicate non-competitive discrimination on the part of lenders. The main innovation of our data and approach is that we use not only information about the features of the loan and the borrower at origination as they affect the contract rate, we also use information about the subsequent fate of these loans in terms of whether and when the borrowers prepay or default. This permits us to infer the actual degree to which borrowers' behavior is associated with their differing demographic characteristics, which then

J. B. Kau · H. J. Munneke (✉)
Department of Insurance, Legal Studies, and Real Estate, University of Georgia, Athens, GA 30602, USA
e-mail: hmunneke@terry.uga.edu

J. B. Kau
e-mail: jkau@terry.uga.edu

D. C. Keenan
THEMA - Université de Cergy-Pontoise, Paris, France
e-mail: dkeenan@uga.edu

lets us determine how much an impersonal competitive market would distinguish among such borrowers in setting contract rates.[1] This in turn allows us to deduce whether actual lender behavior is influenced only by these dispassionate considerations, or whether there is evidence that lenders further differentiate among borrowers, in a manner that would then be inconsistent with competitive behavior, and so we would call discrimination.

Previous Literature

There already exists a voluminous literature on discriminatory practices in mortgage lending; overviews or surveys on the topic include Yinger (1996), Ladd (1998), Lacour-Little (1999), Dynski (2005), and Ross (2006), as well books by Turner and Skidmore (1996) and Ross and Yinger (2002). Most of this literature, however, only examines the possibility of discrimination in regard to loan approval; in particular, great attention has focused on the practice of "red-lining," where a loan may be denied because of the racial composition of the neighborhood in question.[2] Much less attention has been devoted to whether, conditional on the loan being accepted, the contract rate charged itself reflects discriminatory considerations.[3] Using individual but not neighborhood characteristics, Crawford and Rosenblatt (1999) find no evidence of differential treatment by race, in regard to the premium paid over the market rate for conventional loans. Earlier, Duca and Rosenthal (1994), using survey data, had similarly found little connection between race and rates charged. More recent work by D. E. Getter (2006), using the same type survey data, reaches similar results. On the other hand, Boehm et al. (2006), using different survey data, do find that minorities, particularly blacks, pay higher contract rates than do whites. Finally, Nothaft and Perry (2002), using yet another survey data set, but concentrating on neighborhood rather than individual characteristics, find that greater minority representation in a neighborhood raises rates, though here the effect is attributed to hispanics rather than blacks.[4]

This last set of papers, on contract rates, differs most noticeably from our work in that they do not explicitly account for the possibility of prepayment or default, but rather, introduce individual or neighborhood characteristics believed to affect the likelihood of default. Typically, they have very little to say about prepayment. Since we have data on eventual performance of the loans, rather than supposing a connection, we can endogenously determine to what extent observable variables

[1] Such differentiation by race, which is the only type consistent with profit maximizing firms in a perfectly competitive setting is often referred to as "statistical discrimination" in the literature.

[2] See, for instance, Holmes and Horvitz (1994), Munell et al. (1996), or Tootell (1996).

[3] Thus, if our study were to find no discrimination in contract rates charged, it would not follow that there was no overall discrimination in the lending process. On the other hand, if one does find discrimination in contract rates, then this exists regardless of any possible further discrimination in lending practices on other margins.

[4] One particular rate practice that has also been investigated is overages, the difference between the final contract rate and the initially agreed "lock-in" or guaranteed maximum rate. Reporting results from an Office of the Comptroller of the Currency study, Courchane and Nickerson (1997) indicate that, for the three banks examined, individuals who are black or hispanic pay higher overages than do whites. For the bank examined by Black et al. (2003), on the other hand, only hispanics pay any higher overage.

actually are associated with the likelihood of prepayment and default, and then infer from lenders' behavior how their concern for these possible events affect the contract rates they charge. We do not impose the condition that either prepayment or default is more important than the other for contract rates, but instead are able to determine, contrary to the implicit presumption of these previous papers, that prepayment is the predominant consideration. This will in turn be seen to affect the channels by which questions of race are liable to affect contract rates.

The Model

We model both borrowers and lenders as to their behavior with regard to loans. The lender's behavior is exhibited only at the origination of the loan, when the contract rate is set, whereas the borrower's behavior is observed each subsequent month, so long as the loan survives. Note, though, that the only action on the part of the borrower being observed is whether the loan is continued or whether the borrower chooses to terminate the loan through prepayment or default. In determining this, the borrower needs only be concerned with the given terms of the loan, and so has no need to consider the lender's behavior. In contrast to this, in order to appropriately set the terms of the loan at the time of origination, a lender must consider whether the borrower is ever likely to prepay or default, and so does need to have a model of borrower behavior. We assume that the lenders in fact possess the true model of borrower behavior, which we in turn estimate by looking at the actual borrower behavior within our data set. The question we then ask is whether observed lender behavior is consistent with entirely competitive markets. Lenders, in setting contract rates within such a competitive environment, would take account of demographic variables of the borrower, such as race or ethnicity, only to the extent that they are seen to affect how and when the loan is terminated. This termination in turn would be considered only to the extent that it impacts the flow of proceeds from the loan and hence influences the required contract rate.

While the competitive market hypothesis we examine requires that lenders behave according to rather strict standards, we impose far less restrictions on borrower behavior, and instead simply deduce their behavior, using a rather flexible duration-analysis framework. Since the observations on loan survival are made at regular monthly intervals and the borrowers may always choose to continue, to prepay, or to default, we employ a multinomial logit framework to represent the resulting Cox discrete-time competing-risks model (Deng et al. 2000). Thus,

$$\ln\left(p_{jt}/p_{0t}\right) = \delta_j^{'}\mathbf{x}_t + \alpha_{jt} \qquad j = 1, 2 \tag{1}$$

where p_{jt} is the probability of default at time t in the case where j=1, while it is the probability of prepayment at time t in the case where j=2 (with p_{0t} obviously being the probability of continuation). We take time t to denote "mortgage time," which is to say, the age of the mortgage.[5] Thus, the variables α_{jt}, j=1, 2, trace out the

[5] Variables really in calendar time can always be expressed in mortgage time without loss of generality, since the transformation between the two is known for each loan and the coefficients are entirely independent of time.

baseline termination rates for default and prepayment, respectively. The $\mathbf{x}_t$ in turn are various covariates, possibly time varying, that then shift these baselines. Prominent among these covariates are the loan's contract rate, a, together with the ten-year yield, y_t, reflecting the then current term structure. This time horizon is commonly regarded as an appropriate one for predicting behavior with regard to such mortgages.

While most of our covariates vary somewhat over time, the only one we assume the lender considers in any detail is the highly important and volatile term structure, which we take to be driven by a spot interest rate, $r(t)$, of the simple Cox, Ingersoll and Ross (CIR) form, so that

$$dr(t) = \gamma(\theta - r(t))dt + \sigma\sqrt{r(t)}dz(t). \tag{2}$$

Here the first term is deterministic, with θ being a fixed trend and γ the degree of reversion over time toward the trend, whereas the second term represents stochastic movements in the term structure that continually upset the deterministic part, as scaled by volatility σ.[6]

We allow for the possibility that the lender, though not particularly the borrower, may make a sharp distinction between whether a prepayment at time t occurs when the yield $y_t(r(t))$ exceeds that at origination, $y_0(r(0))$, in which case the lender thinks of it as "positive" prepayment, given that it raises expected present value, or whether y_t is less than y_0, in which case it is deemed "negative" prepayment. We can consider (2) as yielding the probability of prepayment or default, $p_j(y_t(r(t)))$, conditional on the then prevailing yield y_t. Thus given the distribution of interest rates $F(r(t) \mid r(0))$ implied by the CIR form, the lender determines the unconditional expectations at origination of such terminations according to

$$\widehat{p}_{1t}(y_0) = \int p_{1t}\, y_t(r(t))\, dF(r(t) \mid r(0)) \tag{3}$$

$$\widehat{p}_{2t}(y_0) = \int_{y_t(r(t)) \leq y_0} p_{2t}\, y_t(r(t))\, dF(r(t) \mid r(0)) \tag{4}$$

$$\widehat{p}_{3t}(y_0) = \int_{y_t(r(t)) \geq y_0} p_{2t}\, y_t(r(t))\, dF(r(t) \mid r(0)) \tag{5}$$

where $\widehat{p}_{1t}$ is then the probability of default at time t as seen at origination, $\widehat{p}_{2t}$ is the probability of negative prepayment, and $\widehat{p}_{3t}$ is then the probability of positive prepayment. We numerically approximate these integrated expectations in the obvious manner by dividing the interest rate range up into a finite number of intervals.

[6] For the other time-varying covariates, we use their actual values when these are known, and otherwise, either extrapolate from the known values or take the average of these values, whichever seems the more appropriate for that particular covariate.

It is not, say, the probability of default $\widehat{p}_{1t}$ at any particular time t with which the lender is ultimately concerned, but rather the total probability that the lender will eventually default. The reason we have bothered to decompose default probability into its component probabilities at various times is partly because we wish to capture the effect of time-varying covariates, but mostly because we have so much right censoring and left truncation in our data. If from the start, our statistical analysis merely classified mortgages by whether they ever defaulted or prepaid during our observation window, the resulting timeless model would significantly underestimate the true eventual probabilities of termination, since we would then not be imputing the many defaults and prepayments that undoubtedly occur but which we did not have occasion to observe. By tracking mortgages over time and assuming that the mortgages when unobserved follow a similar pattern in terms of their maturities to mortgages that are observed, we are then able to take appropriate account of what must remain unseen.

Thus, from the above time-specific probabilities $\widehat{p}_{jt}$ we form the total probabilities $\widehat{P}_j$ that a loan eventually terminates in one of the last three of the now four categories[7]:

$$\widehat{P}_j = \sum_{t=1}^{T} \left(1 + \frac{y_0}{12}\right)^{-t} \widehat{p}_{jt} \prod_{s=1}^{t-1} \left(1 - \sum_{j=1}^{3} \widehat{p}_{js}\right) \qquad j = 1, 3. \tag{6}$$

Notice that we have introduced discounting, using the yield y_0 at origination for convenience, to reflect the idea that terminations later in the time are of less consequence than those earlier in time.[8] We take T to represent a ten-year span, in keeping with the notion that this is the appropriate time horizon considered by the lender. Since these probabilities represent the expectations of the lender, it is entirely appropriate that they only reflect predicted probabilities, and so, do not include any unforeseeable randomness in actual probabilities, as represented by errors of the resulting logit regression. They will therefore act as generated regressors in the estimated equation describing lender behavior.

We thus let the resulting total probabilities $\widehat{P}_j$ enter the determination of the loan's contract rate a as,

$$a = \gamma_0 y_0 + \beta_1 \widehat{P}_1 + \beta_2 \widehat{P}_2 + \beta_3 \widehat{P}_3 + \gamma' \widetilde{\boldsymbol{x}}_{\mathbf{0}} \tag{7}$$

where $\widetilde{\boldsymbol{x}}_{\mathbf{0}}$ is another selection from our available covariates at origination, from which we have separated, merely for emphasis, the ten-year yield y_0. The point of having introduced the further prepayment categories is that anticipated prepayment ought in principle to have different, partly offsetting effects on the contract rate, depending on whether yields at time of prepayment are expected to be higher or lower than when originally set, since the lender will presumably require a higher contract rate when the resulting refinancing is foreseen to be disadvantageous, but a lower one when it is foreseen to be advantageous.

[7] The probability $\widehat{p}_{jt}$ that a loan terminates in category j at time t, conditional on having survived to that time, is multiplied by the probability that it did in fact survive to that time, and then this is discounted and added up over all time periods.

[8] At first sight, it seems peculiar to discount probabilities, but what is being represented is, more properly, expected discounted values, with the amounts at different times being regarded as similar, and so suppressed, leaving just the probabilities and the discount factors.

Also included in $\widetilde{\boldsymbol{x}}_0$ are various demographic variables. We thus examine whether these other covariates play a role in determining contract rates, in addition to contract rate being rationally determined by the then current yield together with the various premia reflecting the possibilities of prepayment and default.

The Statistical Analysis

Our data set consists of a sample of mortgage and borrower characteristics for 2529 30-year fixed-rate conventional loans originated by several different banks within Miami, Florida, between 1975 and 2002. (See Table 1). While we do not have personal characteristics of the individual borrowers, we do know the demographic composition of their neighborhood by census tract, so this is what will be meant by, say, racial composition associated with the loan. Indeed, this usage is consistent with what is implied by the term "redlining," where in making loans, it is by racial composition of the community that banks supposedly discriminate.[9] Loans are observed at monthly intervals ranging from 1986 to 2002, with about ½% defaulting and 36% prepaying within their ten year horizon.[10] As indicated, given such discrete observations, we adopt a multinomial logit framework when performing a duration analysis of these mortgages.

The contract rate a appears in the logit equation (1) and will be endogenous if, for instance, there are many omitted variables in the estimation of it that are correlated with the errors in the estimation of (1). Since the usual two-stage least squares (2SLS) procedure is inappropriate for a non-linear logit model, we follow Petrin and Train (2003) and use a control function method. Thus, a reduced-form estimation of the contract rate is first performed against exogenous variables, including the initial yield y_0, and the residuals of this regression are then separately included in the logit estimations of the termination probabilities p_{jt}. These probabilities p_{jt} are then combined with the estimated CIR term structure model, according to (3–5), and the results aggregated over time, according to (6), to obtain the expected probabilities $\widehat{P}_j$, at origination, that the loan will subsequently end in one of the three termination categories.[11] Since these expectations are in terms of the contract rate a itself, they are endogenous when entered into (7), but this difficulty can be handled by standard 2SLS, given that, unlike (1), equation (7) is linear in the parameters. To obtain a generated instrument for $\widehat{P}_j$, when calculating this value we also calculate the value of $\widetilde{P}_j$ in parallel, where the same procedure is used in determining it, save that the actual contract rate a is replaced throughout by its reduced-form estimate, as previously obtained in the opening step of our

[9] Discrimination based on the race and ethnicity of the individual borrower is assumed not to occur, this being a particularly flagrant illegal practice.

[10] Observed default is thus a rather rare event, with the latent rate of default being of similar magnitude, as will be seen. This proportion of defaulting mortgages; however, is quite similar to those found in nationwide data sets consisting of millions of conventional mortgages over roughly the same time period. One of the points this paper makes is that academic studies traditionally exaggerate the role that default could possibly play in the calculations of a rational lender, while evidently under-appreciating the role played by prepayment.

[11] The estimated CIR model is that of Geyer and Pichler (1999).

Table 1 Descriptive statistics of mortgage sample at origination

Variable	Mean	Std. Dev.
Loan Characteristics:		
Contract Rate	0.0783	0.0094
Points	0.0115	0.5165
Original LTV	0.8185	0.1611
Original Loan Size ($)	104,423	48,992
Adjusted Original Loan Size ($) in constant (yr 2000) dollars via GDP deflator	111,952	47,636
Loan Originated during:		
first quarter of year (=1)	0.2365	0.4250
second quarter of year (=1)	0.2543	0.4355
third quarter of year (=1)	0.2274	0.4192
fourth quarter of year (=1)	0.2819	0.4500
Neighborhood Characteristics:		
Average Household Inc. by Census Tract	59,056	25,408
Proportion of Black by Census Tract	0.1822	0.2575
Proportion of Hispanic by Census Tract	0.5239	0.2667
Proportion of Total Persons below Poverty Level Last yr. by Census Tract	0.1156	0.0778
Unemployment Rate by Census Tract	0.0687	0.0360
Average House Price by Census Tract in constant (yr 2000) dollars via GDP deflator	216,456	238,140
Community Reinvestment Act (= 1)	0.0996	0.2996
Sample Size	2,529	

statistical analysis. The expression $\widetilde{P}_j$ is therefore exogenous and serves as a valid instrument for $\widehat{P}_j$. To control for possible geographic fixed effects, we also demean the loan data by municipality, a slightly larger neighborhood aggregation than tract.[12]

While prepayment and default have some similarities, they also have many differences, and knowledge of how they would behave in a options-based model of "financial" termination leads us to include different market and contractual covariates in their estimations. The effect of the contract rate and the yield, for prepayment in particular, are commonly understood to be in comparison to one another, and so, we choose as a covariate in prepayment the (lagged) *spread*, $s_t \equiv (a - y_{t-2})$.[13] For default, however, any effect of spread is primarily considered

[12] We chose not to apply fixed effects to (1), where we have already in part accounted for omitted variables using the control function method. Controlling for fixed effects would not be entirely useful unless they could be estimated and so included in the prediction of $\widehat{P}_j$, as would be accomplished with a dummy variable procedure. Normally, such an estimation would be impossible due to the incidental parameter problem, but since our situation is more profitably thought of, in the usual language, as "T" growing large with "n" fixed, it appears possible in principle to do such an estimation. Nonetheless, it involves estimating a considerable number of extra parameters, and so, was considered inadvisable.

[13] It is common practice to use a lagged value of the spread, at least for prepayment, to reflect the borrower's use of past information and the time between decision and actual prepayment.

to be a substitution effect between prepayment and default, whereas contract rate is thought to have its own absolute effect. Thus, in addition to spread, contract rate a is entered separately into default. The original loan-to-value ratio, along with house price levels by tract, as adjusted for inflation, were also included in the default regression, whereas adjusted loan size and points were included in prepayment.[14] The primary demographic variables we are interested in are race and ethnicity (*black* and *hispanic*, respectively), but we also include more economic demographic variables—income, unemployment, and poverty—throughout the analysis.

Prior experience on larger data sets shows that the Standard Default Assumption (SDA) schedule serves as an adequate baseline for default, but that the traditional Public Securities (PSA) schedule does a much less adequate job of capturing the more complicated patterns characterizing prepayment.[15] Thus, for default we assumed a scaled SDA schedule for the baseline, whereas for prepayment we imposed little structure on the baseline, using, instead, yearly dummies in mortgage time. We also included quarterly dummies for seasonality in the latter since this is known to greatly influence prepayment, and finally, we also allowed for a linear and quadratic trend term in calendar time.

In the final contract rate equation, we included all the variables that theory would suggest should influence the contract rate in competitive markets, mainly the yield y_0 at origination, the three probabilities of termination, $\widehat{P}_1$, $\widehat{P}_2$, and $\widehat{P}_3$, along with, possibly, points and adjusted loan size, this all together with the demographic variables that would not be separately affecting contract rate in competitive markets—black, hispanic, income, unemployment, and poverty. In addition, we introduced an indicator of the application to a neighborhood or not of the Community Reinvestment Act (CRA), which offers advice to borrowers from certain underprivileged neighborhoods on how to obtain loans, and which might indeed be expected to influence typical contract rates in a neighborhood where it is available. Finally, as with prepayment, seasonal dummies and a trend term in calendar time were also permitted.

The Results

The results for default and prepayment reported in Table 2 are primarily as one might expect. Default is a consequential event for the borrower, and it indeed appears to be driven only by such "financial" considerations as the *loan-to-value ratio* and the *contract rate*, which appear significantly and of the appropriate sign. Correspondingly, none of the demographic considerations are found to be of significance.[16] Against all this support for "financial" default, however, one must note that the more subtle effects of *spread* and neighborhood *house price* are not present. As for prepayment, *spread* and adjusted *loan size* are found to be

[14] Tract house price levels were calculated as a smoothed average of frequently observed home price sales from a much larger set of data subsuming those mortgages used in the current study.
[15] See Kau et al. (2004).
[16] With more observations and so more default, some of these covariates might gain statistical significance, but we doubt they would ever attain much greater economic significance.

Table 2 Default and prepayment estimates—logit estimation

	Default model	Prepayment model
Intercept	−20.6796	4.1184
	(0.0003)	(<.0001)
Contract Rate	58.4676	
	(0.0550)	
Spread: Contract Rate (%) less 10 yr Treasury Constant Maturity Rate lagged 2 periods	16.1052	42.4197
	(0.6354)	(<.0001)
Residual (from reduced-form rate equation)	21.3759	1.2103
	(0.5548)	(0.8427)
Points		−0.0633
		(0.3917)
Original LTV	5.5475	
	(0.0716)	
Adjusted Original Loan Size in constant (yr 2000) dollars via GDP deflator		4.9E-06
		(<.0001)
Proportion of Black by Census Tract (t)	−0.2078	−2.2323
	(0.9476)	(<.0001)
Prop. of Hispanic by Census Tract (t)	2.6229	−0.6518
	(0.3214)	(0.0139)
Avg. Household Income by Census Tract (t) in constant (yr 2000) dollars via GDP deflator	2.3E-05	−1.4E-06
	(0.5416)	(0.4538)
Poverty Rate by Census Tract (t)	4.7673	−1.3046
	(0.5253)	(0.114)
Unemployment Rate by Census Tract (t)	−16.0709	−1.1833
	(0.3540)	(0.5289)
Average House Price by Census Tract (t) in constant (yr 2000) dollars via GDP deflator	−1.0E-05	
	(0.2700)	
Loan Originated in Second Qtr. of Year		0.1502
		(0.1085)
Loan Originated in Third Qtr. of Year		−0.0617
		(0.5271)
Loan Originated in Fourth Qtr. of Year		−0.2000
		(0.0402)
Time Trend (t=1, Jan. 1970, t=2 Feb. 1970…)		−0.9917
		(<.0001)
Time Trend Squared		0.0234
		(<.0001)
Likelihood Ratio	28.9719	941.0266
	(0.0023)	(<.0001)

*Note that *p-values* are reported in the parentheses below the coefficient estimate. The baseline estimates for the default equation and the prepayment equation are not reported, but are available upon request

Table 3 Contract rate estimates—2SLS fixed effect estimation

	Competitive model	Color blind model
Intercept	0.03277	0.03292
	(<.0001)	(<.0001)
Predicted Probability of Default	0.37886	0.76598
	(0.0147)	(<.0001)
Predicted Prob. of Positive Prepayment	−0.04272	−0.04885
	(<.0001)	(<.0001)
Predicted Prob. of Negative Prepayment	0.02743	0.03095
	(<.0001)	(<.0001)
10 yr Treasury Constant Maturity Rate	0.66431	0.65183
	(<.0001)	(<.0001)
Points	−0.00125	−0.00121
	(<.0001)	(<.0001)
Adjusted Original Loan Size ($/100,000) in constant (yr 2000) dollars via GDP deflator	−0.00092	−0.00094
	(0.2640)	(0.5504)
Proportion of Black by Census Tract (t)	0.00362	0.00141
	(0.0437)	(0.6952)
Prop. of Hispanic by Census Tract (t)	0.00269	0.00270
	(0.1568)	(0.4082)
Avg. Household Inc. by Tract (t) ($/100,000) in constant (yr 2000) dollars via GDP deflator	0.00074	0.00130
	(0.5347)	(0.4880)
Poverty Rate by Tract (t)	0.00188	0.00209
	(0.6261)	(0.7905)
Unemployment Rate by census tract (t)	0.00462	0.00730
	(0.6011)	(0.6920)
Community Reinvestment Act	−0.00302	−0.00278
	(<.0001)	(<.0001)
Loan Originated in Second Qtr. of Year	0.00009	0.00007
	(0.8995)	(0.9540)
Loan Originated in Third Qtr. of Year	0.00184	0.00179
	(0.0190)	(0.1863)
Loan Originated in Fourth Qtr. of Year	0.00072	0.00071
	(0.3145)	(0.5575)
Time Trend (t=1, Jan. 1970, t=2 Feb. 1970)	0.00102	0.00115
	(0.8693)	(0.8906)
Adj-R square	0.7097	0.7262

*The dependent variable is the mortgage contract rate. Also note that *p-values* are reported in the parentheses below the coefficient estimate

significant and of the anticipated sign, but *points* turn out insignificant. Whether points should indeed be of consequence is a matter of some dispute: a sunk cost argument would indicate that they should not, though it is often argued that borrowers, anticipating prepayment or not, self-select through points, and so their presence should have a negative effect on prepayment. None of the more economic demographic variables—*income*, *unemployment*, and *poverty*—appear to play a significant role in prepayment, whereas unlike default, *black* and *hispanic* do appear in a statistically significant way, with these groups prepaying less than others.[17] Note that the residuals of the contract-rate reduced-form regression are insignificant in both the default and the prepayment estimations, suggesting that we did not omit any important factors common to both termination and the contract rate.

Turning to the contract rate equation, as reported in the first column of Table 3, the *initial yield*, as well as default and both forms of prepayment appear significantly and of the expected signs, though the coefficient on yield, γ_0, is noticeably less than one.[18] It is particularly striking that good and bad prepayment do appear significantly and of opposite sign, since their creation and the distinction between them is motivated entirely by theoretical considerations; that is, while found to be opposite in effect they are based on the very same prepayment regression. Also, not that during our study period, "financially-motivated" negative prepayment was considerably more likely than the positive prepayment presumably motivated by more "personal" considerations, just as prepayment overall was more likely than default. For instance, taking the mean of our sample, the undiscounted total probability of default is calculated to be nearly 0.5%, with the probability of positive prepayment being about 16.8%, while the probability of negative prepayment is found to be about 78.9%. Observe that while a given probability of default has about a ten times greater impact on the contract rate than does such a probability for negative prepayment, default has just been seen to typically be more than a hundred times less likely to occur, so that despite traditional concerns regarding differences in default attributable to differences in borrower characteristics, it is really differences in their prepayment behavior that is of the greater consequence for contract rates. This is entirely consistent with the prevailing wisdom in the finance community during our study period that default was of negligible concern whereas prepayment was an important issue.[19] Of course, in the previous step, we have determined that race and ethnicity have a statistically insignificant effect on default, so in regard to these characteristics it

[17] That neighborhoods with large minorities prepay less is a result also found in Van Order and Zorn (2002), though Cotterman (2001) gets the effect to be insignificant when enough other variables are included. Using VA loans and FHA loans respectively, with information on individual though not neighborhood racial status, Kelly (1995) and Deng and Gabriel (2006) find that blacks and hispanics prepay less then do whites. Also see Clapp et al. (2001, 2006), and An et al. (2006) for similar conclusions. On the other hand, using American Housing Survey data, Archer et al. (2002) failed to find any difference between the prepayment behavior of these minorities and whites.

[18] All p-values have been calculated using standard-error corrections required by the fact that the termination-probability covariates are generated regressors. The method of correction is that indicated in Appendix 6A of Wooldridge (2002).

[19] Part of the reason why default is not of more consequence is that loans are insured when there is greater than an 80% LTV. We tested for whether there was a difference in contract-rate prediction between loans above and below 80% at origination, but found little indication of it. Part of the reason for this is presumably because the division between the two is really not that stark, given that every loan will drop below 80% in time.

would make little difference what the impact of default was.[20] In contrast as we have seen, race and ethnicity do have a statistically significant impact on prepayment, though one sees that this indirect effect on contract rates if of limited size.

While in the prepayment estimation *points* were insignificant, here, as would be expected, the effect of points are highly significant and negative. Loan size turns out to be insignificant, suggesting that fixed costs of extending a loan may not be an important consideration. None of the more economic demographic variables appear significantly, in accordance with the competitive model, though *CRA*'s presence is significant and positive, as is expected. Most notably, though, the *black* variable, though not the *hispanic* variable, clearly raises contract rates in a statistically significant way. Given a correctly specified model, we can unambiguously conclude that lenders are not behaving as competitive markets would predict, but are instead charging more heavily black neighborhoods higher contract rates than objective evidence would call for. Furthermore, it is on these grounds that they apparently discriminate, and not on other, more economic, demographic characteristics.

Note that a loan from an all-black neighborhood, holding everything equal, including termination behavior, would pay nearly 0.4% more than a corresponding loan from a neighborhood with no blacks.[21] Thus, not only are these effects statistically significant, they are of considerable economic consequence, at least to the borrower. Remarkably, the difference in contract rate that is solely due to being in an entirely black versus an entirely non-black neighborhood is somewhat more than the total amount that the various forms of termination together contribute toward increasing the typical contract rate.[22]

We lack what would undoubtedly be useful data on the individual characteristics of the borrowers within each neighborhood. It is unlikely, however, that such information could alter our primary results. Indeed, suppose that the true model depended not only on neighborhood demographic factors but also on the corresponding unobserved individual characteristics. Under the reasonable assumption that each neighborhood characteristics is a "good" proxy for the corresponding omitted individual characteristic (see Wooldridge 2003), in the sense that the

[20] Previous results have been mixed. Using FHA data, Gabriel (1996) finds no statistically significant effect of neighborhood racial composition on default. Also using such data, Berkovec et al. (1994) conclude that the blackness of a neighborhood is not "strongly and consistently" associated with the likelihood of default, while a greater portion of hispanics does indicate a lower default rate (also see Berkovec et al. 1998). Coleman (2001), once again using FHA data, does find a positive association regarding neighborhood blackness and default, but can no longer locate an effect from hispanics. In contrast, Han (2004), employs similar data to conclude that either group's presence in neighborhoods serves to raise default probabilities. Using a large sample of conventional mortgages, but with less individual and neighborhood characteristics, Van Order and Zorn (2000) conclude that a blacker neighborhood indicates more default, though the effect is not "especially strong." In subsequent work with more explanatory variables, these authors (Van Order and Zorn 2002) then find that the higher the minority composition of a neighborhood, i.e. the higher the proportion of blacks and hispanics together, the greater is the default rate.

[21] In competitive markets, our results indicate that a person from an all-black neighborhood would pay on average over 0.2% less than a corresponding person from a non-black neighborhood, when in fact banks are found to charge the former person over 0.1% more than that paid by the latter person.

[22] If one feels that this comparison is unbalanced, since the one calculation is done between two hypothetical extremes while the other calculation goes only to the mean, it still remains notable that the direct effect of blackness on the contract rate for the typical loan is, ceteris paribus, about a quarter of the amount contributed by all forms of termination to that same contract rate.

expected value of the individual characteristic conditional on all the neighborhood characteristics actually only depends on its corresponding neighborhood characteristic, then the coefficient on neighborhood race we obtain would actually be a combination of the true neighborhood race effect and the missing individual race effect, as it impinges on that neighborhood effect. Since there can be no questions that the correlation between neighborhood and individual race is positive, this means what we have really shown is that the contract rate is raised either as a result of neighborhood racial composition or because of individual race, or because of both. While it would be useful to be able to separate the two effects in some sense it makes little difference, since we know, in all cases, that the lender is discriminating by race beyond what would be seen in neutral, competitive markets.

Possible Explanation of Lender Behavior

Our primary concern has been to see whether lenders behave as competitive markets would predict, and our results have been quite clear and compelling in this regard: they do not behave thus, but instead charge higher contract rates to black neighborhoods than could be justified or sustained in a competitive market.[23] Of course, once we know what lenders are not doing—not behaving competitively—this invites the question of what is it that they are doing. It is quite possible that they are practicing price discrimination, or perhaps simply discriminating in the laymen's sense, possibly under the false impression that loans to such neighborhoods are more costly than the evidence supports.[24] Our model is not really set up to detect the distinctions between these various possibilities. However, an intriguing hypothesis that we can examine is the idea that the lenders have consciously decided not to consider race or ethnicity in any shape or form, in particular that they do not account for its role in termination behavior, though they know full well that it is indeed a factor, at least in prepayment. This could be motivated by fear of legal or social sanctions were they found to be differentiating based even on such rational considerations, or instead their actions might be motivated by the ethical stance that this is simply the right thing to do: to ignore what they know to be true, and thus, disregard the known effects of race and ethnicity on loan behavior.[25] These last

[23] Note that one other reason why competitive markets might distinguish loans by race and ethnicity would be if these factors were associated with a difference in the severity of default. Lacking any data, we did not consider this, though the limited importance of default in general makes it seem unlikely that this could substantially explain the observed pattern of discrimination in contract rates. Van Order and Zorn (2002) do have results indicating that neighborhoods with the highest minority populations experience a slightly higher severity.

[24] If they are practicing price discrimination, it is a little surprising that it is apparently not based on economic differences, but rather on racial ones.

[25] Another possibility addresses why it is that some groups are less likely to prepay than others and suggests that this may be due either to inertia or ignorance. Alternatively, looking at imperfections on the other side of the market, it might be due to the difficulty these groups might have in refinancing. The lender then feels little pressure to reward these borrowers for their behavior, despite it being advantageous to the lender. While this would not necessarily be a sign of racial discrimination in the laymen's sense, it is certainly an indication of a lack of competition in lending. Note that while inertia in positive prepayment calls for higher not lower contract rates, there naturally is much more negative than positive prepayment, and so the inertia is advantageous to the lender overall.

explanations would involve no small amount of irony, since whatever the reason for such purposefully "color-blind" behavior, the already demonstrated result would be to charge black neighborhoods a higher contract rate than would be the case in competitive markets, given that such neighborhoods are associated with unusually low rates of prepayment.

To test the color-blind hypothesis, we ran another analysis where the p_{jt} estimation remain unchanged, but in the calculations of the expected probabilities of termination, the individual roles of *black* and *hispanic* are suppressed by setting those values to the sample mean over all loans, so that these factors no longer contribute to the predicted differences among loans.[26] If lenders were behaving in such a color-blind fashion, our hypothesis would then have to be that there is also no separate role being played by either *black* or *hispanic* in the contract rate equation, since it would seem somewhat peculiar to deliberately ignore the true role of race and ethnicity in termination behavior but then in the end charge different contract rates based on these factors. The results of this analysis are reported in the second column of Table 3. Note that *black*, which previously had been quite significant in determining the contract rate has indeed gone to being quite insignificant and that *Hispanic* has become even more insignificant than before. Other important variables have remained relatively unaffected by the change, as one would conjecture.

Conclusion

One should not conclude that impersonal, competitive markets will never consider the role of neighborhood characteristics in determining contract rates: profit-maximizing lenders will instead incorporate such factors exactly to the extent that they are associated with observable patterns of prepayment and default, which in turn influence the required contract rate. We thus examine whether this observation is adequate to account for the pattern of contract rates actually charged in our sample of mortgages. We are able to conclusively reject the hypothesis that lenders behave in a competitive fashion: instead they charge borrowers in neighborhoods that are predominantly black significantly more than can be rationalized by their subsequent termination behavior.

It would be nice to have more information on individual demographic characteristics and not just neighborhood ones. This might better serve to isolate what it is about the loans in certain neighborhoods that indicate to lenders that higher contract rates will be applied, but seems unlikely to alter our found results that individuals in neighborhoods with a larger proportion of blacks typically pay higher contract rates than can be justified by the actions of such individuals.

Much informal discussion concentrates on default and the possible association of race and ethnicity with the likelihood of such default. Our results suggest that much

[26] We chose not to report the more obvious alternative of rerunning the logit regressions with *black* and *hispanic* absent. In that case, the regression would attempt to account for the missing variables' true effect by attributing as much of it as possible to those covariates still present that are correlated with the missing ones. This, in effect, has the regression attempting to undo the efforts of the lender to remove the effect of the two variables, though one could argue that this is indeed appropriate in the case where the lender is merely seeking to avoid the appearance of discrimination. The results are similar in either case.

of this discussion is misplaced. Firstly, we find no significant effect at all of race and ethnicity on default, and secondly, while we find that, as one would expect, a given probability of default has a greater effect on contract rates than does the same probability of prepayment, the likelihood of default is so much lower than that of prepayment that its effect on contract rates pales in comparison. On the other hand, race and ethnicity of neighborhood do have a statistically significant effect on prepayment, but since they serve to lower the probability of prepayment, not raise it, this lowers rather than raises the contract rates that should be required in neighborhoods that are predominantly black or latino. Lenders consistently fail to provide this compensation, choosing instead to charge markedly higher rates, compared to otherwise similar areas, as a neighborhood becomes increasingly black.

References

An, X., Clapp, J. M., & Deng, Y. (2006). *Omitted mobility characteristics and property market dynamics: application to mortgage termination*. Lusk Center for Real Estate Working Paper, U.S.C.

Archer, W. R., Ling, D. C., & McGill, G. A. (2002). Prepayment risk and lower-income mortgage borrowers. In N. P. Retsinas & E. S. Belsky (Eds.), *Low-income homeownership: Examining the unexamined goal*. Washington: Brookings Institution.

Berkovec, J. A., Canner, G., Gabriel, S., & Hannan, T. (1994). Race, redlining and residential mortgage loan performance. *Journal of Real Estate Finance and Economics, 9*, 263–294.

Berkovec, J. A., Canner, G., Gabriel, S., & Hannan, T. (1998). Discrimination, competition, and loan performance on FHA mortgage lending. *Review of Economics and Statistics, 80*(2), 241–250.

Black, H. A., Boehm, T. P., & DeGenarro, R. P. (2003). Is there discrimination in mortgage pricing? The case of overages. *Journal of Banking and Finance, 27*(6), 1139–1165.

Boehm, T. P., Thistle, P. D., & Schlottmann, A. (2006). Rates and race: an analysis of racial disparities in mortgage rates. *Housing Policy Debate, 17*(1), 109–149.

Clapp, J. M., Goldberg, G. M., Harding, J. P., & LaCour-Little, M. (2001). Movers and Shuckers: interdependent prepayment decisions. *Real Estate Economics, 29*(3), 411–450.

Clapp, J. M., Deng, Y., & An, X. (2006). Unobserved heterogeneity in models of competing mortgage termination risks. *Real Estate Economics, 34*(2), 243–273.

Cotterman, R. F. (2001). *Neighborhood effects in mortgage default risk*. HUD Report.

Courchane, M., & Nickerson, D. (1997). Discrimination resulting from overage practices. *Journal of Financial Services Research, 11*, 133–151.

Crawford, G. W., & Rosenblatt, E. (1999). Differences in the cost of mortgage credit: implications for discrimination. *Journal of Real Estate Finance and Economics, 19*(2), 147–159.

Deng, Y., & Gabriel, S. A. (2006). Risk-based pricing and the enhancement of mortgage credit availability among underserved and higher credit-risk populations. *Journal of Money, Credit, and Banking, 38*(6), 1431–1460.

Deng, Y., Quigley, J. M., & Van Order, R. (2000). Mortgage terminations, heterogeneity and the exercise of mortgage options. *Econometrica, 68*(2), 257–307.

Duca, J. V., & Rosenthal, S. S. (1994). Do mortgage rates vary based on household default characteristics? Evidence on rate sorting and credit rationing. *Journal of Real Estate Finance and Economics, 8*, 99–113.

Dymski, G. (2005). Discrimination in the credit and housing markets: findings and challenges. In W. Rogers (Ed.), *Handbook on discrimination*. (forthcoming).

Gabriel, S. A. (1996). The role of FHA in the provision of credit to minorities. In J. Goering & R. Wienk (Eds.), *Mortgage lending, racial discrimination, and federal policy*. Washington: The Urban Institute.

Getter, D. E. (2006). Consumer credit risk and pricing. *The Journal of Consumer Affairs, 40*(1), 41–63.

Geyer, A. L. J., & Pichler, S. (1999). A state-space approach to estimate and test multifactor Cox-Ingersoll-Ross models of the term structure. *Journal of Financial Research, 22*(1), 107–130.

Han, S. (2004). Discrimination in lending: theory and evidence. *Journal of Real Estate Finance and Economics, 29*(1), 5–46.

Holmes, A., & Horvitz, P. (1994). Mortgage redlining: race, risk and demand. *Journal of Finance, 49*(1), 81–99.

Kau, J. B., Keenan, D. C., & Smurov, A. (2004). *Reduced-form mortgage evaluation*. University of Georgia Working Paper.

Kelly, A. (1995). Racial and ethnic disparities in mortgage prepayment. *Journal of Housing Economics, 4*, 350–372.

Lacour-Little, M. (1999). Discrimination in mortgage lending: a critical review of the literature. *Journal of Real Estate Finance and Economics, 7*, 15–49.

Ladd, H. F. (1998). Evidence on discrimination in mortgage lending. *Journal of Economic Perspectives, 98*(Spring), 41–62.

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

Nothaft, F. E., & Perry, V. G. (2002). Do mortgage rates vary by neighborhood? Implications for loan pricing and redlining. *Journal of Housing Economics, 11*, 244–265.

Petrin, A., & Train, K. (2003). *Omitted product attributes in discrete choice models*. N.B.E.R. Working Paper 9452.

Ross, S. R. (2006). *The continuing practice and impact of discrimination*. University of Connecticut Working Paper.

Ross, S. R., & Yinger, J. (2002). *The color of credit*. Cambridge: MIT.

Tootell, G. M. B. (1996). Redlining in Boston: do mortgage lenders discriminate against neighborhoods? *Quarterly Journal of Economics, 111*(4), 1049–1079.

Turner, M. A., & Skidmore, F. (Eds.). (1996). *Mortgage lending discrimination: A review of existing evidence*. Washington: The Urban Institute.

Van Order, R., & Zorn, P. (2000). Income, location and default: some implications for community lending. *Real Estate Economics, 28*, 385–404.

Van Order, R., & Zorn, P. (2002). The performance of low income and minority mortgages. In N. P. Retsinas & E. S. Belsky (Eds.), *Low-income homeownership: Examining the unexamined goal*. Washington: Brookings Institution.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge: MIT.

Wooldridge, J. M. (2003). *Introductory econometrics, 2nd Edition*. Southwestern, Mason, Ohio.

Yinger, J. (1996). Discrimination in mortgage lending: A literature review. In J. Goering & R. Wienk (Eds.), *Mortgage lending, racial discrimination, and federal policy*. Washington: The Urban Institute.