Mortgage Lending in Boston: Interpreting HMDA Data

By ALICIA H. MUNNELL, GEOFFREY M. B. TOOTELL, LYNN E. BROWNE, AND JAMES MCENEANEY *

The Home Mortgage Disclosure Act was enacted to monitor minority and low-income access to the mortgage market. The data collected for this purpose show that minorities are more than twice as likely to be denied a mortgage as whites. Yet variables correlated with both race and creditworthiness were omitted from these data, making any conclusion about race's role in mortgage lending impossible. The Federal Reserve Bank of Boston collected additional variables important to the mortgage lending decision and found that race continued to play an important, though significantly diminished, role in the decision to grant a mortgage. (*JEL* G14, G21, J71)

Access to credit markets is vital to disadvantaged minorities if they are to overcome the low level of their initial endowments. Any barriers to obtaining credit, particularly mortgage credit, would clearly inhibit the ability of minorities to escape or improve poor neighborhoods. The Home Mortgage Disclosure Act (HMDA) was passed to monitor minority access to the mortgage market. The data collected as a result of HMDA, which include each applicant's race, gender, income, and whether the application was accepted or denied, showed substantially higher denial rates for black and Hispanic applicants than for white applicants. These minorities were two to three times as likely to be denied mortgage loans as whites. In fact, high-income minorities in Boston were more likely to be turned down than low-income whites. The 1991, 1992, and 1993 HMDA data all showed a similar pattern.

This pattern has triggered an intense debate about whether discrimination based on the applicant's race occurs in mortgage lending. Many have argued that these disparities in denial rates are proof of discrimination on the part of lending institutions. Others, including lenders, assert that such conclusions are unwarranted because the HMDA data do not include information on credit histories, debt burdens, loan-to-value ratios, and other factors considered in making mortgage decisions: they argue that these missing pieces of information, not discrimination, explain the high denial rates for minorities. So far, the debate has turned on the ability of the HMDA data to reveal discrimination.

In an attempt to address the limitations of the HMDA data, the Federal Reserve Bank of Boston, with the support of the other federal supervisory agencies, asked lending institutions operating in the Boston Metropolitan Statistical Area (MSA) to specify additional information on the financial, employment, and property characteristics that are relevant to their lending decision.[1] This information, along with other variables that models of

* Member, Council of Economic Advisers; Assistant Vice President and Economist; Senior Vice President and Director of Research; and Assistant Vice President, respectively, Federal Reserve Bank of Boston, Boston, MA 02106. The views expressed are those of the authors, and do not necessarily reflect official positions of the Council of Economic Advisers, the Federal Reserve Bank of Boston, or the Federal Reserve System.

The authors are grateful to Joe Peek, Orley Ashenfelter, Robert B. Avery, Glenn B. Canner, John F. Kain, Mark Sniderman, and two anonymous referees. Special thanks go to interns Janet Feldstein and Virginia Genao, and to research assistants Faith Kasirye and Thomas Miles. Finally, the authors wish to thank the other federal supervisory agencies, including the OCC, the FDIC, OTS, and HUD, and all the financial institutions that responded to the questionnaire, making this study possible.

[1] Boston is the eighth largest MSA in the nation, with a population of 2.9 million. About 15 percent of the Boston MSA population is minority, and many of Boston's neighborhoods are defined along ethnic and racial lines.

mortgage lending indicate should be important to the lending decision, was requested for all applications for conventional mortgage loans made by blacks and Hispanics in 1990 and for a random sample of 3300 applications made by whites. Substantial lender compliance resulted in a very good response rate. A mortgage lending model, where the decision to lend depends on the additional variables collected by the Boston Fed survey as well as Census information on neighborhood characteristics, was employed to test whether race is still a significant factor in the lending decision once financial, employment, and neighborhood characteristics are taken into account.

The results of this study indicate that minority applicants, on average, do have less wealth, weaker credit histories, and higher loan-to-value ratios than white applicants, and that these disadvantages do account for a large portion of the difference in denial rates. Including the additional information on applicant and property characteristics reduces the disparity between minority and white denials from the originally reported 18 percentage points, or a relative rejection ratio of 2.8 to 1, to just over 8 percentage points, or a relative rejection ratio of roughly 1.8 to 1. Put another way, white applicants with the same property and personal characteristics as minorities would have experienced a rejection rate of 20 percent rather than the minority rate of 28 percent.[2] Thus, in the end, a statistically and economically significant gap between white and minority rejection rates remains.

Given the importance of the secondary market buyers of mortgages, and their detailed guidelines, it is not immediately clear how race could play an independent role. The information gathered in this survey provides some insight into how this outcome could emerge: the vast majority of applications—by both whites and minorities—are not "perfect"; they fail to meet at least one of the secondary market standards.[3] Lenders, therefore, have considerable discretion over the extent to which they consider "compensating factors" for these failings. Although discretion to approve applications that fail to meet some standard may be justified because historically default rates on residential mortgages have been very low, discrimination can arise unless the primary market lenders apply this flexibility in an evenhanded manner. The results of this study suggest that, given the same property and personal characteristics, white applicants may enjoy a general presumption of creditworthiness that black and Hispanic applicants do not.

I. Previous Studies of Mortgage Lending

Previous attempts have been made to analyze discrimination in the mortgage market. Harold Black et al., (1978), Thomas A. King (1980), and Robert Schafer and Helen F. Ladd (1981) all found that being a minority increased the probability of being rejected for a mortgage loan. Skepticism endured about these results, however, since the studies examined applications only in selected institutions and omitted several key variables that could be correlated with race. All three works analyzed only certain types of lenders, those under a given regulator, and in one study the survey response was completely voluntary. Furthermore, they all omitted some important variable, such as information on the applicant's credit and employment histories, which raised some doubts about the finding of discrimination. Finally, these results are at least 10 years old; whether discrimination persists is important. This study samples all mortgage lenders, over all regulators, for a much larger array of potentially significant variables, and does so for 1990.

II. A Model of the Mortgage Lending Decision

To determine whether race plays an independent role in the mortgage lending deci-

[2] The effect of race on the mortgage lending decision can be measured in several ways. These different methods will be discussed in detail later in this paper.

[3] For example, more than one half of the applications in this sample exceeded one of the secondary market benchmarks of 28 and 36 percent, respectively, for the ratio of monthly housing expense and total debt burden to total monthly income. Some loans were approved and sold into the secondary market with these ratios in excess of 40 percent and 50 percent, respectively.

sion, it is necessary first to account for all the economic factors that might bear on the financial institution's decision. Omitting a relevant variable that may be correlated with race is a problem with all the preceding work in this area. Thus, this study attempts to control for all the economic variables that may be systematically related to the mortgage lending decision, in order to examine race's role. A brief discussion of how mortgage lenders maximize profits helps to clarify what variables are important.

Mortgage lenders are assumed to maximize their expected profits given their information set at the time of the decision. Maximizing expected profits requires maximizing, through time, the expected difference between the return on the mortgage and the costs of the funds to the lender. The expected return on a mortgage depends on both the interest rate charged for that mortgage and the probability that the borrower defaults. It is assumed that both the mortgage rate and the rate at which lenders borrow are set by competition in the industry. Since the choice variable for the lender is not the interest rate but whether to grant the mortgage at all, mortgage applications are accepted or rejected at the market mortgage rate.[4] As a result, expected profit maximization depends almost exclusively on granting mortgages that minimize the probability of and costs to default.[5]

The risks and costs of default depend on a variety of loan, property, and personal characteristics.[6] The lender uses the information in the applicant's loan file to form an expectation of the risks and costs of default. Few studies have examined the correlation between defaults and the information set of the lender, and surprisingly little is known about the relationship between applicant characteristics and loan performance. Consequently, any empirical study of mortgage lending must include those variables that lenders actually consider when making their decisions, rather than simply what they ought to consider.

As a result, explaining what lenders regard as important in the mortgage decision requires empirically testing all the variables in the lender's information set. The original HMDA data include only a small subset of this information, data about the applicant's race, sex, and income, the property location, and the mortgage

[4] Some evidence (King, 1980) suggests that the terms of the loans offered to minorities are little different from those offered to whites, but the evidence in this area is not conclusive. Any correlation found between interest rates and race has, however, been positive; thus, if minorities tend to get quoted higher rates, the omission of the interest rate from this study would underestimate the effect of race on the mortgage lending decision.

It is also possible that the lender has other choice variables that may affect the probability of default. For example, Anthony M. J. Yezer et al. (1994) assert that loan-to-value ratios may be endogenous. If so, any subsequent bias to the coefficient on race would occur through the estimation of the coefficient on the loan-to-value ratio in the denial regression. There are several ways to check this assertion.

As a sensitivity test of how any endogeny of the loan-to-value ratio affects the coefficient on race, a probit regression was estimated constraining the coefficient on the loan-to-value ratio to be equal to two standard deviations above and below its estimated value in the base regression shown in Table 2. Even over this range, the impact on the size and significance of the race coefficient was minimal. Second, in an ordinary least squares (OLS) regression of the loan-to-value ratio on all of the variables in the denial equation, only a few variables were statistically significant and almost all were of the opposite sign from what one expected if lenders were negotiating the loan-to-value ratio. Third, making the loan-to-value ratio an instrumental variable in the denial equation has little effect on the size or significance of the coefficient on race. See Tootell (1995a) for a more complete discussion of this issue. For both conceptual and empirical reasons, it does not seem likely that a simultaneity problem is biasing the coefficient on race.

[5] Expected profit maximization also depends on the risk of interest rate changes and of prepayment of the mortgage. The risk due to potential changes in interest rates is affected by the loan characteristics of the mortgage, along with exogenous macroeconomic variables, but is unaffected by personal and property characteristics; this risk determines the equilibrium spread between the different loan instruments, not who is accepted or rejected for a loan. The propensity to prepay the loan, however, could be a function of personal characteristics such as age and occupation, for example, which are included in our tests.

[6] There are several reasons why the ability to sell the loan on the secondary market does not eliminate the importance of minimizing the cost and probability of default. First, 50 percent of the loans in our sample were not sold to another institution in the year they were originated. Second, even if the lender sells the loan on the secondary market, default remains a concern, as the purchaser may be able to return the loan to the originator. Third, secondary market buyers will not continue to buy from lenders whose loans frequently default.

amount. Lenders, however, consider many other variables, which they believe determine the risks and costs of a default. Thus, the Federal Reserve Bank of Boston augmented the 1990 HMDA reports by gathering information on 38 additional variables located in the standard loan application form, the credit report, and the lender's worksheet, which constitute practically the entire information set of the lender. These variables were selected based both on expectations of what should be important and on numerous conversations with lenders, underwriters, and others familiar with the lending process about what they believe is important. The next section briefly explains the choice of variables surveyed.

III. The Data

For expositional ease, the variables are grouped into several categories. The first includes variables that predominantly affect the probability of default. The second includes those variables that affect the costs if a default should occur. Finally, the potentially relevant loan and personal characteristics are grouped together.

A. *Probability of Default*

The original HMDA data omit much of the financial information that determines the ability of the applicant to carry the loan. Net wealth, liquid assets, and obligation ratios, both proposed housing expenses relative to income and total debt payment obligations relative to income, are not collected under the provisions of HMDA. Since these variables should be important in the lender's calculation of the applicant's probability of default, and they affect the marketability of the loan to secondary lenders, they were collected in the survey.

The amount of the loan relative to the appraised value of the property was examined, as it measures both the applicant's equity share in the property and the lender's margin of safety if housing prices decline or the loan becomes delinquent. The larger the applicant's equity share, the less likely the applicant will default. Furthermore, if the borrower can no longer maintain the loan, then the larger the down payment, the larger the cushion for the lender to recoup any transactions costs and missed payments.

Lenders also believe that economic fundamentals are not the only important determinants of default risk; the applicant's previous "commitment" to meeting credit obligations is also thought to predict whether someone will default in the future. Although credit projections, not histories, are what should be relevant to the mortgage lending decision, lenders use the applicant's credit history report as an instrument to measure this commitment to debt repayment. Furthermore, loan underwriters tend to view certain elements of the credit report, such as public records of defaults or bankruptcies, as more important than others. This study constructs a concise profile of the prospective borrower's past creditor relationships that provides substantial detail about different aspects of the applicant's credit history.

Because mortgage obligations span a relatively long horizon, the stability of the applicant's income stream is also an important determinant of the risk of default. Specifically, any labor market variable that affects the probability that the applicant may suffer a spell of unemployment would affect the expected profitability of the loan. The mortgage application form contains considerable information concerning the labor force status of the potential borrower. In addition to earnings, the lender gathers information on the applicant's profession, years in that profession, current job seniority, age, gender, level of education, and the industry in which the applicant is employed. This study uses the information in the loan application to calculate an estimate of the probability that the applicant might experience a spell of unemployment. Two basic approaches to measuring this risk are attempted, one based solely on the state of the regional labor market in the applicant's industry and the other derived from all the detailed employment information provided on the applicant's mortgage form.

Since labor markets tend to be local, regional shocks to various industries should be important to the mortgage lending decision. Because unemployment rates are serially correlated and labor sharing is prevalent in some industries, the Massachusetts unemployment rate from the previous year in the industry in which the applicant was employed was one instrument used to measure employment risk. In addition, since the self-employed tend to suffer more income risk and their earnings are more difficult to verify, a dummy variable indicating self-employment was also included in the analysis.

Local unemployment rates by sector may capture industry and regional effects on the mortgage lending decision, but they ignore the consequences of differential treatment in the labor market due to the applicant's age, gender, seniority, education, race, and other personal characteristics. Longitudinal data from the Panel Study of Income Dynamics (PSID) cover the same labor market variables as data contained in the lenders' files. Incorporating only these variables, the PSID data were used to estimate equations predicting the probability that an individual will become unemployed in the next five years. These equations were estimated for each industry separately, each occupation classification separately, and the total sample.[7] The estimated coefficients from these equations were then combined with each applicant's labor market data to derive the lender's expected probability that the applicant would suffer a loss of income due to a spell of unemployment. If minorities are, in fact, the last hired and the first fired in a given industry, or if they are concentrated in more cyclical industries, then these equations would assign them a higher probability of unemployment, all else held constant, and they would be less likely to get a loan. Since unstable labor income could explain the denials that appear attributable to differential treatment by race, the explicit inclusion of a variable representing the probability of becoming unemployed captures the importance in the mortgage lending decision of expectations about future economic conditions, and it distinguishes discrimination in the mortgage market from effects related to race in the rest of the economy.

B. *Costs of Default*

While credit history and employment stability furnish information about the possibility of default, several other variables collected to supplement the HMDA data provide some indication of the magnitude of the loss, should default and foreclosure occur. These variables include whether private mortgage insurance was purchased and neighborhood characteristics that might affect the stability of the value of the mortgaged property.

The acquisition of mortgage insurance decreases the potential cost to the lender if the applicant defaults. The secondary market usually requires private mortgage insurance (PMI) if the loan-to-value ratio is above 80 percent; thus, the marketability of the loan may also be affected by the PMI decision. Consequently, information on whether the applicant applied for and received mortgage insurance was also collected. Because the appropriate treatment of borrowers who applied for mortgage insurance is debatable, this issue will be examined in several different ways.

Risk-averse lenders will avoid loans with the same expected probability and costs of default but higher variability of potential losses. The variability of such losses depends on the variability of housing prices. This study attempts to measure this risk to the lender in a variety of ways. Traditionally, this risk is assumed to be a neighborhood-specific trait rather than a borrower or property-specific trait. Thus, in Robert B. Avery et al. (1994), Census tract dummies were used to capture the riskiness of owning property in each neighborhood. All of the results presented in this paper include dummies for each tract.

Unfortunately, tract dummies do not distinguish between asset price risk and the other neighborhood characteristics that might affect the mortgage lending decision. As a result, this study also examines several sharply drawn,

[7] The variables used to estimate the probability of an applicant in a given industry or occupation suffering a spell of unemployment were the variables in both the PSID and the mortgage application—age, gender, race, education, years on the job, and marital status. The PSID includes all the employment variables that are listed on the mortgage application form. The resulting reduced-form prediction equation is an attempt to mimic the risk assessment performed by the loan officer. In only a few of these industry or occupation equations was race significant. Education, age, marital status, and years on the job were generally statistically significant and of the expected sign, however.

tract-specific instruments to measure the risk associated with the value of the property. The central measure adopted is the ratio of rent to the value of the rental housing stock in the Census tract where the property is located, which can be approximated from Census data.[8] To compensate investors for the higher risk, the same amount of capital invested in an area with greater potential for a capital loss should generate a higher stream of earnings. Other instruments for asset price risk are also examined, including vacancy rates and proportion of boarded-up property in the tract, as well as the percentage change in median house prices in the tract over the past decade.

C. *Loan Characteristics*

Certain characteristics of the loan itself may also be important to the riskiness or costliness of a default. The sample was limited to conventional mortgages, because FHA and VA loans are uncommon in the Boston metropolitan area and because these special programs may have different lending standards from conventional loans.[9] The follow-up survey secured additional information on the terms of the loan: whether, for example, the loan's duration was 15 years or 30 years; whether the interest rate was fixed or adjustable; and whether the application was made under a special loan program designed for low-income individuals. The survey also asked whether the property was a single-family home, a condominium, or a building with two to four units. Finally, varying lending standards at the different institutions could also affect the finding on race. Minorities may be treated the same as whites within any given institution, but they may frequent institutions with tougher lending standards. To ensure that different institutional lending standards did not bias the estimation of the coefficient on race, an indicator variable for each lender was included in the regressions.

D. *Personal Characteristics*

Finally, the original HMDA data included information on the gender and race of the applicant and coapplicant. The follow-up survey requested data on the applicant's age, marital status, and number of dependents. Age could be an indicator of future earnings potential, as the slope of the age-earnings profile changes over the average person's working life. Marital status could affect the stability of the household income. Finally, as the number of dependents increases for any given level of income, the applicant is likely to have less income available to carry the loan.

In summary, the questions in the follow-up survey were designed to secure all the financial, employment, and demographic information that lenders include in their determination to approve or deny a loan application. All told, 38 additional variables were collected.[10]

IV. Survey Design and Results

Because the high denial rates for minorities prompted the survey and because only 1,200 blacks and Hispanics applied for mortgages in Boston in 1990, the survey was designed to collect the additional information on every black and Hispanic applicant. The additional variables were also collected for a random sample of whites. To ensure that any test for discrimination would have sufficient power, the number of whites sampled was chosen to produce roughly equal numbers of white and minority denials; with a 10 percent white rejection rate, a sample of 3,300 applications

[8] Several variations of this variable were examined. The aggregate value of contract rent is given for each tract. Unfortunately, the value of rental property is not collected in the Census but must be estimated. The measure used in this study calculates the aggregate value of the tract's rental property by multiplying the number of rental units by the average value of owner-occupied property in that tract. The validity of this procedure rests on the assumption that rental and owner-occupied property are close substitutes.

[9] In the Boston metropolitan area in 1990, only 4 percent of all home-purchase applications (only 4.5 percent of applications by blacks and 3.5 percent of applications by Hispanics) were for government-backed mortgages. Thus, the conventional mortgage represented the norm for blacks, Hispanics, and whites in Boston.

[10] A copy of the survey is available from the authors upon request.

by whites was required.[11] The sample of applications by whites was selected randomly rather than matched with black and Hispanic applications by institution or key borrower characteristics, because matching would have required prejudging the causes of rejection and precluded an evaluation of the role that variables used in the matching process played in determining rejection rates.

A. *Final Sample*

A high degree of compliance by lenders and considerable follow-up resulted in a very high response rate. However, the failure and closing of several lenders, and the loss by several institutions of some files, caused the survey and the final sample to diverge somewhat.[12] Refinancings, home improvement loans, and some business loans that institutions had mistakenly coded as mortgage originations in their original filings were removed from the sample. Institutions also made corrections to their designation of the race of some applicants. Furthermore, applications with data missing for one of the key variables in the regressions were dropped.[13] As will be explained later, the decision was made to exclude applications that were withdrawn.

B. *Values of Collected Variables*

The means of the variables collected in the follow-up survey are presented in Table 1 for both white applicants and black/Hispanic applicants. These data and all subsequent analyses combine applications by blacks and Hispanics. Both blacks and Hispanics had substantially higher denial rates than whites, and the number of applications by Hispanics was too small to rigorously analyze separately. Moreover, the hypothesis that the relevant independent variables affect the probability of denial for the two minority groups similarly, and the hypothesis that the coefficients on separate black and Hispanic indicator variables are identical, could not be rejected.[14]

The data in Table 1 show that black and Hispanic applicants in the Boston area differ from white applicants in a number of ways. As reported in other surveys, black and Hispanic applicants have considerably less net wealth, liquid assets, and income than whites and they have weaker credit histories. Blacks and Hispanics also make lower down payments and have higher loan-to-value ratios than whites and, therefore, apply more frequently for private mortgage insurance. These differences tend to support arguments that the higher denial rates experienced by minorities are attributable, at least in part, to economic factors.

[11] Practical considerations required limiting the lending institutions surveyed to those that received at least 25 mortgage applications. This reduced the pool of applications only slightly, but decreased the number of lenders to be contacted from 352 to 131. The responses produced almost exactly the same numbers of white and minority rejections.

[12] The pattern of lending by type of institution, however, is very similar to that reported for the original HMDA data. In both the sample and the population, applications are split relatively evenly between depository institutions and mortgage companies; this is true for blacks and Hispanics as well as for whites.

[13] Only around 6 percent of the applications surveyed fell into this missing observations category. Withdrawals were a disproportionately larger percentage of these applications (18 percent, as opposed to 10 percent in the general sample). Rejections, for both whites and minorities, were also disproportionately represented in the applications with missing data. This pattern should be expected, as withdrawals often occur in the middle of the process before all the information can be collected, and applications with unverifiable or missing data are more likely to be rejected. If omitting observations with missing data biased the results, one would expect the patterns of minority and white acceptances and rejections in this group to be different; yet the patterns of acceptances, rejections, and withdrawals for observations with missing data look the same for whites and for minorities. Since the characteristics of these files appeared to be identical for minorities and whites, these omissions were assumed to be random noise. Their exclusion from the regressions should not cause a bias in the results.

[14] A log-likelihood test of the equality of all the coefficients in a black versus a Hispanic equation could not reject at the 5 percent level the hypothesis that the two sets of coefficients were identical. Also, if only the race coefficient for Hispanics and blacks is allowed to differ, one cannot reject the hypothesis that the coefficient on the two different race variables is the same. Because of the few degrees of freedom, the tests were performed without the tract and lender variables. If the tract and lender dummies are included and constrained to be identical across whites and minorities, the hypothesis that the coefficient on the indicator variable for blacks was identical to that on Hispanics could not be rejected.

TABLE 1—VALUES OF VARIABLES COLLECTED ON 1992 FOLLOW-UP SURVEY

Follow-up survey question number	Characteristic	Total	White	Black/Hispanic
	Percent rejected	15	10	28
20	Mean number of units in property purchased	1	1	1
21	Mean age of applicant	37	36	37
21c	Mean age of coapplicant	35	35	35
22	Mean years of school (applicant)	15	16	14
22c	Mean years of school (coapplicant)	15	15	13
23	Percent of applicants married	59.9	61.8	53.6
23c	Percent of coapplicants married	56.1	58.7	47.5
24	Mean number of applicant dependents	1	1	1
25	Mean number of years in line of work (applicant)	20	23	13
25c	Mean number of years in line of work (coapplicant)	8	8	7
26	Mean number of years on job (applicant)	17	19	10
26c	Mean number of years on job (coapplicant)	5	5	5
27	Percent of applicants self-employed	11.7	13.0	7.5
27c	Percent of coapplicants self-employed	7.2	8.4	3.0
30	Mean base monthly income (applicant)	4,109	4,407	3,132
30c	Mean base monthly income (coapplicant)	1,335	1,390	1,155
31	Mean total monthly income (applicant)	4,710	5,074	3,513
31c	Mean total monthly income (coapplicant)	1,453	1,507	1,275
32	Mean proposed monthly housing expense ($)	1,440	1,507	1,223
33	Mean purchase price ($)	186,000	197,000	150,000
34	Percent with other financing	4.7	3.6	8.5
35	Mean value liquid assets ($)	85,000	99,000	41,000
36	Mean value total assets ($)	316,000	373,000	128,000
37	Mean total nonhousing monthly payments ($)	471	485	427
38	Mean value of total liabilities ($)	77,000	90,000	36,000
39	Mean number of commercial credit reports on file	1	1	1
40	Percent meeting credit history guideline for approval	89.8	93.6	77.4
41	Mean number of credit lines on report	13	14	11
42	Percent with more than two late mortgage payments	2.7	1.0	0.9
43	Percent with delinquent consumer credit accounts	16.8	13.9	26.6
44	Percent with public record defaults	7.9	5.9	14.7
45	Mean obligation ratio (housing expense/income)	25.4	25.2	26.2
46	Mean total obligation ratio (total obligations/income)	33.3	32.8	34.8
47	Percent of loans with fixed rates	66.2	67.5	62.0
48	Percent of loans with 30-year rates	86.9	85.9	90.8
49	Percent of loans in special programs (MHFA)	7.3	3.3	20.2
50	Mean appraised value of property ($)	195,000	206,000	157,000
51	Type of property:			
	Percent single family	60.9	67.9	38.1
	Percent condominium	25.6	23.1	33.9
	Percent 2–4 family	13.4	9.0	28.0
52	Percent seeking private mortgage insurance	24.9	21.2	37.1
53	Percent approved for private mortgage insurance	22.3	19.5	31.1
54	Percent with gift or grant account used as part of down payment	17.6	17.2	19.0
55	Percent with cosigner on application	3.7	3.6	4.2
56	Percent with unverifiable information	5.7	4.1	11.0
57	Percent reviewed more than once by underwriter	55.8	56.3	54.0
	Mean of loan to value	0.8	0.7	0.8

Note. Percentage base for each item does not include applicants for whom information was missing.

V. Regression Results

Both ordinary least squares and binomial logit techniques were used to estimate the effects of the collected data on the probability of being denied a mortgage. The logit models produce consistent estimates of the standard errors and efficient estimates of the coefficients, while the ordinary least squares coefficients are more easily interpretable. The coefficients for a base model, which includes the obligation ratios, loan-to-value ratio, credit history, employment stability, property characteristics, tract and lender dummies, and race of the applicant, are presented in columns 1 and 3 of Table 2.[15] The precise definitions of the variables are provided in the Appendix. The logit results reveal that, with the exception of net wealth, all the variables in the equation have a statistically significant impact on the probability of denial, and this impact is in the predicted direction. The coefficient for race is large relative to many of the other coefficients in the model, and it is statistically significant well beyond the 1-percent level. The results from the OLS regression are completely compatible with those produced by the logit equations.

One benefit of OLS estimates is the ease with which one can interpret the coefficients. The OLS coefficient on race implies that the rejection rate of minority applications is 7 to 8 percentage points higher than the rejection rate of applications by whites with similar characteristics. On the other hand, interpreting the logit coefficients is a bit more problematic. Column 2 in Table 2 attempts to measure the impact of each variable on the probability of denial, using the coefficients from the logit estimation. For dichotomous variables, such as race, the figures in the second column represent the percentage point increase in the probability of denial associated with having that particular characteristic. That is, holding the other financial and property characteristics constant, the probability of denial increases 8.2 percentage points for a minority applicant.[16] For continuous variables, such as the total obligation ratio, the figures in the column 2 represent the percentage point increase in the probability of denial associated with a one standard deviation change in that variable. That is, if the total obligation ratio rises one standard deviation (11 percentage points) the probability of denial increases by 4.3 percentage points. The economic impacts predicted by the logit coefficients are similar to those predicted by the OLS estimates.

A. *Risk of Default*

As expected, the results in Table 2 confirm that high obligation ratios increase the probability of having a loan application denied. Because the two obligation ratios tend to move together, that is, an applicant with a high ratio

[15] The coefficients for the tract and lender dummies are not listed because they are so numerous. For the logit equations, the tracts with only acceptances and those with only rejections (only 18 applications fell into the only rejections category) are combined, since estimates of the coefficients of these dummy variables become infinite. Similarly, only separate dummy variables for lenders with both rejections and acceptances are estimated in the logit equation. For the OLS equations, dummies for all but one tract and lender are estimated.

[16] In deriving the impact values reported in Table 2, the first step is to determine the probability of denial for each individual applicant with a particular characteristic, such as being self-employed. This requires determining for each self-employed applicant the probability of denial based on the coefficients of the equation reported in Table 2. These estimated probabilities for each applicant are then averaged to get a single figure for that group. The same probabilities are then calculated for this group when the self-employment coefficient is omitted; that is, when the coefficient for self-employment is not added to the right-hand side of each applicant's equation. The figure reported in column 2 is the percentage point difference between the average probabilities of denial for the self-employed with and without the estimated self-employment effect.

For a continuous variable, for example the total obligation ratio, the first step is to determine the estimated probability of denial for each applicant in the sample, and then average the probabilities. The second step is to add one standard deviation to the total obligation ratio for each applicant, recalculate the estimated probabilities of denial for each applicant, and average the probabilities. As before, the value reported in column 2 is the percentage point difference between these two average probabilities. Note that the impact effects for the dichotomous variables are analogous to the OLS coefficients, but this is not so for the impact effects of the continuous variables, as these effects are calculated assuming a change of one standard deviation in that variable.

TABLE 2—DETERMINANTS OF PROBABILITY OF MORTGAGE LOAN APPLICATION DENIAL

Variable	Logit Base (1)	Logit Percentage point impact (2)	OLS Base (3)
Constant	−13.69		−0.22
	(12.62)		(−1.47)
Risk of default:			
Housing expense/income	0.63	4.6	0.06
	(2.76)		(3.56)
Total debt payments/income	0.08	4.3	0.005
	(7.16)		(7.21)
Net wealth	0.00008	0.4	0.000004
	(0.76)		(0.60)
Consumer credit history	0.51	4.1	0.04
	(9.16)		(9.46)
Mortgage credit history	0.43	1.2	0.03
	(2.27)		(2.16)
Public record history	1.95	19.3	0.19
	(6.50)		(8.09)
Unemployment region	0.11	1.2	0.01
	(2.64)		(1.93)
Self-employed	0.70	3.9	0.05
	(2.31)		(2.71)
Loan/appraised value-low	−0.89	−1.7	−0.12
	(−1.31)		(−2.46)
Loan/appraised value-medium	0.13	0.3	−0.05
	(0.23)		(−1.23)
Loan/appraised value-high	1.40	1.7	0.10
	(3.41)		(2.66)
Cost of default:			
Denied private mortgage insurance	6.16	65.0	0.65
	(8.55)		(16.06)
Loan characteristics:			
Two- to four-family home	0.73	5.7	0.06
	(2.64)		(2.57)
Personal characteristics:			
Race	1.00	8.2	0.07
	(3.73)		(3.34)
Percent correctly predicted	95.3		
Adjusted R^2			0.32
Number of observations	2,925		2,925

Notes: Numbers in parentheses are t statistics. Census tract (part of the cost of default) and lender ID (a loan characteristic) dummy coefficients are not shown because they are so numerous.

of total debt payments to income generally has a high housing expense ratio, it is difficult to sort out precisely the relative importance of the two ratios. Suffice it to say that these measures are crucial to the lending decision.

The net wealth coefficient is not statistically significant. This result is consistent with information provided by lenders consulted in the preliminary phase of this project; they indicated that little weight is placed on wealth because it is so difficult to verify. Liquid assets (not shown) also do not appear to affect the probability of denial, even though they are cited in secondary market guidelines as a compensating factor and were sometimes mentioned by lenders as an important consideration.[17] The reason may be that liquid assets are frequently used for the down payment and, therefore, their effect is captured by the loan-to-value ratio. Prescreening may also exclude people without enough cash to settle.

Credit information was categorized by the severity of the problem in the consumer, mortgage, and public records areas. The results show clearly that the more severe the applicant's past credit problems, the higher the probability that the loan was denied.[18] As one might expect, an earlier default or bankruptcy has the largest negative impact on the chances of obtaining a mortgage loan.

A high loan-to-value ratio also raises the probability of denial. Because of the clear thresholds in the secondary market standards concerning the loan-to-value ratio, a nonlinear formulation of this variable was used. Specifically, the loan-to-value variable was separated into three different variables: loan-to-value ratios less than or equal to 80 percent, greater than 95 percent, and those in between. Thresholds of 80 percent and 95 percent were chosen because the secondary market generally requires mortgage insurance for loans with loan-to-value ratios exceeding 80 percent and usually will not buy loans with loan-to-value ratios in excess of 95 percent. Further, Table 2 reveals that the denial of private mortgage insurance virtually precludes obtaining a mortgage. Few applicants were turned down for private mortgage insurance, but those who were turned down were very unlikely to get a mortgage.

Finally, instability of labor income, as measured by the regional unemployment rate in the applicant's industry and the dummy variable for self-employment, increases the probability of denial.[19] Applicants who are employed in depressed sectors of the regional economy have a lower probability of receiving a loan, all else held constant.

B. *Costs of Default*

In Table 3, the role that other variables may play in the mortgage lending decision is examined in detail. Results from both logit and OLS estimation consistently reinforce the findings in Table 2. Columns 1 and 6 of Table 3 provide the coefficients from regressions that contain almost every variable collected. Tract dummy variables are included since they are the most comprehensive indicator of neighborhood characteristics. Indicator variables for each lender as well as variables describing specific loan and personal characteristics are also included. The coefficient on race rises slightly from its level in the model shown in Table 2 and remains both economically and statistically significant well beyond the 1 percent level. Although the coefficient on the variable indicating whether the interest rate on the loan sought was fixed or adjustable is statistically significant in this equation, its significance is not robust to alternative specifications. All the other variables added to the base regression in Table 2 are insignificant. Almost 300 observations had to be dropped from this regression because only files with

[17] An equation was also estimated including income, liquid assets, the presence of a cosigner, and the ratio of base to total income. None of these variables have a statistically significant effect on the probability of denial, nor do they affect the coefficient on race.

[18] The consumer credit index has six different categories and the mortgage credit index has four. For example, the consumer payments variable equals 1 for an unblemished consumer credit history and 6 for the worst, accounts with 90 or more days past due. This approach assumes that moving from the first rank to the second has the same effect on the probability of getting a loan as moving from the fifth to the sixth. Tests were conducted allowing the coefficients to differ for each rank of all the different credit history categories; this essentially allowed for a nonlinear relationship between the credit history rankings and the probability of denial. A log-likelihood test could not reject the specification of the credit history variables used in the model. Furthermore, allowing a nonlinear relationship in the credit history indices had no effect on the size or significance of the race coefficient.

[19] An equation was estimated that also included years on the job. Since secondary market guidelines request documentation for applicants who have been on the job less than two years, this variable was tested both as a continuous variable and as a dummy variable equal to 1 if the applicant had been on the job for less than two years. The coefficients have the expected signs, but neither variable has a statistically significant effect on the probability of denial or on the coefficient on race.

TABLE 3—ALTERNATIVE DETERMINANTS OF PROBABILITY OF MORTGAGE LOAN APPLICATION DENIAL

	Logit				
Variable	General equation (1)	Tract characteristics (2)	Loan characteristics (3)	Personal characteristics (4)	Probability of unemployment (5)
Constant	−17.87 (−8.58)	−8.71 (−11.25)	−10.46 (−9.82)	−13.85 (−12.06)	−13.64 (−11.97)
Risk of default:					
Housing expense/income	0.72 (2.74)	0.51 (2.97)	0.76 (3.55)	0.70 (3.04)	0.67 (2.75)
Total debt payments/ income	0.11 (7.71)	0.06 (6.58)	0.07 (7.29)	0.08 (7.57)	0.09 (6.78)
Net wealth	0.0003 (1.89)	0.00008 (0.87)	0.00006 (0.79)	0.00008 (0.82)	0.00022 (2.10)
Consumer credit history	0.57 (8.83)	0.38 (9.53)	0.46 (8.93)	0.53 (9.44)	0.51 (8.62)
Mortgage credit history	0.39 (1.81)	0.41 (3.11)	0.43 (2.47)	0.49 (2.52)	0.35 (1.69)
Public record history	1.81 (5.31)	1.45 (6.75)	1.79 (6.55)	1.99 (6.59)	1.85 (5.70)
Unemployment region	0.08 (1.64)	0.07 (2.16)	0.09 (2.29)	0.10 (2.51)	
Self-employed	1.20 (3.48)	0.62 (2.88)	0.50 (1.78)	0.75 (2.49)	
Probability of unemployment	1.03 (1.17)				1.34 (1.95)
Loan/appraised value–low	−1.24 (−1.61)	−0.73 (−1.31)	−1.60 (−2.59)	−0.98 (−1.45)	−1.04 (−1.47)
Loan/appraised value–medium	0.22 (0.36)	−0.13 (−0.29)	−0.35 (−0.71)	0.07 (0.13)	0.18 (0.31)
Loan/appraised value–high	1.56 (3.16)	0.80 (2.17)	0.75 (2.50)	1.36 (3.35)	1.56 (3.47)
Cost of default:					
Denied private mortgage insurance	6.47 (8.44)	4.77 (8.64)	5.99 (8.46)	6.14 (8.64)	6.29 (8.59)
Rent/value in tract		0.67 (3.11)			
Housing units boarded up		−0.02 (−1.28)			
Housing units vacant		0.01 (0.42)			
Housing value appreciation		0.0002 (0.33)			
Loan characteristics:					
Two- to four-family home	0.65 (1.99)	0.62 (3.24)	0.79 (2.96)	0.62 (2.22)	0.68 (2.28)
Fixed-rate loan	0.73 (2.16)		0.56 (2.21)		
Special program (MHFA)	−1.03 (−1.92)		−1.17 (−2.61)		
Term of loan	−0.001 (−0.23)		0.0005 (0.28)		
Gift or grant in down payment	−0.32 (−1.11)		−0.38 (−1.61)		
Cosigner	−0.61 (−1.02)		−0.52 (−1.08)		
Personal characteristics:					
Race	1.10 (3.62)	0.54 (3.13)	0.88 (3.49)	1.01 (3.75)	0.98 (3.43)
Age	−0.0009 (−0.06)			0.01 (0.87)	
Gender (female = 1)	−0.56 (−1.94)			−0.52 (−2.03)	
Number of dependents	−0.001 (−0.01)			0.03 (0.32)	
Marital status (not married = 1)	0.31 (1.19)			0.48 (2.11)	
Percent correctly predicted	96.3	89.2	94.3	95.2	95.5
Adjusted R^2					
Number of observations	2,664	2,907	2,918	2,895	2,698

TABLE 3—*Continued.*

	OLS				
Variable	General equation (6)	Tract characteristics (7)	Loan characteristics (8)	Personal characteristics (9)	Probability of unemployment (10)
Constant	−0.28	−0.20	−0.26	−0.27	−0.18
	(−1.71)	(−3.25)	(−1.67)	(−1.78)	(−1.21)
Risk of default:					
Housing expense/income	0.06	0.05	0.06	0.06	0.06
	(3.62)	(3.71)	(3.69)	(3.63)	(3.43)
Total debt payments/	0.005	0.005	0.005	0.005	0.004
income	(7.19)	(8.26)	(7.48)	(7.53)	(6.78)
Net wealth	0.000009	0.000009	0.000004	0.000003	0.00001
	(1.21)	(1.52)	(0.69)	(0.47)	(1.46)
Consumer credit history	0.04	0.04	0.04	0.04	0.04
	(9.13)	(10.90)	(9.41)	(9.58)	(8.96)
Mortgage credit history	0.02	0.03	0.03	0.03	0.02
	(1.72)	(3.23)	(2.29)	(2.15)	(1.60)
Public record history	0.16	0.20	0.19	0.19	0.18
	(6.34)	(9.09)	(7.92)	(7.67)	(7.04)
Unemployment region	0.004	0.006	0.005	0.01	
	(1.21)	(2.20)	(1.63)	(1.86)	
Self-employed	0.06	0.05	0.06	0.05	
	(2.84)	(3.05)	(2.90)	(2.50)	
Probability of	0.06				0.07
unemployment	(1.08)				(1.54)
Loan/appraised value–	−0.12	−0.10	−0.11	−0.10	−0.14
low	(−2.29)	(−2.38)	(−2.23)	(−2.18)	(−2.73)
Loan/appraised value–	−0.03	−0.05	−0.03	−0.04	−0.06
medium	(−0.78)	(−1.35)	(−0.84)	(−0.95)	(−1.39)
Loan/appraised value–	0.10	0.10	0.09	0.10	0.10
high	(2.60)	(3.19)	(2.63)	(2.85)	(2.58)
Cost of default:					
Denied private mortgage	0.67	0.64	0.66	0.66	0.67
insurance	(15.97)	(17.55)	(16.19)	(16.06)	(15.89)
Rent/value in tract		0.07			
		(2.83)			
Housing units boarded up		−0.002			
		(−1.50)			
Housing units vacant		0.0001			
		(0.11)			
Housing value		0.00002			
appreciation		(0.29)			
Loan characteristics:					
Two- to four-family home	0.06	0.06	0.06	0.05	0.06
	(2.41)	(3.64)	(2.77)	(2.40)	(2.55)
Fixed-rate loan	0.03		0.02		
	(1.45)		(1.43)		
Special program (MHFA)	−0.09		−0.08		
	(−2.60)		(−2.48)		
Term of loan	−0.00009		−0.00002		
	(−0.06)		(−0.16)		
Gift or grant in down	−0.03		−0.03		
payment	(−1.46)		(−1.73)		
Cosigner	−0.04		−0.03		
	(−1.02)		(−1.02)		
Personal characteristics:					
Race	0.07	0.06	0.06	0.06	0.07
	(3.08)	(3.77)	(3.14)	(3.07)	(3.27)
Age	0.0006			0.001	
	(0.74)			(1.36)	
Gender (female = 1)	−0.04			−0.04	
	(−2.21)			(−2.48)	
Number of dependents	0.003			0.003	
	(0.44)			(0.56)	
Marital status (not	0.02			0.02	
married = 1)	(1.03)			(1.27)	
Percent correctly predicted					
Adjusted R^2	0.32	0.32	0.32	0.32	0.32
Number of observations	2,663	2,906	2,917	2,894	2,697

Notes: Numbers in parentheses are t statistics. Census tract and lender ID dummy coefficients are not shown because they are so numerous.

information on every one of these variables could be used; as a result, alternative specifications adding different subsets of the variables measuring the applicant's property, loan, and personal characteristics to the base model were examined in order to maximize the number of observations in each test.

C. *Neighborhood Characteristics*

Columns 2 and 7 of Table 3 provide the coefficients from a regression where specific tract characteristics, rather than tract dummy variables, are added to the model in Table 2. The rent-to-value ratio in the tract is the only tract-specific variable that significantly affects the mortgage lending decision. The previous housing price appreciation in the tract and the vacancy and boarded-up rates in the neighborhood add no information about the mortgage lending decision. The decline of the coefficient on race in this regression is due to the change in the constant when no indicator variables for the tracts are included; the economic impact of race remains roughly the same between the regressions in columns 1 and 2 and columns 6 and 7.

D. *Loan and Personal Characteristics*

Although loan and personal characteristics may be relevant in the determination of the probability of default, the costs of default, or even the probability of prepayment, economic theory does not offer strong priors about their effects. One loan characteristic that turned out to be important in all these equations is whether the applicant was applying for a mortgage for a two- to four-family home. Financial institutions clearly are less willing to make loans on two- to four-family housing that involves rental arrangements.[20]

Columns 3 and 8 of Table 3 examine the importance of other loan characteristics. The duration of the loan, whether the interest rate was fixed or adjustable, whether the loan was applied for under a special program, and whether a gift or grant contributed to the down payment were all included in the base model from Table 2. The coefficients for the variable indicating whether the applicant applied under a special program and whether the loan had a fixed interest rate were both statistically significant in this specification.[21] Applying under a special loan program or applying for an adjustable rate loan both decreased the probability of being rejected for a mortgage; the significance of these two effects, however, was not terribly robust to alternative specifications so these two variables were not included in the base model. The inclusion of these additional variables had some effect on the other coefficients in the model; the coefficient on race declines slightly over 10 percent, although it is still significant well beyond the 1-percent level.

Columns 4 and 9 of Table 3 examine the effects of other personal characteristics on the mortgage lending decision, such as age, gender, number of dependents, and marital status. In this specification, being married or being female tended to increase the chances of getting a loan, although when all the variables are included, as they are in columns 1 and 6, neither coefficient is statistically significant. The importance of the applicant's marital status and gender depends greatly on the exact specification of the model, which is why the base regression omits these variables. The applicant's age and number of dependents had no significant effect on the mortgage lending decision, and the size and significance of the coefficient for race were unchanged with the inclusion of these personal characteristics.

[20] If properly calculated, the obligation ratios should contain the relevant information about the risks associated with the stream of rental earnings. According to Fannie Mae guidelines, the rental income should be discounted by the lender before the obligation ratios are calculated and the loans then treated like any other. The significance of the multiunit variable suggests that lenders avoid these properties for reasons beyond the riskiness of the earnings stream, or that the discounting of the income is somehow biased.

[21] The major special loan program was that of the Massachusetts Housing Finance Authority, which does appear to have different standards from the conventional programs for the applicant's loan-to-value ratio and income. However, whether or not the loan had been applied for under this program did not significantly affect the mortgage decision in the first regression in Table 3. Most other programs cited as special by the lenders were really marketing tools, with no differences in lending standards.

Finally, columns 5 and 10 of Table 3 examine the predicted future unemployment probabilities derived from the PSID equations and the applicant's labor market characteristics.[22] This measure of employment risk approaches statistical significance at the 5 percent level but is not terribly robust to changes in the model's specification. Columns 5 and 10 make clear that whichever measure of employment risk is used, it has no effect on the estimated effect of race in the mortgage lending decision. The labor market does not appear to be the source of any omitted variable bias in the basic equation's results.

The consistently positive and statistically significant coefficient on race suggests that, even after accounting for the applicant's obligation ratios, wealth, credit history, and loan-to-value ratio, and property, neighborhood, and lender characteristics, as well as the stability of the applicant's income and whether he or she received private mortgage insurance, the race of the applicant still plays an important role in the lender's decision to approve or deny the loan. A minority applicant with average minority economic characteristics faces a probability of denial that is roughly 8 percentage points higher than that facing a white individual with the same characteristics.[23]

E. *Specification Issues*

The exact specification of the model is not certain. One potential problem is the appropriate treatment of private mortgage insurance. Insurers could be viewed as outside the direct lending market. To the extent that PMI denials fall disproportionately on minorities, excluding a variable representing denial of mortgage insurance from the equation would ascribe to lenders differential treatment that was occurring elsewhere in the system. For this reason, the denial of mortgage insurance was included in the base equation.

Yet the optimal treatment of PMI is unclear. On the one hand, mortgage insurers consider the same information as the lending institutions, so they could be viewed as just another lender and the mortgage insurance variable could be omitted from the equation.[24] Columns 2 and 5 of Table 4 show that omitting the mortgage insurance variable has little effect on the race coefficient. The estimates of the other coefficients are generally unchanged; the exception, not unexpectedly, is that the loan-to-value ratio takes on slightly greater importance in the absence of the private mortgage insurance variable. On the other hand, since it can be argued that the mortgage insurance decision is not the decision being modeled, columns 3 and 6 contain estimates of the coefficients when private mortgage insurance rejections are removed from the sample; omitting these observations has no effect on the size or significance of the coefficient on race. The stability of the size and significance of the coefficient on race across the equations estimated in Table 4 shows that, at a minimum, the mortgage insurers are not the source of the effect of race on the mortgage lending decision.

Another issue is the treatment of withdrawn applications. Withdrawals may be hidden rejections. That is, during the loan approval process, the lender could encourage applicants to withdraw rather than reject them outright. However, when a multinomial logit is estimated with withdrawals as a third alternative, withdrawals do not look like denials.[25] Most,

[22] Columns 5 and 10 use the occupation equations of the PSID to calculate the applicant's probability of becoming unemployed in the next five years; the results are roughly identical, whether the probability is based on a set of industry equations or on one large equation.

[23] The optimal measurement for the effect of race is not altogether clear. With the equations in Tables 2 and 3, one could either measure how minorities with white characteristics would be treated (adding the minority coefficient to the white sample), which produces a 5.5 percentage point increase in the rejection rate of minorities, or measure how whites with minority characteristics would be treated (subtracting the race coefficient from the minority sample), producing a 8.2 percentage point difference in rejection rates explained by race. The difference is due to the nonlinearity of the logit regression. The latter approach seems superior, because it measures the effect of being a minority within the range of economic characteristics where minorities tend to apply. Additional measures are discussed later in the paper.

[24] In fact, lenders could simply be forcing the mortgage insurer to reject applications they never intended to accept. If so, treating these applications as lender rejections would be appropriate.

[25] Lender and tract dummies had to be omitted from the multinomial regression because there were not enough degrees of freedom to estimate their effects on withdrawals.

TABLE 4—ALTERNATIVE SPECIFICATIONS OF PROBABILITY OF MORTGAGE DENIAL: PRIVATE MORTGAGE INSURANCE

	Logit			OLS		
Variable	Base (1)	Dropping PMI variable (2)	Excluding PMI denials[a] (3)	Base (4)	Dropping PMI variable (5)	Excluding PMI denials[a] (6)
Constant	13.69	13.37	12.96	−0.22	−0.27	−0.22
	(12.62)	(13.17)	(12.91)	(1.47)	(−1.75)	(−1.45)
Risk of default:						
Housing expense/income	0.63	0.67	0.63	0.06	0.06	0.06
	(2.76)	(3.17)	(2.79)	(3.56)	(3.52)	(3.52)
Total debt payments/income	0.08	0.07	0.07	0.005	0.005	0.005
	(7.16)	(7.08)	(6.92)	(7.21)	(7.45)	(7.34)
Net wealth	0.00008	0.00006	0.00006	0.000004	0.000003	0.000004
	(0.76)	(0.56)	(0.59)	(0.60)	(0.39)	(0.60)
Consumer credit history	0.51	0.48	0.50	0.04	0.04	0.04
	(9.16)	(9.46)	(9.21)	(9.46)	(9.45)	(9.33)
Mortgage credit history	0.43	0.44	0.34	0.03	0.03	0.02
	(2.27)	(2.46)	(1.80)	(2.16)	(2.09)	(1.99)
Public record history	1.95	2.01	1.82	0.19	0.22	0.21
	(6.50)	(7.14)	(6.23)	(8.04)	(8.81)	(8.32)
Unemployment region	0.11	0.11	0.12	0.01	0.01	0.005
	(2.64)	(2.92)	(2.89)	(1.93)	(2.03)	(1.57)
Self-employed	0.70	0.69	0.64	0.05	0.05	0.05
	(2.31)	(2.40)	(2.12)	(2.71)	(2.60)	(2.72)
Loan/appraised value-low	−0.89	−0.75	−0.83	−0.12	−0.11	−0.11
	(−1.31)	(−1.13)	(−1.24)	(−2.46)	(−2.19)	(−2.35)
Loan/appraised value–medium	0.13	0.52	−0.19	−0.05	−0.02	−0.05
	(0.23)	(0.99)	(−0.36)	(−1.23)	(−0.41)	(−1.17)
Loan/appraised value–high	1.40	1.64	1.49	0.10	0.13	0.11
	(3.41)	(3.93)	(3.58)	(2.66)	(3.32)	(2.93)
Cost of default:						
Denied private mortgage insurance	6.16			0.65		
	(8.55)			(16.06)		
Loan characteristics:						
Two- to four-family home	0.73	0.81	0.70	0.06	0.07	0.06
	(2.64)	(3.19)	(2.59)	(2.51)	(3.16)	(2.61)
Personal characteristics:						
Race	1.00	0.92	0.93	0.07	0.07	0.06
	(3.73)	(3.63)	(3.54)	(3.34)	(3.06)	(3.12)
Percent correctly predicted	95.3	93.9	94.3			
Adjusted R^2				0.32	0.24	0.22
Number of observations	2,925	2,925	2,850	2,925	2,925	2,850

Notes: Numbers in parentheses are t statistics. Coefficients for census tract and lender ID dummies are not shown because they are so numerous.

[a] Sample excludes applicants denied private mortgage insurance.

though not all, of the independent variables important to the rejection decision are unimportant to the withdrawal alternative. Race is not a significant factor in the withdrawal decision, as the withdrawal rate is basically identical for whites and minorities. By law, withdrawing from the process is the applicant's choice, not the lender's. That choice often depends on factors extraneous to the model; for example, the property might fail an inspection report, or the buyer might simply get cold feet. These data provide little evidence to support modeling withdrawals as a lender, rather than a borrower, choice.

It is also possible that different standards apply to single-family, condo, and multiunit properties. Theoretically, these properties should be examined together, because the secondary market standards are basically equivalent for these three types of loans.[26] Yet some minor differences in the standards of these types of properties might exist. Log-likelihood tests were performed to ensure that condo, single-family, and multiunit properties should be estimated together. One cannot reject the hypothesis that the coefficients for the variables reported in Table 2 are identical for these three property types.[27] Thus, there appears to be no statistical or theoretical reason to separate the sample by property type.

Finally, various interactive terms were examined to test whether a combination of certain variables is important to the mortgage lending decision. Interactions of the loan-to-value ratio with the obligation ratios and the credit history variables, as well as the interplay between the obligation ratios and the credit history measures, were all tested. Only the interactive term for the loan-to-value ratio and the consumer credit history index was statistically significant. The importance of this variable, however, derived solely from its collinearity with the credit history variable; the two are simply too correlated to determine which is the superior specification. None of these interactive terms affected the size or significance of the race coefficient. Finally, allowing for some nonlinearity in the obligation ratios failed to significantly improve the fit of the equation or alter the results for the other variables.

VI. Evaluation of the Results

The significance of the coefficient on race remains robust to the inclusion of different variables and the estimation of different functional specifications. However, several other issues need to be addressed before any conclusion about the role of race can be made. The first relates to the pervasiveness of the behavior captured in the equation; is this result due to one or two isolated institutions or is it a market-wide phenomenon? Another question pertains to the broader issue of whether omitted variables still exist that could be biasing the estimate of the coefficient on race. The possibility that differential treatment is due to statistical discrimination must also be considered; are minority applications being rejected more often, given their economic characteristics, because their loans are somehow less profitable or is the lender discriminating for other than economic reasons? The quality of the data is also discussed. Finally, alternative approaches to measuring the effect of race are examined.

A. *Pervasiveness of the Role of Race*

The pervasiveness of the role of race across many institutions must be determined, in order to formulate the proper policy response. What appears in the aggregate results to be market-wide discrimination might originate from the actions of only a few lenders active in the minority market. To test whether race was an important factor in the mortgage lending decision for more than just a few institutions, the sample was divided into two subsamples, the lenders with the largest numbers of minority loans and the remaining institutions. The former group accounted for 50 percent of the minority applications but only about 5 percent of the institutions. An equation was estimated that allowed the coefficients on the economic variables to differ across these two groups of institutions, while the tract and lender dummies were constrained to be the same in order to conserve on degrees of freedom. The results, provided in Table 5, indicate that race is an important explanatory factor in mortgage lending decisions among both active and inactive lenders in the minority lending market. The importance of the race variable does not derive just from the behavior of several institutions, since the less active lenders to minorities summed together behave like the more active ones. The effect appears to be pervasive in the market.

[26] For example, Fannie Mae guidelines explain that multiunit properties should be treated identically to single-family properties, once the income stream from the multifamily rental property has been properly discounted and incorporated into the obligation ratios.

[27] The coefficients for the tract and lender indicator variables were constrained to be equal across property type to allow the degrees of freedom required for the tests.

TABLE 5—PERVASIVENESS OF THE RESULTS: DETERMINANTS OF PROBABILITY OF DENIAL FOR ACTIVE AND INACTIVE BANKS IN MINORITY MARKET

Variable	Logit: Inactive lenders (1)	Logit: Active lenders (2)	OLS: Inactive lenders (1)	OLS: Active lenders (2)
Constant	−14.64	−9.86	−0.21	−0.28
	(−11.03)	(−6.87)	(−1.37)	(−0.79)
Risk of default:				
Housing expense/income	0.78	0.41	0.05	0.07
	(2.34)	(1.26)	(2.33)	(2.61)
Total debt payments/income	0.07	0.09	0.004	0.005
	(4.57)	(4.96)	(5.68)	(4.25)
Net wealth	0.0002	−0.0004	0.000007	−0.000001
	(1.94)	(−1.01)	(0.99)	(−0.76)
Consumer credit history	0.74	0.38	0.04	0.04
	(7.96)	(5.32)	(6.98)	(6.62)
Mortgage credit history	0.39	0.52	0.03	0.02
	(1.47)	(1.77)	(1.97)	(0.87)
Public record history	1.85	2.03	0.14	0.27
	(4.29)	(4.72)	(4.53)	(7.06)
Unemployment region	0.16	0.07	0.01	0.01
	(2.81)	(1.11)	(1.70)	(1.11)
Self-employed	0.62	0.91	0.03	0.11
	(1.44)	(2.04)	(1.29)	(3.14)
Loan/appraised value–low	−1.01	−1.64	−0.10	−0.12
	(−1.13)	(−1.20)	(−1.87)	(−1.24)
Loan/appraised value–medium	−0.08	−0.35	−0.06	−0.02
	(−0.11)	(−0.32)	(−1.26)	(−0.23)
Loan/appraised value–high	1.52	0.63	0.10	0.10
	(3.37)	(0.62)	(2.33)	(1.35)
Cost of default:				
Denied private mortgage insurance	6.80	6.18	0.73	0.58
	(6.47)	(6.23)	(11.35)	(10.90)
Loan characteristics:				
Two- to four-family home	1.81	−0.05	0.13	−0.01
	(4.35)	(−0.13)	(4.18)	(−0.22)
Personal characteristics:				
Race	0.91	0.98	0.05	0.10
	(2.22)	(2.84)	(1.76)	(3.73)
Percent correctly predicted	95.50			
Adjusted R^2			0.32	
Log-likelihood	514.51			
Number of observations	2,925		2,925	

Notes: Numbers in parentheses are t statistics. Census tract and lender ID dummy coefficients are not shown because they are so numerous.

B. *The Possibility of Omitted Variables*

The size, significance, pervasiveness, and robustness of the coefficient estimate on race do not necessarily identify race's role in the mortgage lending decision. Variables in the lender's information set, related to the lending decision and correlated with race, might still be missing, and this could produce a positive coefficient on race. The extended survey, however, includes every variable mentioned as important in numerous conversations with lenders, underwriters, and examiners, and it contains all of the information on the application form and almost all of the information in the loan file. It does not appear likely that a variable has been omitted that is both systematically related to the mortgage lending decision and in the lender's information set. Any variables that are missing are probably idiosyncratic; a bad credit history may be overlooked, for example, if it occurred because of a temporary medical problem. Yet, it is unclear why such idiosyncratic variables would be systematically correlated with race, particularly given the other variables in these regressions.

The fit of the equation can provide some evidence on the extent of any omitted variables. With a discrete dependent variable, however, the R^2 in the OLS regression is not a meaningful measure of the goodness-of-fit.[28] For qualitative choice models, there is no way to quantify precisely the unexplained variance of the equation, as is done with the R^2 in OLS regressions. The measures of goodness-of-fit that do exist provide only relative comparisons between models; they have no meaning in an absolute sense. One such measure is the percentage of acceptances and denials correctly predicted. The evaluation of the percentage correctly predicted will depend upon the sample frequencies of the alternatives; thus, the most commonly assumed baseline for comparison is a logit with only a constant term, which in this model is equivalent to assuming that every application is accepted. This baseline correctly predicts 85 percent of the applications since 85 percent of the applicants were accepted for a loan. Including the original HMDA data increases the percentage correctly predicted to slightly above 87 percent.[29] The base equation in column 1 of Table 2 correctly predicts over 95 percent of the applications. Using this criterion, the number of misses is diminished by two thirds with the addition of our supplementary variables to the equation.

A major problem with the percentage correctly predicted is the arbitrariness of the definition about when the model accurately predicts an outcome. Another possible approach is examined in Table 6. The first three rows present the predicted mean denial rates for both accepted and rejected applications derived from a logit model containing only the original HMDA data—the race, gender, and income of the applicant, as well as the amount of the loan and the tract and lender dummy variables. Here, a more sophisticated baseline is required because a logit with only a constant would predict the same probabilities of denial for both accepted and rejected applications. The last three rows contain the predicted mean rejection rates for both rejected and accepted applications produced from the base equation in Table 2. The better the predictive power of these additional variables, the wider the difference in the predicted denial rates between the rejected and accepted applications. The equation using only the raw HMDA data does explain some of the differences between rejected and accepted applications. However, inclusion of the extended data adds considerably to the model. The difference in the predicted probability of denial for rejected and accepted applications between the two models basically doubles for the sample as a whole, as well as for the white and minority subsamples, from roughly 29 percentage points to about 56

TABLE 6—PREDICTED REJECTION RATES

	Rejected	Accepted
Original HMDA model	38.9	10.2
Minority	45.6	21.4
White	33.2	7.5
Extended model[b]	62.5	6.3
Minority	68.4	12.4
White	57.6	4.8

[a] Includes race, gender, income of the applicant, and the loan amount, and census tract and lender ID dummies.
[b] Extended model is given in column 1 of Table 2.

[28] The OLS estimate of the R^2 is generally underestimated when the dependent variable is dichotomous; see William H. Greene (1981, 1983).

[29] The null in this case is equivalent to a logit run with only a constant. The model with only the original HMDA data includes race, gender, loan amount, the applicant's annual income, and the lender and tract dummies.

percentage points. The additional variables collected for this study significantly increase the power to predict the probability of rejection, as is shown by the much stronger predictive performance when the extended HMDA data are used.

Even though it is impossible to quantify the unexplained variance of the equation, the predictive power of the equation appears to be strong, and the supplementary variables add significantly to this power. How much of the remaining unexplained variance is due to possible omitted variables and how much to random error cannot be identified. Collecting almost the entire information set of the lender suggests that no variables systematically related to the mortgage lending decision and included in the lender's file have been omitted. To make a compelling case that omitted variables are biasing the estimation of the coefficient on race, one would have to specify the variable, show its relevance in the mortgage lending decision, and provide evidence for the variable's correlation with minority mortgage borrowers.

C. *Statistical Discrimination*

Race could still be an effective signal for default probabilities even after controlling for all the variables in the lender's information set. As discussed in A. Michael Spence (1974), Orley Ashenfelter and Timothy Hannan (1986), Edmund S. Phelps (1972), Shelly J. Lundberg and Richard Startz (1983), and a host of others, statistical discrimination may arise in the case of mortgage lending if mortgage lenders are aware that default rates are correlated with race, even after holding all the other variables in the lender's information set constant. If so, the significant race coefficient is capturing minorities' higher propensity to default, given the other variables in the loan file. For this reason, many have suggested studying mortgage defaults rather than mortgage rejections.

Examining defaults, however, is not an effective way to discern discrimination. For example, some models of statistical discrimination suggest that profit maximization results in equal default rates across races. If noneconomic discrimination were occurring in such models, lenders would hold minority borrowers to stricter standards, and, thus, minority default rates should be lower than white rates. Discrimination based on taste could then be distinguished from statistical discrimination by comparing the default rates of the two groups. This analysis, however, has several serious flaws. Even if taste-based discrimination existed, minority default rates would be lower than white rates only if the mortgage lenders were using the "correct" model in their determination of default risk; if not, minority default rates could be higher, lower, or equal to the white rates. Far more important, however, discrimination occurs at the margin, not the average. Because the average minority applicant has weaker financial characteristics, it is possible that the marginal minority loan that qualifies for a mortgage is of a higher quality than the marginal white loan, while the average creditworthiness of the accepted minorities is lower than that of accepted whites. Because examining average default rates tells us nothing about whether discrimination is occurring at the margin, all the variables related to the mortgage lending decision must be included in any default analysis. In fact, even when every relevant variable is included, default studies provide little information about whether discrimination is occurring at the margin. Tootell (1993, 1995a) and John Yinger (1993) discuss the problems with default analysis in detail.

Just because default analysis is not ideally suited to testing for statistical discrimination does not mean that statistical discrimination is not occurring in this market. For statistical discrimination to exist, however, lenders need to know that minorities default on mortgages more frequently than whites given their economic fundamentals. The studies that attempt to provide this information, Robert Van Order et al. (1993) and James A. Berkovec et al. (1994) produce somewhat mixed results; in general, minorities are found to be either equally likely or more likely to default than whites, although in some areas of the country the reverse is true.[30] Yet, even the best of these studies have omitted important variables.

[30] For example, Van Order et al. (1993) found race negatively related to defaults in some parts of the country.

Van Order et al. do not have the race of the mortgage applicant, only the racial composition of the zip code. Berkovec et al. look only at government-guaranteed loans, not the conventional loans studied here.[31] Neither paper includes the applicant's credit history, a variable important to the mortgage lending decision that is highly correlated with race. The dearth of any evidence that minorities default more frequently, given their economic fundamentals, makes a conclusion of economically rational discrimination problematic.[32]

D. *Quality of the Data*

Another issue pertains to the quality of the data. FDIC examiners reviewed 95 of the rejected applications included in our survey, 62 of which were submitted by minorities. These applications were not selected randomly, but were chosen as the applications with the lowest probability of rejection that were, in fact, rejected.[33] David Horne (1993) states that examiners found serious inaccuracies in the data for many of these files, inaccuracies that may call into question the robustness of the results of the study.

The evidence on the extent of the "errors" is found in Table 1 of Horne (1993), which enumerates the types and frequencies of errors found by the examiners in these 95 files. In an effort to reproduce Horne's results in order to determine the nature of his alterations of the data and the consequences for our results, the "corrections" summarized in Table 1 of Horne were requested. Unfortunately, Horne was unable to replicate his Table 1, and thus he could not provide us with the data.

Horne (1994) alters the data set, based on the examiners' reports, to analyze the effect any errors may have on a regression for the FDIC subsample. To determine the sensitivity of the results to these changes and to determine the nature of the changes, we requested and received the program Horne used to transform the data. After eliminating Horne's programming errors, we found that estimating the base equation using the transformed data reduced the coefficient on race by about 10 percent; however, it remains statistically significant well beyond the 1-percent level.

Why did transforming the data have so little effect? First, many of the data errors were obviously misplaced decimal points, and we too had already made the same corrections as Horne.[34] Second, a large percentage of Horne's changes were to variables that were not used in this analysis. For example, corrections in the HMDA income data are irrelevant to the results of the study as we, and Horne, used the obligation ratios and the monthly income provided in the extended survey; we collected monthly income precisely because we did not trust the original HMDA income data. Finally, most of the remaining changes were quite small; changing monthly income by a dollar or total assets by $100 will not affect the results.

Ironically, about 17 percent of the alterations Horne made were programming errors on his part, most of which were changes to the gender of the applicant when he meant to alter the original HMDA income. Again, HMDA income was not used in the analysis in this paper or in Horne (1994). Furthermore, 10 percent of the corrections are suspicious in that they violate an adding-up constraint or are directly contradicted by the FDIC examiners' write-up provided to us before the publication

[31] Examinations of defaults are based on the assumption of profit-maximizing behavior. Using FHA loans is particularly suspect, since they are government-subsidized and guaranteed loans, just the subsample one would expect to be driven by motivations other than profit maximization.

[32] The minimal evidence on defaults collected for this study suggests that default rates are basically uncorrelated with the racial composition of the tract. To examine the effect of previous default rates on the mortgage lending decision, the base equation was estimated including the default rate of the tract, from 1988 to 1990. The inclusion of lagged default rates has no effect on the other coefficients and is itself insignificant. Previous default experience in the neighborhood does not appear to be correlated with race or to provide useful information in describing the mortgage lending decision.

[33] The institutions selected also had several constraints. First, they had to be FDIC regulated. They also had to have both rejected and accepted minority and white applications. These additional constraints seriously limited the sample further.

[34] We dropped 7 problematic loans from the sample because of apparent errors, and all 7 were among the 95 loans examined by the FDIC.

of Horne (1993).[35] In short, examiners did find some errors in the data, but most of the errors were either small or to variables that were not used in any of the analysis.

The expected effect of any important measurement errors on the results is unclear, which is another possible reason altering the data had such a slight impact on the results. As discussed in W. G. Cochran (1970), if significant errors exist in the independent variables other than race, the multiple correlation of all these variables with race would determine the sign of any potential bias of the estimate of the coefficient on race. One would expect that transcription errors, almost exclusively the errors that Horne discusses, would be uncorrelated with race. As a result, it is not obvious a priori that any bias should be present, let alone what sign it would take.

Horne also claims to have discovered some errors in the dependent variable. Jerry A. Hausman and Fiona M. Scott-Morton (1994) show that measurement error in the dependent variable when the dependent variable is dichotomous can lead to inconsistent estimates of the coefficients. Horne (1994) changes 5 applications from rejections to acceptances. All 5 of these changes appear groundless. For example, one applicant was rejected with a loan-to-value ratio of 95 percent. The loan was resubmitted with a larger down payment and a loan-to-value ratio of 89 percent, and was accepted. Horne made both applications acceptances, arguing that a counteroffer, presumably the demand for a higher down payment, is the same as an acceptance. The information that the 95 percent loan-to-value ratio was unacceptable should be incorporated when estimating the determinants of the mortgage lending decision, as our data set does. Another application rejected by the lender was later overturned by the state mortgage review board, forcing the lender to compensate the borrower by originating the loan; since we are attempting to model lender behavior, not the behavior of the state's mortgage review board, the application should be counted as a rejection, not changed to an acceptance as Horne argues.

His redefinitions of several denials as withdrawals are also not supported by the review documents. For example, in one case the file was rejected, but the lender claimed the applicant called back several weeks after the decision had been made to request that the application be withdrawn; since this application was rejected by the lender, it would be an error to call it a withdrawal. Thus, Horne provides little evidence for the claim that serious errors were made in the categorization of the dependent variable. Moreover, Hausman and Scott-Morton (1994) provide a maximum-likelihood correction which produces consistent coefficient estimates when measurement error exists in a dichotomous dependent variable. Using this correction does not alter the size or significance of the estimated coefficient on race found throughout this paper; the procedure finds little evidence that errors in the dependent variable are a problem.

The errors discussed by Horne (1993, 1994) are basically transcription errors in independent variables other than race, and it is unclear that any such errors that do exist would bias the estimate of the race coefficient in either direction. Given that a large percentage of the corrections were to variables not used in the study, such as the HMDA income data, and that many of the changes were trivial, errors-in-variables do not seem to be a major problem with the data. We were aware that one lending institution had a problem with its data, and we ensured that all the results were unaffected by its inclusion. As this one institution accounted for about one half of all the errors found in the FDIC sample, the FDIC sample is unrepresentative of the sample as a whole. In fact, the regressions Horne runs suggest that measurement errors are not significantly affecting the estimates. Errors-in-variables do not appear to be biasing the results.[36]

[35] For example, monthly nonhousing obligations might increase, the housing expense-to-income ratio would rise, while the total obligations-to-income ratio would not be altered.

[36] Other problems with Horne (1994) are discussed in detail in Tootell (1995b).

Finally, care must be taken about inferring anything about the effects of these corrections on the coefficient on race. Regressions using the altered data tell us little about the effect of errors on the sample as a whole, since the sample of examined files was not selected randomly. Even if the errors were not correlated with race, by disproportionately examining minority applications, 62 of 95, and only examining rejections, Horne biases the race coefficient downward. By disproportionately correcting rejected minority files, the estimate of the race coefficient would tend to decline.

To summarize, the results presented in Table 2 are extremely robust. The estimated effect of race on the mortgage lending decision is relatively stable and always statistically significant. It is difficult to think of variables that have been omitted from the survey that are systematically linked with race and the mortgage lending decision and are in the mortgage file; the overall equation does a very good job of explaining the variation in denial rates. Moreover, the equation is describing widespread behavior, not simply that of a few large institutions or of particular types of institutions. Finally, distinguishing between statistical and noneconomic discrimination is difficult with any data, these included; lenders could be using race as a signal for variables they do not collect, which would produce the positive and significant coefficient on race seen in these regressions. The hypothesis of statistical discrimination begs several questions, however; if such variables exist, why are the lenders not collecting them? And to use race as a signal, the lenders need evidence that race is correlated with outcomes, holding the rest of their information set constant; the default studies have yet to provide much support for this hypothesis.

E. *Alternative Measures of the Effects of Race*

Estimating an equation that includes an explicit measure for race is not the only way to test whether race is an important factor in the mortgage lending decision. If coefficients other than the constant differ, then a preferable measure of discrimination may be obtained from estimating separate white and minority equations. The minority characteristics can then be run through the white equation to get the predicted minority rejection rate if they had been white. Similarly, the white characteristics can be run through the minority equation to get the predicted rejection rate for whites had they been minorities. The resulting discrepancy between the actual minority denial rate and the estimated minority denial rate based on the white equation, or between the actual white denial rate and the white rate predicted from the minority equation, can be interpreted as the effect of race on the mortgage lending decision.

Table 7 presents an equation estimated when the coefficients on the economic variables for white and black/Hispanic applicants are allowed to differ. Again, the coefficients for the tract and lender dummies are constrained to be identical for the two groups to conserve on degrees of freedom. Comparing the first and second equations, a likelihood-ratio test cannot reject the hypothesis that all the coefficients but the constant are the same between the two groups. Comparing columns 2 and 3, it can be rejected that all coefficients including the constant are the same. The rejection rate for blacks and Hispanics predicted by the coefficients from the white equation is 21 percent, rather than the actual 28 percent experienced by minority applicants. In other words, while economic, property, and neighborhood characteristics explain much of the higher minority denial rate, 7 percentage points, slightly less than one half of the difference in the base rejection rates between whites and minorities, remains attributable to race. Furthermore, if whites were treated like minorities their rejection rate would be about 50 percent higher.

Finally, Table 8 provides the results of an equation where race is run interactively with each variable. Only the interactive coefficients on the total obligation ratio and the indicator of whether the applicant was self-employed approach statistical significance in this regression. Thus, the evidence here does not suggest, for example, that the discrimination is coming from different standards for credit history. The significantly larger coefficient on the obligation ratio for minorities is difficult to interpret; taken at face value, perhaps the requirement that minorities devote a smaller share of their

TABLE 7—ALTERNATIVE SPECIFICATIONS OF DENIAL: EQUATIONS BY RACE

		Equation allowing coefficients to differ		
Variable	Base (1)	White (2)	Minority (3)	No race variable (4)
Constant	−13.69	−13.35	−13.28	−13.46
	(−12.61)	(−11.37)	(−7.36)	(−12.48)
Risk of default:				
Housing expense/income	0.63	0.55	0.71	0.64
	(2.76)	(1.78)	(2.01)	(2.81)
Total debt payments/income	0.08	0.07	0.12	0.08
	(7.15)	(4.64)	(5.41)	(7.20)
Net wealth	0.00008	0.00008	−0.00035	0.00007
	(0.73)	(0.77)	(−0.61)	(0.65)
Consumer credit history	0.51	0.53	0.51	0.51
	(9.15)	(7.22)	(6.06)	(9.32)
Mortgage credit history	0.43	0.38	0.64	0.48
	(2.27)	(1.64)	(1.68)	(2.54)
Public record history	1.95	2.03	1.94	2.02
	(6.50)	(5.05)	(4.19)	(6.78)
Unemployment region	0.11	0.11	0.14	0.10
	(2.64)	(2.31)	(1.70)	(2.47)
Self-employed	0.70	1.09	−0.34	0.67
	(2.31)	(3.10)	(−0.54)	(2.24)
Loan/appraised value−low	−0.89	−1.06	−2.00	−1.02
	(−1.31)	(−1.29)	(−1.11)	(−1.53)
Loan/appraised value−medium	0.13	0.31	−1.30	0.08
	(0.23)	(0.47)	(−0.90)	(0.16)
Loan/appraised value−high	1.40	1.47	0.39	1.38
	(3.41)	(3.44)	(0.30)	(3.36)
Cost of default:				
Denied private mortgage insurance	6.16	6.91	5.25	6.09
	(8.55)	(6.91)	(5.20)	(8.48)
Loan characteristics:				
Two- to four-family home	0.73	1.10	0.66	0.77
	(2.64)	(2.85)	(1.76)	(2.82)
Personal characteristics:				
Race	1.00			
	(3.73)			
Log-likelihood	−527.91	−518.96		−534.95
Percent correctly predicted	(95.3)	(95.5)		(95.2)

Notes: Numbers in parentheses are t statistics. Census tract and lender ID dummy coefficients are not shown because they are so numerous.

discretionary income to housing expenditures implies lender concern about the ability of minorities to control their nonhousing expenses. Yet, the correlation between the interaction variable of race and the total obligation ratio and the race indicator variable is over 0.95. In general, the evidence seems to point to differences in the constant term rather than to different treatment of certain characteristics between the races.

VII. Conclusions

This study examined one avenue through which differential treatment could affect minorities' access to credit and opportunities for

TABLE 8—EQUATIONS WITH RACE RUN INTERACTIVELY WITH ALL VARIABLES

Variable	Logit (1)	OLS (2)
Constant	13.35	−0.19
	(11.37)	(−1.24)
Risk of Default:		
Housing expense/income	0.55	0.04
	(1.78)	(1.99)
Total debt payments/income	0.06	0.004
	(4.66)	(5.28)
Net wealth	0.00008	0.000004
	(0.79)	(0.55)
Consumer credit history	0.53	0.03
	(7.22)	(6.60)
Mortgage credit history	0.38	0.02
	(1.64)	(1.46)
Public record history	2.03	0.17
	(5.05)	(5.38)
Unemployment region	0.11	0.01
	(2.31)	(1.55)
Self-employed	1.09	0.06
	(3.10)	(2.94)
Loan/appraised value–low	−1.06	−0.09
	(−1.29)	(−1.71)
Loan/appraised value–medium	0.31	0.10
	(0.47)	(2.66)
Loan/appraised value–high	1.47	−0.02
	(3.44)	(−0.48)
Cost of default:		
Denied private mortgage insurance	6.91	0.73
	(6.91)	(13.28)
Loan characteristics:		
Two- to four-family home	1.10	0.06
	(2.85)	(2.36)
Personal characteristics:		
Race	−0.06	−0.17
	(−0.03)	(−1.20)
Interactive term:		
Housing expense/income	0.16	0.05
	(0.34)	(1.29)
Total debt payments/income	0.05	0.005
	(1.99)	(3.01)
Net wealth	−0.0004	−0.000002
	(−0.74)	(−0.38)
Consumer credit history	−0.02	0.02
	(−0.19)	(1.94)
Mortgage credit history	0.26	0.04
	(0.59)	(1.30)
Public record history	−0.10	0.06
	(−0.16)	(1.16)

TABLE 8—*Continued.*

Variable	Logit (1)	OLS (2)
Interactive term:		
Unemployment region	0.03	0.01
	(0.30)	(0.88)
Self-employed	−1.43	−0.05
	(−2.00)	(−0.87)
Loan/appraised value–low	−0.93	−0.10
	(−0.48)	(−0.66)
Loan/appraised value–medium	−1.61	−0.05
	(−1.02)	(−0.42)
Loan/appraised value–high	−1.08	0.09
	(−0.80)	(0.74)
Denied private mortgage insurance	−1.67	−0.19
	(−1.17)	(−2.40)
Two- to four-family home	−0.44	−0.01
	(−0.85)	(−0.20)
Percent correctly predicted	95.5	
Adjusted R^2		0.32
Number of observations	2,925	2,925

Notes: Numbers in parentheses are *t* statistics. Census tract and lender ID dummy coefficients are not shown because they are so numerous.

homeownership. It found that black and Hispanic mortgage applicants in the Boston area were more likely to be turned down than white applicants with similar characteristics.

It is important to clarify the limited focus of this analysis; it abstracts from discrimination that may occur elsewhere in the economy. For example, if minorities are subject to discrimination in education or labor markets, they will have lower incomes and their applications may reflect higher obligation ratios, greater loan-to-value ratios, or poorer credit histories. The lower creditworthiness of minorities, on average, is one measure of this effect. Denial of a mortgage loan application on the basis of these economic characteristics would not be considered discriminatory for the purposes of this study. Similarly, if blacks and Hispanics are discouraged from moving into predominantly white areas, they will limit their search to neighborhoods sanctioned for minorities. These tend to be older central cities with high-density housing, such as two- to four-family homes. The increased probability of denial associated with purchasing a multiunit property would further handicap minority borrowers. Furthermore, differential treatment could occur at many stages in the lending process. For example, minorities may be discouraged from even applying for a mortgage loan as a result of a prescreening process. All these effects may cause the regressions to understate the role of race in mortgage lending.

On the other hand, the positive race coefficient in the denial regression could represent different outcomes by race. Although there is no good way to test this hypothesis, an omitted variable not collected by the lender but correlated with race and mortgage performance could account for the differential treatment found in this paper's results. This explanation has several problems, however. First, it is difficult at this point to advance a variable that is related to the mortgage performance, correlated with the minority status of mortgage applicants, and not collected by the survey; lenders complained to us that their mortgage files contained too many variables, not that they were missing important ones. Further, studies of mortgage outcomes have yet to show that, all else equal, minorities do default more often, or that their defaults are more costly. Finally, in our discussions with the lenders, none of them indicated that race was a relevant variable; they did not appear to be using race as a signal. The significance of the coefficient on race may be consistent with profit maximization, but the lenders have hesitated to make that claim and no study has provided strong support for it.

The original HMDA data suggested that differential treatment by race was occurring in the mortgage market. Almost everyone agreed that important variables related both to loan performance and to minority status were missing from the HMDA data. We collected almost all of the information in the lender's information set, and all of the information the lenders we spoke with said they used in the decision-making process, and found that race still played a significant role in the mortgage lending decision. The lack of hard evidence suggesting that race may be an accurate signal for loan performance, given the other information which could be collected by the lender, suggests that a serious problem may exist in the market for mortgage loans. As a result, lenders, community groups, and regulators must work together to guarantee that minorities are treated fairly. Since current examination methods cannot ensure that compensating factors are being used consistently, or that any differences in treatment are statistically significant, statistical analysis is probably the best procedure for monitoring this problem.

APPENDIX—SUMMARY OF VARIABLE DEFINITIONS WITH MEANS AND STANDARD DEVIATIONS

Summary of variable definitions	
Housing expense/income	1 if greater than .30 0 otherwise
Total debt payments/income	Value of question 46
Net wealth	Value of question 36 less question 38
Consumer credit	1 if no "slow pay" account (code 1 in question 43) 2 if one or two slow pay accounts (code 2) 3 if more than two slow pay accounts (code 3) 4 if insufficient credit history for determination (code 0) 5 delinquent credit history with 60 days past due (code 4) 6 serious delinquencies with 90 days past due (code 5)
Mortgage credit	1 if no late payments (code 1 in question 42) 2 if no payment history (code 0) 3 if one or two late payments (code 2) 4 if more than two late payments (code 3)
Public record	1 if any public record of credit problems (codes 1, 2, 3, 4 in question 44) 0 otherwise
Probability of unemployment	1989 Massachusetts unemployment rate for applicant's industry
Self-employed	1 if self-employed 0 otherwise
Loan/appraised value	Value of loan amount divided by question 50
Denied private mortgage insurance	Derived from question 53
Rent/value in tract	Rental income divided by estimate of value of rental property from Census
Two- to four-family homes	0 if purchasing a single-family or a condo 1 if purchasing a two- to four-family home
Race	1 if applicant was black or Hispanic 0 otherwise

APPENDIX—*Continued.*

Variable	Mean	Standard deviation
Total debt payments/income	33.29	11.28
Net wealth (S)	238.67	1000.87
Consumer credit history	2.17	1.70
Mortgage credit history	1.74	0.53
Probability of unemployment	3.81	2.07
Loan/appraised value	0.75	0.18
Rent/value in tract	0.09	0.23

Note: The follow-up survey questions are given in Table 1.

REFERENCES

Ashenfelter, Orley and Hannan, Timothy. "Sex Discrimination and Product Market Competition: The Case of the Banking Industry." *Quarterly Journal of Economics*, February 1986, *101*(1), pp. 149–73.

Avery, Robert B.; Beeson, Patricia and Sniderman, Mark. "Cross Lender Variation in Home Mortgage Lending." *Economic Review* (Federal Reserve Bank of Cleveland), 1994 Quarter 4, *30*(4), pp. 15–29.

Berkovec, James A.; Canner, Glenn B.; Gabriel, Stuart A. and Hannan, Timothy H. "Discrimination, Default, and Loss in FHA Mortgage Lending." Working paper, Federal Reserve Board of Governors, November 1994.

Black, Harold; Schweitzer, Robert L. and Mandell, Lewis. "Discrimination in Mortgage Lending." *American Economic Review*, May 1978, *68*(2), pp. 186–91.

Cochran, W. G. "Some Effects of Errors of Measurement on Multiple Correlation." Journal of the American Statistical Association, March 1970, *65*(329), pp. 22–34.

Greene, William H. "On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model." *Econometrica*, March 1981, *49*(2), pp. 505–15.

______. "Estimation of Limited Dependent Variable Models by Ordinary Least Squares and the Method of Moments." *Journal of Econometrics*, February 1983, *21*(2), pp. 195–212.

Hausman, Jerry A. and Scott-Morton, Fiona M. "Misclassification of a Dependent Variable in a Discrete Response Setting." Working Paper No. 94-19, Department of Economics, Massachusetts Institute of Technology, June 1994.

Horne, David. "Evaluating the Role of Race in Mortgage Lending." *FDIC Banking Review*, 1993, pp. 1–15.

______. "Mortgage Lending, Race, and Model Specification." Mimeo, Division of Research and Statistics, FDIC, 1994.

King, Thomas A. "Discrimination in Mortgage Lending: A Study of Three Cities." Working Paper No. 91, Office of Policy and Economic Research, Federal Home Loan Bank Board, Washington, DC, 1980.

Lundberg, Shelly J. and Startz, Richard. "Private Discrimination and Social Intervention in Competitive Labor Markets." *American Economic Review*, June 1983, *73*(3), pp. 340–47.

Phelps, Edmund S. "The Statistical Theory of Racism and Sexism." *American Economic Review*, September 1972, *62*(4), pp. 659–61.

Schafer, Robert and Helen F. Ladd. *Discrimination in mortgage lending*, MIT-Harvard Joint Center for Urban Studies. Cambridge, MA: MIT Press, 1981.

Spence, A. Michael. *Market signaling: Informational transfers in hiring and related screening processes*. Cambridge, MA: Harvard University Press, 1974.

Tootell, Geoffrey M. B. "Defaults, Denials, and Discrimination." *New England Economic Review*, September/October, 1993, pp. 45–51.

______. "Can Studies of Application Denials and Mortgage Defaults Uncover Taste-Based Discrimination." Unpublished manuscript, Federal Reserve Bank of Boston, July 1995a.

——. "Turning a Critical Eye on the Critics." Unpublished manuscript, Federal Reserve Bank of Boston, May 1995b.

Van Order, Robert; Westin, Ann-Margaret and Zorn, Peter. "Effects of Racial Composition of Neighborhoods on Defaults, and Implications for Racial Discrimination in Mortgage Markets." Paper presented at the Allied Social Science Associations meetings in Anaheim, CA, January 1993.

Yezer, Anthony M. J.; Phillips, Robert F. and Trost, Robert P. "Bias in Estimates of Discrimination and Default in Mortgage Lending: The Effects of Simultaneity and Self-Selection." *Journal of Real Estate Finance and Economics*, November 1994, *9*(3), pp. 197–215.

Yinger, John. "Discrimination in Mortgage Lending: A Literature Review and Recommendation for Future Research." Manuscript for Discrimination in Mortgage Lending: Research and Enforcement Conference, Housing and Urban Development, May 1993.