An Evaluation of the Federal Reserve Bank of Boston's Study of Racial Discrimination in Mortgage Lending

Dennis Glennon and Mitchell Stengel

Office of the Comptroller of the Currency
Economic & Policy Analysis

Working Paper 94 -2
April 1994

Abstract: In 1992, the Federal Reserve Bank of Boston (Boston Fed) released a now well-known study of mortgage lending practices in the Boston Metropolitan Statistical Area (MSA). Using an econometric model to examine extensive mortgage loan data collected from 131 financial institutions in the Boston MSA, the authors of the Boston Fed study tried and failed to find explanations other than racial discrimination for the significant disparities observed in the rejection rates for white and minority loan applicants.

The Boston Fed study attracted considerable attention from Congress, the banking industry, the civil rights community, bank regulators, and the news media. But three follow-up studies have raised a variety of problems with the Boston Fed study relating to data and methodology.

In this paper, we discuss and evaluate the problems cited by critics of the Boston Fed study. We focus our attention on three broad areas of concern: model specification; data errors; and differences in characteristics of the groups being compared.

Our principal findings can be summarized as follows:

(i) Several alternative model specifications perform better than the Boston Fed model in terms of various econometric performance measures; however, the race of the applicant continues to have a large and highly significant effect on the outcome of the lending application process.

(ii) The results of the Boston Fed model are affected only slightly when some of the more obvious and easily correctable data errors are corrected.

(iii) Allowing for different coefficients for whites and minorities, our analysis supports the Boston Fed's conclusion that approximately half the difference in denial rates can be attributed to differences in the financial characteristics of the borrowers and the neighborhood characteristics of the property; the remaining half can be attributed to differences in treatment by race.

We conclude with a qualified confirmation of the results of the Boston Fed study. However, we believe there are still several important specification issues that cannot be investigated and several data problems that cannot be corrected using the data provided by the Boston Fed. Additional research at both the MSA and the individual bank levels is warranted.

The opinions expressed are those of the authors and do not necessarily represent those of the Office of the Comptroller of the Currency. The authors would like to thank Jeffrey Brown, Michael Carhill, Julia Lane, Kevin Jaques, Daniel Nolle, and Mark Winer for helpful comments. Any errors are the responsibility of the authors.

Please address questions of substance to the authors at the Economic & Policy Analysis Department, Office of the Comptroller of the Currency, 250 E Street, Washington, D.C., 20219, (202) 874-5240.

I. Introduction

The Federal Reserve Bank of Boston's study of mortgage lending practices in the Boston MSA (Munnell et al. 1992) has redefined the debate over the existence of racial discrimination in mortgage lending. Using a much more extensive data set than any previous study, the Boston Fed analyzed differences in denial rates across races after controlling for wealth, credit and employment histories. The authors conclude that ". . . even after controlling for financial, employment, and neighborhood characteristics, Black and Hispanic mortgage applicants in the Boston metropolitan area are roughly 60 percent more likely to be turned down than whites" (Munnell et al. 1992, p. 2). They attribute the difference to discrimination on the part of the lending institutions. For many, the Boston Fed study seemed to answer – once and for all – the question of whether racial discrimination exists in mortgage lending. Moreover, the results of the study were a major impetus in the renewed and intensified efforts of the federal government to detect, punish, and eradicate such discrimination.

In recent months, however, three follow-up studies have identified several data and methodological problems that appear to call into question the Boston Fed's central finding of substantial discrimination against minorities (Horne 1994; Liebowitz 1993; Zandi 1993). In this paper, we discuss several of the issues raised by the follow-up studies. We focus our attention on three broad areas of concern, model misspecification, data errors, and differences in initial endowments. We find many of the criticisms of the Boston Fed's method of specifying and assessing the accuracy of their model to be legitimate. There are reasonable arguments supporting alternative model specifications that may better capture the underwriter's decision-making process. Moreover, we find that there are unquestionably numerous data entry and data reporting errors in the information supplied by the institutions that participated in the study. Nevertheless, using data provided by the Boston Fed, we test several alternative specifications of the model and find their results – including the magnitude and significance of the race variable – to be quite robust with regard both to alternative specifications and to the correction of the more obvious and easily correctable data errors. Moreover, we find that approximately half of the difference in observed denial rates can be attributed to differences in the initial endowments (*i.e.*, the basic characteristics of the applicants or the properties) associated with individual racial groups, and that the remaining difference in the denial rates represents a discriminatory differential of roughly the same magnitude as that found by the original study.

The follow-up studies also raise several other methodological and sample design problems that, if valid, could substantially alter the Boston Fed's conclusion. For the most part, however, those issues cannot be tested using the Boston Fed data and therefore are speculative at best. We assess the validity of these criticisms as applied to the Boston Fed study, and where possible, discuss the likely impact they would have on the results.

The paper is structured as follows. In the next section, we summarize the Boston Fed's methodology and findings. Section III introduces the issues of goodness-of-fit and separate regression models for whites and minorities. Section IV explores the question of differences in endowments by race. In section V, we discuss several alternative model specifications that we believe better represent the underwriting procedure employed in the mortgage lending market. In section VI, a summary and discussion of some of the more troubling data errors are presented and issues of model misspecification and omitted variables are also explored. We conclude with an overall assessment of the validity of the Boston Fed's findings with regard to the existence of discrimination, and a discussion of future research efforts.

II. The Boston Fed model

The Boston Fed found that Home Mortgage Disclosure Act (HMDA) data consistently show that non-Asian minority home mortgage application denial rates are two to three times higher than those for whites. This led Munnell et al. to argue:

> [t]his pattern has triggered a resurgence of the debate on whether discrimination exists in home mortgage lending. Some people believe that the disparities in denial rates are evidence of discrimination on the part of banks and other lending institutions. Others, including lenders, argue that such conclusions are unwarranted, because the HMDA data do not include information on credit histories, loan-to-value ratios, and other factors considered in making mortgage decisions. These missing pieces of information, they argue, explain the high denial rates for minorities (Munnell et al. 1992, p. 1).

The Boston Fed sought to resolve this controversy, at least in the case of a single MSA, by conducting a follow-up study. With the voluntary participation of the lending institutions, they collected information on the creditworthiness of applicants from all HMDA-reporting institutions in the Boston MSA that received at least total 25 mortgage applications in 1990. A final sample of 3,062 files (2,340 white and 1,013 Hispanic and Black) was chosen from 18,838 total loan applications.[1] An expanded HMDA data form was then sent to each institution, requesting

---

[1] The initial sample design developed by the Boston Fed identified all 1210 minority (Black and Hispanic) applicants in the Boston MSA in 1990 for inclusion; however, for practical reasons, the Boston Fed chose to survey only institutions that had received at least 25 mortgage applications from borrowers of all races. Sixty-seven minority applications were lost from the survey for this reason. Information then was requested from the 1,143 applicants for conventional mortgage loans made to Blacks and Hispanics and from a random sample of the 3,300 white applicants filed in the Boston MSA in 1990. Bank failures, inability to locate all requested loan files, and corrections to earlier submissions reduced the final sample size to 3,062 (2,340 white, and 1,013 minority applicants).

By design, the Boston Fed sample included a larger proportion of minorities than whites; the final sample contained 59.7 percent of all minority applications from the MSA in 1990, and 14.6 percent of the whites. Disproportionate sampling of this type is quite common and often well-justified in logit regression studies. It is generally done in order to guarantee that the sample contains adequate numbers of the groups least represented in the total population, in this case minorities in general and denied minorities in particular. While a common practice, disproportionate sampling does pose certain technical problems in terms of the validity of the regression results, and different authors have proposed various corrective techniques. (See Maddala 1983, pp. 90-91, for discussion.) We applied two such techniques, a weighted regression and an intercept adjustment, to the Boston Fed model. Neither had any substantial effect on information on 38 additional variables the researchers thought could influence the lending decision for each loan application included in the sample.

The expanded HMDA sample data show that the minority applicants in the Boston MSA in 1990, on average, had less wealth, higher loan-to-value ratios, and poorer credit histories than white applicants. (*See* Appendix 1 for details.) As stated by the Boston Fed, "These differences tend to support arguments that the higher denial rates experienced by minorities are attributable, at least in part, to financial characteristics, credit histories, and other economic factors" (Munnell et al. 1992, p. 25). Minorities were also more likely to have applied under special loan programs and to have applied for private mortgage insurance. While minorities had lower incomes, they also applied to purchase less costly homes, so their obligation ratios were similar to those of white applicants.

The centerpiece of the Boston Fed study consisted of the development of an econometric model to assess the importance of these differences – and of the applicant' s race – on the outcome of the lending decision. The model attempts to replicate the mortgage lending decision-making process by estimating the probability of denial for each mortgage loan application, as a function of the financial characteristics and credit and employment history of the applicant, the characteristics of the home being purchased, and the neighborhood where the property is located. A dummy variable representing minority status is intended to test for discrimination. That is, after controlling for the relevant financial, credit, etc. characteristics, is there an unexplained portion of the difference in denial rates that is correlated with the applicant' s race?

Since the outcome of the application process can take on only two possible values – approval or denial – a logit limited dependent variable model is used. The dependent variable (referred to as the logit, or the log odds ratio) takes the form

$$g(x) = \ln[p/(1-p)]$$

where $p =$ probability of denial.[2]

---

the signs, magnitudes, or significance levels of the estimated coefficients, including the race variable; nor was there any appreciable impact on the various measures of goodness-of-fit.

[2] In the estimation procedure,

$$g(x) = \beta x$$

where x = independent variables, and
β = estimated parameters.

Since the dependent variable can take on only one of two values (1 if denied, 0 if approved), iterative regression techniques are used to estimate the model parameters.

Table 1. Results of regression on the likelihood of denial: Full Boston sample

VARIABLE	COEFFICIENT (P-VALUES)	
	BOSTON I[A]	BOSTON II[B]
CONSTANT	-6.61 (0.0001)	-6.52 (0.0001)
ABILITY TO SUPPORT LOAN		
HOUSING EXPENSE RATIO	0.47 (0.0014)	0.46 (0.0023)
TOTAL DEBT RATIO	0.04 (0.0001)	0.05 (0.0001)
NET WEALTH	0.00008 (0.2714)	0.00009 (0.1627)
RISK OF DEFAULT		
CONSUMER CREDIT HISTORY	0.33 (0.0001)	0.36 (0.0023)
MORTGAGE CREDIT HISTORY	0.35 (0.0027)	0.31 (0.0001)
PUBLIC RECORD HISTORY	1.2 (0.0001)	1.2 (0.0001)
PROBABILITY OF UNEMPLOYMENT	0.09 (0.0010)	0.08 (0.0028)
SELF-EMPLOYED	0.52 (0.0051)	0.46 (0.0133)
POTENTIAL DEFAULT LOSS		
LOAN-TO-VALUE RATIO	0.58 (0.0014)	0.61 (0.0014)
DENIED PRIVATE MORTGAGE INSURANCE	4.7 (0.0001)	4.6 (0.0001)
RENT/VALUE IN TRACT	0.68 (0.0005)	NA
LOAN CHARACTERISTICS		
PURCHASING 2- TO 4-FAMILY HOME	0.58 (0.0003)	0.55 (0.0008)
PERSONAL CHARACTERISTICS		
RACE	0.68 (0.0001)	0.71 (0.0001)
NUMBER OF OBSERVATIONS	3062	2932
PERCENT CORRECT PREDICTIONS (P>0.5)	89	89
HOSMER-LEMESHOW TEST (P-VALUE)	NA	0.16

[a] As reported in Munnell et al. (1992, Table 5); the t-statistics were converted to p-values for ease of interpretation.
[b] Data provided to the public by the Boston Fed contains fewer observations and the "Rent/value in tract" variable has been deleted.

The results of the Boston Fed statistical analysis are summarized as Boston I in Table 1. (*See* Appendix 2 for a list of variable definitions.) Munnell et al. found that, after controlling for the creditworthiness of the applicants, a minority applicant was still 56 percent more likely to be turned down than a white applicant. This result is reflected in the large, positive, and significant coefficient on the race variable. The authors tried numerous alternative specifications (*see* Munnell et al. 1992, Appendix B, Tables 1 through 12), and found the magnitude and significance of the coefficient on the race variable to be highly robust across all specifications.

We estimated the same model using the partial data set (2,932 observations) made available to the public by the Boston Fed following the publication of the lending discrimination study.[3] The results are presented as Boston II in Table 1. All coefficients and significance levels and the percent correct predictions are nearly identical to those of the original study.

III. Issues in evaluating the Boston Fed model

Goodness-of-fit

While certainly important, the robustness of the Boston Fed results in and of itself does not validate the model or their finding of discrimination in mortgage lending in the Boston MSA. It does suggest that they are reasonably satisfied the model includes the "correct" variables, specified with the appropriate functional form. It does not, however, tell us how effective the model is in predicting outcomes. Like any statistical model, it must also be tested for goodness-of-fit (*i.e.*, to see if the model provides reasonably accurate estimates of the probability of denial).
Unfortunately, assessing the goodness-of-fit of a limited dependent model is difficult, since the conventional method of assessing model reliability (using the distance between the observed and the predicted values of the dependent variable) is complicated by the binary specification of the dependent variable (1 if denied, 0 if approved). Therefore, unlike least squares regression models, limited dependent variable models have no generally accepted goodness-of-fit measures.[4]

One approach to the goodness-of-fit used by the Boston Fed is to measure the number of "correct" predictions as a percent of the total number of observations. A prediction is said to be correct if the estimated probability of denial is greater than 0.5 (*i.e.*, $\text{prob}(y=1 \mid \mathbf{x}) > 0.5$) for an application that is actually denied, or if the estimated probability of denial is 0.5 or less (*i.e.*, $\text{prob}(y=1 \mid \mathbf{x}) \leq 0.5$) for an application that actually is approved. By this measure, the Boston Fed models consistently obtain an overall correct prediction percentage of roughly 90 percent (with approximately 98 percent of approvals correctly predicted, but only 35 percent of denials

[3] In addition to the reduced number of observations, in order to protect the confidentiality of the applicants in the Boston sample several variables were deleted by the Boston Fed before the data was released to the public.

[4] See Judge et al. (1990) for a critique of some of the more common measures of goodness-of-fit for limited dependent variable models.

correctly predicted).

This correct predictions approach, however, can be misleading. First, despite the obvious intuitive appeal, the choice of 0.5 – or any other value for that matter – as the cut-off that distinguishes predicted approvals from predicted denials is arbitrary. Second, this approach can inflate the predictive power of a model in the very common situation where there are many more occurrences of one outcome (in this case, approvals) than the other (denials). For example, using the Boston Fed data, 85.5 percent of all applications in the sample were approved. So, a naive "model" that simply says "approve every single application" (*i.e.*, the estimated probability of denial is set to zero for all observations) would correctly identify 85.5 percent of the outcomes. Even though the overall predictive power of the naive "model" appears good (85.5 percent correct), it incorrectly predicts approval for every denied application. Also, the predictive accuracy of the Boston model, at 90 percent correct, is much more modest when compared to the baseline of 85.5 percent for the naive "model."[5]

We believe this method of evaluating a model's goodness-of-fit is inappropriate in a more fundamental sense as well, since the objective of a logit model is not to correctly predict the outcome for each individual observation, but to provide a reasonable estimate of the likelihood an outcome will occur (for a given set of values for the independent variables). We know by the very nature of a probabilistic statement that some applications with very low predicted probabilities of denial will in fact be denied and some with very high probabilities will in fact be approved. For example, if the probability of denial estimated by an accurately specified model is 10 percent for each of 50 applications, we would expect 45 actual approvals and 5 actual denials for the group. The five denials are fully consistent with the nature of a discretionary decision-making process, as accurately captured by the logit model. They are not "incorrect" in any meaningful sense. Yet the correct predictions approach would classify them as incorrect and would rank a model that predicts approval for all 50 of these applications as a better-fitting model; such findings are clearly inappropriate. For this reason, we believe a more accurate method of assessing the goodness-of-fit of a logit model would rely on a measure that compares the expected number of occurrences of a particular outcome (denials in this case) to the actual number of such occurrences.

Hosmer and Lemeshow (1989) have developed such a measure.[6] Their test statistic is derived by

---

[5] See Horne (1994) for further discussion of this argument.

[6] Munnell et. al. (1992) develop a crude version of this approach in their Table 6 where they compare predicted denial rates to actual rates for three groups of applicants (grouped according to debt-to-income ratio). The "predicted denial rates" in this case are based on the expected number of denials for each group (which is equal to the sum of the estimated probabilities from the logit model for all the applications in that group) rather than an arbitrary cut-off such as the 50 percent used in calculating the percent of correct predictions. The predicted denial rates shown in their Table 7 are quite close to the actual rates, and very substantially closer than the rates predicted by two alternative models.

Their analysis, however, employs no formal statistical test of the model's ability to predict. Rather "calculating the Pearson chi-square statistic from the 2 by g table of observed and estimated expected frequencies" (Hosmer and Lemeshow 1989, p. 141), where g is the number of groups. We calculated the Hosmer-Lemeshow test statistic for the Boston II model estimated using the publicly available data. The p-values presented in Table 1 are for a chi-square test with g - 2 degrees of freedom. Each p-value indicates the probability that differences between the expected and observed values as great as, or greater than, those derived from the model's estimated probabilities, are due solely to random chance.[7] That is, if, Boston II is the true model of mortgage loan decision-making, there is only a 16 percent probability of observing the pattern of differences between the actual and expected numbers of denials that we see emerge from the results of the Boston II model. This relatively low p-value suggests the model is marginally adequate.[8]

Separate models by race

Table 2 reports on separate models for whites and minorities.[9] Two interesting results emerge from the separate models specification: First, only six of the 12 coefficients in the minority model are significantly different from zero; as compared to 11 of 12 in the white model. Second, though many of the variables in the minority model do not contribute to explaining the denial decision, the model performs extremely well in predicting outcomes (as measured by the Hosmer-Lemeshow test statistic). Those results suggest that (i) minorities with characteristics identical to whites are treated differently, as reflected in the difference in the estimated regression coefficients across equations (*i.e.*, different underwriting standards are applied – discrimination exists); or (ii) the difference in the estimated parameters may only reflect the sensitivity of the estimation procedure to the differences in initial endowments of minority applicants (*e.g.*, higher loan-to-value ratios, poorer credit history, and lower wealth), on average; or (iii) both i and ii.

---

it uses a crude, casual comparison of the numbers of expected and actual denials. Moreover, they divide the sample into a smaller number of unequal groups and perform the grouping according to the debt-to-income ratio rather than by estimated probabilities. Though the number of groups is somewhat discretionary, Hosmer and Lemeshow favor using at least seven, with equal (or as close to equal as possible) numbers of observations in each group. See Hosmer and Lemeshow (1989, pp. 140-145), for a more thorough discussion of their test statistic. (The SAS software used for the present analysis uses ten equal groups.)

[7] The null hypothesis is that the number of expected denials is equal to the number of actual denials in each of the groups.

[8] While there are no absolute standards for this test, a critical value of p = .05 (or p = .10) would be consistent with generally accepted levels of significance. The Hosmer-Lemeshow test statistic for the Boston II model indicates that there is room for improvement in the model specification in terms of the model's ability to replicate the mortgage lending decision.

[9] The results reported for the separate minority and white models in Table 2 are very similar to those reported in Appendix B, Table 13, in Munnell et al. (1992). The results may differ slightly due to the smaller data set made available to the public by the Boston Fed. See note 3, above.

Munnell et al. support the first conclusion. They reject the hypothesis of a difference in treatment due to differences in factors unrelated to race. This follows directly from their analysis of the difference in the estimated coefficients across race equations. In their Appendix B, they report the results of an analysis of the difference in estimated parameters between the minority and white models. They initially report that there exists a statistically significant difference in the estimated parameters across equations. They argue, however, that the difference can be explained by the introduction of race as an explanatory variable (a simple shift parameter). They therefore conclude that the qualifying standards applied across the races are the same; that is, that the weights assigned to measure the relative importance of each decision variable on the underwriter's decision are, in a statistical sense, equal.

Munnell et al. do not test for the possibility that the difference in treatment across equations is associated with factors other than race. Differences may arise, however, for several reasons: (i) the differences in initial endowments (financial and credit characteristics, etc.) between whites and minorities, on average, may be large enough to distort the relationship between the likelihood of denial and race; (ii) there may be omitted variables or relationships correlated with the race variable;[10] (iii) there may be significant non-linearities embedded in the relationship between the explanatory variables and the logit variable which have a disproportionately adverse effect on minorities; or, (iv) there may be a more complex two- (or more) step procedure involved, which would cause a single-step analysis to inadvertently assign too much weight to the race variable and thus inappropriately implying discrimination where none exists.

In the discussion below we address the first three of these four possibilities. We address the first issue by testing the hypothesis that the difference in treatment is due to the significant differences in the average values of the independent variables (*i.e.*, initial endowments) by racial group. Our primary hypothesis is that the significantly lower wealth/liquid assets, poorer credit history, higher loan-to-value ratios (LTV), and higher debt ratios of the minority group can account for a significant portion of the difference in denial rate. We test this hypothesis using a Blinder-Oaxaca procedure. This procedure decomposes the differences in the dependent variable between the white and minority equations into two components. The first is a measures of the difference in the dependent variable associated with the difference in initial levels of the endowments by groups; the second measures the difference due to differences in the estimated parameters across equations (*i.e.*, that associated with differences in treatment – discrimination).

[10] They dismiss this criticism as a possible explanation for the difference in denial rates. They argue their data set contains all the important information necessary to develop a model to mimic the underwriters' decision process. We discuss the issue of model specification in more detail in Section V, below.

Table 2. Results of Boston regression model, by race

Variable	Coefficient (p-value)	
	Boston II Minority	Boston II White
Constant	-6.84 (0.0001)	-6.19 (0.0001)
Ability to support loan		
Housing expense ratio	0.43 (0.0829)	0.46 (0.0182)
Total debt ratio	0.07 (0.0001)	0.04 (0.0001)
Net wealth	-0.00022 (0.6391)	0.00009 (0.1415)
Risk of default		
Consumer credit history	0.32 (0.0001)	0.30 (0.0001)
Mortgage credit history	0.52 (0.0361)	0.31 (0.0227)
Public record history	1.0 (0.0003)	1.4 (0.0001)
Probability of unemployment	0.08 (0.1854)	0.09 (0.0075)
Self-employed	0.06 (0.8707)	0.60 (0.0043)
Potential default loss		
Loan-to-value ratio	0.80 (0.2218)	0.58 (0.0033)
Denied private mortgage insurance	4.0 (0.0001)	4.9 (0.0001)
Rent/value in tract	NA	NA
Loan characteristics		
Purchasing 2- to 4-family home	0.37 (0.1116)	0.73 (0.0014)
Personal characteristics		
Number of observations	685	2247
Percent correct predictions (p= 0.5)	80	91
Hosmer-Lemeshow test (p-value)	0.80	0.31

We address the second and third possibilities by testing several alternative model specifications that include additional variables not investigated in the Boston Fed study and that incorporate specific non-linear relationships between the dependent variable and the independent variables. More specifically, we use both higher order expressions for several of the independent variables that are likely to have greater influence the greater their deviation from industry standards, and step function relationships (*e.g.*, different underwriter criteria are used if the LTV is .80 or higher) that more closely approximate the procedures generally used in the industry.

The fourth issue is much more complex. It suggests that there may exist a two- (or more) step process that initially evaluates the creditworthiness of the borrower using a reduced set of "key" qualifying variables instead of the broad spectrum of variables identified in the Boston Fed study.[11] If a large proportion of the minority rejections take place at this level, for example because a higher proportion of applicants have insufficient funds to close or poor credit histories, a single step analysis may overemphasize the importance of race on the approval decision. An analysis of this sort is currently outside the scope of this paper.

IV. The importance of the difference in endowments by race

An analysis of the sample statistics, by race, shows the financial characteristics of minorities in the Boston MSA differ significantly from those of whites (*see* Appendix 1). Minorities tend to have less wealth, less income, a higher percentage of credit problems, and a lower percentage of liquid assets in excess of closing costs. Further, they generally borrow a larger percentage of the value of the property and more often apply for private mortgage insurance (PMI). By themselves, lower wealth and income need not imply lower qualifications, since borrowers with lower incomes and wealth tend to purchase less expensive homes. Though the median loan-to-income ratio of minorities is higher (2.45) than whites (2.03), monthly housing payments-to-income ratios of minorities, in general, meet the secondary market guidelines. The relationship between white and minority applicants with respect to the other variables, however, supports the argument that higher denial rates can be attributed to lower minority qualifications.[12]

The methodology used in the Boston Fed's model (*i.e.*, the difference in treatment is confined to an analysis of the difference in the intercept term through a simple dummy variable specification) implicitly assumes that the groups have a similar endowment distribution. Accordingly, the difference in denial rates unexplained by the augmented model must be attributed to discrimination. They find that approximately 56 percent of the difference in the denial rates remains after controlling for wealth, employment, and credit history.[13] This remaining difference

---

[11] Munnell et al. outline just such an underwriting procedure (Munnell et al. 1992, pp. 10-12). However, they do not explicitly incorporate the two-stage aspect in their econometric model.

[12] This is also supported by the higher percentage of minority applicants who apply under special programs (51 percent compared to 13.4 percent for whites). Moreover, the large percentage of minorities evaluated under special programs seems to suggest that the underwriting standards are different for this group; just as we would expect the underwriting standards on an Federal Housing Administration (FHA) or Veterans Administration (VA) loan to be different from those used to evaluate a conventional home purchase loan. This leads us to believe that the difference in endowments may explain at least part of the difference in denial rates.

[13] This is derived from the results reported in Table 8 (Munnell et. al. 1992). After controlling for endowments using the black/Hispanic characteristics but white experience, the minority denial rate falls 7.9 percentage points, from 28.1 percent to 20.2 percent. This represents a 44.4 percent decline, due to differences in endowments, in the initial 17.8 percentage point difference in denial rates.

Elsewhere in their paper, Munnell et al. give different estimates, using different methodologies, of the explained and unexplained shares of the difference in denial rates. We use the 44 percent figure here is attributed to discrimination.

Because the Boston Fed model restricts the difference between races to the single shift term, however, it is possible that a portion of the difference attributed to discrimination could be due to differences in the distribution of endowments by groups; a likely result given the substantial differences in endowments. For this reason, we performed the Blinder-Oaxaca procedure (Berndt 1991; Blinder 1973), which decomposes the difference in the dependent variable into differences associated with initial endowments and those associated with discrimination.

The procedure relies on the property that the fitted regression line passes through the regression means. This property holds for the logit of a multiple logistic regression model (*i.e.*, the logit is the log odds ratio, and is defined as: $g(\mathbf{x}) = \ln[p/(1-p)] = ß\mathbf{x}$). The decomposition of the difference in the log odds ratios, by race, evaluated at the mean values of the independent variables, is derived as follows:

$$g(X_w^a) - g(X_m^a) = b_w (X_w^a - X_m^a) + X_m^a \Delta b \qquad (1)$$

where

$g(X_i^a)$ = the mean log odds ratios, $\ln[p(y=1 \mid X_i^a)/(1-p(y=1 \mid X_i^a))]$ for i = m (minority), w (white), are calculated using the average values of the independent variables, X_i^a, for each group;

$p(y=1 \mid X_i^a)$ = probability of denial, given the mean values of the independent variables;

b_w = estimated coefficient from Table 2, Boston II, White;

X_i^a = average value of the independent variables (i = m (minority), w = (white)); and

Δb = difference in the estimated coefficient between the Boston II, White and Boston II, Minority results reported in Table 2.

The first term on the right hand side of the equation represents the amount of the difference in the log odds ratios associated with the differences in average endowments, and the second term is that associated with differences in the estimated parameters. A test of the null hypothesis of equality of parameters across equations (whites only versus minority only) reveals that the coefficients used in the Boston Fed model are indeed different, supporting the hypothesis that the initial endowments are important determinants of the denial rate.[14]

The results suggest that 44 percent of the *difference* in denial rates is associated with

because the methodology employed to derive it is most directly comparable to the results of the Blinder-Oaxaca procedure.

[14] A chi-square test of the hypothesis of the equality of parameter estimates rejected the null hypothesis at the 1 percent level ($\chi 2$-stat = 58.55, $\chi^2_{01,12} = 26.22$). Similar results were reported in Munnell et al. (1992, Appendix B).

discrimination. In other words, the "average" minority applicant (derived using the average values of the independent variables), had an estimated probability of denial of 24.5 percent, while the "average" white applicant's estimated probability of denial was 6.8 percent. Controlling for differences in the average initial endowments lowers the estimated probability of denial of the "average" minority applicant to 14.6 percent, thus explaining 9.9 percentage points, or 56 percent of the initial difference in average estimated denial rates. The remaining 44 percent of the difference is attributed to differences in treatment by race.

Our results suggest that the degree of discrimination is lower than that found in the Boston Fed study; a result we consider more accurate because the Blinder-Oaxaca procedure does not assign the differences in denial rates exclusively to the intercept term, but also allows for variation in the estimated slope coefficients. However, and more importantly, our results show that the difference in denial rates cannot be explained entirely by the difference in endowments. This substantive agreement with the Boston Fed's central conclusion should not be overshadowed by the relatively minor difference in magnitude (56 percent versus 44 percent) between our results and those of Munnell et al.

An important caveat to this procedure, however, is that the model must include all relevant decision variables. If the model is misspecified, the second term on the right-hand side of equation (1), $X_m^a \Delta b$, will overstate the degree of discrimination. It is this issue we address in the next section.

V. Specification of the Boston model

One of the most important issues related to the Boston model, and one that has been the subject of considerable comment and criticism, is the question of the precise specification of the regression model. That is, which variables are included in the regression equation, and in what form(s)? Since the exact specification of the regression equation is not derived directly from a theoretical model of mortgage lending, and since there is no well-established, standard model in the previous literature, Munnell et al. had considerable leeway in their choice of specification.

The Boston Fed collected data for a list of variables identified as the primary decision variables through "numerous conversations with lenders, underwriters, and others familiar with the lending process" (Munnell et al. 1992, p. 13). A number of alternative specifications of the model were explored (see Appendix B, Munnell et al. 1992) before the authors settled on the particular form of the relationship between the independent variables and the decision to approve or deny a mortgage application. In this process, the authors chose both the variables to include and their mathematical forms (e.g., linear or nonlinear, dummy or continuous) used in the final model.

While Boston Fed staff tried a large number of different specifications, there are reasons to believe that the final model presented in the paper fails to capture all relevant aspects of the relationship between the independent variables, including the race of the applicant, and the outcome of the mortgage lending decision. Therefore, it is fruitful to explore alternative specifications they did not try – or, at least, did not report. These alternative variables and alternative forms of variables, are suggested by the model used in the Department of Justice case against Decatur Federal, by the academic literature in the field, and by discussions with Office of the Comptroller of the Currency (OCC) examiners and fair lending and compliance staff.[15]

The data

Table 1, above, presents our replication of the Boston Fed's basic model using the publicly available dataset. The Boston Fed also made available to each federal banking agency a set of the raw data for participating institutions for which the agency is the primary federal regulator. The OCC and the Federal Deposit Insurance Corporation (FDIC) then agreed to exchange their data sets (after removing all fields that could identify individual institutions or applicants). This yielded a combined data set of 1,603 observations, or slightly over half of the original sample used by the Boston Fed. These data were used for most of the analysis reported below.[16]

Appendix 1 presents a comparison of the median values of a select group of series for the OCC/FDIC sample and the full sample used in the published study. As shown, the smaller OCC/FDIC sample is very similar in almost all characteristics, with a few exceptions. In the OCC/FDIC sample denied applicants (both white and minority) had less net wealth

[15] For some of the more suggestive studies from the discrimination and the closely-related default and redlining literatures, see Berkovec, Canner, Gabriel and Hannan (1993); King (1980); Perle, Lynch, and Horner (1993); Schill and Wachter (1993); Schafer and Ladd (1981); Siskin and Cupingood (1993); and Van Order (1993).

[16] The OCC/FDIC data set was used in the analysis rather than the larger (2,932) data set released to the general public, because the smaller data set contains several variables that were deleted from the larger data set. Of particular importance was the census tract variable, which permitted matching of the Boston sample with census tapes and, in turn, made it possible to use of a variety of neighborhood demographic characteristics.

Table 3. Regression results: Full Boston sample and OCC/FDIC sample

Variable	Coefficients (p-values)	
	OCC/FDIC[a]	Boston II[b]
Constant	-6.8 (0.0001)	-6.52 (0.0001)
Ability to support loan		
Housing expense ratio	0.48 (0.0271)	0.46 (0.0023)
Total debt ratio	0.04 (0.0001)	0.05 (0.0001)
Net wealth	0.00008 (0.2615)	0.00009 (0.1627)
Risk of default		
Consumer credit history	0.34 (0.0001)	0.36 (0.0023)
Mortgage credit history	0.49 (0.0024)	0.31 (0.0001)
Public record history	1.3 (0.0001)	1.2 (0.0001)
Probability of unemployment	0.07 (0.1078)	0.08 (0.0028)
Self-employed	0.38 (0.1446)	0.46 (0.0133)
Potential default loss		
Loan-to-value ratio	0.51 (0.0045)	0.61 (0.0014)
Denied private mortgage insurance	4.5 (0.0001)	4.6 (0.0001)
Rent/value in tract	0.12 (0.7978)	NA
Loan characteristics		
Purchasing 2- to 4-family home	0.56 (0.0164)	0.55 (0.0008)
Personal characteristics		
Race	0.94 (0.0001)	.71 (0.0001)
Number of observations	1603	2932
Percent correct predictions	89	89
Hosmer-Lemeshow test (p-value)	0.82	0.16

[a] Data provided by the Boston Fed to the OCC and FDIC. [b] Boston II results from Table 1 are reproduced for ease of comparison.

and somewhat lower monthly incomes; denied applicants (both white and minority) were more likely to have applied for private mortgage insurance; and minority applicants (both approved and denied) were more likely to be applying under special programs of some kind.

Table 3 compares the logit regression results, applying the same Boston Fed model specification to the OCC/FDIC sample as to the publicly available sample. Most of the estimated coefficients are quite similar. Of those that show substantial change for the smaller sample, mortgage credit history and race are larger, while self-employment and rent/value are smaller. Three of the explanatory variables (probability of unemployment, self-employed, and rent/value) lose their significance at the 5 percent level in the OCC/FDIC sample; all the others retain the same significance, or lack thereof, as in the full sample regression. Interestingly, the goodness-of-fit of the model increases substantially for the smaller sample, suggesting that the model performs better as a predictor of loan decisions for the sample of non-member state-chartered and national-chartered banks than for the full sample.

Alternative specifications

Alternative variables and alternative forms of the variables were explored to improve the specification and goodness-of-fit of the Boston Fed model. The alternative variables and forms that were significant, individually or in combination, or that improved the overall performance of the model are discussed below.

One general issue that merits discussion relates to the nature of the mortgage decision-making process, and how some of the most important underwriting standards are applied. In particular, it appears, from discussions with examiners, underwriters, and others familiar with mortgage lending, that there are significant discontinuities and/or nonlinearities in the way underwriters look at factors like the loan-to-value and debt ratios. For example, loan-to-value ratios may significantly affect the decision-making process only above certain levels; applications with very high loan-to-value ratios may be automatically disqualified, regardless of how strong the rest of the file may be. Certainly a change in the loan-to-value ratio from 85 percent to 90 percent will exert much more influence on the outcome of the underwriting process than a change from 50 percent to 55 percent.

Considerations of this type seem to be of particular importance in a model of the lending process, yet the Boston model is conspicuous for the absence of variable forms that attempt to capture such factors. In our efforts to develop alternative specifications, we devoted considerable attention to experimenting with forms that might capture these characteristics of the underwriting process, in particular with regard to the loan-to-value, monthly housing expense-to-income and the total debt-to-income ratios.

Loan-to-value ratio. While the loan-to-value ratio is highly significant in the Boston model, the form in which it is cast in the model is subject to question. In particular, the loan-to-value ratio itself enters into the model, implying a linear relationship between the ratio and the dependent variable (the logit).[17] Judging from the results reported in the Boston Fed study (Munnell et al. 1992, both text and Appendix B), it appears that no attempt was made to capture the kinds of nonlinear or threshold effects discussed above.

In order to better reflect the nonlinearities associated with the loan-to-value ratio, that variable was replaced by a set of three variables designed to capture the threshold effect. These are: (i) the excess above 80 percent (if 80 percent or less, the variable is set equal to zero), (ii) the excess above 80 percent squared, and (iii) a dummy variable

---

[17] Obviously, this implies a nonlinear relationship between LTV, or any other independent variable, and the probability of denial (as opposed to the dependent variable itself). However, in the Boston Fed specification, if one applicant has an estimated probability of denial of, say, 10 percent and an LTV of 50 percent, and another has the same 10 percent probability of denial but an LTV of 90 percent (because of more favorable debt ratios, for example), then a 5 percentage point increase in LTV (from 50 percent to 55 percent in the first case, and from 90 percent to 95 percent in the second) will have the same impact on the probability of denial (an increase to 10.26 percent) for both applicants.

set equal to 1 if the ratio is above 90 percent. The 80 percent level is generally regarded as an important threshold, since the secondary market (Fannie Mae and Freddie Mac) requires private mortgage insurance for all loans above 80 percent. The squared term allows for nonlinearities in the relationship between LTV and the dependent variable, and the dummy variable differentiates loans with exceptionally high LTVs.

As shown in Table 4, for Model I all three alternative forms of the LTV specification are significant at the 5 percent level. The signs indicate, plausibly, that increasing LTV above the 80 percent level increases the probability of denial, although at a decreasing rate, and that increases above the 90 percent level increase the probability of denial even further. The estimated coefficient for the variable that measures the impact of the race of the applicant is somewhat smaller, suggesting that some of the unexplained difference in probability associated with race is captured by the threshold effect of the loan-to-value ratio. However, the race coefficient remains highly significant. The various measures of goodness-of-fit and predictive ability all show marginal improvement over the Boston model – except for the percent of correctly predicted denials, which shows a more substantial improvement. Moreover, the log likelihood ratio shows substantial improvement.[18] All of the remaining coefficients and significance levels change only slightly.

Debt ratios. The two standard debt ratios are used in the Boston Fed model to measure the ability of the borrower to support the loan payments. The ratio of proposed monthly housing expenses to income was entered in the form of a dummy variable that was set equal to 1 if the ratio exceeded 30 percent. The total debt-to-income ratio was entered directly into the equation.

The debt ratios are another area where discussions with examiners and underwriters, as well as the logic of the underwriting process, suggest that there may be significant nonlinear and/or discontinuous effects. While the Boston model captures some of these with the dummy variable formulation for the housing debt ratio, other studies have used other specifications that seemed worth exploring.

Using the widespread secondary market standards (28 percent for the housing expense ratio and 36 percent for the total debt ratio), it is hypothesized that the likelihood of denial increases as a function of the degree to which the ratios exceed these threshold levels. Specifically, a set of three variables was substituted for the Boston formulation: (i) the excess of the housing debt ratio above 28 percent squared, (ii) the excess of the

[18] The log likelihood ratio (-2 log likelihood) is a common diagnostic statistic for all maximum likelihood estimators, including logit models. It is used in tests of model significance. In this case, when comparing logit models estimated for the same sample, a substantial difference in the likelihood ratios indicates that the model with the smaller ratio does a better job at explaining the observed pattern of outcomes.

total debt ratio above 36 percent, and (iii) the excess above 36 percent squared.

As indicated in Table 4, for Model II all three of the new debt ratio variables are highly significant. The estimated coefficients imply that the probability of denial increases exponentially at higher levels of housing debt; it also increases at higher levels of total debt, though at a decreasing rate. The impact of race is slightly larger in magnitude than in the Boston Fed specification and remains highly significant.

The goodness-of-fit statistic, although still consistent with a well-specified model, shows a considerable decline relative to the Boston model, while the measures of predictive ability show improvement, and the log likelihood ratio is virtually identical to that of Model I. All of the remaining coefficients and significance levels change only slightly.

Employment and education. The Boston model used two variables – (i) the Massachusetts unemployment rate for the industry where the applicant was employed, and (ii) a dummy variable indicating if the applicant was self-employed – to represent employment status, history, and stability. The estimated coefficients were both positive, indicating that higher levels of employment instability, as proxied by the unemployment rates and self-employed status, are associated with higher probabilities of denial. Although both coefficients were significant at the 2 percent level or better as reported for the full sample, when estimated for the combined OCC/FDIC sample, neither was significant, even at the 10 percent level.

There are two problems with the use of the unemployment rate variable in the Boston model. First, the model does not make use of potentially useful information related to employment in the loan file, for example the number of years in the current occupation.[19] Further, education is also a useful proxy for employment and income stability; applicants with more years of education are more likely to find and retain jobs, and have better potential for advancement and income growth. Second, the employment variable used in the Boston model assigns the same probability of unemployment to everyone in the same industry (by two-digit standard industry code (SIC)), whether clerk or chief executive officer.

The potential weakness of the Boston approach is illustrated by the case of a loan file where both the applicant and coapplicant were security guards with less than one year in the occupation; the applicant had 1.5 years on the current job, the coapplicant less than one year. Because the applicant worked for a bank, however, the unemployment

---

[19] As reported by Munnell et al. 1992 in their Appendix B, the Boston Fed tried several alternative specifications including the number of years on the current job and a dummy variable indicating more than two years on the current job.

rate for the industry was relatively low, and, partially as a result of this factor, the application had one of the lowest estimated probabilities of the entire sample. Yet the very short employment tenures of both the applicant and coapplicant constitute a significant negative factor that should plausibly have contributed to a higher probability of denial. This was clearly not captured by the industry unemployment rate specification used in the Boston model.

In an attempt to better model employment history and income stability, two variables were substituted for the probability of unemployment: (i) the number of years of education for the person (applicant or coapplicant) with the higher employment income, and (ii) a dummy variable set equal to one if the applicant's number of years in the current line of work was greater than five. The self-employed dummy variable was retained.

Table 4. Results of alternative regression models: OCC/FDIC sample

Variable	OCC/FDIC[a]		Model I		Model II		Model III		Model IV		Model V	
	β	p-value	β	p-value	β	p-value	β	p-value	β	p-value	β	p-value
Constant	**-6.8**	0.0001	**-6.4**	0.0001	**-5.48**	0.0001	**-5.29**	0.0001	**-6.35**	0.0001	**-2.48**	0.0052
Ability to support loan												
Housing expense ratio	**0.48**	0.0271	**0.54**	0.0153			**0.35**	0.1127	**0.46**	0.0494		
Excess above 28%					**0.0003**	0.0176					**0.0007**	0.3430
Total debt ratio	**0.04**	0.0001	**0.04**	0.0001			**0.05**	0.0001	**0.06**	0.0001		
Excess above 36%					**0.13**	0.0001					**0.15**	0.0001
Excess above 36% sqd					**-0.0008**	0.0001					**-0.001**	0.1345
Net wealth	**.00008**	0.2615	**.00009**	0.1959	**.00006**	0.3563	**.00008**	0.2699	**.00009**	0.3103	**.00008**	0.2445
Risk of default												
Cons. credit history	**0.34**	0.0001	**0.44**	0.0081	**0.34**	0.0001	**0.37**	0.0001	**0.34**	0.0001	**0.38**	0.0001
Mtg. credit history	**0.49**	0.0024	**0.35**	0.0001	**0.52**	0.0018	**0.52**	0.0020	**0.53**	0.0023	**0.49**	0.0099
Public record history	**1.3**	0.0001	**1.3**	0.0001	**1.18**	0.0001	**1.26**	0.0001	**1.41**	0.0001	**1.34**	0.0001
Prob. unemployment	**0.07**	0.1078	**0.06**	0.1871	**0.06**	0.1296			**0.07**	0.1288		
Self-employed	**0.38**	0.1446	**0.50**	0.0644	**0.29**	0.2837	**0.59**	0.0248	**0.28**	0.3429	**0.46**	0.1389
Yrs. of education							**-0.08**	0.0066			**-0.10**	0.0033
Over 5 yrs. in occ.							**-0.48**	0.0184			**-0.35**	0.1292
Potential default loss												
Loan-to-value ratio	**0.51**	0.0045			**0.46**	0.0122	**0.52**	0.0046	**0.50**	0.0064		
Excess above 80%			**1.73**	0.0302							**1.76**	0.0360
Excess above 80% sqd			**-0.23**	0.0333							**-0.23**	0.0354
Above 90 percent			**0.92**	0.0001							**0.90**	0.0008
Denied PMI	**4.5**	0.0001	**4.4**	0.0001	**4.55**	0.0001	**4.42**	0.0001	**4.92**	0.0001	**5.02**	0.0001
Rent/value in tract	**0.12**	0.7978	**0.18**	0.7155	**0.24**	0.6044	**0.19**	0.6828				

Variable	OCC/FDIC[a]		Model I		Model II		Model III		Model IV		Model V	
	β	p-value	β	p-value	β	p-value	β	p-value	β	p-value	β	p-value
% moved in pre-1985									**-0.02**	0.067	**-0.02**	0.0471
Loan characteristics												
2- to 4-family home	**0.56**	0.0164	**0.55**	0.0201	**0.47**	0.0487	**0.41**	0.0822	**0.54**	0.0246	**0.32**	0.2248
Personal characteristics												
Race	**0.94**	0.0001	**0.80**	0.0001	**0.98**	0.0001	**0.86**	0.0001	**0.91**	0.0001	**0.71**	0.0017
No. of observations	1603		1603		1603		1588[b]		1378[c]		1367[d]	
% correct predictions	89		90		89.6		89		89		91	
% correct approvals	97.6		97.7		97.6		97.7		97.9		97.4	
% correct denials	38.8		43.1		42.2		40.4		40.3		51.5	
-2 log likelihood	863.8		833.4		833.3		838.4		723.4		501.2	
Hosmer-Lemeshow test (p-value)	0.82		0.83		0.49		0.66		0.72		0.50	

[a] OCC/FDIC model from Table 3 is reproduced for ease of comparison.
[b] Years of education was missing for 15 observations.
[c] 1990 Census tapes could be matched with loan applications for only 1378 observations. See footnote 22.
[d] Of the 1378 observations used for Model IV, years of education was missing for 11 observations.

As shown in Table 4, for Model III the dummy variable for self-employed applicants grows both larger and considerably more significant. The two new variables are both negative, as expected, indicating that the probability of denial is lower for applicants with more education and with longer experience in their current occupation; both estimated coefficients are significant. The race coefficient is slightly lower and still highly significant. The goodness-of-fit statistic is less than that associated with the Boston Fed specification; however, it still suggests the specification of Model III is quite good. The measures of correct predictions show modest improvement.

The estimated coefficients for the remaining variables, in general, show only slight changes in magnitude. There was, however, a substantial change in significance level of the coefficient on the housing debt ratio. The dummy variable indicating a housing debt-to-income ratio above 30 percent went from about 3 percent significance to about 11 percent.

Neighborhood effects. It is widely believed that the location of the property being purchased enters into the underwriting decision. In particular, if the home is in an area where it can reasonably be expected that property values may show little growth, or in fact decline, then a loan is less likely to be approved, all other things being equal, because the probability of default is higher and the position of the mortgagee is less secure.

Measuring this location effect, however, has proven to be difficult. The Boston Fed study uses a rather unusual proxy variable, the ratio of rental income to the value of the rental housing stock in the Census tract where the property is located. The estimated coefficient was positive, indicating that higher rent-to-value ratios are associated with higher probabilities of denial. The estimate is highly significant in the published results for the full sample, but insignificant in the combined OCC/FDIC sample.

The Boston Fed's choice was based on the presumption that landlords demand a higher return on riskier properties. While there is some logic to this line of reasoning, the authors do not present any evidence to support the contention that this ratio is in fact higher in riskier or deteriorating areas, or to prove that any such differential returns that might exist are, in fact, due to differences in risk.[20] While the authors did try several other measures of neighborhood effects, as reported in the Appendix,[21] there are

---

[20] In addition, it is not at all clear how the rent-to-value ratio was derived. The authors state that it can be derived from Census tract data, but the Census reports housing values only for owner-occupied housing units.

[21] The number of units boarded up, the number vacant, a measure of housing value appreciation in recent years, the rate of foreclosure, a dummy variable indicating minority population share over 30 percent, and a set of separate dummy variables for each Census tract.

many other possible candidates that have been used by other authors and that are readily available from the Census tapes.

After matching the OCC/FDIC sample with 1990 Census tapes,[22] the proportion of households who had moved into their residences prior to 1985 was substituted for the rent-to-value ratio. The pre-1985 variable can be interpreted as a proxy for neighborhood stability, with the presumption that the more households that stay in their residences for a long time, the more stable the property values in an area and the lower the default rate and collateral risks to the lender.

As shown in Table 4, for Model IV the estimated coefficient for the new variable, percent moved in pre-1985, is negative, indicating that a location in a more stable neighborhood lowers the probability of

[22] Matching the data proved to be problematic. The 1990 HMDA data reported Census tracts based on the 1980 tract definitions, and many tract boundaries were changed for the 1990 Census. Therefore, loan files could not be matched with the 1990 Census tapes in those cases where the 1980 tract designation was changed. Since tract-based neighborhood variables could not be defined in such cases, they were dropped from the analysis. This process resulted in the loss of 225 cases, leaving a sample of 1,378 applications. While this could potentially lead to biased results, a comparison of the descriptive statistics for the full OCC/FDIC sample and the reduced sample showed remarkable similarity of means and standard deviations.

Table 5. Results of regression model: Boston Fed and alternative specification

Variable	Boston I.A[a]		Model V	
	Coefficient	p-value	Coefficient	p-value
Constant	**-7.2**	0.0001	**-2.48**	0.0052
Ability to support loan				
Housing expense ratio	**0.49**	0.0371		
Excess above 28%			**0.0007**	0.3430
Total debt ratio	**0.05**	0.0001		
Excess above 36%			**0.15**	0.0001
Excess above 36% sqd			**-0.001**	0.1345
Net wealth	**.00009**	0.2843	**.00008**	0.2445
Risk of default				
Cons. credit history	**0.35**	0.0001	**0.38**	0.0001
Mtg. credit history	**0.51**	0.0036	**0.49**	0.0099
Public record history	**1.4**	0.0001	**1.34**	0.0001
Prob. unemployment	**0.07**	0.1402		
Self-employed	**0.24**	0.4094	**0.46**	0.1389
Yrs. of education			**-0.10**	0.0033
Over 5 yrs. in occ.			**-0.35**	0.1292
Potential default loss				
Loan-to-value ratio	**0.51**	0.0051		
Excess above 80%			**1.76**	0.0360
Excess above 80% sqd			**-0.23**	0.0354
Above 90%			**0.90**	0.0008
Denied PMI	**4.87**	0.0001	**5.02**	0.0001
Rent/value in tract	**0.23**	0.6856		
% moved in pre-1985			**-0.02**	0.0471
Loan characteristics				
2- to 4-family home	**0.56**	0.0210	**0.32**	0.2248
Personal characteristics				
Race	**0.98**	0.0001	**0.71**	0.0017
No. of observations	1367		1367	
% correct predictions	89		91	
% correct approvals	97.5		97.4	
% correct denials	40.7		51.5	
-2 log likelihood	723.4		501.2	
H-L test (p-value)	0.50		0.50	

[a] Results reported are for the same specification as Boston I, but for a reduced sample size.

denial, as expected.

The estimate is significant at the 10 percent level only, which is marginal, yet far better than the virtually total lack of significance for the rent-to-value variable. In other respects as well, comparison of the alternative to the Boston specification produces mixed results. The goodness-of-fit statistic is almost identical. The overall percent correct predictions is slightly higher, but that is due entirely to a better performance predicting approvals, while the percent of correct denials is lower. The race coefficient is slightly lower and still highly significant. The remaining coefficients and significance levels showed only slight changes.

Other variables. A number of alternative specifications were also attempted in various other categories such as gender, age, and marital status; property characteristics (e.g., condominium and non-owner occupied); types of institutions (e.g., national banks, state banks, mortgage subsidiaries); and a dummy variable for each separate lending institution. Since these specifications did not substantially alter the outcomes, they are not reported in detail here.

Alternative specification: Summary. Table 4 presents, as Model V, regression results for a specification combining all of the alternative variables discussed in the preceding sections. The signs of the new variables are all the same as when they were introduced separately, and the magnitudes are approximately the same as well. The significance of three of the new variables, however, disappears when they are used in combination: The excess of the housing debt ratio above 28 percent squared, the excess of the total debt ratio above 36 percent squared, and the dummy variable indicating more than five years in the same line of work. The coefficient on the race variable is lower than that in the preceding four models reported in Table 4, and also is lower than the results when the Boston specification is run on the same sample. This indicates that the new specification is capturing some variation in outcomes that was previously attributed to the race of the applicant.[23] The coefficient for race is still large and positive, however, and

[23] The intercept is also considerably smaller than almost all previous specifications, indicating that the new variables are capturing some of the effects that were previously combined in the constant term.

remains highly significant.

Table 5 reproduces the results from Model V and from the Boston Fed specification estimated for the same sample (1,367 observations) for purposes of comparison. The two specifications have virtually identical goodness-of-fit statistics. The log likelihood ratio for the alternative specification, however, shows very substantial improvement, and the alternative specification does better than the basic model in terms of the percent correct predictions. Of particular note is the large improvement in the percent of denials correctly predicted, from 40.7 percent to 51.5 percent. Finally, there is little change in the signs, magnitudes, or significance levels of any of the other variables.

An examination of the estimated probabilities of denial from the two specifications for individual applicants reveals that the difference in results from the two models can be substantial in many cases. For about half of the observations, the two estimated probabilities are within two percentage points; for 13 percent of the observations, however, the difference is greater than ten percentage points and,for a few cases it exceeds 40 percentage points. The probabilities estimated by the original Boston specification exceed those from the alternative specification for about 60 percent of the observations; they are smaller for the remaining 40 percent. Further, of those applications that were actually denied and had estimated probabilities of denial of 0.5 or less in the Boston model but more than 0.5 in the alternative specification, the amount of increase was generally quite substantial. Thus, the increase in the percent of denials correctly predicted, discussed in the previous paragraph, is not due to probabilities that went from just slightly below 50 percent to just slightly above that level.

Because of the form of the logit model, it is difficult to deduce the actual impact of changes in the independent variables on the probability of denial directly from the regression results. Therefore, Table 6 presents the results of the alternative specification discussed above in a more readily comprehensible format.

Table 6. Impact on probability of denial

Variable	Impact on estimated probability of denial (percent)	
	Boston I	Model V
Ability to support loan		
Housing expense ratio	33.9	3.8
Total debt ratio	33.0	67.8
Net wealth	4.5	5.0
Risk of default		
Consumer credit history	37.2	32.8
Mortgage credit history	11.4	12.9
Public record history	113.7	107.2
Probability of unemployment	11.4	
Self-employed	35.1	22.9
Years of education		-13.1
Over 5 years in occupation		-13.9
Potential default loss		
Loan-to-value ratio	11.5	78.4
Denied PMI	596.0	609.6
Rent/value in tract	9.3	16.7
% moved in pre-1985		-9.4
Loan characteristics		
2- to 4- family home	42.5	16.7
Personal characteristics		
Race	56.0	49.3

The table shows the percentage change in the average estimated probability of denial as a result of a change in the value of each independent variable. For the dummy variables, the table shows the impact of the presence of the particular attribute (e.g., self-employed) relative to an otherwise identical application without the attribute (not self-employed). For the continuous variables, such as net wealth, the table shows the impact of a one standard deviation change.[24]

Most of the variables common to both specifications show little change in their impacts on the estimated probability of denial; the one exception is self-employment, which increases the average probability only 22.9 percent in the alternative specification, down from 35.1 percent in the original study. As expected, the most significant changes arise in the cases of the variables that have been transformed in the alternative specification: The total impact of a one standard deviation increase in the loan-to-value ratio has increased dramatically; the housing debt ratio has much less impact, while that of the total debt ratio is approximately double what it was before.[25] Most importantly, the impact of minority status, although somewhat diminished, is still quite large: A Black or Hispanic applicant would have, on average, a 49 percent greater probability of being denied than an otherwise identical white.

Conclusion: Robustness of the race coefficient

The efforts to develop alternative specifications for the model were fruitful in terms of the results for individual variables and groups of variables that seemed, a priori, to

---

[24] The methodology employed in constructing the table is the same as that used by the Boston Fed.

> . . . the first step is to determine the probability of denial in the absence of a particular characteristic, such as being self-employed. This requires determining for each non-self-employed applicant the probability of denial based on the [estimated] coefficients . . . . These estimated probabilities for each applicant are then averaged to get a single figure for the group. The second step is to add to each non-self-employed applicant's probability of denial the impact of being self-employed (the coefficient . . . multiplied by 1). These new probabilities are averaged. The figure reported in the [table] . . .
>
> is the percent difference between the average probability of denial for the non-self-employed with the self-employment effect and the probability for the non-self-employed without it.
>
> For a continuous variable, such as [net wealth] . . . the procedure is slightly different. In this case, 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 [net wealth] . . . for each applicant, recalculate the estimated probabilities of denial, and average the probabilities. As before, the value reported in the [table] . . . is the percent difference between these two average probabilities (Munnell et al. 1992, p. 29).

This procedure was modified slightly for the alternative specification in the case of the loan-to-value ratio, where there are three separate variables based on the ratio (excess above 80 percent, excess above 80 percent squared, and a dummy variable indicating if over 90 percent). The loan-to-value ratio was increased by one standard deviation, and all three variables were recalculated; then the new probability of denial was recalculated for each observation based on the recalculated loan-to-value variables, and the new average was calculated. A similar procedure was followed for the total debt ratio.

[25] Based on discussions with underwriters, this is a very plausible outcome. Lenders pay much more attention to the borrower's total debt burden (relative to income) than to its composition.

better approximate the mortgage underwriting process, as well as in terms of improved predictive ability. Nonetheless, the one result of the Boston study that has attracted the most attention and that has the strongest public policy implications, the estimated coefficient for the race variable and its significance, remains highly robust across specifications. Through many dozens of alternative specifications, both those reported above and many others not reported, the coefficient is always positive, within a narrow range (most often between 0.7 to 0.9), and very highly significant. At least with the alternative specifications that could be explored with the data provided by the Boston Fed,[26] we could not refute the Boston Fed's conclusion that mortgage lenders in the Boston MSA, collectively, treated Black and Hispanic applicants differently from whites in 1990, and that, on average, a Black or Hispanic applicant had an approximately 50 percent greater probability of being turned down than a white applicant with otherwise identical characteristics.

Of course these results are valid only insofar as the data truly reflects the underlying population. Several recent articles have suggested that the Boston Fed data is contaminated by numerous data errors and inconsistencies. They argue these errors could significantly influence the outcome of the Boston Fed study. In the next section, we address many of these issues.

VI. Data entry/reporting errors

The Boston Fed states that they subjected the data to "careful visual inspection and computer edits and repeatedly called back lenders to verify that data items that seemed unusually large or small actually reflected the contents of the lenders' files" (Browne 1993). There are, however, numerous instances in the partial data sets provided to the OCC and FDIC by the Boston Fed in which the data remain suspect. We discuss several examples of data errors we found in the OCC/FDIC data and the implications those errors may have for the Boston Fed's results.

Analyzing the impact of data errors and inconsistencies

Duplicate files and coding errors. Several of the more obvious data entry errors were easily corrected. For example, duplicate files were dropped from the data set, and the debt obligation ratios entered as decimal values were converted to percentages. There are relatively few of these errors, and correcting them resulted in no substantive

[26] For example, we had serious reservations about the particular way that credit history was modeled by the Boston Fed. However, it was not possible to explore alternative formulations of this important variable, because the credit history variables contained in the data provided by the Boston Fed were already coded in the precise form used in the Boston study. Therefore, any effort to model different kinds of credit history effects will have to await collection of new data for future studies. Also, see discussion in Section VI under Model misspecifications and omitted variables in the text for other specification issues that could not be investigated with the data made available by the Boston Fed.

change in the Boston Fed's conclusions. It is not surprising these errors had so little impact on the results given that there is relatively low leverage associated with extreme values in logit models (a statistical property of the estimating procedure), so that even large changes in these values (i.e., converting the debt ratio from a decimal to a percent) can have very small effects on the parameter estimates.

Correcting for other data entry problems, however, is more problematic. Often it is difficult to know if the correction itself introduces more error; and deleting the observation may bias the sample if the errors are non-random across groupings. Moreover, for errors of inconsistency it is difficult to determine which data are false; therefore, the method used to correct the errors may substantially influence the results. We discuss several of these problems below.

Inconsistencies that suggest the debt ratios are underreported. It is likely, given the large number of files with monthly non-housing debt payments far below those consistent with the level of outstanding liabilities, that the total debt obligation ratio is *under*reported in many files.[27] Moreover, the proposed monthly housing payments for another 3 percent of the applicants were below that required on a loan with a 6 percent mortgage interest rate (at a time when market rates were above 10 percent). This suggests that both the monthly housing expense ratio and the total debt ratio were underreported for these observations as well.[28]

In these cases, the applicant may appear qualified according to the model (that is, assigned a low estimated probability of denial), yet appear unqualified to the underwriter, based on accurate information concerning housing and non-housing debt payments. Errors of this type reduce the reliability of the model in general, and the importance of the housing expense and total debt obligation ratios specifically as indicators of loan disposition. There is no obvious way these errors can be corrected short of re-examining the loan files; a solution we believe is impractical.[29]

Including unverified data. Two types of data verification errors are introduced in the

---

[27] For example, one approved applicant with an annual income of $89,544 had negative net wealth of almost $2 million, yet reported total non-housing debt payments of only $156 per month. This is an unreasonably low monthly debt payment. Using a monthly payment rule-of-thumb of .005, this person should have a non-housing monthly debt payment of approximately $10,000.

[28] Even under the assumption that the rates used to calculate the proposed monthly housing payment were "teaser" rates on adjustable-rate loans (though many of them are coded as fixed-rate mortgages), the underwriting decision should have been based on a "qualifying" rate that reflects the then-current market rate on fixed-rate mortgages. Day and Liebowitz (1993) also discuss this apparent data entry error.

[29] One suggested solution would involve deleting these files from the data set. However, we believe this is inappropriate because the remaining sample may well be unrepresentative of the underlying population, thus introducing sample selection bias that will distort the estimates in an unknown manner.

Boston Fed study: (i) several banks reported data from the unverified, initial application,[30] and (ii) several applications contained information provided by the borrower that was not (or could not be) verified by the bank. Both of these data verification problems are especially troublesome for it is difficult to know, a priori, how sensitive the coefficient of the race variable is to these errors.

For example, including unverified data from the original, rather than the final, application form may significantly affect the model's results if the unverified data tend to overstate the qualifications of marginal borrowers (i.e., if the financial and credit history characteristics of the denied applicants appear to approach those of the approved applicants).[31] The overstated qualifications of marginal borrowers reduce the model's ability to accurately identify the importance of each explanatory variable in the lending decision. If the unverified data from the original applications enter disproportionately from a single race category, a model that relies on a single (categorical) variable to proxy for the race effect may wrongly attribute the difference in denial rates to race. The difference in denial rates will be more accurately reflected by differences in financial and credit history characteristics of the applicant if the data used in the model more accurately reflect the differences in qualifications. The greater the difference between the actual (verified) data and the reported data, the more likely the results of the statistical model will be misleading. Verification errors of this type are impossible to correct without re-examining the loan files.

Data verification errors introduced by failing to incorporate information about the ability of the bank to verify all (relevant) information on the application are less troublesome to correct. The Boston Fed data set contains an indicator variable (N56 - unverifiable information) that may reduce the bias associated with overstated unverified data. This variable identifies loan applications containing credit, employment, income, residence, or "other" information that could not be (or was not) verified by the loan officer. Interestingly, unverified information was not necessarily a fatal derogatory characteristic (see Table 7). However, minorities with unverified information were 70 percent more likely to be denied than whites with unverified information.

---

[30] FDIC examiners found that several institutions reported information from the original (unverified) application, not from the final application, even though the original data are not necessarily used to make the loan decision (Horne 1994).

[31] In many cases, income, assets, and debt obligations are modified during the verification period. This is especially true of income. Expected raises, part-time income, child support, interest income, annual bonus and over-time income, a sudden change in income, etc. that are normally reported at the time of initial application may often be either reduced or not counted (though they may be used as a compensating factor) by the underwriter. Moreover, the non-housing expenses listed on the initial application may be incomplete; they are often adjusted using the information from the applicant's credit report.

Table 7. Loan Files with Unverified Data by Race and Loan Disposition

	Approved	Denied	Percent denied
Non-minority	26	24	48
Minority	9	41	82
Total	35	65	65

The Boston Fed data set identifies four categories of information that could not be verified: credit references, employment, income, and residence. We tried several methods of incorporating the unverified information dummy variable into the Boston Fed model. We found the results to be sensitive to the method used (see Table 8); however, the magnitude and significance of the race coefficent are not affected. Including the unverified information variable increases the explanatory power of the model, suggesting either (i) the probability of denial increases if vital information is not verified by the bank, or (ii) loan officers tend not to verify information if the initial qualifications of the applicant are so poor that approval seems unlikely. It is not clear from the data if either or both of these are driving the result. It is interesting to note, however, that roughly 37 percent of the minority denials have unverified data, as compared to only 21 percent of the majority denials.[32] We return to this issue below.

[32] It is also interesting to note that of the 100 files that are identified as having unverified information, half are minority applicants. That is, only 4.1 percent of the white applications have unverified information, while nearly 14 percent of the minorities do. Moreover, 26 of the 50 white applicants with unverified information (52 percent) were approved, but only 9 of the 50 minorities (18 percent) were.

Table 8. Results of regression model: Unverified Information

Variable	Coefficient (p-value)	
	Model VI	Model VII
Constant	-6.82 (0.0001)	-6.6 (0.0001)
Housing expense ratio	0.41 (0.0804)	0.35 (0.1241)
Total debt ratio	0.042 (0.0001)	0.035 (0.0001)
Unverified Income and "Other" x Debt Ratio		0.073 (0.0001)
Net wealth	0.0001 (0.1723)	0.0001 (0.1534)
Consumer credit history	0.34 (0.0001)	0.33 (0.0001)
Mortgage credit history	0.41 (0.0209)	0.48 (0.0047)
Public record history	1.37 (0.0001)	1.41 (0.0001)
Probability of unemployment	.071 (0.1017)	0.065 (0.1306)
Self-employed	0.47 (0.0867)	0.41 (0.1306)
Loan-to-value ratio	0.46 (0.0136)	0.51 (0.0050)
Denied private mortgage insurance	4.42 (0.0093)	4.43 (0.0001)
Rent/value in tract	0.59 (0.0167)	0.62 (0.0107)
Purchasing 2- to 4-family home	0.22 (0.6508)	0.18 (0.7036)
Race	0.85 (0.0001)	0.90 (0.0001)
Unverified information	2.45 (0.0001)	
Number of observations	1601[a]	1603
Percent correct predictions	91.3	90.6
Hosmer-Lemeshow test (p-value)	0.87	0.62

[a] The unverified information variable was missing for two observations.

Sufficient funds to close. Having sufficient "funds to close" (that is, to cover the downpayment, closing costs, and two months' payments) is a virtually universal underwriting requirement, and lack of sufficient funds is often cited as a reason for denial. The handling of this criterion in the Boston Fed model, however, is problematic, and may not fully capture its impact on the lending decision.

In the Boston Fed model, the downpayment amount is reflected in the loan-to-value ratio, which proves to be a consistently important and significant variable. In addition, the model attempts to capture the sufficiency of funds to close through the use of net wealth and, alternatively, liquid assets. Neither of these latter variables, however, turns out to be significant, a result which, according to the authors, appears "to support lenders' claims that they do not place much weight on wealth . . . . Pre-screening may also exclude people without enough cash to settle" (Munnell et al. 1992, p. 30).

These explanations notwithstanding, it still seems curious that some measure of funds to close does not have significant explanatory power in the final model, in particular in light of the fact that coming up with a downpayment has been found to be the primary obstacle to homeownership for millions of American households (Apgar et al. 1990, Horne 1994). On closer examination, however, the treatment of this issue in the Boston model may be inadequate for several reasons. First, no attempt is made to measure available assets relative to the amount of money needed to close. Second, the various fees and points that must be paid by the buyer at closing are not taken into account; these can often constitute a substantial amount, especially for a marginal applicant or first-time buyer who is struggling to come up with the necessary funds. And finally, the way that equity in a present home and gifts or grants were handled raises questions about the ability of the model to fully capture the "funds to close" standard.

While the amount of equity in a present home should, in principle, be captured in net worth, it is likely that the net worth variable is misleading and inaccurate in this context. Net worth includes, among other things, real estate owned (which may include properties other than the current home), net worth of businesses owned, autos, and furniture and personal property. From the total assets figure it is impossible to distinguish the amount of equity in the current home (which is generally to be sold to move into the new home) from other assets the applicant may neither desire nor be able to liquidate in order to cover the closing costs of the new home.[33]

There are similar questions with regard to gifts or grants used to meet part of the downpayment and closing costs. The data questionnaire in the Boston study included an item, "Did a gift or grant account for any part of the down payment?" and a dummy variable reflecting the answer was included in the model. The estimated coefficient indicated a lower probability of denial for applicants who had received gifts or grants, but it was not significant. To the extent that gifts actually received are accurately reflected in liquid assets, net wealth, downpayment amount, and the loan-to-value ratio, this result, by itself, is not a cause for concern. In fact, it is consistent with lenders' contentions, as cited in the Boston Fed study (Munnell et al. 1992, pp. 14 and 30), that wealth is not a significant factor in their decision-making.

However, it is possible that the net wealth and liquid assets data do not accurately account for gifts received. Often, for example, gifts are received just prior to closing, and the amount of the anticipated gift may not be included in the liquid assets reported on the application form. This problem is even more likely to occur in those cases where data for the study was taken from the initial, rather than the final, application form (Horne 1994, p. 7). FDIC examiners also found instances where gifts were indicated in the loan file, but not verified or received (Horne 1994, p. 11). These findings raise

[33] In addition, the value listed for furniture and personal property is notoriously arbitrary and inflated.

two possibilities: that a gift actually received was not reflected in reported liquid assets, or that a gift included in reported liquid assets was ultimately never received. In either case, the Boston Fed model would fail to accurately capture the effect of the gift, or its absence, and the results of the model may be biased as a consequence.[34]

The inability of the Boston model to accurately measure the true amount of assets available to cover closing costs may explain the insignificant results obtained for net wealth and liquid assets. Further, it means that denials that arise entirely, or in large part, because of insufficient funds may be attributed by the model to other factors. As examples, if minorities are disproportionately less likely to be current homeowners or if, as homeowners, they tend to have less equity in their homes, or if they are disproportionately more likely to report anticipated gifts that fail to materialize, these problems would result in an upwardly biased estimate of the race coefficient.

We attempted to better capture the issue of sufficient funds to close, with a variable representing the ratio of liquid assets to closing costs (defined for this purpose as the downpayment plus two months' proposed housing costs); the results were not significant. However, this may well have been due to the measure used and to the problems with the Boston data. To explore this issue adequately, it would be necessary to collect data on the aspects of financial ability not included in the Boston Fed's questionnaire: verification and actual receipt of gifts or grants, equity in a current home, and the amount of other closing costs (fees, points, etc.)

Model misspecification and omitted variables

The model has also been criticized for failing to include several relevant decision variables that are correlated with race, which implies that the model has omitted variable bias.[35] Of particular interest, given the Boston Fed data, is the inclusion of information on special programs and program guidelines.[36]

---

[34] Without knowing how the occurrence of the two situations cited in the text is correlated with race, it is not possible to speculate, a priori, on the sign or magnitude of the effect that of any such bias might have on the estimated coefficient of the race variable.

[35] This type of misspecification can seriously affect the results of a regression analysis, resulting in biased and inconsistent estimators (even if the omitted variables are uncorrelated with race). This implies that, not only are the magnitudes of the coefficients wrong (i.e., biased), but the usual tests of significance are no longer valid. See Lee (1982) and Kmenta (1986).

[36] We separate the analysis of these two variables from the earlier discussion in Section V, Alternative specifications, because of the measurement problems associated with these variables. The special loan programs series has numerous errors. Adjustable-rate, FHA, VA, conventional, jumbo, balloon, and construction loan are incorrectly identified as special programs by several of the banks. Moreover, loans made to finance the purchase of other real estate owned by the bank (OREO) were also included in this category. Since all of these applications are obviously misclassified as special loan programs, we excluded

Many mortgage lenders are developing special loan programs to assist first time homebuyers and "other" specially qualified applicants (such as individuals who are unable to make a nominal downpayment in spite of strong financial/credit histories); or are participating in programs developed by the Federal National Mortgage Association (Fannie Mae), the Federal Home Loan Mortgage Corporation (Freddie Mac), or state or local government agencies. The presumption is that the underwriting criteria applied to these loan applications are different from those applied to conventional loan applications. The Boston Fed data set pools applications without respect to whether they participate in a special program. This may result in a comparison of loan applications with very different underwriting standards. Accordingly, we would have expected the model to include a special loan program variable which would capture the change in the weights applied to those applications made under a special loans program.

The results of the statistical model may be biased if the special loan program information is ignored, especially if a disproportionate share of one racial group participated in special programs. (This is true with the Boston Fed's sample, where over 50 percent of minority applicants – 51 percent of approvals and 55 percent of the denials – applied under a special loan program, as compared to only about 13 percent of the white applicants – 12.8 percent of approvals and 18.8 percent of denials.) Because the Boston Fed model does not include a special loan program variable, the race coefficient may capture the effect associated with the difference in underwriting standards, not just the difference in treatment due to race.[37]

A special program indicator by itself may not adequately capture the impact of less stringent underwriting standards. Interaction variables are used instead of a single indicator variable to separate out the effects of the special loan programs' underwriting standards on debt ratios, credit history, and loan-to-value ratio. The special loan program variables are, in general, of the right sign and are statistically significant (see Table 9 - Model VIII). All else being equal, a borrower who applies under a special program will tend to have a lower probability of denial. However, the sign on the total

---

them from our sample of applications made under a special loan program.

The program guidelines series has been criticized as inconsistent across institutions by the Boston Fed because the variable was not adequately defined. There may be some merit to their concern. However, although the instructions to the banks were ambiguous, the data suggest that banks did not systematically report that all loans denied failed to meet program guidelines. We suspect most banks cited failure to meet guidelines if the application was denied for credit history problems. Still, it is unknown whether the variable is derived using general (i.e., secondary market) guidelines, bank-specific guidelines, or program-specific guidelines.

[37] The Boston Fed reports (Munnell et al. 1992, pp. 52 and 63) that they tried, unsuccessfully, to incorporate a dummy variable for special programs into their model. However, their lack of success is most likely attributable to the data problems associated with this variable, as discussed in the previous footnote.

debt-to-income interaction variable suggests underwriters are more concerned about a borrower's total debt burden when the borrower is applying under a special program. An applicant applying under a special loan program with high debt levels has a higher probability of denial. (This result is consistent with special loan programs found at many lending institutions today, such as first-time homebuyers programs, that allow very high loan-to-value ratios but generally are strict on debt ratios and recent credit history.)

Though the inclusion of controls for special programs increases the explanatory power of the model relative to the basic Boston specification (see Table 4), the impact on the race variable is small. Moreover, the goodness-of-fit falls dramatically. Though the coefficients of the model are plausible, the low p-value for the Hosmer-Lemeshow test suggests the accuracy of the model is only marginal.

A related model misspecification issue concerns the failure to incorporate information on the credit guidelines used by the bank to evaluate the loan application. Misspecification may exist due to (i) differences in underwriting standards across banks (some banks may be more flexible

Table 9. Results of regression model: Special Loans Programs and Program Guidelines

Variable	Model VIII		Model	
	Coefficient	p-value	Coefficient	p-value
Constant	**-6.88**	0.0001	**-1.00**	0.0413
Ability to support loan				
Housing expense ratio	**0.59**	0.0099	**0.68**	0.0124
Total debt ratio	**0.035**	0.0001	**0.027**	0.0045
Net wealth	**0.00008**	0.2568	**-0.00001**	0.8924
Risk of default				
Consumer credit history	**0.36**	0.0001		
Mortgage credit history	**0.48**	0.0038		
Public record history	**1.25**	0.0001		
Probability of unemployment	**0.06**	0.1235	**0.04**	0.3932
Self-employed	**0.40**	0.1278	**0.66**	0.0301
Potential default loss				
Loan-to-value ratio	**1.00**	0.0070	**0.55**	0.0597
Denied private mortgage insurance	**4.56**	0.0001	**4.41**	0.0001
Rent/value in tract	**0.59**	0.0121	**0.73**	0.0084
Loan characteristics				
Purchasing 2- to 4-family home	**0.12**	0.8109	**0.22**	0.7071
Total debt ratio x spec. pgm.	**0.079**	0.0005	**0.11**	0.0006
Mort. credit history x spec. pgm.	**-0.10**	0.2974	**-0.18**	0.1073
Housing exp. ratio x spec. pgm.	**-0.07**	0.0581	**-0.11**	0.0070
Loan-to-value ratio x spec. pgm.	**-0.97**	0.0373	**-0.48**	0.2145
Personal characteristics				
Race	**0.92**	0.0001	**0.74**	0.0030
Meets guidelines			**-4.12**	0.0001
Number of observations	1603		1601[a]	
Percent correct predictions	89.5		92.9	
Hosmer-Lemeshow test (p-value)	0.15		0.12	

[a] The "meets guidelines?" question was not answered – either yes or no – for two observations.

than others, especially if they anticipate selling the mortgage), (ii) banks rejecting well qualified applicants that are overqualified under special program guidelines (see Horne 1994), and (iii) banks approving marginally to poorly qualified applicants under the less stringent, special program qualifying standards.

Table 10. Results of regression model: Unverified data and special loan programs

Variable	Model X		Model XI	
	Coefficient	p-value	Coefficient	p-value
Constant	**-1.42**	0.0046	**-1.07**	0.0293
Ability to support loan				
Housing expense ratio	**0.59**	0.0375	**0.64**	0.0192
Total debt ratio	**0.029**	0.0028	**0.028**	0.0034
Net wealth	**0.000007**	0.9396	**-0.000009**	0.9192
Risk of default				
Probability of unemployment	**0.046**	0.3532	**0.04**	0.4317
Self-employed	**0.74**	0.0155	**0.707**	0.0193
Potential default loss				
Loan-to-value ratio	**0.54**	0.0716	**0.54**	0.0651
Denied private mortgage insurance	**4.27**	0.0001	**4.34**	0.0001
Rent/value in tract	**0.75**	0.0094	**0.78**	0.0057
Loan characteristics				
Purchasing 2- to 4-family home	**0.23**	0.6847	**0.15**	0.7934
Total debt ratio x spec. pgm.	**0.094**	0.0094	**0.09**	0.0113
Mort. credit history x spec. pgm.	**-0.14**	0.2204	**-0.12**	0.3053
Housing exp. ratio x spec. pgm.	**-0.09**	0.0478	**-0.094**	0.0442
Loan-to-value x spec. pgm.	**-0.65**	0.1232	**-0.69**	0.0987
Personal characteristics				
Race	**0.68**	0.0094	**0.42**	0.1248
Meets guidelines	**-3.91**	0.0001	**-4.03**	0.0001
Unverified information	**2.01**	0.0001		
Unverified info. x minority			**2.52**	0.0001
Number of observations	1601		1601	
Percent correct predictions	93.1		93.1	
Hosmer-Lemeshow test (p-value)	0.026		0.26	

Ideally, loan applications should be matched directly with the guidelines used to evaluate the credit quality of the borrower. Discrimination would exist if the guidelines were applied unfairly across racial groups. Unfortunately, this information was not collected for the Boston Fed study. The Boston Fed did, however, request that the banks indicate if the applicants' credit histories met the bank's loan policy guidelines for approval.[38] We use the "meets guideline" variable to partially address the types of misspecification discussed in (i) and (iii).[39] The results are reported in Table 9 as Model IX. The coefficient on the race variable is significantly reduced but is not eliminated after controlling for credit history guidelines.[40] This model, however, performs poorly.

We combined the effects of unverified information, special programs, and program guidelines and report the results in Table 10. The coefficients for both the unverified information and the meet guidelines variables are statistically significant. However, the model performs poorly as indicated by the very low Hosmer-Lemeshow statistic (see Model X). Modifying the unverified information variable by replacing it with an interaction term between race and unverified information significantly improves the model's goodness-of-fit (see Model XI). More interestingly, the coefficient on the race variable is no longer significantly different from zero. This suggests minority applicants are treated the same as majority applicants, except those with unverified information. The difference in the denial rates is explained by the differences in housing expense ratio, total debt ratio, credit history, loan-to-value ratio, denied PMI, and meets guidelines.

The results of Model XI certainly do not prove that discrimination did not exist in the Boston mortgage market in 1990. The fact that Model XI is the only specification, out of dozens attempted, where race is not significant is probably more important than the lack of significance itself; this is particularly true given the numerous problems, discussed above, with all three of the variables (unverified information, special program, and meets guidelines) used in this model.

Given the problems with the variables, and in the context of all the other specifications where race was so consistently significant and robust, it is probably more accurate to

---

[38] It is important to note that the survey question applied to credit history only, not to all the guidelines for the loan program under which the application was submitted. Moreover, it is unclear exactly what this variable measures, even in the case of credit history. We expect the variable is used in most cases to indicate if the application was denied for credit history problems. It is unknown whether the variable is derived using general (i.e., secondary market) guidelines, bank-specific guidelines, or program-specific guidelines. That is, were all loan applications subject to the same underwriting standards, or were some applications given special consideration under a special loan program?

[39] Applications rejected as overqualified for special programs cannot be identified solely from the data and, therefore, cannot be dealt with in the same manner.

[40] Day and Liebowitz (1993) include the "meets guidelines" variable in their model. They also found it explains a large portion of (but does not eliminate) the difference in denial rates across racial groups. They proceed to partition the sample into finer groups, by property type. Although we believe there may exist differences in underwriting standards by property type, we are concerned that further partitioning the data creates subsamples that no longer represent the population and, therefore, introduce additional selection bias. We believe, however, that additional research is needed to address this issue.

characterize Model XI as demonstrating that it is always possible, with sufficient effort, to find some formulation that eliminates the significance of virtually any variable in virtually any model. Moreover, Model XI may even suggest an alternative hypothesis that discrimination may still exist, but in a different form. That is, given that a large percentage of the denied minority applicants had unverified information in their files, there may be a racial disparity in the "quality of assistance" given to minority applicants. Loan officers may spend more time working with white applicants than with minorities to assure that they submit the documentation necessary to verify all the information in their files. If that is the case, the lack of verified information may be masking this more subtle form of discrimination. In any case, the strongest conclusion that can be drawn from Model XI is that this is an area that needs to be looked into more carefully in future research.

VII. Conclusions

Was the Boston Fed right? Most of the commentary and controversy over the Boston Fed study has focussed on the central conclusion. Was there really significant discrimination against minorities in the mortgage lending market in Boston in 1990? Or, do the data and methodological shortcomings of the study so undermine the validity of the results as to cast doubt on the existence of discrimination?

The results of our analysis of the Boston Fed's methodology and data permit us to answer the first question with only a qualified yes. To the extent that we could correct the most obvious data problems, explore alternative specifications, and subject the model to additional econometric tests – that is, going as far as was possible using the Boston data – the model overall and the race coefficient in particular were remarkably robust. Our results could not refute the findings of the Boston Fed concerning differences in treatment between minorities and whites. The coefficient on the race variable remains statistically significant and of approximately the same magnitude as that found in the original study.

However, as pointed out above, there were numerous data errors that could not be readily corrected and alternative specifications that could not be tested with the available data. Thus, it is not possible to definitively reaffirm, or to repudiate, the results of the Boston Fed study. The only way to do so would be to go back to the original loan files at all 131 participating institutions to verify the original data, correct data errors, and collect data on additional variables (or different forms of some variables). However, given the intense publicity and controversy generated by the release of the Boston Fed study, as well as the federal and state enforcement actions that have ensued as a direct consequence of the study, it is virtually certain that such a follow-up effort will never take place.

Where do we go from here? While setting the record straight once and for all about the

Boston Fed study would certainly be important for its own sake, the impossibility of doing so does not necessarily affect the research agenda in this area. Even if the Boston study were completely beyond reproach in the unanimous opinion of the research community, it still represents only one study, in one city, at one point in time. As such, it calls out for replication in other times and places. We should be careful not to draw conclusions about the existence or pervasiveness of discrimination in mortgage lending until and unless we find consistent and systematic evidence in repeated studies.

Moreover, in the process of carrying out other studies, researchers can extend and modify the Boston Fed methodology in two important respects. First, we can learn from their problems and shortcomings and can make every attempt to avoid the data problems, for example, and to collect data on additional variables. Second, we can explore the possibility of adapting the econometric methodology to requirements of the examination procedures employed by the federal bank regulators, in order to develop a statistical tool that can assist bank examiners in detecting the existence of discrimination in any single lending institution.

These tasks will constitute the principal objectives of the next phase of our research.

Appendix 1. Key characteristics of mortgage applicants, by race and loan disposition
Full Boston sample (3,062 observations) and OCC/FDIC sample (1,603 observations)

Variable	White				Black/Hispanic			
	Approved		Denied		Approved		Denied	
	Boston	OCC/FDIC	Boston	OCC/FDIC	Boston	OCC/FDIC	Boston	OCC/FDIC
Ability to support loan								
Housing expense/income (%)[a]	26.0	26.0	26.6	26.0	26.0	26.0	28.0	29.0
Total debt/income (%)[a]	33.0	32.5	37.0	37.0	34.0	34.0	38.0	38.0
Net wealth ($000)[a]	93	94	75	59	39	37	33	29
Monthly base income (applicant plus coapplicant) ($)[a]	4,666	4,691.5	4,471	4,260	3,333	3,412	3,600	3,254
Liquid assets ($000)[a]	38	40	28	23	19	18.75	15.5	13.85
Risk of default								
Percent with poor credit history[b]	14.6	12.7	38.9	41.9	23.4	21.4	51.5	50.4
Probability of unemployment[a]	3.2	3.2	3.2	3.6	3.2	3.2	3.2	3.2
Percent self-employed	12.0	12.9	22.4	21.4	7.5	7.6	7.4	6.9
Potential default loss								
Loan-to-value (%)[a]	77.3	75.1	83.1	83.5	85.0	84.3	90.0	89.7
Rent/value in tract (%)[a]	4.6	4.6	4.9	5.1	7.3	7.2	8.9	9.1
Percent applied for PMI	21.6	22.2	17.1	23.9	42.2	42.0	26.6	36.5
Percent denied PMI[c]	0.7	0.8	75.0	82.0	1.3	1.8	82.5	73.8

Variable	White				Black/Hispanic			
	Approved		Denied		Approved		Denied	
	Boston	OCC/FDIC	Boston	OCC/FDIC	Boston	OCC/FDIC	Boston	OCC/FDIC
Loan characteristics								
Percent purchasing 2- to 4-family home	7.7	7.8	18.3	19.6	24.8	24.8	34.4	31.3
Percent fixed-rate loans	68.6	60.7	62.8	56.4	60.6	46.6	69.6	63.5
Percent 30-year fixed-rate loans	85.9	87.5	83.3	82.9	91.1	94.3	91.3	95.6
Percent in special loan programs	12.6	12.8	16.1	18.8	40.6	50.0	40.3	54.8
Personal characteristics								
Age[a]	34.0	34.0	35.0	35.0	36.0	36.0	36.0	37.0
Percent married	63.0	61.7	53.2	54.7	53.7	57.2	55.0	50.0
Percent with dependents	37.6	35.5	39.9	42.7	52.6	55.0	52.2	46.1

[a] Median value.

[b] Poor credit history is defined as more than two late mortgage payments or delinquent consumer credit histories (more than 60 days past due) or bankruptcies or other public record defaults.

[c] Based on number applying for private mortgage insurance.

Appendix 2. Variable Definitions

Housing expense ratio	Equals 1 if proposed monthly housing payment to income ratio is greater than .30; otherwise equals 0.
Excess above 28 percent	Equals proposed monthly housing payment to income ratio minus .28 if the result is positive; otherwise equals 0.
Total debt ratio	Equals total monthly debt payment divided by monthly income.
Excess above 36 percent	Equals total debt ratio minus .36 if the result is positive; otherwise equals 0.
Excess above 36 percent squared	Equals total debt ratio minus .36 if the result is positive, squared; otherwise equals 0.
Net wealth	Equals total assets minus total liabilities.
Consumer credit history	Equals 1 if no "slow pay" accounts, 2 if one or two slow pay accounts, 3 if more than two slow pay accounts, 4 if insufficient credit history, 5 if account shows delinquent credit (60 days past due), and 6 if account shows seriously deliquent credit (90 days past due).
Mortgage credit history	Equals 1 if no late payments, 2 if no payment history, 3 if one or two late payments, and 4 if more than two late payments.
Public record history	Equals 1 if any public record of credit history; otherwise equals 0.
Probability of unemployment	Equals the 1989 Massachusetts unemployment rate for the applicant's industry.
Self-employed	Equals 1 if the applicant is self-employed; otherwise equals 0.
Loan-to-value ratio	Equals the loan amount divided by the appraised value.

Excess above 80 percent	Equals the loan-to-value ratio minus .80 if the result is positive; otherwise equals 0.
Excess above 80 percent squared	Equals the loan-to-value ratio minus .80 if the result is positive, squared; otherwise equals 0.
Excess above 90 percent	Equals 1 if loan-to-value ratio is greater than .90; otherwise equals 0.
Denied private mortgage insurance	Equals 1 if the applicants applied for and were denied private mortgage insurance; otherwise equals 0.
Rent/value in tract	Equals rental income divided by value of rental housing stock in the Census tract in which the property is located.
Purchasing 2- to 4-family home	Equals 1 if the applicant is purchasing a two- to four-family home; otherwise equals 0.
Race	Equals 1 if the applicant is Black or Hispanic; otherwise equals 0.
Years of education	Equals the number of years of education for the applicant (if there is a coapplicant, equals the number of years of education for the person with the greater base monthly income from wages.
Over five years in occupation	Equals 1 if the applicant has been in the recorded line of work more than five years; otherwise equals 0.
Percent moved in pre-1985	Equals the number of households in the Census tract that moved in before 1985 divided by the total number of households in that tract that report a date moved in.

Unverified information Income Employment Credit References Residence Other	Equals 1 if reported information could not be verified by the underwriter (an aggregate and individual verificiation series were created for income, employment, credit, residence, and other); otherwise equals 0.
Special program	Equals 1 if the applicant applied under a special loan program; otherwise equals 0.
Meets guidelines	Equals 1 if the applicants' credit history met the lender's loan policy guidelines for approval; otherwise equals 0.

References

Apgar, William C., Denise DiPasquale, Jean Cummings, and Nancy McArdle. 1990. The State of the Nation's Housing. Cambridge, Massachusetts: Joint Center for Housing Studies, Harvard University.

Berkovec, James, Glenn Canner, Stuart Gabriel, and Timothy Hannan. 1993. "Race, Redlining, and Residential Mortgage Defaults: Evidence from the FHA-Insured Single-Family Loan Program." U.S. Department of Housing and Urban Development, Home Mortgage Lending & Discrimination: Research and Enforcement Conference. May 18-19.

Berndt, Ernst R. 1991. The Practice of Econometrics: Classic and Contemporary. Reading, Massachusetts: Addison-Wesley Publishing Company.

Blinder, Alan S. 1973. "Wage Discrimination: Reduced Form and Structural Estimates." The Journal of Human Resources. vol. 18, pp. 436-455.

Browne, Lynn E. 1993. "Boston Fed Study Shows Race Bias." The Wall Street Journal. September 21.

Day, Ted, and S.J. Liebowitz. 1993. "Mortgages, Minorities, and Discrimination." School of Management, University of Texas at Dallas. Unpublished manuscript.

Horne, David K. 1994. "Evaluating the Role of Race in Mortgage Lending." FDIC Banking Review. Forthcoming.

Hosmer, David W., Jr. and Stanley Lemeshow. 1989. Applied Logistic Regression. New York: John Wiley & Sons.

Judge, George G., W.E. Griffith, R. Carter Hill, Helmet Lütkepohl, and Tsoung-Chao Lee. 1990. The Theory and Practice of Econometrics, 2nd ed. New York: John Wiley & Sons.

King, A. Thomas. 1980. "Discrimination in Mortgage Lending: A Study of Three Cities." Federal Home Loan Bank Board, Research Working Paper No. 91. February.

Kmenta, Jan. 1986. Elements of Econometrics. New York: Macmillian Publlishing Company.

Lee, L.F. 1982. "Specification Error in Multinomial Logit Models: Analysis of the Omitted Variable Bias." Journal of Econometrics, vol. 20, pp. 197-209.

Liebowitz, Stan. 1993. "A Study That Deserves No Credit." The Wall Street Journal. September 1.

Maddala, G.S. 1983. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge, England: Cambridge University Press.

Munnell, Alicia H., Lynn E. Browne, James McEneaney, and Geoffrey M.B. Tootell. 1992. "Mortgage Lending in Boston: Interpreting HMDA Data." Federal Reserve Bank of Boston, Working Paper No. 92-7. October.

Perle, Euguene D., Kathryn Lynch, and Jeffrey Horner. 1993. "Model Specification and Local Mortgage Market Behavior." American Real Estate and Urban Economics Association Meeting, Washington, D.C. June 1.

Schafer, Robert and Helen F. Ladd. 1981. Discrimination in Mortgage Lending. MIT-Harvard Joint Center for Urban Studies. Cambridge, Massachusetts: Massachusetts Institute of Technology Press.

Schill, Michael H. and Susan M. Wachter. 1993. "A Tale of Two Cities: Racial and Geographic Disparities in Home Mortgage Lending in Boston and Philadelphia." American Real Estate and Urban Economics Association Meeting, Washington, D.C. June 1.

Siskin, Bernard R. and Leonard A. Cupingood. 1993. "Use of Statistical Models to Provide Statistical Evidence of Discrimination in the Treatment of Mortgage Loan Applicants: A Study of One Lending Institution." U.S. Department of Housing and Urban Development, Home Mortgage Lending & Discrimination: Research and Enforcement Conference. May 18-19.

Van Order, Robert, Ann-Margret Westin, and Peter Zorn. 1993. "Effects of the Racial Composition of Neighborhoods on Default and Implications for Racial Discrimination in Mortgage Markets." Federal Home Loan Mortgage Corporation. March.

Zandi, Mark. 1993. "Boston Fed's Bias Study Was Deeply Flawed." The American Banker. August 19.