The Pricing of Home Mortgage Loans to Minority Borrowers: How Much of the APR Differential Can We Explain?

Author Marsha J. Courchane

Abstract

The public releases of the 2004 and 2005 HMDA data have engendered a lively debate over the pricing of mortgage credit and its implications regarding the treatment of minority mortgage borrowers. This research uses aggregated proprietary data provided by lenders and an endogenous switching regression model to estimate the probability of taking out a subprime mortgage, and annual percentage rate (APR) conditional on getting either a subprime or prime mortgage. The findings reveal that up to 90% of the African American APR gap, and 85% of the Hispanic APR gap, is attributable to observable differences in underwriting, costing, and market factors that appropriately explain mortgage pricing differentials. Although any potential discrimination is problematic and should be addressed, the analysis suggests that little of the aggregate differences in APRs paid by minority and non-minority borrowers are appropriately attributed to differential treatment.

Home Mortgage Disclosure Act (HMDA) reporting requirements were changed in 2004 to require lenders for the first time to provide information on the difference between the annual percentage rate (APR) and a comparable Treasury rate for all loans above a rate spread-reporting threshold.[1] The Federal Reserve Board publicly released these initial data in September 2005. Similar data for loans originated in 2005 were released in September 2006.

The 2004 and 2005 HMDA data show that loans to minority borrowers have a substantially higher probability of exceeding the rate-spread reporting hurdle, suggesting that, on average, minority borrowers pay higher APRs for their mortgages than do non-minority borrowers. This observation has engendered a lively debate over its possible implications regarding differential treatment.[2] Partly on the basis of these data, regulatory agencies have referred several pricing cases to the Department of Justice. In a separate action, Countrywide, one of the nation's largest lenders, reached a settlement with the New York Office of the Attorney General based on investigation of Countrywide's pricing practices.[3]

This paper provides a unique perspective on this debate because of the access to data not generally available to the public. Specifically, a number of lenders have allowed pooling of their records into an anonymous data set that includes detailed loan-level information on their 2004 and 2005 originations. These data allow exploration of the pricing of mortgage loans across minority and non-minority borrowers at a deeper level of detail than is possible with HMDA data alone. In particular, the data allow rough approximation of the work of regulators as they analyze apparently unexplained pricing differentials uncovered in the HMDA data.

Regulators use HMDA data to allocate efficiently their investigative activities—attention is focused typically on lenders with relatively larger pricing differentials that cannot be fully explained by HMDA data alone. However, once lenders are targeted for additional reviews, regulators significantly modify how they conduct their investigations. First, rather than restricting their attention to rate-spreads for loans above the rate-spread reporting hurdle, regulators explicitly focus on mortgage prices (generally APRs) for all loans.[4] Second, regulators dramatically increase the control variables used in their statistical analyses, including variables used in underwriting loans, as well as variables used to account for market factors that can affect pricing outcomes.

The data uniquely allow approximation of this second-level analysis of regulators, albeit at a national level using pooled data rather than separately for individual lenders.[5] Specifically, the data include APR, typical underwriting variables such as loan-to-value ratios (LTV), debt-to-income ratios (DTI), and FICO scores, as well as other variables that likely affect mortgage pricing, such as documentation status, the existence of prepayment penalties, and sourcing channel. Identification of whether the lenders in the sample operate primarily in the subprime or prime markets is also possible.[6]

The data show that there are large pricing differentials to explain. For example, the difference between the mean APR of African American borrowers and the mean APR of White non-Hispanic borrowers is 120 basis points in 2004 and 128 basis points in 2005, and the equivalent difference between Hispanic and White non-Hispanic borrowers is 56 basis points in 2004 and 74 basis points in 2005. Not surprisingly, APRs are substantially higher in the subprime market, and much of the explanation for why minority borrowers tend to have higher APRs than non-minority borrowers is because minority borrowers disproportionately take out subprime loans.

An endogenous switching regression model is used to estimate APR separately in the prime and subprime markets because the supply and, arguably, the demand functions in the two markets are quite different.[7] A full set of explanatory variables is used to largely explain the differential tendency of minority borrowers to obtain subprime loans. Up to all but about 10 basis points of the difference in mean APRs between African American, Hispanic, and White non-Hispanic borrowers can also be explained.

The explanatory variables included in the models are by necessity parsimonious, because commonality is required in all the data elements across lenders. Additional data including a complete set of appropriate explanatory variables might even more fully explain differences in minority and non-minority mortgage outcomes. This suggests that widespread and consistent differential treatment of African American and Hispanic mortgage borrowers is highly unlikely to be the primary cause of the observed raw differences in mean APR paid by minority and non-minority borrowers.

Previous Research

The uniqueness of the data means that no previous researchers have precisely considered the questions addressed in this paper. However, there have been several studies that have looked at the question of whether minority borrowers pay differentially higher prices for mortgage loans. For example, two recent Federal Reserve Board Bulletin articles provide a comprehensive discussion of the newly expanded 2004 and 2005 HMDA data.[8] The authors use HMDA and a secondary data source to conclude that much, but not all of the minority and non-minority differences in rate-spread reporting incidence can be explained by observable factors associated with underwriting and market dynamics. The authors' data constrains their analysis to focus on the incidence of rate-spread reportable mortgages, and the severity of the rate-spread conditional on incidence, rather than directly analyzing mortgage prices or APRs.

Two other recent papers analyze the newest HMDA data. The Center for Responsible Lending (CRL) released a recent study that measures the effect of race and ethnicity on the price of subprime mortgages.[9] The CRL study uses 2004 HMDA data merged with a proprietary data set, resulting in a sample of about 177,000 subprime loans. Their analysis includes all the HMDA variables, as well as such credit-related variables as FICO scores, LTVs, documentation status, and term to maturity. The CRL study finds that both African American and Hispanic borrowers are more likely to obtain higher-priced loans. Again, however, the focus is on the incidence of reported rate-spreads.

The Consumer Federation of American (CFA) also studied mortgage pricing, releasing in December of the past two years analyses that find that women, particularly African American women, are more likely to obtain subprime loans, and thus pay higher rates.[10] The CFA study relies solely on HMDA, and therefore includes no lender- or borrower-specific underwriting or mortgage information. Nor do they have available APRs.

Several papers in the past few years have also focused on borrowers ''choice'' of prime versus subprime mortgages. Courchane, Surette, and Zorn (2004) find that this choice is influenced significantly by non-financial factors, and that there exists considerable ability to transition between prime and subprime markets as life

events or credit events vary. Pennington-Cross (2003) finds that credit risk factors are the most important determinants of market choice. LaCour-Little (2006) focuses on the home purchase preferences of low- and moderate-income borrowers, and finds that individual credit characteristics and financial factors drive mortgage product choice. None of these three studies focuses on pricing or APR.

This paper combines and extends these two, related strains of research. Using an endogenous switching regression framework, the probability of borrowers getting mortgages in the prime or subprime markets, and their APRs conditional on the market in which they obtain mortgage loans, are simultaneously estimated. The data include APR, as well as an unusually rich set of explanatory variables that likely affect lenders' underwriting and pricing decisions.

Data

Despite its recent enhancement, regulators recognize that the HMDA data need extensive supplementation in order to accurately assess whether or not similarly situated borrowers are treated similarly. As part of fair lending investigations, the regulatory agencies have the right to examine all of the relevant underwriting and pricing data pertinent to mortgage lending decisions, as well as all the necessary information about borrower characteristics. The data from fair lending examinations are protected by examination privilege and, typically, do not make it into the research arena.

Lenders generated the proprietary data used here in the analysis for fair lending purposes. Restrictions on the use of the data are designed to shield the identity of borrowers and lenders, but the data do include loan-level pricing and underwriting factors rarely available to researchers. These data provide for the first time, therefore, an opportunity to conduct a large scale and relatively complete study of the potential differential treatment of minority mortgage borrowers.

Lenders' motivations for collecting and providing these data vary, but range from proactive concerns regarding the fair lending integrity of their business operations to the necessity of responding to fair lending inquiries by federal regulators. The data include over one million loan-level records of originations in 2004 and 2005, from a total of 22 lenders/subsidiaries (nine prime and nine subprime lenders/subsidiaries for mortgages originated in 2004, and eight prime and eight subprime lenders/subsidiaries for mortgages originated in 2005). The lenders and subsidiaries in the data self-identify as either prime or subprime, so markets can be differentiated at a lender- or subsidiary-level rather than at the mortgage loan-level more typical of most recent HMDA analysis. The choice in this matter reflects the continuing difference in prime and subprime market dynamics, and the fact that lenders or their subsidiaries almost always specialize in one or the other of the two markets.[11] As a consequence, all mortgages originated by a self-identified subprime lender are considered to be subprime mortgages while all

mortgages originated by a self-identified prime lender are considered to be prime mortgages.

The analyses are limited to first lien, single-family, conventional purchase-money and refinance mortgages.[12] The analyses are also limited to African American, Hispanic, and White non-Hispanic borrowers. To ensure that these categories are mutually exclusive, and to allow for the analysis of marginal impacts of race/ethnicity, African American borrowers are restricted to those that have a non-Hispanic ethnicity and exclude from Hispanic borrowers those with a non-White race.

The proprietary lender data include all HMDA data elements, as well as a variety of loan specific underwriting and pricing characteristics. Specifically, data are available for APR, borrower race/ethnicity, borrower gender, FICO score, LTV, DTI, loan amount, contract rate type (fixed or adjustable), loan purpose (purchase money or refinance), occupancy status (owner-occupied or not owner-occupied), documentation status (full documentation or not full documentation), existence of a prepayment penalty, loan term to maturity, and origination channel (wholesale or retail). In addition, the data are appended to select census and aggregated HMDA variables at a state-county-tract level, including the tract-level Herfindahl-Hirschman Index (HHI), tract-level denial rate, tract-level educational distribution, tract-level percentage owner-occupied, and tract-level percentage of subprime originations in the previous year.[13]

The data come from a ''convenience'' sample rather than a random sample. Despite the non-random sampling of the data construction, the distributions of key variables are reasonably consistent with those of the full HMDA data in the equivalent years.[14] Exhibit 1 shows these raw comparisons.[15] Note, for example, that the percentage minority is within about four percentage points of that in HMDA, although African American borrowers tend to be over-represented in the subprime population while Hispanic borrowers are generally somewhat under-represented. The loan purpose and occupancy status distributions are very similar in both data, generally within two percentage points of each other. The income and loan amount distributions are also reasonably similar, but with wider variance across the two data. There is a somewhat larger difference, however, in the rate-spread distributions. Specifically, subprime lenders are more likely than the HMDA subprime lender population to originate rate-spread reportable loans, while prime lenders originate fewer rate-spread reportable loans than do average HMDA prime lenders. This likely results from having subprime lenders in the data that are more specialized in subprime lending than is generally the case in the HMDA population.

Post-sampling weights are constructed to increase the confidence with inference is drawn to the HMDA population, and then they are applied to the proprietary lender data throughout the analyses. To create these weights, 576 mutually exclusive cells were created by interacting discrete categories of variables common to both sets of data; specifically, two market (prime and subprime), three race/

Exhibit 1 | Comparison of HMDA and Proprietary Lender Data

		2004				2005			
		Subprime Loans		Prime Loans		Subprime Loans		Prime Loans	
Variable Name	Variable Value	HMDA	Lender Data	HMDA	Lender Data	HMDA	Lender Data	HMDA	Lender Data
Race/Ethnicity	African American	17.11%	19.43%	5.83%	4.32%	19.68%	24.24%	6.55%	5.76%
	Hispanic	14.77%	12.38%	7.82%	5.98%	19.45%	19.52%	9.77%	8.39%
	White non-Hispanic	68.12%	68.19%	86.36%	89.70%	60.88%	56.24%	83.68%	85.85%
Income	Income $<$ \$20,000	1.95%	1.76%	1.43%	0.93%	1.38%	0.92%	1.21%	1.21%
	\$20,000 $\leq$ Income $\leq$ \$200,000	93.64%	91.40%	86.72%	79.54%	94.25%	95.60%	85.41%	88.02%
	\$200,000 $<$ Income	2.65%	2.69%	6.35%	8.04%	3.08%	3.41%	7.43%	6.83%
	Unknown Income	1.76%	4.15%	5.50%	11.49%	1.29%	0.08%	5.95%	3.93%
Loan Amount	Loan Amount $<$ \$100,000	30.65%	27.64%	25.96%	23.63%	24.49%	20.72%	21.16%	21.94%
	\$100,000 $\leq$ Loan Amount $<$ Loan Limit	61.72%	63.87%	64.18%	64.03%	66.46%	68.72%	67.41%	67.04%
	Loan Limit $\leq$ Loan Amount	7.63%	8.49%	9.86%	12.34%	9.05%	10.56%	11.43%	11.02%
Loan Purpose	Purchase Money	32.98%	34.63%	45.12%	47.66%	39.02%	46.53%	49.80%	48.26%
	Refinance	67.02%	65.37%	54.88%	52.34%	60.98%	53.47%	50.20%	51.74%
Occupancy Status	Owner-Occupied	90.31%	90.99%	87.64%	89.86%	91.10%	91.79%	85.58%	86.58%
	Not Owner-Occupied	9.69%	9.01%	12.36%	10.14%	8.90%	8.21%	14.42%	13.42%
Reported Rate Spread	No	49.33%	38.34%	96.10%	98.75%	17.47%	7.40%	91.61%	96.45%
	Yes	50.67%	61.66%	3.90%	1.25%	82.53%	92.60%	8.39%	3.55%

ethnicity (African American, Hispanic, and White non-Hispanic), four income (less than $20,000, between $20,000 and $200,000, greater than $200,000, and unknown), three loan amounts (less than $100,000, between $100,000, and the conforming limit, and greater than the conforming limit), two loan purposes (purchase and refinance), two owner-occupancy (owner-occupant and investor), and two reported rate-spreads (yes and no) categories.[16] The distributions of the proprietary lender and HMDA mortgages over each of the mutually exclusive buckets are then calculated and used to create post-sampling weights by dividing the percentage of the HMDA data in each cell by the percentage of the proprietary data in the equivalent cell. Finally, the appropriate post-sampling weight for each of the individual buckets is applied to all mortgages in that bucket.[17]

The first data step is to use these post-sampling weights to provide the basic comparisons that motivate the study. Unlike HMDA, the data allow direct comparison of the APRs of all mortgages originated by the included lenders. Therefore, four market outcomes (mean APR, percentage with a subprime loan, mean subprime APR, and mean prime APR) are shown for each of the borrower race/ethnicity categories, and the African American and Hispanic outcomes are compared to those of White non-Hispanic borrowers.

Exhibit 2 provides these comparisons, and clearly shows that minority borrowers pay higher APRs than non-minority borrowers. In 2004, the mean APR of African American borrowers is 120 basis points above White non-Hispanic borrowers and the mean APR of Hispanic borrowers is 56 basis points above White non-Hispanic borrowers.[18] In 2005, these differences grow to 128 basis points above White non-

Exhibit 2 | Outcome Comparisons Without Controls

Year	Borrower Race/Ethnicity	Mean APR	Percentage with Subprime Loans	Mean Subprime APR	Mean Prime APR
	African American	7.22	43.60%	8.48	6.25
	Hispanic	6.58	33.26%	7.86	5.94
	White non-Hispanic	6.03	17.22%	7.83	5.65
2004	African American minus White non-Hispanic	1.20	26.38%	0.65	0.60
	Hispanic minus White non-Hispanic	0.56	16.04%	0.03	0.29
	African American	7.99	44.54%	9.57	6.72
	Hispanic	7.45	34.73%	9.24	6.50
	White non-Hispanic	6.71	16.28%	9.07	6.25
2005	African American minus White non-Hispanic	1.28	28.26%	0.50	0.47
	Hispanic minus White non-Hispanic	0.74	18.44%	0.17	0.25

Hispanic borrowers for African American borrowers, and 74 basis points above White non-Hispanic borrowers for Hispanic borrowers.

These APR differences can be decomposed into two factors. Mean APRs are substantially higher in the subprime market—generally on the order of about 200 basis points in 2004 and about 275 basis points in 2005—and minority borrowers disproportionately obtain loans in the higher priced market. For example, African American and Hispanic borrowers are 26.38 and 16.04 percentage points, respectively, more likely than White non-Hispanic borrowers to obtain a subprime mortgage in 2004, and 28.26 and 18.44 percentage points more likely in 2005. This disproportionate propensity for minority borrowers to obtain loans in the higher priced subprime market explains roughly one-half the APR difference between minority and non-minority borrowers in 2004, and roughly two-thirds the difference in 2005.[19]

The higher APRs paid by minority borrowers in both the subprime and prime markets is the other factor explaining the overall higher APRs paid by minority borrowers. For example, in 2004, the mean APR of African American and Hispanic subprime borrowers is 65 basis points and 3 basis points, respectively, above that of White non-Hispanic borrowers, and in the prime market it is above White non-Hispanic borrowers by 60 basis points and 29 basis points, respectively. In 2005, these differences decline for African American borrowers but increase for Hispanic borrowers. Specifically, in 2005 the subprime market differences are 50 basis points and 17 basis points, respectively, for African American and Hispanic borrowers, and 47 basis points and 25 basis points, respectively, in the prime market.

The substantially higher mortgage prices paid by minority borrowers are troubling, regardless of why they occur. Crafting an appropriate public policy response, however, depends critically on causality. The extent to which differential treatment of minority borrowers may contribute to these outcomes is discussed in the remainder of the paper.

Modeling and Econometric Issues

Simplistically stated, fair lending laws applicable to mortgage lending require that similar borrowers be treated similarly (concerns of disparate treatment), and that there be a valid business justification for any similar treatment that differentially impacts minority borrowers (concerns of disparate impact).[20] Outcomes such as those observed in Exhibit 2 raise potential concerns about both issues.

Determining the potential role of disparate impact in explaining these results requires comparing APRs to business performance measures such as defaults, losses or profitability, and assessing whether the higher APRs paid by minority borrowers are justified by appropriate business necessity.[21] A full disparate impact assessment, therefore, requires business performance variables not included in the

data. As a consequence, the analysis cannot definitively check for all forms of potential discrimination.

Moreover, although the data are loan-level, they combine in an anonymous fashion loans from many lenders. The analysis, therefore, focuses on aggregate market impacts and not on potential discriminatory actions or outcomes of individual lenders or borrowers. While discrimination is problematic even if it affects only a single borrower, the task of assessing such individual impacts is left to others.

An endogenous switching regression framework is used to model the determination of mortgage market type (subprime or prime), and the determination of APR conditional on market. Applying these models to the data, equations are simultaneously estimated for the probability of getting a subprime mortgage and for APR conditional on getting a subprime or prime mortgage. These estimates are used to predict expected APR for minority and non-minority borrowers, controlling for observable differences in borrower and loan characteristics. Differences in mean mortgage outcomes between minority and non-minority borrowers explained by the analysis are unassociated with discrimination. The residual unexplained differences in APR, therefore, provide an upper bound for the extent that possible differential treatment explains the higher prices paid by minority mortgage borrowers.

An endogenous switching regression framework is used to account for the likelihood that unobserved factors affecting whether borrowers obtain a subprime or prime mortgage also affect the APRs borrowers receive in each market (such factors might include, for example, borrowers' financial literacy or propensity to shop for credit, and market segmentation or ''steering'' on the part of lenders). Specifically, the endogenous switching regression framework appropriately accounts for the possibility that the error term in the subprime/prime market choice equation is correlated with the error terms in the subprime and prime APR equations.[22]

More concretely, borrowers' tendency to get subprime mortgages is assumed to be characterized by an underlying response variable, y^*, and that the following regression relationship describes this underlying response:

$$y^* = MODEL_1 + u, \qquad (1)$$

where $MODEL_1$ is an unspecified function and u is an $N(0,1)$ distributed error.[23] In practice, y^* is not observed; we observe only whether borrowers take out mortgages in the prime or subprime market. This process is modeled through the use of an indicator variable, y, having the characteristic that $y = 1$ (borrowers take out a subprime loan) when $y^* > 0$, and $y = 0$ (borrowers take out a prime

loan) when $y^* \leq 0$. The probability of borrowers taking out subprime or prime mortgages can therefore be stated, respectively, as follows:

$$P(subprime) = P(y = 1) = P(y^* > 0) = P(u > -MODEL_1) = 1 - \Phi(-MODEL_1), \quad (2)$$

$$P(prime) = P(y = 0) = P(y^* \leq 0) = P(u \leq -MODEL_1) = \Phi(-MODEL_1), \quad (3)$$

where $\Phi(-MODEL_1)$ is the standard normal cumulative density function evaluated at $-MODEL_1$.

It is further assumed that there are separate underlying functions for the determination of APR in the subprime and prime markets, and that the regression relationships for these underlying functions, APR_2 and APR_3 respectively, are given by the following equations:

$$APR_2 = MODEL_2 + \varepsilon_2, \quad (4)$$

$$APR_3 = MODEL_3 + \varepsilon_3, \quad (5)$$

where $MODEL_2$ and $MODEL_3$ are unspecified functions, and ε_2 and ε_3 are $N(0, \sigma_2^2)$ and $N(0, \sigma_3^2)$ distributed errors, respectively.

In practice, APR_2 is only observed for borrowers taking out subprime mortgages and APR_3 for borrowers taking out prime mortgages. As noted above, the error term in the subprime/prime market determination equation we assume to be correlated with the error terms in the APR equations, and define σ_{2u} and σ_{3u} as the respective covariances of ε_2 and ε_3 with u. As a consequence of these correlations, the expected values of APR for borrowers in the subprime and prime markets are defined, respectively, as follows:

$$E(APR_2|u > -MODEL_1) = MODEL_2 + E(\varepsilon_2|u > -MODEL_1) = MODEL_2 + \sigma_{2u}\frac{\phi(-MODEL_1)}{1 - \Phi(-MODEL_1)}, \quad (6)$$

and

$$E(APR_3|u \leq -MODEL_1) = MODEL_3 + E(\varepsilon_3|u \leq -MODEL_1)$$
$$= MODEL_3 - \sigma_{3u}\frac{\phi(-MODEL_1)}{\Phi(-MODEL_1)}, \quad (7)$$

where $\phi(-MODEL_1)$ is the Standard Normal density function evaluated at $-MODEL_1$.

Note that the assumed correlation of the errors in Equations (4) and (5) with the error in Equation (1) adds the familiar inverse Mills ratio terms to the expected APR functions of Equations (6) and (7). Recall further that, by assumption, $\sigma_u = 1$, so the covariances σ_{2u} and σ_{3u} can be decomposed into $\sigma_2\rho_{2u}$ and $\sigma_3\rho_{3u}$, respectively, where ρ_{2u} and ρ_{3u} are the respective correlation coefficients. The impacts of the inverse Mills ratio terms, therefore, depend directly on the extent of the correlation in the errors. At the extreme, if the errors are uncorrelated (i.e., if $\rho_{2u} = \rho_{3u} = 0$), Equations (6) and (7) reduce to $MODEL_2$ and $MODEL_3$, respectively, and the subprime and prime APR equations can appropriately be estimated independently of the market determination equation.

Finally, borrowers' expected unconditional APR can be written as a weighted average of borrowers' expected conditional APRs in the prime and subprime markets, as follows:

$$E(APR) = E(APR_2|u > -MODEL_1)P(subprime)$$
$$+ E(APR_3|u \leq -MODEL_1)P(prime)$$
$$= \left[MODEL_2 + \sigma_{2u}\frac{\phi(-MODEL_1)}{1 - \Phi(-MODEL_1)}\right]$$
$$(1 - \Phi(-MODEL_1))$$
$$+ \left[MODEL_3 - \sigma_{3u}\frac{\phi(-MODEL_1)}{\Phi(-MODEL_1)}\right]\Phi(-MODEL_1). \quad (8)$$

Estimating this system requires specifying the functional relationships of the three models. There are four key issues affecting determination of these specifications. The first is the fact that there is in each part of the APR determination process the potential for differential treatment, although it cannot directly be observed if any differential treatment occurs. The strategy, therefore, is to model each outcome as a function of observable characteristics in a manner that allows for the

possibility of differential treatment. It is assumed that any differential treatment toward minority borrowers can appropriately be captured solely by dummy variables identifying borrower race/ethnicity in an estimation over all borrowers. Ross and Yinger (2002), among others, have argued that differential treatment can manifest itself in relatively complex interactions with race/ethnicity. Appropriately accounting for this possibility suggests that estimations assessing differential treatment should be more generally structured so as to allow all coefficients to vary over borrower race/ethnicity, not just the constant term. However, the qualitative results are unchanged by this more general specification, so for ease of exposition, the more simplified approach is presented.[24]

The second modeling issue is that the explanatory variables included in the true outcome models are unknown. This presents a particular challenge because it is also reasonable to assume that many, if not most, explanatory variables typically considered to be members of the true mortgage outcome models are correlated with race/ethnicity, which creates an unusually explicit tension between over-excluding and over-including variables in the specifications of the models being estimated. For example, omitting variables that are part of the true model could increase the size of the estimated parameters on the race/ethnicity dummies (if the omitted variables are correlated with race/ethnicity), and so overstate the potential for differential treatment. On the other hand, including variables not part of the true model could reduce the size of the race/ethnicity dummy variables (if the included variables are correlated with race/ethnicity), and so understate the potential for differential treatment. The qualitative results turn out to be relatively insensitive to the inclusion or exclusion of variables likely subject to the concerns above.[25] As a consequence, this paper presents only a single model that is relatively inclusive in its choice of explanatory variables.

The third modeling issue concerns system identification. Identification of the switching regression system is assured by the non-linearity introduced by the standard normal function in the market determination probabilities. Identification is enhanced, however, by also including variables in $MODEL_1$ that are excluded in both $MODEL_2$ and $MODEL_3$.[26]

The fourth concern is the potential bias in the estimated coefficients caused by the likely endogeneity of some of the explanatory variables. Specifically, rather than the explanatory variables ''causing'' APR as the relationship is modeled here, APR may determine the explanatory variables. As has been noted by Barth, Cordes, and Yezer (1980) and then explored more deeply in Yezer, Phillips, and Trost (1994), unless explicitly accounted for, the presence of this sort of endogeneity will bias parameter estimates. The robustness tests suggest, however, that any such bias has relatively little practical impact on the conclusions.[27]

Consistent with the above discussion, models are specified for each of the $i = 1, 2, 3$ regression relationships and the $j = AA, H, WNH$ race/ethnicity subgroups as follows:

$$MODEL_{1j} = \alpha_{1j} + X_j\beta_1 + Z_j\Gamma_1, \quad (9)$$

$$MODEL_{2j} = \alpha_{2j} + X_j\beta_2, \quad (10)$$

$$MODEL_{3j} = \alpha_{3j} + X_j\beta_3, \quad (11)$$

where α_{ij} is a constant term for i^{th} regression relationship and j^{th} race/ethnicity subgroup, β_i is a parameter vector in the i^{th} regression relationship (identical across the j race/ethnicity subgroups) for variables common across all three regression relationships, Γ_1 is a parameter vector for variables unique to $MODEL_1$ (identical across the j race/ethnicity subgroups), X_j is a matrix of explanatory variables common across all models with observations restricted to the j^{th} race/ethnicity subgroup, and Z_j is a matrix of explanatory variables unique to $MODEL_1$ with observations restricted to the j^{th} race/ethnicity subgroup.

The endogenous switching regression system is estimated using the full information maximum likelihood procedure incorporated into the ''movestay'' command of Stata.[28] Robust standard errors are calculated to account for possible heteroscedasticity in the error terms.

The next step after obtaining the switching regression estimates is to assess differences in mortgage outcomes across borrower race/ethnicity while controlling for observable differences across borrowers. The approach to this process is to use the estimation results to predict average outcomes for African American and Hispanic borrowers, and to separately compare these predicted outcomes to those of (hypothetical) White non-Hispanic borrowers who have identical observable characteristics as the African American and Hispanic borrowers.[29]

Noting that White non-Hispanic is the omitted borrower race/ethnicity category, and therefore that $\alpha_{1WNH} = \alpha_{2WNH} = \alpha_{3WNH} = 0$, the regression relationships for White, non-Hispanic borrowers with African American and Hispanic characteristics ($j = AA, H$) can be specified as follows:

$$MODEL_{1WNH|X_j,Z_j} = X_j\beta_1 + Z_j\Gamma_1, \quad (12)$$

$$MODEL_{2WNH|X_j} = X_j\beta_2, \quad (13)$$

$$MODEL_{3WNH|X_j} = X_j\beta_3. \quad (14)$$

Comparing expected APRs of minority and non-minority borrowers in the subprime and prime markets, therefore, reduces to a comparison of the race/ethnicity dummy variable coefficients, plus differences in the inverse Mills ratio terms of Equations (6) and (7). For example, when comparing expected subprime APRs of African American and Hispanic borrowers to those of White non-

Hispanic borrowers with identical characteristics, $MODEL_{2AA} - MODEL_{2WNH|X_{AA}}$ and $MODEL_{2H} - MODEL_{2WNH|X_H}$ reduce, respectively, to α_{2AA} and α_{2H}. Moreover, the differences in the nonlinear inverse Mills ratio components of each of these comparisons reduce to zero if observable characteristics fully explain the raw differences in the probabilities of minority and non-minority borrowers taking out subprime mortgages (i.e., when $\alpha_{1AA} = \alpha_{1H} = \alpha_{1WNH} = 0$). When this is the case, therefore, a comparison of differential APRs in the subprime and prime markets can be conducted independently of a market determination analysis.

Empirical Analysis

The proprietary lender data allows inclusion of an unusually complete set of underwriting and mortgage pricing variables. Specifically, the explanatory variables include the key underwriting factors of FICO score, LTV, and DTI. These variables are assumed to directly affect credit risk (lower FICO scores and higher LTVs and DTIs are assumed to be associated with higher credit risk), and therefore affect market determination (higher credit risk is hypothesized to be associated with a greater probability of taking out a subprime mortgage) and APR conditional on market (higher credit risk is hypothesized to be associated with higher APRs). Additional variables contract rate type, loan purpose, occupancy status, and documentation type, which often are said to be secondarily associated with mortgage underwriting decisions. The hypothesis is that adjustable rate mortgages, refinance mortgages, non-owner-occupied properties, and less than full documentation all are associated with higher credit risk, and so result in higher APRs and a higher probability of taking out subprime mortgages.

In addition to underwriting variables, several variables that are commonly assumed to affect mortgage pricing are included. Both loan origination costs and default losses, for example, are assumed to have large fixed cost components, and so loan amount is included as an explanatory variable. In addition, smaller loan amounts are assumed to be associated with a greater probability of taking out subprime mortgages and higher APRs. On the other hand, prepayment penalties, shorter loan terms, and retail (as opposed to wholesale) originations are assumed to be associated with lower costs and therefore lower APRs.

Several other factors that measure local market dynamics are also included. A Herfindahl- Hirschman Index (HHI) is constructed using tract-level HMDA data to measure the extent of market competition in the neighborhoods in which borrowers' properties are located.[30] Greater competition (lower HHI values) is expected to be associated with lower APRs.[31] A tract-level denial percentage is also calculated from the HMDA data. The expectation is that the costs of origination are higher in neighborhoods with higher denial percentages, and so APRs will be higher as well. In addition, state dummy variables and monthly calendar dummy variables are included to account for variations in state laws and fluctuations in the yield curve that could affect market dynamics and/or pricing.

Several variables are also included in the market determination model that are excluded from the APR models in an effort to aid identification.[32] Specifically, it is assumed that the more financially knowledgeable are the borrowers, the less likely are they to take out a subprime mortgage, all things equal. Financial knowledge/literacy is measured using two proxy variables, both of which are created from the 2000 census and measured at the tract level. The first is the population distribution of educational attainment (percentage of population with less than a high school degree, percentage completed high school, percentage attended some college, percentage completed college) and the second is the percentage of owner-occupied households in the tract. Borrowers are assumed to interact with and seek advice from their neighbors, and that formal education and mortgage market experience both are associated with greater financial knowledge and literacy. As a result, it is hypothesized that the higher the tract population educational attainment and the greater the percentage of homeowners in the tract, the lower the probability of taking out a subprime mortgage.

The percentage of the previous years' originations in each tract that are subprime mortgages is also included as an additional measure of mortgage market dynamics in the market determination model. The hypothesis is that subprime lenders tend to market by neighborhood characteristics that remain relatively constant year-to-year, and that borrowers tend to follow the example and advice of their neighbors (roughly consistent with the financial knowledge argument above). As a result, there is expected to be a persistence to subprime lending by neighborhood, so an increase in the tract percentage of subprime originations in the previous year is hypothesized to increase borrowers' probability taking out subprime mortgages.

As noted previously, the data are restricted to borrowers in three mutually exclusive race/ethnicity categories—African American non-Hispanic borrowers (African American), White Hispanic borrowers (Hispanic), and White non-Hispanic borrowers. In addition, four borrower gender categories—male only, female only, joint male and female, and unknown gender—are also included.

Continuous variables are included in the analysis in one of three forms, based on relatively simplistic semi-parametric explorations of their relationships to the dependent variables. First, LTV and DTI are transformed into categorical variables, reflecting their highly modal distributions and inconsistently monotonic relationships. LTV is divided into seven mutually exclusive categories: less than or equal to 70%, greater than 70% and up to 80%, greater than 80% and up to 85%, greater than 85% and up to 90%, greater than 90% and up to 95%, greater than 95% and up to 100%, and greater than 100%, while DTI is divided into five mutually exclusive categories: less than or equal to 28%, greater than 28% and up to 36%, greater than 36% and up to 50%, greater than 50%, and unknown. Second, FICO score and loan amount are entered as linear spline functions, reflecting their largely monotonic but nonlinear relationship with the dependent variables. In particular, FICO score is treated as a continuous spline with knot points at 600 and 700, while loan amount has both knot points and intercept adjustments at $100,000, the conforming loan limit, and $500,000. Third, the remaining continuous variables are entered as simple linear functions.

Descriptive statistics for the categorical and splined independent variables are presented in Exhibit 3 for 2004 and Exhibit 4 for 2005.[33] Distributions across variable categories are provided for all mortgages combined, and separately for subprime and prime mortgages. In an effort to illustrate the simple bivariate relationships between the dependent and independent variables, for all mortgages the percentage of subprime mortgages is added for each category, and for subprime and prime mortgages, the mean APR is provided. Exhibit 3 shows, for example, that in 2004, 8.16% of all loans were made to African American borrowers, and that 43.60% of African American borrowers take out subprime loans. As a consequence, therefore, 17.06% of subprime loans are made to African American borrowers, at mean APRs of 8.48%, while only 5.82% of prime loans are made to African American borrowers, at mean APRs of 6.25%.

The bivariate relationships in Exhibits 3 and 4 are generally as hypothesized. As noted earlier, for example, minority borrowers are clearly more likely to obtain a subprime mortgage and to have higher APRs in both the subprime and prime markets. In addition, FICO score is strongly related to the probability of taking out a subprime mortgage. In 2004, for example, 74.87% of borrowers with FICO scores below 600 took out subprime mortgages, while only 6.33% of borrowers with FICO scores above 700 did so. The bivariate relationships with LTV and DTI are less monotonic. Generally higher LTV and higher DTI are associated with higher APR, but not consistently, and both LTV and DTI appear to have complex relationships with the probability of taking out a subprime mortgage. The most obvious exceptions to the hypothesized relationships occur for occupancy status, which shows lower APR for non-owner-occupants, and documentation status and prepayment penalty, which sometimes show full-documentation mortgages and mortgages with prepayment penalties having higher APRs.

Descriptive statistics for the remaining variables are shown in Exhibit 5 for 2004 and Exhibit 6 for 2005. Here the mean value of each variable is shown separately for subprime and prime mortgages, and the correlation with APR is provided only for the variables included in the APR models. As hypothesized, there is a positive correlation with APR for both HHI and mortgage denial percentage. In addition, subprime borrowers are more likely to live in tracts with populations that have a lower educational attainment, fewer owner-occupants, and a higher percentage of subprime originations in the previous year.

Exhibits 7 and 8 look more closely at the distributions of FICO scores for 2004 and 2005, respectively. Panel A of each exhibit shows the FICO score distributions by borrower race/ethnicity for the subprime market, and Panel B shows equivalent distributions for the prime market. The exhibits clearly show that within each market, White non-Hispanic borrowers tend to have higher FICO scores than Hispanic borrowers, who in turn tend to have higher FICO scores than African American borrowers. This echoes the relationship in mean APRs across the three borrower groups, suggesting that differences in FICO scores likely are an important factor in explaining APR differentials. Exhibits 8 and 9 also show that subprime borrowers tend to have lower FICO scores than do prime borrowers, but

Exhibit 3 | 2004 Discrete Variable Descriptive Statistics

		All Loans		Subprime Loans		Prime Loans	
Variable Name	Variable Value	Distribution	P(Subprime)	Distribution	Mean APR	Distribution	Mean APR
Race/Ethnicity	African American	8.16	43.60	17.06	8.48	5.82	6.25
	Hispanic	9.23	33.26	14.72	7.86	7.78	5.94
	White non-Hispanic	82.61	17.22	68.21	7.83	86.40	5.65
FICO Score	$500 \leq$ FICO $<$ 600	7.98	74.87	28.66	8.80	2.53	7.76
	$600 \leq$ FICO $<$ 700	34.45	32.53	53.74	7.88	29.37	5.94
	$700 \leq$ FICO $\leq$ 800	56.94	6.33	17.29	6.76	67.39	5.53
Loan-to-Value	LTV $\leq$ 70%	32.72	11.42	17.92	7.61	36.63	5.51
	70% $<$ LTV $\leq$ 80%	43.03	16.62	34.30	7.68	45.33	5.53
	80% $<$ LTV $\leq$ 85%	3.91	56.24	10.55	8.22	2.16	6.19
	85% $<$ LTV $\leq$ 90%	8.63	42.55	17.62	8.20	6.27	6.48
	90% $<$ LTV $\leq$ 95%	5.77	31.22	8.63	8.22	5.01	6.35
	95% $<$ LTV $\leq$ 100%	5.73	38.33	10.53	8.44	4.46	7.11
	100% $<$ LTV	0.21	45.17	0.45	8.86	0.14	6.68
Debt-to-Income	DTI $<=$ 28%	22.95	12.34	13.58	7.87	25.42	5.58
	28% $<$ DTI $\leq$ 36%	25.00	15.10	18.10	7.92	26.82	5.63
	36% $<$ DTI $\leq$ 50%	40.35	28.55	55.25	8.04	36.42	5.76
	50% $<$ DTI	8.98	22.61	9.74	7.76	8.78	5.95
	Unknown DTI	2.73	25.53	3.34	7.42	2.57	6.22
Loan Amount	Loan Amount $<$ \$100,000	28.20	23.39	31.64	8.62	27.30	6.12
	\$100,000 $\leq$ Loan Amount $<$ \$334,000	62.45	20.28	60.74	7.70	62.90	5.63
	\$334,000 $\leq$ Loan Amount $<$ \$500,000	5.74	23.36	6.42	7.23	5.55	5.14
	\$500,000 $\leq$ Loan Amount	3.61	6.88	1.19	6.65	4.25	4.91
Contract Rate Type	Fixed	49.28	10.92	25.81	7.40	55.46	5.92
	Adjustable	37.46	37.04	66.55	8.17	29.80	5.06
	Unknown Contract Rate Type	13.26	12.02	7.64	7.84	14.74	6.21

Exhibit 3 | (continued)

2004 Discrete Variable Descriptive Statistics

		All Loans		Subprime Loans		Prime Loans	
Variable Name	Variable Value	Distribution	P(Subprime)	Distribution	Mean APR	Distribution	Mean APR
Loan Purpose	Purchase Money	42.61	16.18	33.06	7.89	45.13	5.78
	Refinance	57.39	24.32	66.94	7.97	54.87	5.65
Occupancy Status	Owner-Occupied	88.18	21.35	90.27	7.94	87.63	5.67
	Not Owner-Occupied	11.82	17.17	9.73	8.01	12.37	5.94
Documentation Type	Full Documentation	39.14	26.90	50.50	8.07	36.15	5.74
	Not Full Documentation	42.60	11.44	23.37	7.88	47.67	5.51
	Unknown Documentation Type	18.25	29.84	26.12	7.77	16.18	6.20
Prepayment Penalty	No Prepayment Penalty	73.36	5.74	20.20	7.75	87.36	5.63
	Prepayment Penalty	9.87	85.06	40.28	8.40	1.86	6.10
	Unknown Prepayment Penalty	16.77	49.15	39.51	7.59	10.77	6.28
Loan Term	Loan Term $\leq$ 5 years	0.54	25.78	0.67	8.17	0.51	5.50
	5 years $<$ Loan Term $\leq$ 15 years	16.99	11.20	9.13	8.42	19.06	5.54
	15 years $<$ Loan Term $\leq$ 20 years	3.24	19.85	3.08	8.32	3.28	5.95
	20 years $<$ Loan Term $\leq$ 30 years	79.23	22.93	87.11	7.88	77.15	5.74
	30 years $<$ Loan Term $\leq$ 40 years	0.01	53.76	0.01	10.77	0.00	5.84
Channel	Wholesale	29.96	42.76	61.43	7.92	21.66	5.89
	Retail	70.04	11.48	38.57	7.99	78.34	5.66
Gender	Male	27.69	25.59	33.98	7.96	26.03	5.81
	Female	21.60	26.18	27.12	8.07	20.15	5.86
	Joint Male / Female	50.60	15.89	38.56	7.85	53.77	5.60
	Unknown Gender	0.12	60.95	0.34	8.06	0.06	6.20

Notes: The number of loans (weighted) for all loans = 794,204; the number of loans (weighted) for subprime loans = 165,610; and the number of loans (weighted) for prime loans = 628,594.

Exhibit 4 | 2005 Discrete Variable Descriptive Statistics

		All Loans		Subprime Loans		Prime Loans	
Variable Name	Variable Value	Distribution	P(Subprime)	Distribution	Mean APR	Distribution	Mean APR
Race/Ethnicity	African American	9.32	44.54	19.68	9.57	6.55	6.72
	Hispanic	11.81	34.73	19.45	9.24	9.77	6.50
	White non-Hispanic	78.87	16.28	60.88	9.07	83.68	6.25
FICO Score	$500 \leq$ FICO $<$ 600	9.17	73.86	32.12	9.99	3.04	7.88
	$600 \leq$ FICO $<$ 700	37.40	31.27	55.43	8.98	32.58	6.47
	$700 \leq$ FICO $\leq$ 800	52.70	4.86	12.15	8.17	63.55	6.14
Loan-to-Value	LTV $\leq$ 70%	25.54	10.43	12.63	8.95	29.00	6.06
	70% $<$ LTV $\leq$ 80%	44.95	21.11	44.99	9.11	44.95	6.19
	80% $<$ LTV $\leq$ 85%	4.04	54.37	10.40	9.22	2.33	6.50
	85% $<$ LTV $\leq$ 90%	10.82	34.30	17.59	9.58	9.01	6.75
	90% $<$ LTV $\leq$ 95%	5.61	17.76	4.72	9.30	5.85	6.65
	95% $<$ LTV $\leq$ 100%	8.90	22.69	9.57	9.24	8.72	6.95
	100% $<$ LTV	0.13	13.96	0.09	7.90	0.15	6.82
Debt-to-Income	DTI $<=$ 28%	17.72	12.66	10.63	9.16	19.62	6.25
	28% $<$ DTI $\leq$ 36%	20.13	16.53	15.77	9.17	21.29	6.28
	36% $<$ DTI $\leq$ 50%	45.06	30.84	65.87	9.21	39.49	6.33
	50% $<$ DTI	13.07	10.98	6.80	9.52	14.75	6.33
	Unknown DTI	4.03	4.85	0.93	7.65	4.86	6.39
Loan Amount	Loan Amount $<$ \$100,000	21.86	23.63	24.49	9.57	21.16	6.56
	\$100,000 $\leq$ Loan Amount $<$ \$360,000	67.21	20.86	66.46	9.10	67.41	6.25
	\$360,000 $\leq$ Loan Amount $<$ \$500,000	6.11	22.58	6.54	8.91	5.99	6.27
	\$500,000 $\leq$ Loan Amount	4.82	11.01	2.51	9.07	5.44	6.12
Contract Rate Type	Fixed	24.67	17.84	20.86	7.76	25.69	6.09
	Adjustable	22.14	65.47	68.72	9.71	9.69	6.41
	Unknown Contract Rate Type	53.19	4.13	10.42	8.76	64.62	6.38

Exhibit 4 | (continued)

2005 Discrete Variable Descriptive Statistics

		All Loans		Subprime Loans		Prime Loans	
Variable Name	Variable Value	Distribution	P(Subprime)	Distribution	Mean APR	Distribution	Mean APR
Loan Purpose	Purchase Money	47.53	17.32	39.02	9.40	49.80	6.39
	Refinance	52.47	24.52	60.98	9.08	50.20	6.22
Occupancy Status	Owner-Occupied	86.75	22.16	91.10	9.16	85.58	6.28
	Not Owner-Occupied	13.25	14.16	8.90	9.61	14.42	6.47
Documentation Type	Full Documentation	30.38	39.88	57.44	9.21	23.15	6.20
	Not Full Documentation	22.60	33.37	35.75	9.37	19.09	6.35
	Unknown Documentation Type	47.02	3.06	6.81	8.25	57.77	6.33
Prepayment Penalty	No Prepayment Penalty	51.25	9.48	23.02	9.53	58.80	6.17
	Prepayment Penalty	16.27	91.00	70.17	9.19	1.86	5.92
	Unknown Prepayment Penalty	32.48	4.42	6.81	8.25	39.34	6.52
Loan Term	Loan Term $\leq$ 5 years	3.85	6.84	1.25	7.45	4.54	6.24
	5 years $<$ Loan Term $\leq$ 15 years	8.80	3.60	1.50	7.91	10.75	6.03
	15 years $<$ Loan Term $\leq$ 20 years	2.72	5.61	0.72	7.92	3.25	6.32
	20 years $<$ Loan Term $\leq$ 30 years	83.58	23.46	92.95	9.22	81.07	6.34
	30 years $<$ Loan Term $\leq$ 40 years	1.06	71.09	3.58	10.19	0.39	6.59
Channel	Wholesale	52.23	36.48	90.31	9.32	42.05	6.31
	Retail	47.77	4.28	9.69	8.09	57.95	6.31
Gender	Male	32.75	25.83	40.10	9.29	30.78	6.44
	Female	23.39	26.54	29.42	9.35	21.77	6.41
	Joint Male / Female	43.73	14.65	30.37	8.94	47.30	6.17
	Unknown Gender	0.14	16.06	0.11	9.46	0.15	6.26

Notes: The number of loans (weighted) for all loans = 367,714; the number of loans (weighted) for subprime loans = 77,575; and the number of loans (weighted) for prime loans = 290,139.

Exhibit 5 | 2004 Continuous Variable Descriptive Statistics

	Subprime Loans			Prime Loans		
Variable Name	Mean	Std. Error	Correlation with APR	Mean	Std. Error	Correlation with APR
Tract Herfindahl-Hirschman Index	416.16	299.31	0.08	489.58	388.00	0.04
Tract Mortgage Denial Percent	26.99	9.03	0.32	22.58	8.31	0.35
Tract Percentage Less than High School	19.49	12.63	—	14.31	10.62	—
Tract Percentage Completed High School	29.13	8.95	—	25.81	9.90	—
Tract Percentage Some College	29.36	7.25	—	29.69	6.95	—
Tract Percentage Completed College	22.01	14.15	—	30.19	17.41	—
Tract Percentage Owner-Occupied	65.83	18.56	—	67.74	19.01	—
Tract Percentage Subprime Originations in Previous Year	13.74	9.13	—	9.08	6.65	—

Notes: The number of loans (weighted) for subprime loans = 165,610; and the number of loans (weighted) for prime loans = 628,594.

that there is a fairly substantial overlap in these distributions. This suggests that FICO scores likely are important factors in explaining market determination, but that they likely are not the only factor.

The estimation results from the endogenous switching regression models are provided in Exhibit 9 for 2004 and Exhibit 10 for 2005. The majority of the estimated coefficients are statistically significant with *p*-values less than 0.0001. Moreover, the data strongly reject the null hypothesis of no endogeneity, implying that a switching regression framework is appropriate for this analysis. In particular, the estimated error correlation coefficients are statistically significant in both years.[34]

The positive sign of the correlation coefficient for the 2004 subprime APR equation implies that borrowers taking out subprime mortgages pay higher APRs in the subprime market than would random borrowers from a pool of both prime and subprime borrowers. Equivalently, the negative correlation coefficients estimated in the prime equations for both years imply that borrowers taking out prime mortgages pay higher APRs in the prime market than would random borrowers. In other words, some borrowers that could do better in the prime

Exhibit 6 | 2005 Continuous Variable Descriptive Statistics

	Subprime Loans			Prime Loans		
Variable Name	Mean	Std. Error	Correlation with APR	Mean	Std. Error	Correlation with APR
Tract Herfindahl-Hirschman Index	358.67	243.86	0.00	422.26	307.21	0.06
Tract Mortgage Denial Percent	28.81	8.87	0.15	25.23	8.54	0.21
Tract Percentage Less than High School	20.59	12.78	—	16.19	11.39	—
Tract Percentage Completed High School	29.77	8.58	—	26.87	9.98	—
Tract Percentage Some College	28.77	7.19	—	29.72	7.15	—
Tract Percentage Completed College	20.88	13.55	—	27.23	16.54	—
Tract Percentage Owner-Occupied	65.04	19.11	—	67.12	18.33	—
Tract Percentage Subprime Originations in Previous Year	26.63	12.44	—	21.30	11.22	—

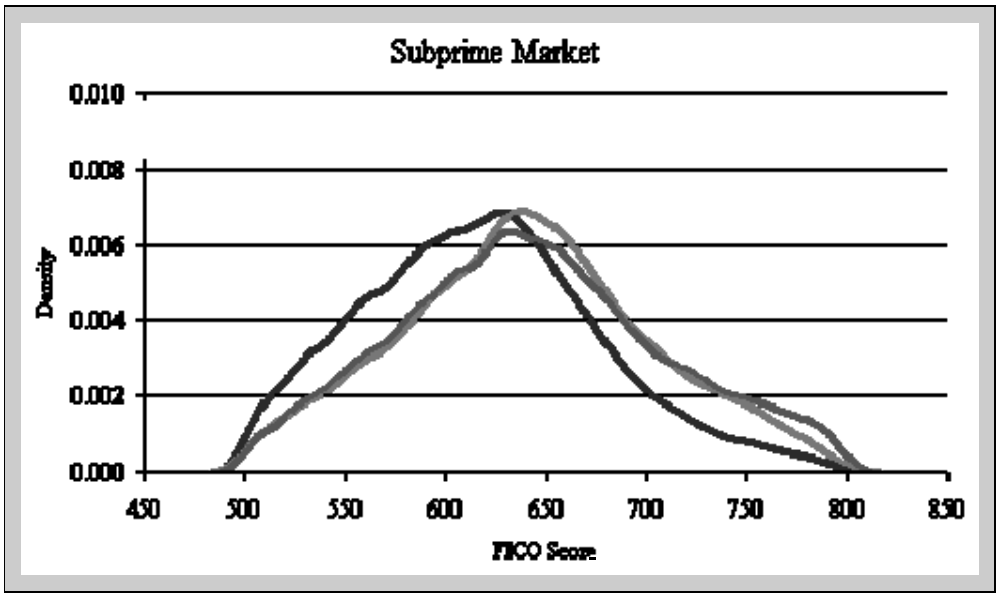
Notes: The number of loans (weighted) for subprime loans = 77,431; and the number of loans (weighted) for prime loans = 289,874.

market nonetheless take out a subprime mortgage, and some borrowers that could do better in the subprime market nonetheless take out a prime mortgage. A possible partial explanation for this apparently poor borrower matching in the subprime market is this market's generally faster approval rates.[35] This result is also consistent with the possibility that there are omitted variables in the market determination and subprime APR models that are correlated with credit risk and known to lenders, which leads to higher-risk borrowers getting a subprime mortgage at a higher APR. This latter explanation, however, does not well fit the prime market results.[36] More broadly, these relationships appear consistent with unobserved market segmentation or steering on the part of lenders in an effort to maximize revenues, although there is no way to verify this from the data. Interestingly, the subprime coefficient switches sign in 2005, albeit with a very small absolute value.

The generally small absolute value of the correlation coefficients suggests that endogeneity plays a relatively minor part in determining borrowers' expected APRs. On average, for example, the inverse Mills ratio terms shift expected APRs

Exhibit 7 | 2004 FICO Score Distributions

Panel A: FICO score distributions by borrower race/ethnicity for the subprime market



Panel B: FICO score distributions by borrower race/ethnicity for the prime market

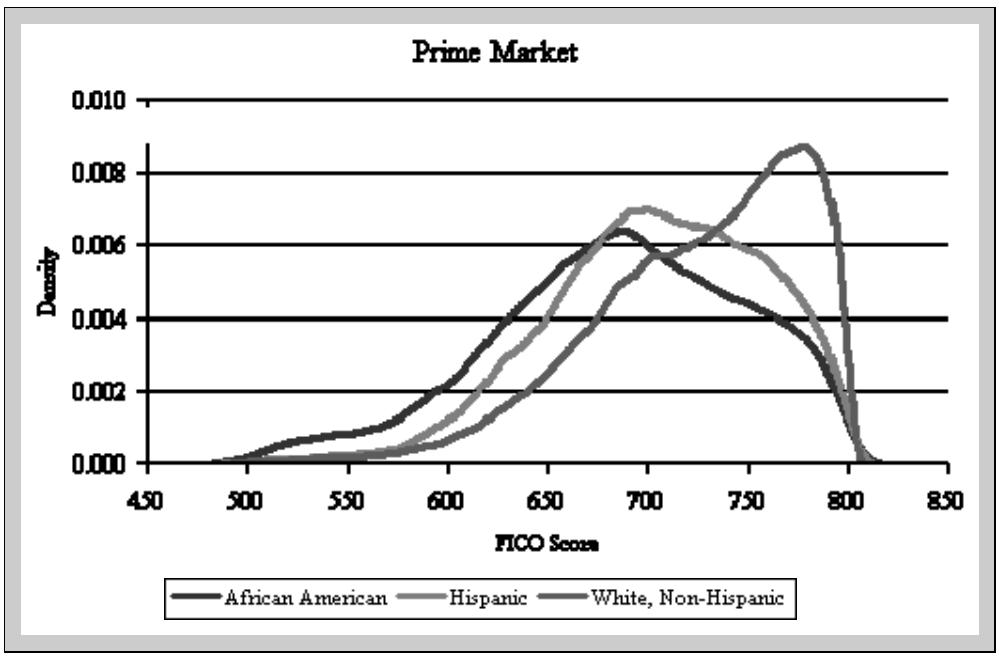
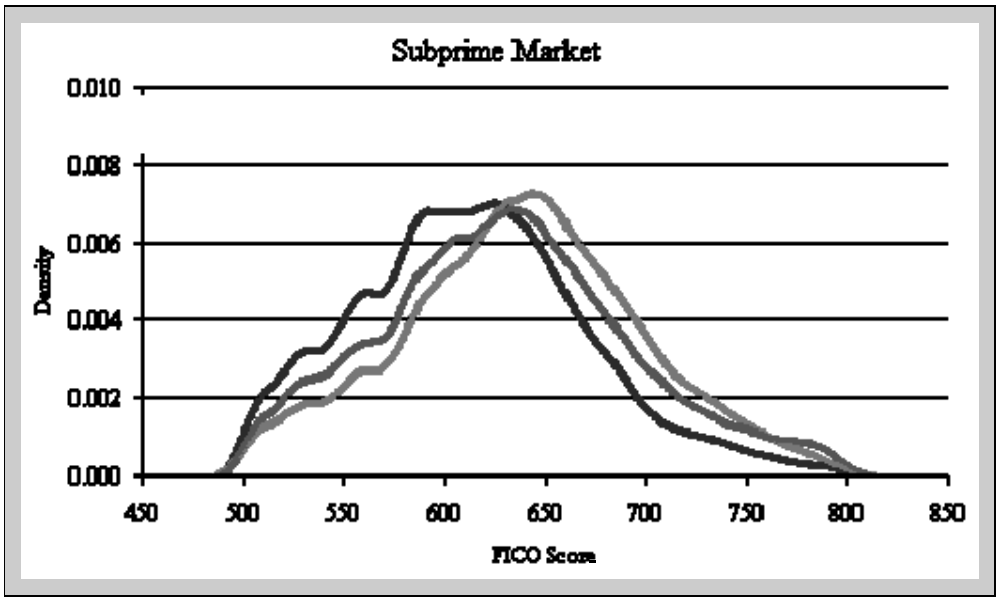


Exhibit 8 | 2005 FICO Score Distributions

Panel A: FICO score distributions by borrower race/ethnicity for the subprime market



Panel B: FICO score distributions by borrower race/ethnicity for the prime market

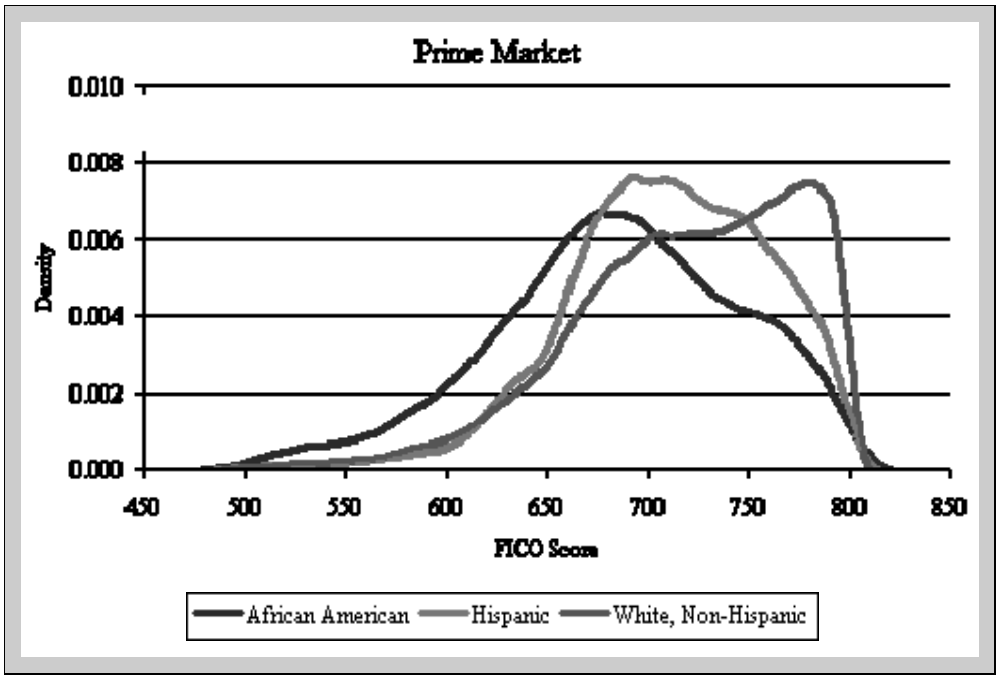


Exhibit 9 | 2004 Endogenous Switching Regression Estimates

		P(Subprime)		APR|Subprime		APR|Prime	
Variable Name	Variable Value	Estimate	Prob (z)	Estimate	Prob (z)	Estimate	Prob (z)
Intercept		3.06	<0.0001	22.21	<0.0001	28.06	<0.0001
Race/Ethnicity	African American	0.09	<0.0001	0.10	<0.0001	0.08	<0.0001
	Hispanic	0.03	0.0130	0.13	<0.0001	0.04	<0.0001
	White non-Hispanic	0	—	0	—	0	—
FICO Score Splines	500 ≤ FICO < 600	−0.52	<0.0001	−1.44	<0.0001	−3.45	<0.0001
(100s)	600 ≤ FICO < 700	−1.03	<0.0001	−1.01	<0.0001	−0.54	<0.0001
	700 ≤ FICO ≤ 800	−0.36	<0.0001	−0.72	<0.0001	0.01	0.0180
Loan-to-Value	LTV <= 70%	−1.10	<0.0001	−1.77	<0.0001	−0.65	<0.0001
	70% < LTV ≤ 80%	−0.89	<0.0001	−1.63	<0.0001	−0.60	<0.0001
	80% < LTV ≤ 85%	−0.46	<0.0001	−1.41	<0.0001	−0.32	<0.0001
	85% < LTV ≤ 90%	−0.64	<0.0001	−1.21	<0.0001	−0.01	0.9330
	90% < LTV ≤ 95%	−0.68	<0.0001	−0.86	<0.0001	−0.02	0.8000
	95% < LTV ≤ 100%	−0.88	<0.0001	−0.33	<0.0001	0.43	<0.0001
	100% < LTV	0	—	0	—	0	—
Debt-to-Income	DTI ≤ 28%	0	—	0	—	0	—
	28% < DTI ≤ 36%	0.06	<0.0001	0.06	<0.0001	−0.01	<0.0001
	36% < DTI ≤ 50%	0.22	<0.0001	0.15	<0.0001	0.03	<0.0001
	50% < DTI	0.02	0.1950	−0.04	<0.0001	0.04	<0.0001
	Unknown DTI	0.59	<0.0001	−0.16	<0.0001	0.28	<0.0001
Loan Amount Splines	Loan Amount < $100,000	3.06	<0.0001	−19.73	<0.0001	−8.92	<0.0001
($1,000,000s)	$100,000 ≤ Loan Amount < $334,000	0.07	0.2910	−2.94	<0.0001	−1.04	<0.0001
	$334,000 ≤ Loan Amount < $500,000	−1.40	<0.0001	−0.23	0.1120	0.02	0.9190
	$500,000 ≤ Loan Amount	−2.12	<0.0001	0.04	0.8090	−0.13	<0.0001

Exhibit 9 | (continued)

2004 Endogenous Switching Regression Estimates

		P(Subprime)		APR|Subprime		APR|Prime	
Variable Name	Variable Value	Estimate	Prob (z)	Estimate	Prob (z)	Estimate	Prob (z)
Loan Amount	Loan Amount ≤ $100,000	0	—	0	—	0	—
Indicator	$100,000 < Loan Amount ≤ $334,000	−0.15	<0.0001	0.04	<0.0001	0.06	<0.0001
	$334,000 < Loan Amount ≤ $500,000	−0.02	0.5170	0.20	<0.0001	0.15	<0.0001
	$500,000 < Loan Amount	0.00	0.9700	0.13	0.0010	0.13	<0.0001
Contract Rate Type	Fixed	0	—	0	—	0	—
	Adjustable	0.78	<0.0001	0.37	<0.0001	−0.96	<0.0001
	Unknown Contract Rate Type	−2.02	<0.0001	−0.29	<0.0001	−0.24	<0.0001
Loan Purpose	Purchase Money	0	—	0	—	0	—
	Refinance	0.34	<0.0001	−0.03	<0.0001	0.05	<0.0001
Occupancy Status	Owner-Occupied	0	—	0	—	0	—
	Not Owner-Occupied	0.05	<0.0001	0.22	<0.0001	0.27	<0.0001
Documentation Type	Full Documentation	0	—	0	—	0	—
	Not Full Documentation	0.11	<0.0001	0.16	<0.0001	−0.07	<0.0001
	Unknown Documentation Type	0.75	<0.0001	−0.32	<0.0001	0.13	<0.0001
Prepayment Penalty	No Prepayment Penalty	0	—	0	—	0	—
	Prepayment Penalty	2.78	<0.0001	0.42	<0.0001	0.01	0.4490
	Unknown Prepayment Penalty	2.10	<0.0001	−0.68	<0.0001	0.08	<0.0001
Loan Term	Loan Term ≤ 5 years	−0.74	<0.0001	−1.78	<0.0001	−0.05	0.8240
	5 years < Loan Term ≤ 15 years	−1.29	<0.0001	−2.25	<0.0001	−0.26	0.2810
	15 years < Loan Term ≤ 20 years	−1.02	<0.0001	−2.45	<0.0001	0.16	0.4960
	20 years < Loan Term ≤ 30 years	−1.16	<0.0001	−2.72	<0.0001	0.26	0.2720
	30 years < Loan Term ≤ 40 years	0	—	0	—	0	—

Exhibit 9 | (continued)

2004 Endogenous Switching Regression Estimates

		P(Subprime)		APR|Subprime		APR|Prime	
Variable Name	Variable Value	Estimate	Prob (z)	Estimate	Prob (z)	Estimate	Prob (z)
Channel	Wholesale	0.90	<0.0001	−0.08	<0.0001	0.04	<0.0001
	Retail	0	—	0	—	0	—
Gender	Male	0	—	0	—	0	—
	Female	−0.02	0.0090	−0.01	0.0150	0.00	0.7300
	Joint Male/Female	−0.08	<0.0001	0.04	<0.0001	−0.03	<0.0001
	Unknown Gender	1.54	<0.0001	0.16	<0.0001	0.07	0.4610
Tract HHI (10,000s)		−1.75	<0.0001	0.12	0.1780	−0.11	0.0350
Tract Mortgage Denial Percentage (100s)		0.66	<0.0001	0.64	<0.0001	0.63	<0.0001
Tract-Level Education	Percentage Less than High School	0	—	—	—	—	—
	Percentage Completed High School	0.03	0.7410	—	—	—	—
	Percentage Some College	0.15	0.0060	—	—	—	—
	Percentage Completed College	−0.40	<0.0001	—	—	—	—
Tract Percentage Owner-Occupied (100s)		14.39	<0.0001	—	—	—	—
Tract Percentage Subprime Originations in Previous Year (100s)		0.65	<0.0001	—	—	—	—
State Control Dummies		yes		yes		yes	
Calendar Month Control Dummies		yes		yes		yes	
Standard Error of Residual		—	—	0.97	<0.0001	0.59	<0.0001
Correlation Coefficient		—	—	0.03	<0.0001	−0.11	<0.0001

Notes: The number of observation is 794,204. The Wald Chi-Squared Test of Independent Equations with 2 df (p-value) is 939.08 (<0.0001).

Exhibit 10 | 2005 Endogenous Switching Regression Estimates

		P(Subprime)		APR|Subprime		APR|Prime	
Variable Name	Variable Value	Estimate	Prob (z)	Estimate	Prob (z)	Estimate	Prob (z)
Intercept		1.47	<0.0001	16.88	<0.0001	21.33	<0.0001
Race/Ethnicity	African American	0.09	<0.0001	0.03	<0.0001	0.12	<0.0001
	Hispanic	0.04	0.1210	0.07	<0.0001	0.12	<0.0001
	White non-Hispanic	0	—	0	—	0	—
FICO Score Splines	500 ≤ FICO < 600	−0.27	<0.0001	−1.44	<0.0001	−2.33	<0.0001
(100s)	600 ≤ FICO < 700	−1.20	<0.0001	−0.81	<0.0001	−0.49	<0.0001
	700 ≤ FICO ≤ 800	−0.33	<0.0001	−0.36	<0.0001	−0.08	<0.0001
Loan-to-Value	LTV <= 70%	−1.21	<0.0001	0.35	0.0030	−0.61	<0.0001
	70% < LTV ≤ 80%	−1.03	<0.0001	0.48	<0.0001	−0.51	<0.0001
	80% < LTV ≤ 85%	−0.43	<0.0001	0.49	<0.0001	−0.22	0.0090
	85% < LTV ≤ 90%	−0.71	<0.0001	0.87	<0.0001	0.02	0.7930
	90% < LTV ≤ 95%	−0.78	<0.0001	1.06	<0.0001	0.02	0.7920
	95% < LTV ≤ 100%	−0.93	<0.0001	1.33	<0.0001	0.19	0.0220
	100% < LTV	0	—	0	—	0	—
Debt-to-Income	DTI ≤ 28%	0	—	0	—	0	—
	28% < DTI ≤ 36%	−0.06	0.0070	−0.03	0.0060	−0.01	0.1920
	36% < DTI ≤ 50%	0.17	<0.0001	−0.02	0.0300	0.00	0.8450
	50% < DTI	−0.23	<0.0001	0.04	0.0020	−0.01	0.2610
	Unknown DTI	0.17	0.0190	−0.28	0.0020	−0.05	0.0010
Loan Amount Splines	Loan Amount < $100,000	−9.69	<0.0001	−13.55	<0.0001	−9.24	<0.0001
($1,000,000s)	$100,000 ≤ Loan Amount < $334,000	−1.35	<0.0001	−1.73	<0.0001	−0.65	<0.0001
	$334,000 ≤ Loan Amount < $500,000	−0.01	0.9930	−0.10	0.6660	0.01	0.9770
	$500,000 ≤ Loan Amount	−1.29	<0.0001	0.49	0.0020	0.01	0.4080

Exhibit 10 | (continued)

2005 Endogenous Switching Regression Estimates

		P(Subprime)		APR|Subprime		APR|Prime	
Variable Name	Variable Value	Estimate	Prob (z)	Estimate	Prob (z)	Estimate	Prob (z)
Loan Amount	Loan Amount ≤ $100,000	0	–	0	–	0	–
Indicator	$100,000 < Loan Amount ≤ $334,000	0.13	<0.0001	−0.07	<0.0001	0.09	<0.0001
	$334,000 < Loan Amount ≤ $500,000	0.45	<0.0001	−0.01	0.7500	0.23	<0.0001
	$500,000 < Loan Amount	0.49	0.0010	0.07	0.0850	0.16	0.0020
Contract Rate Type	Fixed	0	–	0	–	0	–
	Adjustable	0.69	<0.0001	1.71	<0.0001	0.35	<0.0001
	Unknown Contract Rate Type	−1.03	<0.0001	1.63	<0.0001	0.13	<0.0001
Loan Purpose	Purchase Money	0	–	0	–	0	–
	Refinance	0.38	<0.0001	−0.15	<0.0001	0.00	0.5820
Occupancy Status	Owner-Occupied	0	–	0	–	0	–
	Not Owner-Occupied	0.10	<0.0001	0.53	<0.0001	0.22	<0.0001
Documentation Type	Full Documentation	0	–	0	–	0	–
	Not Full Documentation	0.32	<0.0001	0.41	<0.0001	0.23	<0.0001
	Unknown Documentation Type	−1.18	<0.0001	−0.81	<0.0001	0.08	<0.0001
Prepayment Penalty	No Prepayment Penalty	0	–	0	–	0	–
	Prepayment Penalty	3.04	<0.0001	−0.16	<0.0001	−0.01	0.3090
	Unknown Prepayment Penalty	1.45	<0.0001	0	–	0.35	<0.0001
Loan Term	Loan Term ≤ 5 years	0.82	<0.0001	−1.21	<0.0001	−0.43	<0.0001
	5 years < Loan Term ≤ 15 years	−0.33	<0.0001	−0.59	<0.0001	−0.40	<0.0001
	15 years < Loan Term ≤ 20 years	−0.18	0.0020	−0.48	<0.0001	−0.08	0.0010
	20 years < Loan Term ≤ 30 years	−0.23	<0.0001	−0.27	<0.0001	−0.16	<0.0001
	30 years < Loan Term ≤ 40 years	0	–	0	–	0	–

Exhibit 10 | (continued)

2005 Endogenous Switching Regression Estimates

		P(Subprime)		APR|Subprime		APR|Prime	
Variable Name	Variable Value	Estimate	Prob (z)	Estimate	Prob (z)	Estimate	Prob (z)
Channel	Wholesale	1.24	<0.0001	0.31	<0.0001	0.23	<0.0001
	Retail	0	—	0	—	0	—
Gender	Male	0	—	0	—	0	—
	Female	−0.02	0.2590	−0.02	0.0090	−0.02	0.0250
	Joint Male/Female	−0.03	0.0450	−0.03	<0.0001	−0.08	<0.0001
	Unknown Gender	−0.15	0.2160	0.18	0.0150	−0.01	0.8110
Tract HHI (10,000s)		−1.01	<0.0001	0.93	<0.0001	−0.18	0.0750
Tract Mortgage Denial Percentage (100s)		−0.55	<0.0001	0.62	<0.0001	0.26	<0.0001
Tract-Level Education	Percentage Less than High School	0	—				
	Percentage Completed High School	−0.97	<0.0001				
	Percentage Some College	−0.54	<0.0001				
	Percentage Completed College	−0.97	<0.0001				
Tract Percentage Owner-Occupied (100s)		2.76	0.5310				
Tract Percentage Subprime Originations in Previous Year (100s)		0.94	<0.0001				
State Control Dummies		yes		yes		yes	
Calendar Month Control Dummies		yes		yes		yes	
Standard Error of Residual				0.82	<0.0001	0.64	<0.0001
Correlation Coefficient				−0.09	<0.0001	−0.13	<0.0001

Notes: The number of observation is 367,714. The Wald Chi-Squared Test of Independent Equations with 2 df (p-value) is 222.23 (<0.0001).

only by about one basis point. Endogeneity can, however, have larger impacts for some borrowers. In 2004, for example, endogeniety increases expected APRs by more than five basis points for 5% of subprime borrowers.

The race/ethnicity coefficients are of greatest import in the analysis. Starting with the market determination models, the Hispanic coefficients are not significantly different from zero at the 5% level in either year. Observable characteristics, therefore, fully explain the difference in the probability of taking out a subprime mortgage between Hispanic and White non-Hispanic borrowers. As discussed above, this implies that the endogeniety of the switching regression model will have no impact in explaining differences in APR between Hispanic and White non-Hispanic borrowers.

The African American coefficient in the market determination model is small but statistically significant for both 2004 and 2005. This means that observable characteristics cannot fully explain the differential subprime probabilities for African American and White non-Hispanic borrowers; however, the unexplained differences are quite small.

The African American and Hispanic coefficients are statistically significant in APR estimations of the subprime and prime markets in both 2004 and 2005. Because, as noted above, the correlation coefficients and the race/ethnicity dummies in the market determination model are both small, the race/ethnicity coefficients in the APR models closely approximate differences in expected APRs. In 2005, for example, the direct effect from the African American coefficient suggests that, holding constant other observable characteristics, African American borrowers have APRs that are 3 basis points above White non-Hispanic subprime borrowers, and 12 basis points above White non-Hispanic prime borrowers. This is substantially less than the Exhibit 2 uncontrolled differentials of 50 and 47 basis points, respectively, suggesting that observable characteristics explain much of the difference in mortgage prices paid by African American and White non-Hispanic borrowers.

The results for Hispanic borrowers are similar, although somewhat smaller in size. Exhibit 2 shows uncontrolled Hispanic differentials of 17 basis points and 25 basis points, respectively, in the subprime and prime markets in 2005. Controlling for observable characteristics reduces these differences to 7 basis points in the subprime market and 12 basis points in the prime market.

The remainder of this section briefly discusses the other coefficients in the switching regression equations. Before doing so, however, note that the primary use of the estimations is for prediction. The model specifications, therefore, were not developed to capture accurately the marginal impact of each of the remaining included explanatory variables. As a consequence, the individual coefficients of these variables should not be given undue weight.

FICO scores performed as expected—the higher the FICO score the lower the probability of taking out a subprime mortgage. The greatest marginal impact on

market determination occurs in the FICO score range of 600 to 700, which is, arguably, the FICO score boundary between subprime and prime lending. Higher FICO scores also are associated with lower APRs, but the strength of this relationship declines significantly in the higher FICO scores ranges. LTV is significant in explaining market determination, but without a clear, monotonic relationship. This reflects, the bimodal competing tendencies of equity and high-risk lending in the subprime market. Higher LTVs are consistently and monotonically related to higher APRs in the subprime market. That same relationship holds for LTVs below 90% in the prime market. At greater LTVs, the strength and consistency of the relationship declines, however, perhaps because mortgage insurers in the prime market share credit risk for LTVs at higher levels.

DTI also is significant in explaining the probability of taking out a subprime mortgage, with the highest probability associated with unknown DTIs and DTIs in the 35% to 50% range. The former likely is associated with the higher frequency of ''no income, no asset'' loans in the subprime market. As expected, lower loan amounts are associated with higher APRs in both the prime and subprime markets, consistent with costs being driven by large fixed origination and loss severity components.

Occupancy status and loan term both perform as expected—non-owner-occupied is associated with higher APRs in both the subprime and prime markets, as are longer loan terms. Surprisingly, adjustable rate mortgages have higher APRs in the 2004 subprime market, and in both markets in 2005, perhaps because of a flattening of the yield curve. Equally surprising is that refinance mortgages have lower APRs in all estimations except the 2004 prime market. As expected, less than full documentation loans are associated more closely with subprime than prime lending, and except for the 2004 prime market, also are associated with higher APRs. Prepayment penalties are much more associated with subprime than prime lending.[37] Despite widespread evidence from rates sheets that prepayment penalties lower contract rates, the estimations capture such an impact on APR only for the 2005 subprime market. Likely this is due to correlation between prepayment penalties and the other explanatory variables, and does not reflect a failure to deliver rate reductions to borrowers with prepayment penalties.

Subprime lending is more highly related to the wholesale channel, and wholesale (as opposed to retail) originations have higher APRs everywhere except the 2004 subprime market. There is no substantial difference in the mortgage outcomes of male and female borrowers, but there is a slight tendency for joint male/female borrowers to take out prime mortgages and to have lower APRs.

As expected, higher market concentrations (higher HHIs) are associated with higher APRs in the 2005 subprime market, but not consistently elsewhere. Further, higher HHIs are associated with a lower probability of taking out a subprime mortgage, suggesting that, at least in 2004 and 2005, subprime is the least concentrated of the two markets. Tract-level mortgage denial percentages are statistically significant, but inconsistently related to the probability of taking out

a subprime mortgage. Denial rates, however, are strongly related to APRs in the subprime and prime markets of both years, with higher denial rates (reflecting higher origination costs) associated with higher APRs.

Tract-level education performs about as expected, with higher educational attainment roughly associated with a reduced probability of taking out a subprime mortgage. Tract percentage owner-occupied performs counter to expectations, however, with higher percentage associated with higher subprime probabilities. Finally, tract percentage of subprime originations in the previous year is associated with a higher probability of taking out a subprime mortgage, as expected.

Conclusion

Exhibit 11 provides a summary of the comparisons between minority and non-minority mortgage outcomes, as derived from the endogenous switching regression model estimates. It follows the format of Exhibit 2 but adds rows for White non-Hispanic borrowers with minority borrower characteristics, calculated as described above.[38]

Exhibit 11 clearly demonstrates that the vast majority of APR differentials between minority and non-minority mortgage borrowers can be explained by observable characteristics appropriately associated with underwriting and pricing outcomes. For example, the uncontrolled difference in the 2004 mean APR of African American and White non-Hispanic borrowers is 116 basis points.[39] Controlling for observable differences between borrowers, however, reduces this difference to only 10 basis points. For Hispanic borrowers in 2004, uncontrolled differences of 55 basis points are reduced to 8 basis points after controls. Results for 2005 are similar—African American borrower differentials go from 127 basis points to 9 basis points, and Hispanic borrower differentials go from 75 basis points to 11 basis points.

One of the primary reasons the underwriting and pricing controls so significantly narrow differences in mean predicted APR is the success of the control variables in predicting the probability that borrowers will take out subprime mortgages. Specifically, the market determination model explains nearly all of the race/ethnicity differentials in the probability of taking out a subprime mortgage. Minority borrowers have a dramatically greater probability of taking out a mortgage in the higher priced subprime market—28.38 percentage points higher for African American borrowers in 2005, for example. The control variables in the market determination model reduce these unexplained race/ethnicity differentials to around one percentage point or less—the unexplained probability differential of taking out a subprime mortgage is only 0.70 percentage points for African American borrowers in 2005, for example. Because control variables explain over 95% of the market determination differential, any remaining unexplained race/ethnicity differentials in APR that might be suggestive of differential treatment must be largely associated with unexplained differentials within either the subprime or prime markets.

Exhibit 11 | Outcome Comparisons with Controls

Year	Borrower Race / Ethnicity	Unconditional Mean Predicted APR	Predicted Probability of Having a Subprime Loan	Mean Predicted Subprime APR	Mean Predicted Prime APR
2004	African American	7.18	44.08%	8.47	6.23
	Hispanic	6.57	33.12%	7.85	5.93
	White non-Hispanic	6.02	17.26%	7.82	5.64
	White non-Hispanic with African American Characteristics	7.08	42.84%	8.38	6.16
	White non-Hispanic with Hispanic Characteristics	6.49	32.79%	7.72	5.89
	African American minus White non-Hispanic	1.16	26.82%	0.65	0.59
	African American minus White non-Hispanic with African American Characteristics	0.10	1.24%	0.10	0.08
	Hispanic minus White non-Hispanic	0.55	15.85%	0.04	0.29
	Hispanic minus White non-Hispanic with Hispanic Characteristics	0.08	0.32%	0.13	0.04
2005	African American	7.99	44.94%	9.58	6.71
	Hispanic	7.46	34.92%	9.26	6.49
	White non-Hispanic	6.71	16.55%	9.10	6.25
	White non-Hispanic with African American Characteristics	7.89	44.23%	9.55	6.59
	White non-Hispanic with Hispanic Characteristics	7.35	34.66%	9.19	6.37
	African American minus White non-Hispanic	1.27	28.38%	0.49	0.47
	African American minus White non-Hispanic with African American Characteristics	0.09	0.70%	0.03	0.12
	Hispanic minus White non-Hispanic	0.75	18.37%	0.16	0.25
	Hispanic minus White non-Hispanic with Hispanic Characteristics	0.11	0.25%	0.07	0.12

Conditional APR differentials are also significantly explained by observable characteristics appropriately associated with underwriting and pricing outcomes, although these reductions are not as dramatic as for overall (unconditional) APR. For example, in 2004, uncontrolled African American borrower differentials are 65 basis points in the subprime market and 59 basis points in the prime market, and these gaps decline to 10 basis points and 8 basis points respectively after accounting for observable characteristics. In 2005, African American borrower gaps of 49 basis points and 47 basis points for the subprime and prime markets, respectively, are narrowed to 3 basis points and 12 basis points, respectively, after controls.

The trends for Hispanic borrowers generally follow those of African American borrowers. In 2005, gaps of 16 basis points and 25 basis points for the subprime and prime markets, respectively, are reduced to 7 basis points and 12 basis points, respectively, after controls. The 2004 results for Hispanic borrowers show an interesting twist. The prime market gap reduces from 29 basis points to 4 basis points with controls. In the subprime market, however, the gap increases from a raw value of 4 basis points to 13 basis points after controlling for underwriting and pricing characteristics. This later outcome implies that, at least for the subprime market in 2004, Hispanic borrowers have slightly better characteristics (lower risk/cost) than White non-Hispanic borrowers in general.

The significantly higher (uncontrolled) APRs paid by minority borrowers are problematic, and need to be addressed. The empirical results suggest, however, that relatively little of the aggregate differences in APRs paid by minority and non-minority borrowers are attributable to the differential treatment of borrowers. Instead, in both years, up to 90% of the African American APR gap and 85% of the Hispanic APR gap is due to observable differences in underwriting, costing, and market dynamic factors that appropriately explain mortgage pricing differentials.[40]

Aggregate analyses like those in this study, of course, cannot assess whether there is differential treatment of individual minority mortgage borrowers. Nor should any such differential treatment that might occur be ignored. The analysis does suggest, however, that public policies aimed at remediating APR differentials would achieve a far greater return through the elimination of race/ethnicity differentials in FICO scores, income, wealth used to lower LTV ratios, and, arguably, financial literacy, than they would through the elimination of any possible differential treatment.

Finally, the analysis of mortgage pricing outcomes is unique in its use of an endogenous switching regression framework. The data strongly reject the null hypothesis that the estimating equations are independent, supporting the need to explicitly address the potential endogeneity inherent in the setting of APRs. As a practical matter, however, this approach can rarely be applied in regulatory reviews of a single lender, since loans available for analysis will often be either entirely subprime or entirely prime. Moreover, as noted above, the findings reveal a

relatively small correlation in the error terms and a small unexplained minority borrower differential in the probability of taking out a subprime mortgage, suggesting that explicitly accounting for endogeneity will generally have a relatively minor impact on the assessment of APR differentials between minority and non-minority borrowers.

Endnotes

[1] The rate spread reporting threshold is 3 percentage points above a comparable Treasury rate for a first lien loan and 5 percentage points above a comparable Treasury for a second lien loan. In calculating the rate spread, the lender uses the Treasury yield for securities of a comparable maturity as of the fifteenth day of a given month. Lenders use the fifteenth day of a given month for any loan on which the interest rate was set on or after that day through the fourteenth day of the next month. The APR used in the calculations is the one calculated and disclosed to the consumer under section 226.18 of Regulation Z (12 C.F.R. pt. 226).

[2] See, for example, *Inside Mortgage Finance*, 2006, issues 42 and 43, and Consumer Federation of America 2005 and 2006.

[3] See http://www.oag.state.ny.us and select December 2006 press releases.

[4] As Jack Guttentag notes, "APR is a good idea badly executed. . ." (http://www.mtgprofessor.com). The value of APR as a measure of price is that it incorporates both upfront fees and monthly payments. The disadvantage is that in so doing, APR assumes that loans run their full term to maturity. In fact, however, most borrowers prepay their mortgages well before they mature. This means that upfront fees tend to be undervalued in the APR calculation. That said, APR is the best, consistently applied measure of the all-in costs of taking out a mortgage.

[5] The point is that the authors have access to data that typically are available only to regulators conducting a second-stage analysis, not that the precise model structure or variable specifications used by these regulators are employed. In particular, no regulators use endogenous switching regression models, nor do they consistently rely as heavily on market-level controls. Further, regulators do not pool data across lenders but analyze each lender separately. Any unexplained race/ethnicity differentials in the models may result from this pooling.

[6] There is no consensus on how best to distinguish between subprime and prime—specifically, whether the distinction should be made at the lender or the loan level. Users of HMDA data have historically relied on lender designations provided by the Department of Housing and Urban Development. The introduction of loan-level rate-spread variables in the 2004 HMDA, however, has allowed researchers to make the subprime versus prime designation at the loan level. Because lenders in this study self-designate as either prime or subprime, and because lenders are believed to concentrate on serving higher- versus lower-risk borrowers who employ substantially different business strategies and practices, the analysis relies on lender designations to distinguish between subprime and prime.

[7] See Halvorson (1985) for an early application of switching regression models under similar circumstances.

[8] See Avery, Canner, and Cook (2005) and Avery, Brevoort, and Canner (2006). For a detailed history of the evolution of HMDA, see Kolar and Jerison (2005).

[9] See Bocian, Ernst, and Li (2006).

[10] See Consumer Federation of America (2005, 2006).

[11] For example, subprime lending is generally characterized by more reliance on direct marketing by lenders and shopping for an ''accept'' on the part of borrowers, while prime lending is more typically characterized by a softer sell on the part of lenders and price shopping on the part of borrowers. This reflects the substantially different business environments of the two markets.

[12] Minorities are disproportionately represented in FHA lending. However, as a consequence of limiting the data to conventional loans, there are no FHA mortgages in the analysis, so no potential pricing differentials in that market were explored.

[13] Additional variables are available from some, but not all, of the lenders contributing to the analysis.

[14] See Berkovec and Zorn (1996) for a discussion of the representativeness of the HMDA data.

[15] For the HMDA data comparisons, prime lenders in 2004 are defined as those with less than 30% of their mortgage originations with reported rate spreads, and in 2005 as those with less than 45% of their mortgage originations with reported rate spreads. The HMDA data are restricted to first lien, purchase and refinance, single-family, conventional originations only. If borrower race/ethnicity is unknown, the loans were excluded.

[16] Out of the total of 576 cells, there were 13 in 2004 with a zero loan count in the HMDA data, and in 2005 there were 8 with a zero loan count. The proprietary lender data had an additional 81 cells with a zero loan count in 2004, and an additional 130 cells with a zero loan count in 2005. To address the fact that the proprietary lender data did not fully span the support for the HMDA data, the 'nearest neighbor' cells were collapsed until there were no cells with zero loan counts in the proprietary lender data that did not also have zero loan counts in the HMDA data. This results in a total of 482 post-sampling weights in 2004 and 438 post-sampling weights in 2005. In the vast majority of instances, only the highest or lowest income cells were collapsed with the middle-income cell (e.g., combine the less than $20,000 cell with the between $20,000 and $200,000 cell). Occasionally, the loan purpose cells also were combined.

[17] The use of post-sampling weights does not affect the qualitative results. The unexplained APR differentials in Exhibit 11 range from 8 to 11 basis points. Without the post-sampling weights, the unexplained APR differentials range from 7 to 14 basis points.

[18] With a sample size of over one million, all differences in Exhibit 2 are statistically significant at p-values less than 0.0001.

[19] The larger importance of this component in 2005 is primarily due to the greater average difference in subprime and prime APRs in 2005.

[20] For a discussion of these issues see, for example, Mahoney (1998). As Mahoney notes, fair lending laws regarding disparate impact also require the search for a less discriminatory alternative, but the discussion and analysis of this point is beyond the scope of the current research.

[21] For a further discussion of these points, see Ross and Yinger (2006).

[22] For a discussion of switching regression models and their estimation see, for example, Maddala (1983).

[23] The beginning of a unit variance for the error term is assumed because the estimation technique (Probit) identifies $MODEL_1$ parameters only up to a normalization with respect to the standard error of the error terms.

[24] The dummy variable approach results in unexplained APR differentials (presented in Exhibit 11) ranging from 8 to 11 basis points. The unexplained APR differentials range from 8 to 13 basis points when all model coefficients are allowed to vary over borrower race/ethnicity.

[25] When the explanatory variables exclude tract mortgage denial percentage, tract percentage owner-occupied, and tract percentage subprime originations in the previous year, unexplained APR differentials range from 10 to 15 basis points. Adding to these exclusions the Herfindahl-Hirschman Index further increases the unexplained differentials to a range of 12 to 21 basis points. However, even with all four exclusions, over 80% of the raw APR differential for African American borrowers is explained by the models and over 70% of the raw differential for Hispanic borrowers is explained. A less radical approach to addressing the possible over-inclusion of variables is to subtract race/ethnicity-specific means from these four potentially problematic variables in an effort to 'purge' any correlation with borrower race/ethnicity. Doing so results in unexplained APR differentials that range from 12 to 17 basis points, and explain over 85% of the raw APR differential for African American borrowers and over 80% of the raw differential for Hispanic borrowers.

[26] The results are robust to these identification exclusions. Specifically, a system where the additional variables in $MODEL_1$ are also included in $MODEL_2$ and $MODEL_3$ is also estimated, along with a system where the variables are excluded in all three models. As well, we separately estimate systems where each of our excluded variables is included one at a time in $MODEL_1$. In all instances, the estimated correlation coefficients are very similar to those presented in Exhibits 9 and 10, and the APR differentials are very similar to those presented in Exhibit 11.

[27] LTV, DTI, and loan amount are the variables most generally considered to raise endogeniety concerns. To address endogeniety concerns, a more reduced form version of the model that excludes these three variables while adding borrower income as a percentage of area median income is estimated. The resulting APR differentials of this alternative specification are very similar to those presented in Exhibit 11.

[28] For a discussion of the movestay command, see Lokshin and Sajaia (2004).

[29] This is the decomposition first proposed by Blinder (1973) and Oaxaca (1973).

[30] The authors thank Glenn Canner for his suggestion to include the HHI and the tract denial percentage.

[31] The HHI is constructed as the sum of squared market shares of firms in a tract. As such, the index ranges from 10,000 in the case of 100% market concentration to near zero in the case of many firms with equally small market shares.

[32] The nonlinear nature of the inverse Mills ratio terms in Equations (6) and (7) ensure system identification. Identification is enhanced, however, by including variables in the probit model of market determination that are excluded in the models of subprime and prime APR determination as expressed in Equations (4) and (5). As noted previously, the results are robust to the inclusion of these variables.

[33] In these exhibits, splined variables are treated as categorical, with the categories defined by each segment of their linear spline.

[34] As a robustness test, the subprime and prime observations were also combined and a simple OLS regression of APR was run. The unexplained APR differential for African American borrowers is 15 basis points for both 2004 and 2005, and for Hispanic borrowers it is 9 basis points in 2004 and 15 basis points in 2005.

[35] See, for example, LaCour-Little (2007) on this point.

[36] The unexpected result in the prime market may be due to the way in which APR calculations treat discount points. Specifically, subprime borrowers are more likely to be income constrained, and so may be more likely to lower their mortgage contract rates by paying points and financing them through an increased mortgage balance. APR calculations discount points over the full length of the loan term (typically 30 years), while the market typically uses the expected life of the loan (2 to 10 years) in establishing the pricing trade-off between points and contract rate. The result is that the APR calculation undervalues points relative to the market. All things equal, therefore, borrowers taking out points will tend to have lower APRs than borrowers who do not. Since subprime borrowers are more likely to take out points, borrowers actually taking out prime mortgages likely will tend to have higher APRs than would random borrowers given a prime mortgage.

[37] Some critics have argued that subprime borrowers are forced/required to take out prepayment penalties, suggesting that this variable is more the ''result of'' rather than the ''cause of'' taking out a subprime mortgage. Similar concerns have been expressed regarding contract rate type. To address these concerns, estimations excluding these variables from the subprime probability model were also run while continuing to include them in the subprime and prime APR models. Not surprisingly, excluding these variables reduces the predictive power of the subprime model, and so results in somewhat larger unexplained APR differentials. The change is biggest for Hispanic borrowers, but even here the model explains roughly 75% of the raw APR differential. For African American borrowers, the reduced model explains roughly 90% of the raw APR differential.

[38] There are slight differences between the values in Exhibit 2 and equivalent values in Exhibit 11, since Exhibit 11 values incorporate adjustments for endogeneity and nonlinear components to their predictions.

[39] With the exception of differences in subprime probabilities for Hispanic and White non-Hispanic borrowers with Hispanic characteristics, all differences in Exhibit 11 are statistically significant.

[40] There is a component of art in determining the variables to appropriately include on the estimations. The results of many individual tests demonstrate the robustness of the findings to alternative variable specifications. A final relatively comprehensive such test is to: (1) address concerns that the possible over-inclusion of variables correlated with borrower race/ethnicity may downward-bias unexplained APR differentials by subtracting race/ethnicity-specific means from the set of potentially problematic variables (tract HHI, tract mortgage denial percentage, tract percentage owner-occupied, and tract percentage subprime originations in previous year) to ''purge'' their correlation, and (2) restrict from the market determination model the prepayment penalty and contract rate type variables because of concerns that they are ''caused'' by the determination of subprime versus prime market rather than the reverse. Not surprisingly, these adjustments increase the unexplained APR differentials, primarily because of the resulting model's reduced ability to explain the probability of taking out a subprime mortgage. Nonetheless, this arguably over-simplified model explains over 80% of the raw APR differential for African American borrowers, and roughly 70% of the APR differential for Hispanic borrowers.

References

Avery, R.B., G.B. Canner, and R.E. Cook. New Information Reported Under HMDA and Its Implications in Fair Lending Enforcement. *Federal Reserve Bulletin*, Summer, 2005.

Avery, R.B., K.P. Brevoort, and G.B. Canner. Higher Priced Home Lending and the 2005 HMDA Data. *Federal Reserve Bulletin*, Summer, 2006.

Barth, J., J. Cordes, and A. Yezer. Financial Institution Regulations, Redlining and Mortgage Markets. In *The Regulation of Financial Institutions*, conference volume 21, Federal Reserve Bank of Boston, 1980, April, 101–43.

Berkovec, J. and P. Zorn. How Complete is HMDA?: HMDA Coverage of Freddie Mac Purchases. *Journal of Real Estate Research*, 1996, 11:1, 39–56.

Blinder, A. Wage Discrimination: Reduced Forms and Structural Estimates. *Journal of Human Resources*, 1973, 8:4, 436–55.

Bocian, D.G., K.S. Ernst, and W. Li. Unfair Lending: The Effect of Race and Ethnicity on the Price of Subprime Mortgages. Center for Responsible Lending Working Paper, May, 2006.

Consumer Federation of America. Subprime Locations: Patterns of Geographic Disparity in Subprime Lending. Unpublished report, 2005.

——. Subprime Locations: Patterns of Geographic Disparity in Subprime Lending. Unpublished report, 2006.

Courchane, M., B. Surette, and P.M. Zorn. Subprime Borrowers: Transitions and Outcomes. *Journal of Real Estate Finance and Economics*, 2004, 29:4, 365–92.

Halvorsen, A. Switching Regression Estimates of a Sequential Production Process: The Case of Underground Coal Mining. *The Review of Economics and Statistics*, 1985, 67:10, 161–65.

Kolar, J. and J. Jerison. The Home Mortgage Disclosure Act: Its History, Evolution, and Limitations. Buckley Kolar presentation, March, 2005.

LaCour-Little, M. The Home Purchase Mortgage Preferences of Low- and Moderate-Income Households. *Real Estate Economics*, 2007, 35:3, 265–90.

Lokshin, M. and Z. Sajaia. Maximum Likelihood Estimation of Endogenous Switching Regression Models. *The Stata Journal*, 2004, 4:3, 282–89.

Maddala, G.S. *Limited-Dependent and Qualitative Variables in Economics*. Cambridge University Press, 1983.

Mahoney, P. The End(s) of Disparate Impact: Doctrinal Reconstruction, Fair Housing and Lending Law and the Antidiscrimination Principle. *Emory Law Journal*, 1998, 47:2, 409–525.

Oaxaca, R. Male-Female Wage Differentials in Urban Labour Markets. *International Economic Review*, 1973, 14, 251–70.

Pennington-Cross, A. Credit History and the Performance of Prime and Non-Prime Mortgages. *Journal of Real Estate Finance*, 2003, 27, 279–301.

Ross, S. and J. Yinger. *The Color of Money: Mortgage Discrimination, Research Methodology, and Fair Lending Enforcement*. MIT Press, 2002.

——. Uncovering Discrimination: A Comparison of the Methods Used by Scholars and Civil Rights Enforcement Officials. *American Law and Economics Review*, 2006, 562–614.

Yezer, A., R. Phillips, and R. Trost. Bias in Estimates of Discrimination and Default in Mortgage Lending: The Effects of Simultaneity and Self-Selection. *Journal of Real Estate Finance and Economics*, 1994, 9:3, 197–217.

The views expressed in this paper are those of the author and are not necessarily those of CRA International. Sincere thanks are due Peter Zorn, whose support to me

has been exceptional, and to Adam Gailey for his outstanding research assistance on this project, and the lenders, who must remain anonymous, for sharing their data with us. I also acknowledge the helpful comments from colleagues, as well as those from Michael LaCour-Little, Roberto Quercia, Stephen Ross, and Tony Yezer. Any errors or omissions in the paper are mine alone.

Marsha J. Courchane, CRA International, Boston, MA 02116-5092 or mcourchane@crai.com.