Evaluating Statistical Models of Mortgage Lending Discrimination: A Bank-Specific Analysis

Mitchell Stengel and Dennis Glennon

Office of the Comptroller of the Currency
Economic & Policy Analysis

Working Paper 95-3
May 1995

Abstract: The Boston Fed's study on mortgage lending discrimination has prompted a considerable rise in the use of statistical methods to evaluate banks' lending behavior. However, several issues concerning the validity, accuracy, and reliability of statistical models remain unresolved. In this paper, we attempt to address several of those issues within the context of a bank-specific analysis. Using data gathered from three nationally chartered banks, we extend the "standard" model outlined in the Boston Fed study. We conclude that a model specification that incorporates the specific underwriting guidelines of the individual bank is more appropriate than a broadly-defined, generic specification applied across all banks. We also show that the process of incorporating the bank-specific guidelines is itself difficult and can be complicated further by the lack of accurate data upon which to build a valid representative sample. We discuss several additional methodological issues concerning model validity, specification, sample design, and data accuracy that must be addressed if a statistical approach is to become a useful and accurate supervisory tool.

The authors would like to thank Jeffrey Brown, Michael Carhill, Marsha Courchane, Robert DeYoung, and Mark Winer for helpful comments, Melissa Barnhart and Robert Hostler for research support, and Amy Millen for editorial assistance. Any errors are the responsibility of the authors.

Please address questions to the authors at the Economic & Policy Analysis Department, Office of the Comptroller of the Currency, 250 E Street SW, Washington, D.C. 20219, (202) 874-5700.

Questions regarding subscriptions to the OCC Economic & Policy Analysis Working Paper Series should be addressed to Communications Division, Comptroller of the Currency, Washington, DC 20219.

I. Introduction

The Boston Fed's study on mortgage lending discrimination (Munnell et al., 1992) has prompted a considerable rise in the use of statistical methods to evaluate banks' lending behavior. The banking agencies and a growing number of banks are using, developing, or at least considering those models to test for unfair lending practices by their mortgage underwriting departments.

Most of these models employ some variant of the regression methodology used by the Boston Fed. However, several methodological issues concerning the Boston Fed approach are unresolved. Those issues were raised in follow-up studies (Glennon and Stengel, 1994; Horne, 1994; Carr and Megbolugbe, 1993; and Day and Liebowitz, 1993) and in the work reported in this paper. Until they are resolved, the widespread acceptance and application of this approach may not be warranted. These concerns are particularly important when moving from the market-level approach of the Boston Fed to the bank-specific focus of the regulators' and banks' models.

In this study, we discuss issues of model validity, specification, and accuracy, and of sample design and data collection, all of which may affect the usefulness and accuracy of statistically-based methods of identifying lending discrimination. We also explain the specific conditions under which we believe the single-equation methodology employed by the Boston Fed is valid.

We address those issues in the context of the results of our efforts to develop bank-specific models for three nationally chartered banks. We start with a specification that closely resembles that developed by the Boston Fed to analyze each bank's underwriting decision process. We discuss several of the still-unresolved issues that arise when moving from a market-level to a bank-specific analysis and other modifications and reformulations we introduced into the models to better capture the mortgage decision-making process.

Our results show that, by incorporating the specific underwriting guidelines of the individual bank and by making other changes in the model specification, the alternative, bank-specific approach significantly improves the ability of the model to explain the outcomes of the mortgage lending decision process. That approach uniformly outperforms a more broadly defined, generic specification typified by the Boston Fed model (hereafter referred to as the "market-level" model) in terms of the significance of most individual variables, the overall goodness-of-fit of the model, and tests for the accuracy of the model specification.

With regard to the ability of the models to identify patterns of differences in treatment by race, however, the results of our comparison of model specifications using data from the three separate banks are mixed. In one bank, the market-level model erroneously attributes to the race of the applicant differences in outcomes that can be explained by other factors in a more bank-specific model. At the second bank, the two approaches are generally in agreement; the market-level model suggests statistically significant racial differences in outcomes for two of three minority groups, and our alternative model specification finds differences in treatment for all three groups. However, this result was not supported by examiner reviews either of a large sample of applications at the same bank or of a smaller number of specific denied minority files identified by the regression model.

At the third bank, the results of regressions under both approaches suggest no significant difference in the treatment of one of the two minority groups studied. However, for the second minority group at this bank, the hypothesis of no difference in treatment was quite sensitive to alternative model specifications. We suspect that these results are due to anomalies in the sample data and are not necessarily the result of unfair treatment. Moreover, the lack of robustness to model specification led us to uncover sample-design problems inherent in a design based solely on the variables reported in HMDA -- problems that may undermine the validity of the model's results.

The evaluation of the accuracy and validity of the statistical results involved comparing them with those of an examiner-based file review at each bank. The experimental nature of our research to date complicates the assessment of the sensitivity of our results to the possibility of sampling and model specification flaws without a comparison using an alternative methodology. We believe the results of the statistical model are accurate only insofar as the sample reflects the true population distribution of the portfolio and the specification of the model captures the salient factors used by the bank in formulating its underwriting decisions. However, these conditions may not be achieved fully because of errors in the analysis.[1] A comparison with the results of other methodologies allows us to examine the accuracy of the model independent of the modeling technique.

For this reason, we argue that, at least at this stage of our research, a statistical model should be used only to *suggest* that an institution may use race as a decision variable. That is, a statistical model should be used only as a screening device and its results should be verified by a judgmental review of selected loan files. Although such a review is imperfect and may suffer from other methodological deficiencies, the results of the two approaches must be compared to develop a useful statistical model and supervisory tool.

The OCC conducted judgmental (comparative file review) fair lending examinations at each bank in which the statistical model employed in this study was tested. The results helped to evaluate the accuracy of the statistical model. In addition, comparison of the judgmental and statistical results suggested areas that require additional research.

---

[1] For example, (i) loan applications that are evaluated using different underwriting guidelines (e.g., conforming and non-conforming loans; or standard loan products and special programs for low-income borrowers) may be combined into a single sample; or (ii) the major determinants of the loan decision may be mis-specified, which can result in omitted variables or errors-in-variables (parameter) bias.

The remainder of the paper is structured as follows: Section II discusses briefly the circumstances under which the Boston Fed methodology may be appropriate. Section III describes the sampling methodology and data collection procedures used in the present study. The market-level model specifications and regression results are presented in Section IV. Section V presents the alternative, bank-specific model and discusses the results obtained and the issues encountered in applying this model to the three national banks. Section VI examines questions for further research and issues about the usefulness of statistical models as tools for bank regulators in the fair lending area.

II. Using statistical models to identify unfair lending behavior

As several different types of discriminatory practices can occur at different stages in the mortgage application and lending process, we must first delineate the specific focus of this study. In this paper, we are concerned solely with the particular type of post-application discrimination known as disparate treatment.

Post-application discrimination refers to unfair practices in the development or application of underwriting standards based on the race or gender of the applicant(s). This differs from other forms of discrimination not included in the post-application category. For example, a lender may refuse to lend against property located in targeted areas (identified by the racial or ethnic composition of the population), a practice commonly known as redlining (i.e., geographic discrimination). Alternately, at the time of first contact, lenders may discourage potential borrowers from submitting applications or steer them to other institutions, based solely on the applicant's race. That practice is known as pre-application discrimination.

In post-application discrimination, disparate treatment must be distinguished from disparate impact. The former refers to the uneven or unfair application of lending standards to different groups of applicants. For example, a bank might routinely require an unmarried woman to obtain a co-signer on a loan application, but not require one for an unmarried man; or, a lender may systematically include non-wage income in the calculated qualifying income for white applicants, but exclude it, or consider it only as a compensating factor, for minority applicants. Disparate impact refers to the use of a lending standard which, even if applied without explicit consideration of group membership, would have a disproportionately adverse effect on one or more particular groups of applicants, such as racial minorities. For example, a bank might impose a minimum loan amount restriction that discourages all requests for loans below the stated minimum, irrespective of the applicant's race. Such a minimum loan amount rule is likely to have a disproportionately adverse effect on minority applicants. (For a more detailed discussion, see HUD et al., 1994; Rachlis and Yezer, 1993.)

Because different types of discrimination can occur at the various stages in the mortgage application and lending process, different methodologies are required to test for the specific types of unfair lending practices. For example, a type of market-share analysis or evaluation of an institution's attempts to solicit business from specific target areas might be used to identify banks that practice redlining of predominately minority census tracts (Avery and Buynak, 1981; Benston and Horsky, 1991; Bradbury et al., 1989; Department of Justice, 1993; Schafer and Ladd, 1981); matched-pairs testing programs could be used to identify banks that practice pre-application exclusion or steering of minority applicants (Galster, 1992; Lawton, 1993; Leeds, 1993; Smith and Cloud, 1993).

Since a fundamental difference exists in what constitutes discrimination in the two post-application categories, disparate treatment and disparate impact, different statistical techniques should be used when testing for them. To illustrate, a bank's underwriting guidelines could be applied consistently across racial or gender groups, but could have been designed in a way that excludes minorities and/or women applicants. A test designed to capture disparate treatment could fail to recognize the embedded difference in impact. More importantly for this analysis, the development of a single statistical model to test simultaneously for both disparate treatment and disparate impact introduces serious econometric problems that may influence the credibility of the estimates.[2] However, given the institutional framework of most banks, underwriters generally evaluate applicants within the context of their pre-determined underwriting guidelines. In this case, a single-equation model like the one employed in the present study is a valid technique for evaluating the consistency of the application of the stated guidelines.[3]

[2] A full discussion of these issues is outside the scope of this paper. See Yezer, et al. (1994) and Rachlis and Yezer (1993) for more detailed discussions of the effect of selection and simultaneity biases on the estimated parameters.

[3] Rachlis and Yezer (1993) and Yezer, Phillips, and Trost (1994) show that, under certain assumptions about the decision behavior of lenders and borrowers, single-equation statistical models introduce simultaneity and selectivity problems that lead to bias and inconsistent parameter estimates. It is likely that these conditions are relevant when studying the issue of disparate impact, since in that case the model must determine the (true) relationship between the borrower's and lender's behaviors and the profitability of making a loan (usually associated with the likelihood of default). Intuitively, Yezer et al. argue that the model must determine the optimal underwriting guidelines that can differentiate poor from better credit-risk applicants based on non-race-based information. The simultaneity and selectivity problems are of less concern when studying the more narrowly defined (and more straightforward) issue of the systematic application of *pre*-determined underwriting guidelines -- since the issue here is not whether the guidelines are the appropriate (or optimal) standards upon which the approve/deny decision is determined. Instead, the model need only determine whether the bank's lending standards (as stated in the underwriting policy guidelines and confirmed through underwriter interviews) are applied fairly and evenly to all applicants.

III. Data and sampling methodology

The data used in the present study were gathered by OCC staff from three nationally chartered banks.[4] Stratified random samples of one-to-four-family, conventional, non-purchased home purchase mortgage loan applications were drawn from the 1993 HMDA Loan Application Register (LAR) of each of the three banks.[5] The samples were drawn from their respective populations, stratified by both race and disposition. The size of each random sample was derived from the distributions of loan amounts (as reported in the HMDA-LAR). The selection criteria were set so as to generate a representative sample of loan applications with a mean loan amount within (plus or minus) 10 percent of the population mean (i.e., 10 percent tolerance) with a 95 percent probability (i.e., 5 percent precision).

The minority applicants consisted of all applications in which at least one of the applicants (either the applicant or co-applicant) was Native American, Asian, black, or Hispanic.[6] We oversampled all race/disposition categories, except white approvals, to avoid small sample problems in low-outcome categories. All minority denied files were included in the sample (a 100 percent coverage rate) at two banks; 50 percent of minority denials were included at the third bank. Coverage rates for minority approvals and for whites varied by bank and were sensitive to the distribution of within-category loan amounts.

The regression sample size for each bank is reported in the tables in Sections IV and V. Some data attrition occurred at each bank because of mis-classified files and missing files or data points. In addition, unusual or idiosyncratic information affecting the disposition of the application was found in several cases. These include applications in which the borrowers were overqualified under the guidelines of the special loans programs under which they applied; files in which the applicants rejected the bank's counteroffer; and files in which the applicants appeared qualified, but either failed to provide required documentation or experienced construction delays that resulted in the loan offer expiring before completion of a new home. These observations proved problematic, because their inclusion in the regression may mask or distort the true underlying relationship between the bank's application of its policy guidelines across racial groups and the disposition of the loan applications. We address this issue in more detail in the following discussion of the results.

---

[4] The only criterion for the selection of those banks was that they had a sufficient number of denied minority applicants to allow for reasonable sample sizes and that they were scheduled for a fair lending exam during the period of analysis. There was no attempt to select the banks based on any *a priori* beliefs about their mortgage lending practices or their treatment of minorities.

[5] See Snedecor and Cochran (1989), and Jaeger (1984) for detailed discussions of the methods used to construct a stratified random sample.

[6] Not all banks received a sufficient number of applications from all four protected classes. Only those minority classes with sufficient observations were included in the population from which the sample was drawn at each bank.

Together with our research on developing statistical models for examination purposes, the white approved and minority denied files used in our model were also evaluated by examiners using the OCC's non-statistical, comparative file procedures.[7] The results of the two methods of analysis at each bank were compared; we explored further cases in which opposite conclusions were reached about treatment by race to understand the problems of developing a valid statistical model of loan outcomes.

IV. Testing for disparate treatment: the market-level approach

The statistical technique chosen is a standard logit model, commonly used to study the class of discrete choice problems, such as the mortgage application approve/deny decision.[8] To test model specification, we begin by following the procedures outlined in the market-level model of Munnell et al. (1992) and estimate the probability of denial for each mortgage loan application as a function of the broadly defined categories of decision variables: (i) ability to support the loan; (ii) risk of default; (iii) potential default loss; (iv) loan characteristics; and, (v) personal characteristics. A dummy variable representing the race of the applicant was included as a personal characteristic to test for discrimination. That is, the race variable addresses the question of whether, after controlling for the relevant financial and other characteristics, any unexplained difference remained in denial rates that was correlated with the applicant's race. Using this methodology, a positive and significant estimated coefficient for the race variable means that the null hypothesis of no adverse difference in the treatment of minorities can be rejected, suggesting the possibility of discriminatory treatment against minorities in the mortgage application process. We use this approach to test both market-level and bank-specific statistical models.

We begin our modeling exercise by first testing the feasibility of using a "generic" or market-level model. Our objective is to test the hypothesis that a generic model, based on the standard underwriting criteria discussed in the Boston Fed study, can be developed and applied across all banks. The hypothesis is based on the underlying premise that a generic model will be reasonably accurate as a first approximation of an individual bank's behavior. This approach, however, implicitly assumes that the secondary market underwriting guidelines are adopted uniformly, with little or no variability, as the policy guidelines of all banks in the market.[9]

---

[7] Examiners draw samples of denied minorities and approved whites and review each file in detail. Particular attention is focused on the reasons for denial listed by the lender. For example, if a particular minority applicant was denied because of high debt ratios, the examiners will search for approved whites with equally high -- or higher -- debt ratios and look for strengths or compensating factors in the white approved files and/or weaknesses in the denied minority file that could plausibly justify the different outcomes. For more detail, see Office of the Comptroller of the Currency (1993).

[8] The two outcomes in the logit model were approval and denial; withdrawals were excluded from the sample by design.

[9] If this hypothesis is correct, it could be used to support suggestions that the mandatory HMDA data be augmented to include a standard set of "key" variables that reflect the (secondary market) *de facto* underwriting standards.

To test this hypothesis, we begin by defining the decision variables as closely as possible to those used in the Boston Fed study (the variables are defined in Munnell et al. (1992)). We could not replicate the model specification completely for any of the three banks, because of data limitations. For example, we found that: (i) the number of applicants denied private mortgage insurance was too small for some banks to include a separate variable indicating PMI rejection; (ii) the census tract identifiers were missing for a substantial percentage of the applicants, making it impractical to include a rent-to-value (within the census tract) ratio as an explanatory variable; and (iii) it was not possible to construct the variable representing the applicant's probability of unemployment, since we could not accurately identify, from the information provided on the mortgage application, the SIC (standard industrial classification) codes for the industries in which most of the applicants were employed.

The results of applying the market-level model specification separately to each of the three banks are presented in Table 1.[10] These results (see Market-Level Model I, Banks A-C) show that the hypothesis of no difference in the treatment of minorities can be rejected for two of the three banks -- Banks A and B -- using this specification. Bank A received a sufficient number of applications (for statistical purposes) from only one minority group. The coefficient on the single minority variable is positive and significant at roughly the one-half of 1.0 percent level. Bank B received a sufficient number of applications for three of the four minority categories identified under HMDA. The coefficients on two of the three minority variables (Minority I and Minority III) are also positive and significantly different from zero at the one-half of 1.0 percent level or better. These results for Banks A and B suggest that minorities faced a higher probability of denial than white applicants who are similar in all other respects. In contrast, the results for Bank C show that the hypothesis of no difference in treatment of minorities could not be rejected at the 10 percent level for either of the two minority groups included in the sample. In short, these results suggest disparate treatment against minorities may have existed at both Bank A and Bank B.[11]

Table 1: Market-Level Model Specification[a]
Unrestricted Data Sets

[10] Institutions will be referred to as Bank A, B, and C to protect their identity and the confidentiality of the data. For the same reasons, the minority groups will be referred to as Minority I, Minority II, and Minority III.

[11] For an idea of the magnitude of differential treatment by race implied by these regression results, consider, for example, a white applicant at Bank A with an estimated probability of denial of 0.25 (approximately one-quarter of the white applicants in the Bank A sample were rejected). The same applicant, if treated as a minority (that is, if the Minority I dummy variable were set equal to 1), would now have an estimated probability of denial of 0.38 according to Market-Level Model I, or an increase in the estimated probability of denial of 50 percent.

For Bank B, the estimated coefficients imply that, if a white applicant had an initial estimated probability of denial of 0.33 (approximately one-third of the white applicants in the Bank B sample were rejected), the estimated probability would increase to 0.507 or 0.504 if the same applicant were treated as a member of minority group I or III, respectively. These levels represent increases in the probability of denial of approximately 52 percent.

	Market-Level Model I Bank A		Market-Level Model I Bank B		Market-Level Model I Bank C	
Variables	Estimated coefficient	p-value	Estimated coefficient	p-value	Estimated coefficient	p-value
Intercept	**-6.882**	.0001	**-4.670**	.0001	**-5.907**	.0001
Hsg exp > .30	**0.548**	.1530	**0.209**	.2887	**0.965**	.0002
Debt ratio	**0.024**	.0075	**4.903**	.0001	**0.009**	.1861
Net wealth (000)	**-.0002**	.6347	**-.00006**	.8251	**-.00006**	.9689
Consumer credit	**0.269**	.0388	**0.098**	.0642	**0.289**	.0070
Mortgage credit	**0.105**	.8047	**0.222**	.1577	**0.027**	.8994
Public record	**1.411**	.0001	**0.767**	.0002	**2.006**	.0001
LTV	**4.619**	.0001	**1.789**	.0047	**0.044**	.0001
PMI rejected			**0.831**	.2597	**1.519**	.0909
Self-employed	**-0.192**	.6828	**0.565**	.0278	**0.841**	.0177
2-4 Family	**0.340**	.4907	**-1.266**	.0745	**0.746**	.0327
Minority I	**0.590**	.0062	**0.723**	.0029	**0.309**	.2970
Minority II			**0.410**	.2306	**0.088**	.7617
Minority III			**0.711**	.0008		
Number of obs.	766		729		552	
-2 Log L	684.0	.0001	817.2	.0001	518.4	.0001
Hosmer-Lemeshow		.1168		.2412		.1274
% correct (prob= 0.5)	77.8		69.3		78.4	

[a] See Munnell, et al. (1992) for a detailed discussion of the definitions of the variables used in these models.

Table 1 (cont.): Market-Level Model Specification Restricted Data Sets[b]			
	Market-Level Model II Bank A	Market-Level Model II Bank B	Market-Level Model II Bank C

Variables	Estimated coefficient	p-value	Estimated coefficient	p-value	Estimated coefficient	p-value
Intercept	**-6.725**	.0001	**-4.828**	.0001	**-6.932**	.0001
Hsg exp > .30	**0.588**	.1239	**0.204**	.3133	**0.515**	.0636
Debt ratio	**0.023**	.0085	**5.238**	.0001	**0.045**	.0001
Net wealth (000)	**-.0002**	.6396	**-.0002**	.6202	**.00004**	.8217
Consumer credit	**0.263**	.0445	**0.121**	.0255	**0.235**	.0332
Mortgage credit	**0.029**	.9471	**0.267**	.1043	**-0.087**	.6990
Public record	**1.397**	.0001	**0.770**	.0003	**2.032**	.0001
LTV	**4.422**	.0001	**1.624**	.0119	**0.048**	.0001
Self-employed	**-0.158**	.7363	**0.510**	.0556	**0.722**	.0601
2-4 Family	**0.337**	.5030	**-1.205**	.1366	**0.606**	.0968
Minority I	**0.575**	.0089	**0.605**	.0161	**0.232**	.4483
Minority II			**0.596**	.0982	**-0.016**	.9584
Minority III			**0.688**	.0015		
Number of obs.	754		698		534	
-2 Log L	666.7	.0001	769.8	.0001	483.6	.0001
Hosmer-Lemeshow		.2465		. 4004		.0054
% correct (prob= 0.5)	79.4		71.1		80.0	

[b] The data used in Model II have been modified as follows: Files in which the applicant rejected the bank's counteroffer and files denied PMI were deleted from the full data sets in Model I; for Banks A and B, files in which the applicants were overqualified for the special loans programs under which they applied were also deleted. (There were no such overqualified files for Bank C.)

Generally, the market-level models perform poorly as predictors of denial probabilities in these single-bank regressions. A large percentage of the explanatory variables in each model are statistically insignificant. The Hosmer-Lemeshow goodness-of-fit statistics are marginal for Banks

A and C.[12] If the data sets are adjusted (see Table 1, Market-Level Model II, Banks A-C) by deleting a number of denied files with certain ambiguous characteristics (i.e., applicants that either rejected the bank's counteroffer, were overqualified for the special loans programs under which they applied, or were rejected for private mortgage insurance), the goodness-of-fit improves for Banks A and B, even for those two banks, however, the variables included in the model still perform poorly as individual explanatory variables (over half of the explanatory variables -- excluding the constant term and the race variables -- are insignificant). This suggests that the race variables may capture inappropriately some of the variation in the probability of denial through their correlation with the other explanatory variables, both included and, possibly, excluded.

---

[12] There is no single, universally accepted measure of goodness-of-fit for logit regression comparable to R^2 for ordinary least squares. Of the many different measures found in the applied literature, such as the McFadden R^2, the percent correct predictions, and the percent concordant/discordant, we find the Hosmer-Lemeshow goodness-of-fit statistic (Hosmer and Lemeshow, 1989, pp. 140-145) to be most intuitively appealing and the best suited to the probabilistic nature of the logit model. Their test statistic is derived by "calculating the Pearson chi-square statistic from the 2 by g table of observed and estimated expected frequencies" (p. 141), where g is the number of groups. We calculated the Hosmer-Lemeshow test statistic for each market-level model; the corresponding p-values are presented in Table 1. 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.

The intuition underlying this statistic is as follows: Since a logit model is not intended to predict correctly 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 that some applications with very low predicted probabilities of denial will be denied and some with very high probabilities will be approved. For example, if the probability of denial estimated by an accurately specified logit 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, a prediction-based measure of goodness-of-fit (such as a percent correct prediction, or percent concordant/discordant) 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 the Hosmer-Lemeshow statistic is a more accurate method of assessing the goodness-of-fit of a logit model.

Although no absolute standards exist for this test statistic, higher p-values indicate better goodness-of-fit for a particular model specification. The test statistics for the market-level models in Table 1 indicate that the models' ability to replicate the mortgage lending decision could be improved.

Although we believe that the Hosmer-Lemeshow measure is the most meaningful and useful test statistic, we also report two other goodness-of-fit measures: (i) the percent correct predictions (the number of applicants with estimated probabilities of denial greater than or equal to 50 percent who were denied, plus the number with estimated probabilities less than 50 percent who were approved, as a percentage of the total sample); and (ii) -2 Log Likelihood. Both of the alternative measures are consistent with the results indicated by the Hosmer-Lemeshow statistic.

In the next section we develop an alternative specification that improves on the performance of the market-level model by addressing some of the bank-specific omitted-variable problems. Our results show that, when moving from a market to a bank-specific model, one must incorporate the individual bank's underwriting policy guidelines. We also demonstrate, however, that the process is often difficult and can be complicated further by the lack of accurate data upon which to build a valid, representative sample of the bank's portfolio of loan applications.

V. The bank-specific model

The alternative modeling methodology presented in this section resembles that outlined previously. However, its purpose is to test for consistency, or equality, of treatment across racial groups by an *individual* bank. Since the lending standards are developed and applied by individual banks, the model must be designed to test explicitly each bank's application of its own stated guidelines, instead of a more generic set that reflects secondary market requirements. We found considerable differences in the underwriting policies and guidelines of the three banks included in the present study. Therefore, we modified the specification of the regression model for each to incorporate its stated policy guidelines from the time the sample applications were drawn.

Information on the detailed, bank-specific underwriting guidelines in effect during the sample period was gathered from interviews and discussions with bank underwriters and other personnel, from bank mortgage lending manuals and other internal documents from that period, and from discussions with OCC examiners familiar with each bank's mortgage lending operations. This information was used to construct the independent variables in the statistical model. The major categories of bank-specific underwriting guidelines incorporated into the model specifications are summarized in Table 2.

Table 2: A Summary of Bank-specific Policy Guidelines[a]

1. Loan-to-value (LTV)
    - PMI generally required if LTV exceeds 80%.
        - LTV threshold may vary for high-income applicants, non-conforming loans, or special (low- and moderate-income) program loans where bank may self-insure.

2. Housing expense-to-income ratio
    - Most banks use secondary market threshold of 28% as policy guideline.
    - Limit may go as high as 33-35% -- or occasionally even higher -- for special loan programs and/or low LTV loans.

3. Total monthly debt-to-income ratio
    - Most banks use secondary market threshold of 36%.
    - Most banks put more weight on total debt ratio than housing expense ratio.
        - Higher total debt ratio thresholds (up to and even exceeding 40%) may be acceptable for special programs, high income applicants, and/or low LTV loans.

4. Funds to close
    - Cash reserves equal to two months' housing expense generally are required.
        - May be required only for LTV above specified threshold value.
        - May be waived for special programs.
        - Larger reserves may be required for self-employed and/or high LTV loans.

5. Credit history
    - Credit history over most recent 18-24 months given greatest weight.
        - One or two late payments in most recent period may be overlooked if no other derogatories in credit history.
            - Some banks are more lenient with medical late payments and certain retailers who are known for excessive late payment reports.
        - Minor derogatory incidents (generally defined as late payments of 30 or 60 days) more than two (or three) years old generally are ignored.
        - Special loan programs may be less flexible and often do not discount late payments as readily.
    - Bankruptcies more than two years old often are not considered as fatal if unblemished credit record has been reestablished over the most recent 24 months.
    - Consumer judgments, collections, and losses often are not fatal if paid off in full or currently being paid as agreed.

6. Compensating factors that can offset other negatives in file
    - Reduction in monthly housing expense.
    - Large cash reserves beyond minimum requirement.
    - Low LTV.

7. Income and employment stability
    - Self-employed often must have been in business at least two years.
    - All applicants generally must have at least two years in same occupation (except new graduates).

8. Banking relationship
    - Significant pre-existing banking relationship:
        - Private banking customer.
        - Large depositor/investor (but not retail checking/savings or small business).
        - Large business account.
    - Bank employee.

[a] The information in this table is illustrative and does not necessarily apply to all three banks. It is presented without reference to the specific banks included in the present study to ensure their anonymity.

V.a. Alternative model design

Two major differences emerge between the alternative and the "generic" specifications. First, we include variables specific to each bank's guidelines omitted from the earlier specification, and second, we introduce non-linearity and discontinuities that characterize the relationship between some of the decision variables and the likelihood of denial.

With respect to the omitted variable issue, we found some degree of commonality in the choice of decision variables in the three banks. For example, loan-to-value, housing expense and debt ratios, and credit history are used in similar forms in those banks. However, we also found that each bank maintains an array of bank-specific decision variables considered fundamental to its decision, but not considered -- or at least not in the same manner -- by the other banks. Examples would include special loan programs, such as those designed to promote mortgage loans to low- and moderate-income households, private banking relationships, and employee loans -- under all of which the normal underwriting standards are modified or suspended. Moreover, a bank may have specific policies and procedures under which an applicant qualifies for more favorable terms. Not only will these programs vary across banks, but a single bank may have several such programs with various levels of exceptions.

One factor that was treated differently by the three banks -- and that was omitted entirely from the generic model -- is a direct measure of the applicant's ability to cover the closing costs of a home purchase.[13] The ability to pay closing costs is an important factor in almost all lenders' underwriting standards, and the inability to do so is often listed by lenders in the HMDA data as a reason for denial. The market-level model includes net wealth as a proxy for this measure. Net wealth, however, may be grossly inaccurate for this purpose. Applicants have an incentive to overestimate their assets, especially personal property, and underestimate the value of their liabilities. Moreover, the importance of net worth can be distorted by the inclusion of assets the applicant has neither the intention nor the ability to liquidate.[14] In the present study, we attempt to rectify this deficiency by collecting detailed data on both the closing costs on the purchase of a home and on the assets available to meet them. This allows us to construct several alternative variables that measure the applicants' ability to cover all expenses associated with the home purchase transaction.

---

[13] Closing costs consist of the downpayment and various taxes, escrows, and fees. In addition, lenders often require that applicants have a cushion of liquid assets sufficient to cover a number (typically about two) of months' mortgage payments.

[14] Liquid assets are sometime used in place of net wealth. However, liquid assets suffer from the same measurement problems as net wealth. Even if these problems could be solved, the use of liquid assets provides information only on the amount of funds available; it completely ignores the amount of funds needed to close. There is no intuitive reason why an applicant with more liquid assets should be better qualified, unless the amount of *net* funds is also larger.

In addition, we concentrated on threshold values for such factors as the loan-to-value and debt ratios when developing model specifications that incorporate bank-specific underwriting guidelines. As we argue in Glennon and Stengel (1994), the decision to approve or deny will be more sensitive to the relationship between the reported and the threshold values for a particular variable than to the reported value itself.[15] Moreover, we would expect the likelihood of denial to increase the larger the excess above the threshold value. This suggests that the relationship between the size of many of the decision variables and the probability of denial is not a simple linear -- or even necessarily a continuous -- one, and that the more complex relationship should be incorporated explicitly into the regression model, using the specific bank policy guidelines. The non-linearity of the relationship between the decision variables and the probability of denial has been used in other studies of individual bank behavior (see, for example, Siskin and Cupingood, 1993). We expand on this concept, using the specific threshold values obtained from each bank's underwriting policy guidelines.

V.b. Bank-specific models: results and evaluation

The specifications of the individual bank models were modified, as discussed in the preceding section, to introduce bank-specific guidelines, several alternative variables to measure the ability to cover closing costs, and threshold effects and non-linearities of several variables.

There is a significant increase in the explanatory power of all three bank-specific models relative to the market-wide specification, as reflected in the decrease in the log likelihood measure and the increased p-values of the Hosmer-Lemeshow goodness-of-fit statistics. All of the individual, non-race (minority) explanatory variables have the correct sign and virtually all are significant.

In addition to the signs of the coefficients and the tests for significance and goodness-of-fit, we also perform more formal comparative tests of the models' explanatory power. The standard methods used to compare model specifications (e.g., a comparison of log likelihood statistics) are inappropriate here, because the bank-specific and market-level specifications for each bank are non-nested; that is, the bank-specific specification includes some variables that are absent from

[15] For example, we would expect to find a strong and direct relationship between the housing expense ratio and the probability of denial. However, it is unlikely that two applicants with similar credit history, employment stability, and wealth variables would be evaluated differently if they had different housing expense ratios both below the secondary market threshold values. That is, everything else the same, an applicant with a housing expense ratio of 25 percent would not be evaluated much differently than an applicant with a 20 percent housing expense ratio. In contrast, a third applicant with similar characteristics, but a housing expense ratio in excess of 30 percent or 35 percent (i.e., above the secondary market threshold value) is likely to be evaluated more harshly. In addition, the LTV may significantly affect the decision only above certain levels; applications with very high LTV's may be automatically disqualified regardless of the overall strength of the file. Certainly a change in the LTV from 80 percent to 85 percent will exert much more influence on the outcome of the underwriting process than a change from 50 percent to 55 percent.

the market-level specification and vice versa. However, the relative performance of the non-nested specifications may be tested using an artificial regression procedure for limited dependent variable models proposed by Davidson and MacKinnon (1984, 1993).[16] When this procedure is applied, the bank-specific models consistently outperform the market-level specification.

With regard to the race variables, some of the estimated coefficients are significant and positive, indicating possible disparate treatment. In some cases, these results are consistent with those derived from the market-level specification; in other cases, the two approaches diverge.

Bank A. As shown in Table 3, the explanatory power of the models using the bank-specific approach increased significantly as compared with the earlier models. Both the Hosmer-Lemeshow statistics and the log likelihood measures suggest that the bank-specific models fit the data well and have greater explanatory power than the market-level model. Moreover, the individual explanatory variables, including those measuring insufficient funds to close, the threshold effects (excess debt ratio and excess LTV), and other bank-specific standards (e.g., decline in monthly housing expense and banking relationship) are generally significant and of the correct sign.

[16] See Appendix for a more detailed discussion of this approach.

Table 3: Regression Results for Bank A [a]				
	Bank-specific Model I.A		Bank-specific Model II.A[b]	
Variables	Estimated coefficient	p-value	Estimated coefficient	p-value
Intercept	**-3.116**	.0001	**-3.155**	.0001
Excess debt ratio	**0.106**	.0001	**0.106**	.0001
Excess LTV (adj PMI)	**19.943**	.0001	**19.430**	.0001
Consumer credit	**0.028**	.0385	**0.030**	.0275
Public record	**1.145**	.0001	**1.144**	.0001
Insufficient funds	**0.935**	.0001	**1.002**	.0004
Banking relationship	**-1.176**	.1036	**-1.575**	.0613
% decline in monthly housing expense	**-1.198**	.0647	**-1.296**	.0455
Minority	**0.473**	.0791	**0.445**	.1067
Number of obs.	766		754	
-2 Log L	492.0	.0001	477.3	.0001
Hosmer-Lemeshow		.3609		.1786
% correct (prob= 0.5)	88.9		89.3	

[a] The variables in the OCC models are defined as follows: Excess debt ratio is the maximum of the difference between the debt-to-income ratio and the threshold value of 36 percent and zero. Excess LTV (adj PMI) is the maximum of the difference between the LTV and the threshold value (80 percent, 75 percent, or 70 percent, depending on the loan size) and zero (value set to zero if applicant obtained PMI). Consumer credit is the number of derogatory entries in the credit report over the past 24 months (fewer than two derogatory entries of less than 30-60 days in the previous 12 months or four in the previous 24 months are ignored). Public record is a dummy variable that equals 1 if there are any judgments, collections, losses, bankruptcies or foreclosures in the applicant' s credit history, and zero otherwise. Insufficient funds is a dummy variable that equals 1 if the sum of liquid assets, equity in current house, gifts and grants, and other financing is less than the sum of downpayment, points and fees, and required monthly reserves, 0 otherwise. Banking relationship is a dummy variable that equals 1 if the applicant is a bank employee or a private banking customer. % decline in monthly housing expense is the minimum of the relative change in monthly housing expense (proposed minus current divided by current) and zero.

[b] For Model II, files in which the applicant rejected the bank' s counteroffer and files in which the applicants were overqualified for the special loans programs under which they applied were dropped from the full data set used in Model I.

With regard to racial differentials, incorporation of bank-specific information and the introduction of threshold effects into the model for Bank A has reduced the significance of the relationship that appeared in the market-level specification between the likelihood of denial and the race of the applicant. Under the bank-specific model specification, we fail to reject the hypothesis of no difference in the treatment of minorities (at the 8 percent level of significance for Model I.A and at the 11 percent level for Model II.A). This conclusion is robust over several alternative methods of including the bank-specific information. Moreover, in Table 3.a, the results of the artificial regressions show that for both data sets (Model I.A and II.A), the bank-specific model ($F_2(x_s\beta_s)$) dominates the market-level model ($F_1(x_m\beta_m)$).

Table 3a: Artificial Regression Results for Bank A[a]					
H_c: $y = (1-\alpha) F_1(x_m\beta_m) + \alpha F_2(x_s\beta_s)$		t-test of hypothesis H_o: $a = 0$			
		Model I.A		Model II.A	
Non-nested hypothesis		t_a	p-value	t_a	p-value
If $a = 0$, then the compound model collapses to $y = F_1(x_m, \beta_m)$	**Market vs Bank-specific** H_1: $E(y\mid\Omega) = F_1(x_m, \beta_m)$ H_2: $E(y\mid\Omega) = F_2(x_s, \beta_s)$	13.88	.0001	13.86	.0001
If $a = 0$, then the compound model collapses to $y = F_2(x_s, \beta_s)$	**Bank-specific vs Market** H_1: $E(y\mid\Omega) = F_2(x_s, \beta_s)$ H_2: $E(y\mid\Omega) = F_1(x_m, \beta_m)$	0.26	.7926	0.19	.8469

[a] The test of the significance of the coefficient a from the artificial regression (see Appendix equation A.2) is a test of the hypothesis that α is significant in the compound model H_c. In this table, the combined results of a significant t-statistic (t_a) -- reflected in a low p-value -- in the first row and an insignificant t-statistic (t_a)-- a high p-value -- in the second imply that the bank-specific specification $F_2(x_s, \beta_s)$ fits the data better than the market-level specification $F_1(x_m, \beta_m)$. Conversely, the combination of an insignificant t_a -- a low p-value -- in the first row and a significant t_a -- a low p-value -- in the second row implies that the market-level specification fits the data better than the bank-specific specification. That is, a particular specification dominates only if t_a is significant when the model is specified under the (non-nested) alternative hypothesis and insignificant when the model is specified under the (non-nested) null hypothesis. If the t-statistics are insignificant under both non-nested hypothesis tests, or both significant, the results of the tests suggest either that both specifications fit the data equally well or that neither specification is satisfactory. (See Davidson and MacKinnon, 1993, for a more detailed discussion of these techniques.)

These results suggest that, as a practice, it is unlikely that race affects application outcomes in a systematic way at Bank A. They also demonstrate the deficiencies of using a general, market-level model, rather than an institution-specific approach. Although the market-level model specification indicated significant disparate treatment, the finding was not supported for this bank once the specification was modified to incorporate the specific underwriting guidelines of the individual lender.

The results from the bank-specific statistical models for Bank A were consistent with the conclusion drawn from the judgmental examination at the same bank. That is, both methods reached the conclusion that Bank A applied its guidelines fairly and evenly across racial groups.[17]

[17] Note that, given the problems discussed in Section II, we cannot conclude, nor can we test for, the possibility that the underwriting guidelines are unfair or have a disproportionately adverse impact on

In this case, the modification of the model specification to reflect the individual bank's stated policy guidelines was critical in reaching this conclusion, because the more generic, market-level specification reached the opposite conclusion with regard to racial differentials.

Bank B. As was the case for Bank A, the alternative model specification performs better than the market-level model in terms of goodness-of-fit and the significance of several of the explanatory variables; all the coefficients have the expected signs and almost all are significant at the 1 percent level (see Table 4). Moreover, in Table 4.a, the results of the artificial regression analysis support the conclusion that the bank-specific model dominates the market-level model.

The relationship between the race of the applicant and the likelihood of denial, however, becomes *stronger* under the bank-specific model specification. White applicants appear to receive more favorable treatment relative to all three minority groups using the restricted data set (instead of only two of the minority groups under the market-level model specification).[18] These results suggest that even after controlling for the bank's specific guidelines, the difference in treatment identified under the market-level model specification persists for Bank B.

The rejection of the hypothesis of no difference in treatment across racial groups is robust over several methods of introducing the bank-specific guidelines. These results, however, differed with the conclusion of the comparative file exam undertaken at the same time at Bank B. The OCC examiners found no evidence of systematic disparate treatment against any minority racial group. However, the OCC examiners, in their initial review of the files, identified several applications that raised questions about possible differences in treatment by race. Many of these particular files were also among the cases identified by the regression model as outliers or as observations with the largest impacts on the race coefficients. Only after further review of these files, and discussions with bank personnel, did the examiners conclude that there was no disparate treatment.

---

minorities (i.e., disparate impact). The design of our analysis can be used only to test whether the standards are applied in a fair and even manner (i.e., disparate treatment).

[18] The following indicates the approximate magnitude of racial differentials implied by the model results, when expressed in terms of estimated probabilities of denial: If a hypothetical white applicant had an estimated probability of denial of .33 (the approximate white denial rate for Bank B), and that same white applicant were treated as a member of Minority group I, the estimated probability of denial would increase to .49, or an increase of 47 percent. The comparable figures for the other groups are: Minority II .51 (a 54 percent increase) and Minority III .56 (a 68 percent increase).

Table 4: Regression Results for Bank B[a]				
	Bank-specific Model I.B		Bank-specific Model II.B[b]	
Variable	Estimated coefficient	p-value	Estimated coefficient	p-value
Intercept	**-3.530**	.0001	**-3.751**	.0001
Debt ratio	**5.503**	.0001	**5.703**	.0001
Excess LTV (adj. PMI)	**10.463**	.0001	**10.042**	.0001
Insufficient funds	**0.560**	.0119	**0.703**	.0020
Insufficient funds (phase II)	**0.511**	.0074	**0.498**	.0116
Gift	**-0.998**	.0042	**-1.071**	.0036
Consumer credit	**0.115**	.0575	**0.142**	.0203
Public record	**0.879**	.0002	**0.843**	.0004
Banking relationship	**-1.533**	.0024	**-1.476**	.0038
Special program	**-4.432**	.0001	**-4.330**	.0001
Minority I	**0.658**	.0108	**0.555**	.0382
Minority II	**0.748**	.0520	**0.840**	.0386
Minority III	**0.935**	.0001	**0.926**	.0001
Number of obs.	729		698	
-2 Log L	690.2	.0001	651.2	.0001
Hosmer-Lemeshow		.9739		.9164
% correct (prob= 0.5)	76.3		76.6	

[a] The variables in the OCC models are defined as follows: Excess LTV (adj PMI) is the excess of LTV over the threshold for the loan type and program, set equal to zero if LTV is below the threshold or if the applicant obtained PMI. Debt ratio is the total debt-to-income ratio. Insufficient funds is a dummy variable that equals 1 if verified funds are less than the sum of the downpayment, closing costs, and required reserves, and zero otherwise. Insufficient funds (phase II) is a dummy variable that captures the change in underwriting standards in the second half of the sample period. Gift is a dummy variable that equals 1 if the applicant receives a gift or grant, and zero otherwise. Consumer credit is a dummy variable that equals 1 if the applicant has late payments in the past 24 months that are considered serious under the bank's guidelines, and zero otherwise. Public record is a dummy variable that indicates any judgments, collections, losses, bankruptcies, or foreclosures in the applicant's credit history, and zero otherwise. Banking relationship is a dummy variable that equals 1 if the applicant is a bank employee or a private banking customer. Special program is a dummy variable that equals 1 if the application was submitted under the bank's special program for low- and moderate-income applicants.

[b] For Model II, files in which the applicant rejected the bank's counteroffer, files in which the applicants were overqualified for the special loans programs under which they applied, and files denied PMI were dropped from the full data set used in Model I.

Table 4a: Artificial Regression Results for Bank B[a]					
H_c: $y = (1-\alpha)\, F_1(x_m\beta_m) + \alpha\, F_2(x_s\beta_s)$		t-test of hypothesis H_o: *a*= 0			
		Model I.B		Model II.B	
Non-nested hypothesis		t_a	p-value	t_a	p-value
If *a*= 0, then the compound model collapses to $y = F_1(x_m, \beta_m)$	**Market vs Bank-specific** H_1: $E(y \vert \Omega) = F_1(x_m, \beta_m)$ H_2: $E(y \vert \Omega) = F_2(x_s, \beta_s)$	12.21	.0001	11.32	.0001
If *a*= 0, then the compound model collapses to $y = F_2(x_s, \beta_s)$	**Bank-specific vs Market** H_1: $E(y \vert \Omega) = F_2(x_s, \beta_s)$ H_2: $E(y \vert \Omega) = F_1(x_m, \beta_m)$	0.20	.8414	0.30	.7681

[a] See footnote to Table 3a.

Several possible reasons exist for the apparent inconsistency between the results of the statistical model and those of the judgmental, comparative file exam. Although the general methodology used to evaluate the lending practices of Bank B was the same as that for Bank A, we encountered several problems unique to that institution. First, the bank's underwriting guidelines were more complex, requiring more intricate definitions of the model variables. Second, many of the underwriting guidelines changed -- sometimes more than once -- during the sample period, because of changes in bank management and its policies and objectives for the mortgage banking operations. The actual changes in underwriting, however, often occurred only gradually, according to bank officials, and only after new training for loan officers and underwriters.[19]

Third, the sample contained a number of files with peculiar characteristics and outcomes that did not fall neatly into either approvals or denials. For example, OCC examiners conducting the judgmental fair lending review at Bank B found that it had classified a number of files as denials for HMDA reporting purposes, when the customers had requested clearly that their applications be withdrawn. Since the model employed for the regression analysis was a binomial logit, withdrawn applications had been excluded from the samples at all banks.

To adjust for this problem, all denied files were reviewed a second time, and 41 such withdrawals were detected and deleted from the full Bank B data set used in OCC Model I.B. Table 5 presents the results of reestimating the model on the reduced data set. The estimated

---

[19] Consideration of these two problems greatly increased the complexity of modeling the underwriting decision process. For example, the guidelines setting the acceptable maximum total debt ratio were a function of LTV, income, self-employed status, and participation in special programs; required reserves were a function of both LTV and self-employed status. With regard to the changes in underwriting standards implemented during the sample period, observable changes in bank practice did not take place on the day that new policy manuals were issued, but were spread out over weeks and months, and may well have occurred at different rates for different bank offices.

coefficients of all three minority race variables remain positive and significant; all three have increased in magnitude relative to the Bank B models in Table 4 and are now significant at the 1 percent level.[20] Of the non-race variables, the signs of the coefficients are the same; the magnitudes of several variables have changed somewhat, but little change occurred in the significance levels of the coefficients. Moreover, the bank-specific model still dominates the market-level model in the artificial regression tests of explanatory power (Table 5a).

Table 5: Additional Regression Results for Bank B[a]

	Bank-specific Model III.B	
Variable[b]	Estimated coefficient	p-value
Intercept	**-4.224**	.0001
Debt ratio	**6.186**	.0001
Excess LTV (adj. PMI)	**11.192**	.0001
Insufficient funds	**0.611**	.0105
Insufficient funds (phase II)	**0.574**	.0055
Gift	**-0.937**	.0128
Consumer credit	**0.151**	.0176
Public record	**0.826**	.0008
Banking relationship	**-2.182**	.0011
Special program	**-3.659**	.0001
Minority I	**0.707**	.0105
Minority II	**1.045**	.0084
Minority III	**1.173**	.0001
Number of obs.	688	
-2 Log L	607.4	.0001
Hosmer-Lemeshow		.8665
% correct (prob= 0.5)	78.5	

[a] Files determined, on second review, to be withdrawals, have been deleted from the data set.
[b] Variable definitions are identical to those in Table 4.

The review of the loan files undertaken to identify withdrawals also revealed a number of files with unusual characteristics and ambiguous outcomes that appeared to defy classification as

[20] For a white applicant with an estimated probability of denial of .33, these coefficients imply a new probability of .50 if treated as Minority I (a 51 percent increase), .59 if treated as Minority II (a 76 percent increase), and .62 as Minority III (an 85 percent increase).

clear-cut approvals or denials;[21] in addition, the bank had subsequently reopened several denied files (predominantly minorities) and approved those loan applications. Various methods were attempted to adjust for these peculiarities, including deleting categories of the anomalous files and reversing the outcomes of the applications the bank later approved. However, there was little change in the results of the regression model in general or for the minority variables in particular.

Table 5a: Artificial Regression Results for Bank B
(Additional regression)[a]

H_c: $y = (1-\alpha)\, F_1(x_m\beta_m) + \alpha\, F_2(x_s\beta_s)$		t-test of hypothesis H_o: $a = 0$	
		Model III.B	
Non-nested hypothesis		t_a	p-value
If $a= 0$, then the compound model collapses to $y = F_1(x_m, \beta_m)$	**Market vs Bank-specific** H_1: $E(y\mid\Omega) = F_1(x_m, \beta_m)$ H_2: $E(y\mid\Omega) = F_2(x_s, \beta_s)$	12.14	.0001
If $a= 0$, then the compound model collapses to $y = F_2(x_s, \beta_s)$	**Bank-specific vs Market** H_1: $E(y\mid\Omega) = F_2(x_s, \beta_s)$ H_2: $E(y\mid\Omega) = F_1(x_m, \beta_m)$	0.87	.3895

[a] See footnote to Table 3a.

Thus, the results of the initial specification held up, even after modifying the model to remedy some of these problems. The positive and significant coefficients for the minority variables, with their implication of differential treatment by race, were robust to many specification and data modifications. The contradiction between the conclusions of the statistical and judgmental analyses remained unresolved.[22]

---

[21] This group consisted primarily of cases in which the applicants appeared to be qualified -- and were often given preliminary approval -- but failed to provide required documents (e.g., verifications of employment or income, termite inspections), were unable to sell their present homes, or were unable to complete construction of a new home by the time the loan offer expired.

[22] After making these corrections, we also tested the hypothesis that the importance of some or all factors included in the model changed after the introduction of new policy guidelines (as noted in footnote 19). We estimated the model, using the same specification as in Tables 4 and 5, but splitting the sample into two periods -- pre- and post-issuance of new policy manuals. The results were not robust to this division of the sample. The Minority I variable was no longer significant in either period; the Minority II variable was significant only in the first period (at the 5 percent level); and the Minority III variable was significant only in the second period (at the 1 percent level). This suggests the possibility that the single/pooled model may be picking up differences in treatment due to the comparison of files across the

An in-depth file review of individual denied minority files was undertaken in a further effort to verify the finding of discrimination -- or to determine whether there might be some undetected flaw in the statistical model as applied to individual banks in general or this bank in particular. We identified those files whose deletion would have the largest impact on the estimated minority coefficients. There were 22 such files in the three minority categories combined. These files were reviewed one by one; each file was also compared with several white approvals with generally similar characteristics and qualifications.

The detailed file reviews and comparisons, however, failed to corroborate the existence of disparate treatment against any minority group. This apparent contradiction was explained by two factors. First, several applications had rather unusual characteristics, subject to special underwriting guidelines that were not captured by the model specification. Examples include an applicant who could not verify the source of a cash downpayment; a downpayment that came entirely from gifts; an assumption of an existing loan; nonowner-occupied properties; and properties with more than one housing unit. Second, there were several cases of data errors, including credit history items, debt ratios, and special program status.

The first factor, applications with unusual characteristics, indicates possible limitations in the model specification. Although we built many bank-specific underwriting guidelines and thresholds into the variable definitions, mortgage lending is a complex process, and Bank B's underwriting manual contained dozens of detailed limits and guidelines, some of them applicable only rarely. A statistical model cannot possibly capture every single provision; special cases will appear as outliers or poor fits, and the deviations from the more general guidelines will be reflected in the error term. The results for Bank B suggest that, at least for that bank, the effect of these deviations was not random, but, rather, was related to minority racial status.[23]

Some data errors, the second factor, are inevitable in any research design that involves collecting data from primary sources. The problem is more acute in this type of study, however, because loan files, at virtually any large bank, often contain numerous (sometimes confusing and

---

shift in policy, not due to differences in treatment with respect to the underwriting standards.

Though the results still suggest some difference in treatment, they also reinforce the difficulties of trying to model the changes in underwriting guidelines within the sample period. In particular, the fact that the changes in actual underwriting *practice* did not occur on one single date, but, rather, were spread over a lengthy period and occurred at different times in different underwriting centers means that splitting the sample in the manner described previously may not adequately capture the exact structural change. For these reasons, we have continued to rely on the combined, single-sample model, incorporating the guideline changes by way of the definitions of the individual variables rather than by splitting the sample.

[23] It is generally assumed that such cases are distributed randomly over all race categories. Our data could contain a disproportionate number of minority files with these characteristics solely because of the sampling procedure employed (i.e., oversampling of minority applicants), and not because the bank uses these unusual characteristics as justification for disproportionately denying minority applicants.

mutually inconsistent) documents. It can be difficult to find and verify all the necessary data. Although a quality control check of the data collection effort at Bank B revealed that, overall, the level of accuracy was high, the results of the file reviews demonstrated that important data errors, though infrequent, affected the choice of individual files identified as having a large impact on the estimated minority coefficients.

Thus, the judgmental file reviews found no evidence to support the conclusion of differences in treatment by race at Bank B. As bank-specific statistical modeling is still in the early stages of development, more attention must be paid to the issue of how to proceed when a statistical model and the judgmental approach reach opposing conclusions and when the differences cannot be explained easily or readily. In this study, we used one method of identifying denied minority files for follow-up judgmental review and matched them with approved whites using general characteristics like loan-to-value and the debt ratios. Other methods for both identifying denied minorities and matching them with approved whites for comparative file reviews should be explored.

Bank C. Although the market-level model suggests that no difference exists in the treatment of minorities at this bank, we nonetheless attempted to analyze the lending decision process there against the specific standards stated in its policy guidelines in the same manner as for Banks A and B. This exercise was extremely informative, for we unmasked an additional source of concern in the construction of a statistically valid procedure -- a source of concern embedded in the sample design.[24]

Unlike for Banks A and B, an unexpectedly large proportion of the sample at Bank C contained applications from three groups with unusual characteristics that were not identified in the HMDA data used to generate the sample: (i) borrowers purchasing bank-owned properties (OREO);[25] (ii) private banking customers; and (iii) borrowers applying under CRA-type loans programs intended for low- and moderate-income applicants. Altogether, over one-third of the sample (and a disproportionately larger fraction of minority applications) fell into at least one of these categories. Since these unusual characteristics may significantly affect the relationship between the outcome of the approve-deny decision and the creditworthiness measures of the applicant(s), we attempted to modify the model specification to consider them.

Initially, we introduced dummy variables into the market-level model specification for Bank C; these were intended to capture the differences in underwriting standards used to evaluate

---

[24] These problems probably would not have surfaced had we accepted the results of the market-level model outright. Although these problems seem to have a major effect on the results for Bank C, they should not affect the overall validity of the approach. We suspect the types of problems encountered at this bank are unusual and will occur rarely at other banks. Modifying the sampling procedure to allow for a pre-test would greatly reduce the likelihood that these problems will have as important an effect at other banks as they did in this particular sample.

[25] Referred to as other real estate owned, these properties were acquired by the bank through foreclosure proceedings on prior loans.

OREO and private banking applications. This partial modification of the specification had a dramatic impact on the coefficient of the Minority I race variable. The estimated coefficient increased in magnitude and became significant at the 1.5 percent level (results are not reported). The effect of these changes improved the goodness-of-fit of the model (though the H-L statistic remained below 10 percent). When we added a third variable, which reflects the borrower's ability to finance closing costs (insufficient funds), the goodness-of-fit continued to improve, and the coefficient on the Minority I variable remained significant. These results suggest that a generic model specification, based on broadly defined secondary market requirements (i.e., the market-level model), may fail to capture differences in treatment when the underwriting standards of the bank deviate from the secondary market norms, because of special loan programs or other unusual borrower characteristics like those encountered in the Bank C sample. Therefore, we further modified the model specification and developed a fully bank-specific model for Bank C to account for the differences in the way the bank treated applicants with these three unusual characteristics.

Bank C's guidelines allow for some flexibility in determining the appropriate threshold values for the debt payments-to-income ratios, LTV, and the allowable number of "derogatory" credit lines. Moreover, the bank's policy guidelines reveal that different threshold values are used for private banking customers, OREO properties, and CRA loan programs. It was not clear, however, if there is a corresponding shift in emphasis placed on specific requirements because of these factors (e.g., everything else the same, is a history of credit problems considered a more serious derogatory factor when evaluating a loan file under CRA loan programs than under a conventional program?).

We attempted to adjust the model in two ways to take account of the applications with one or more of the three unusual characteristics: (i) by adjusting the threshold values in the model specification for the debt-to-income ratios and the LTV ratio, and (ii) by creating dummy variables, interaction dummy variables, and systematically excluding from the sample the various groups of files with the unusual characteristics.

The complexity of this exercise is reflected in Table 6, which reports the results of three methods of adjusting the model by using different threshold values and dummy variables. Modification of the specification to include the bank's guidelines improved the model's fit. Though the H-L (p-value) statistic is sensitive to the model specification, the log likelihood ratio improves significantly relative to the market-level model specifications reported in Table 1. Moreover, the results in Table 6a shows the bank-specific models, in general, provide a better fit to the data than the market level model. All three bank-specific models dominate the market-level model over both data sets (i.e., Model I and II). Model C.c, however, only weakly dominates the market-level specification over the unrestricted (Model I) data set.

Table 6: Regression Results for Bank C[a]						
	Bank-specific Model I.C.a		Bank-specific Model I.C.b		Bank-specific Model I.C.c	
Variables	Estimated coefficient	p-value	Estimated coefficient	p-value	Estimated coefficient	p-value
Intercept	**-2.000**	.0001	**-2.285**	.0001	**-5.709**	.0001
Hsg exp > .30			**0.684**	.0113		
Hsg exp > X					**1.192**	.0085
Exc debt ratio - A	**0.014**	.0099	**0.011**	.0193		
Exc debt ratio - B					**1.615**	.0001
Excess LTV > .80	**0.115**	.0001	**0.101**	.0001	**0.047**	.0001
Adj exc LTV > .80					**1.145**	.0239
LTV < .70			**-1.141**	.0019		
Insufficient funds	**1.989**	.0001	**2.036**	.0001	**1.827**	.0001
Consumer credit	**0.054**	.0737	**0.350**	.0035	**0.252**	.0399
Public record	**2.757**	.0001	**2.672**	.0001	**2.683**	.0001
Self-empl (app)	**1.153**	.0010	**1.084**	.0054		
2-4 family			**1.036**	.0110	**1.136**	.0051
Private banking	**-3.349**	.0006	**-3.119**	.0001	**-2.897**	.0004
OREO	**-2.451**	.0001	**-2.769**	.0001	**-2.615**	.0001
Spcl loan pgm	**0.670**	.1066				
Minority I	**0.952**	.0065	**0.683**	.0622	**0.465**	.2217
Minority II	**0.363**	.2434	**0.084**	.7641	**0.088**	.7887
Number of obs.	552		552		552	
-2 Log L	442.4	.0001	416.3	.0001	392.8	.0001
Hosmer-Lemeshow		.3431		.7027		.5025
% correct (prob= 0.5)	81.7		81.5		83.3	

[a] The variables in the OCC model for Bank C are defined as follows: Housing expense > .30 is a dummy variable that equals 1 if proposed monthly housing expense exceeds 30 percent of monthly income, 0 otherwise. Housing expense > X is similarly defined, except the threshold value may exceed 30 percent based on LTV and loan amount (as stated in the bank's guidelines). Excess debt ratio (A) is the excess of the total monthly debt obligation ratio above 36 percent of monthly income; (B) is a dummy variable that equals 1 if total monthly debt obligation exceeds the bank's variable policy guideline threshold values, 0 otherwise. Excess LTV is the excess of LTV above 80 percent, however, it is set to 0 if the applicant obtains PMI. Adjusted excess LTV is a dummy variable that equals 1 if the LTV exceeds 80 percent and the applicant obtains PMI. LTV < .70 is a dummy variable that equals 1 if the LTV is below 70 percent, 0 otherwise. Insufficient funds is a dummy variable equal to 1 if closing costs exceed verified funds-to-close, 0 otherwise. Consumer credit is an indicator variable that takes on values from 1 to 6 depending on the number and duration of late payments in the past 24 months that are considered serious under the bank's guidelines, 0 otherwise; except in Models I.C.a and II.C.a, where consumer credit is the number of late payments in the past 12 months in excess of 2 of duration less than 60 days past due. Public record is a dummy variable equal to 1 if any judgments, foreclosures, etc. exist in the applicant's credit history. Self-employed is a dummy variable equal to 1 if the applicant is self-employed, 0 otherwise. 2-4 family property, Private banking relationship (beyond a deposit relationship), OREO property being sold by the bank, and Special loan program are all dummy variables equal to 1 if the condition holds, 0 otherwise.

Table 6: Regression Results for Bank C[b] (cont.)						
	Bank-specific Model II.C.a		Bank-specific Model II.C.b		Bank-specific Model II.C.c	
Variables	Estimated coefficient	p-value	Estimated coefficient	p-value	Estimated coefficient	p-value
Intercept	**-2.171**	.0001	**-2.233**	.0001	**-5.714**	.0001
Hsg exp > .30			**0.362**	.2171		
Hsg exp > X					**1.174**	.0109
Exc debt ratio - A	**0.071**	.0001	**0.059**	.0004		
Exc debt ratio - B					**1.669**	.0001
Excess LTV > .80	**0.122**	.0001	**0.106**	.0002	**0.047**	.0002
Adj exc LTV > .80					**1.382**	.0134
LTV < .70			**-1.120**	.0038		
Insufficient funds	**1.998**	.0001	**2.050**	.0001	**1.920**	.0001
Consumer credit	**0.046**	.0917	**0.274**	.0296	**0.218**	.0881
Public record	**2.493**	.0001	**2.536**	.0001	**2.588**	.0001
Self-empl (app)	**1.031**	.0060	**0.896**	.0324		
2-4 family			**0.971**	.0207	**1.018**	.0154
Private banking	**-2.994**	.0015	**-2.879**	.0004	**-2.817**	.0005
OREO	**-2.408**	.0001	**-2.684**	.0001	**-2.669**	.0001
Spcl loan pgm	**0.653**	.1238				
Minority I	**0.801**	.0294	**0.576**	.1302	**0.453**	.2494
Minority II	**0.158**	.6427	**-0.036**	.9164	**0.068**	.8420
Number of obs.	534		534		534	
-2 Log L	402.8	.0001	388.0	.0001	376.0	.0001
Hosmer-Lemeshow		.0814		.3832		.6955
% correct (prob= 0.5)	83.0		82.6		83.5	

[b] The data used in Model II have been modified as follows: Files in which the applicants were denied PMI and files in which the applicants rejected the bank's counteroffers were deleted from the full data set used in Model I.

Table 6a: Artificial Regression Results for Bank C[a]					
H_c: $y = (1-\alpha) F_1(x_m\beta_m) + \alpha F_2(x_s\beta_s)$		t-test of hypothesis H_o: $a= 0$			
		t_a	p-value	t_a	p-value
Non-nested hypothesis		Model I.C.a		Model II.C.a	
If $a= 0$, then the compound model collapses to $y = F_1(x_m, \beta_m)$	**Market vs Bank-specific** H_1: $E(y\vert\Omega) = F_1(x_m, \beta_m)$ H_2: $E(y\vert\Omega) = F_2(x_s, \beta_s)$	14.32	.0001	8.19	.0001
If $a= 0$, then the compound model collapses to $y = F_2(x_s, \beta_s)$	**Bank-specific vs Market** H_1: $E(y\vert\Omega) = F_2(x_s, \beta_s)$ H_2: $E(y\vert\Omega) = F_1(x_m, \beta_m)$	1.181	.2375	0.80	.4242
		Model I.C.b		Model II.C.b	
If $a= 0$, then the compound model collapses to $y = F_1(x_m, \beta_m)$	**Market vs Bank-specific** H_1: $E(y\vert\Omega) = F_1(x_m, \beta_m)$ H_2: $E(y\vert\Omega) = F_2(x_s, \beta_s)$	12.01	.0001	9.49	.0001
If $a= 0$, then the compound model collapses to $y = F_2(x_s, \beta_s)$	**Bank-specific vs Market** H_1: $E(y\vert\Omega) = F_2(x_s, \beta_s)$ H_2: $E(y\vert\Omega) = F_1(x_m, \beta_m)$	0.67	.5063	0.30	.7634
		Model I.C.c		Model II.C.c	
If $a= 0$, then the compound model collapses to $y = F_1(x_m, \beta_m)$	**Market vs Bank-specific** H_1: $E(y\vert\Omega) = F_1(x_m, \beta_m)$ H_2: $E(y\vert\Omega) = F_2(x_s, \beta_s)$	9.89	.0001	17.52	.0001
If $a= 0$, then the compound model collapses to $y = F_2(x_s, \beta_s)$	**Bank-specific vs Market** H_1: $E(y\vert\Omega) = F_2(x_s, \beta_s)$ H_2: $E(y\vert\Omega) = F_1(x_m, \beta_m)$	2.15	.0318	1.29	.1984

[a] See footnote to Table 3a.

The results in Table 6 also show that the Minority I race coefficient can be very sensitive to the method used to capture the flexibility in the underwriting guidelines, going from below a 1 percent significance level in Model I.C.a to insignificance (at nearly the 25 percent level) in Model I.C.c. Moreover, this sensitivity persisted after adjusting the data set for observations in which the applicant rejected the bank's counteroffer and for applicants denied PMI (see Table 6, OCC Models II.C.a - II.C.c).

In addition to testing the bank-specific models against the market-level model, we test the relationship between Model C.a (the only model for which the Minority I coefficient is consistently significant) against each of the other two models (Models C.b and C.c). The results reported in Table 6b suggest that both C.b and C.c dominate C.a. However, the test results provide only weak support for the conclusion that C.c dominates C.a over the restrictive (Model II) data set. These results seem to suggest that race is not a significant factor in Bank C's rejection decision.

<table>
<tr><th colspan="6">Table 6b: Artificial Regression Results for Bank C[a]</th></tr>
<tr><td colspan="2" rowspan="2">H_c: $y = (1-\alpha) F_1(x_i\beta_i) + \alpha F_2(x_j\beta_j)$</td><td colspan="4">t-test of hypothesis H_o: $a = 0$</td></tr>
<tr><td>t_a</td><td>p-value</td><td>t_a</td><td>p-value</td></tr>
<tr><td colspan="2">Non-nested hypothesis</td><td colspan="2">Model I</td><td colspan="2">Model II</td></tr>
<tr><td>If $a= 0$, then the compound model collapses to $y = F_1(x_a,\beta_a)$</td><td>Model C.a vs Model C.b
H_1: $E(y|\Omega) = F_1(x_a,\beta_a)$
H_2: $E(y|\Omega) = F_2(x_b,\beta_b)$</td><td>5.65</td><td>.0001</td><td>0.98</td><td>.3253</td></tr>
<tr><td>If $a= 0$, then the compound model collapses to $y = F_2(x_b,\beta_b)$</td><td>Model C.b vs Model C.a
H_1: $E(y|\Omega) = F_2(x_b,\beta_b)$
H_2: $E(y|\Omega) = F_1(x_a,\beta_a)$</td><td>0.82</td><td>.4108</td><td>0.46</td><td>.6480</td></tr>
<tr><td colspan="6"></td></tr>
<tr><td>If $a= 0$, then the compound model collapses to $y = F_1(x_a,\beta_a)$</td><td>Model C.a vs Model C.c
H_1: $E(y|\Omega) = F_1(x_a,\beta_a)$
H_2: $E(y|\Omega) = F_2(x_c,\beta_c)$</td><td>8.57</td><td>.0001</td><td>89.52</td><td>.0001</td></tr>
<tr><td>If $a= 0$, then the compound model collapses to $y = F_2(x_c,\beta_c)$</td><td>Model C.c vs Model C.a
H_1: $E(y|\Omega) = F_2(x_c,\beta_c)$
H_2: $E(y|\Omega) = F_1(x_a,\beta_a)$</td><td>1.49</td><td>.1351</td><td>2.12</td><td>.0341</td></tr>
</table>

[a] See footnote to Table 3a.

The sensitivity of the Minority I coefficient to alternative model specifications, and the inability of the non-nested hypothesis test to provide strong support for one of the alternative models ultimately undermined our confidence in the ability of the models to accurately fit the data or to identify any differences in the treatment of Minority I applicants.[26] Moreover, these results suggest the data set may not be very informative. A closer examination of the Bank C sample leads us to suspect that this failure of the econometric modeling to achieve stronger results is primarily because of the three unusual characteristics of the sample previously discussed. In particular, this disparity occurs because of the disproportionate incidence of these characteristics

[26] The results for the other racial group, Minority II, were unambiguous across all specifications and data sets, indicating no difference in treatment by race -- and were in agreement with the market-level model results.

among the Minority I applicants, rather than because of the modeling procedure (i.e., the complexity associated with incorporating the bank's specific underwriting guidelines into the model specification).

Although roughly one-third of the total sample drawn for Bank C had at least one of the three unusual characteristics previously discussed, over half of the relatively small Minority I group (15 percent of the total sample) filed under either the OREO, private banking, or CRA loan programs. More specifically, Minority I applicants were very highly overrepresented among the OREO group (over two-fifths of all the Minority I applications, and nearly two-thirds of the Minority I approvals) and among the CRA program applicants (over 11 percent of Minority I applicants applied under the CRA program as compared with 7 percent of all other applicants). However, Minority I applicants were highly *under*represented (only 1 percent of all the Minority I applications) among the private banking customers. In addition, although CRA loan programs generally are intended to improve the chances of approval for low- and moderate-income applicants -- and might therefore be expected to have lower rejection rates -- we found that at Bank C the rejection rates were considerably higher for these loans than for other applicants, for both minority groups and for whites. This is especially true for Minority I applicants -- 80 percent of the denied Minority I applicants had filed under the CRA loans program.

The results of our analysis of Bank C's behavior reinforce our concerns about drawing representative samples based solely on HMDA data, which does not include information about special programs or relationships under which a bank may suspend its normal underwriting guidelines. This can lead to serious data problems under a stratified sampling design with oversampling of a subset of strata (i.e., the minority strata). For banks like Bank A that have few exception categories (and receive relatively few applications under these exceptions), this problem is of little concern. However, some banks, like Bank C, at least during specific sample periods, may have a large proportion of their applicants (or applicants within specific strata) that fall under the exception guidelines. Without pre-sampling, it is impossible to know *ex ante* if the sample properly reflects the population with respect to these exceptions.

One possible solution to this dilemma would be to drop all applications filed under the CRA programs, OREO, or private banking. However, given the relatively large proportion of these files in the Bank C sample, and the especially large proportion of approved Minority I applicants, we would have little confidence in the results of such a model and therefore do not undertake such a solution.

The only seemingly valid way to resolve this statistical problem would have been to draw a larger sample to ensure that a sufficient number of "normal" observations, or a better mix of applicants with similar "unusual characteristics," are included in each cell. However, since there is no simple or direct way to identify files with the unusual characteristics in advance under current HMDA reporting requirements, drawing a second sample and repeating the analysis would not be feasible within the constraints of the present research effort.[27]

VI. Conclusion

In this paper, we have attempted to address several issues that have emerged in the literature since the Boston Fed study and, more importantly, the problems encountered in the process of moving from a market-level to a bank-specific analysis. We have raised additional concerns that must be addressed if a statistical approach can become a useful and accurate examination tool.

We conclude with four general comments about our analysis of individual banks' lending practices. First, the methodology used in this study is applicable only when evaluating a bank's application of its pre-determined underwriting guidelines across racial groups (disparate treatment). It cannot be used to address the issue of disparate impact, racial steering, or redlining.

Second, a broadly defined model based on secondary market criteria appears to be inappropriate at the level of the individual bank. In particular, the results of the statistical tests for differences in treatment of minorities are sensitive to the failure to incorporate the individual bank's policy guidelines. We have found that considerable differences can exist in the underwriting standards used by banks, including the threshold values for variables such as the debt ratios and loan-to-value. Since the statistical procedure is designed to test for equal application of the bank's underwriting policies by focusing on the estimated denial probabilities, a poorly designed model may inadvertently show a difference in treatment by race when the observed difference in fact results from incorporating inappropriate decision criteria into the model specification.

This point is especially germane to the issue of applying a fixed underwriting policy model (i.e., a market-level model) across several banks. Such a model would have difficulty uncovering whether any difference in treatment by race results from disparities in the application of an individual bank's guidelines (i.e., discrimination) or in the guidelines at different banks (i.e., where each bank applied its own guidelines fairly). We attempt to avoid this problem by explicitly modeling each bank's guidelines. However, the process of incorporating the bank-specific underwriting guidelines is often difficult, and additional work in this area is clearly warranted.

Third, the process of incorporating bank-specific guidelines can be complicated further by the lack of accurate data upon which to build a valid, representative sample of a bank's application

[27] The non-statistical (judgmental) exam by OCC examiners, using a larger sample, found no difference in treatment across racial groups at this bank.

pool. The information provided by banks under the Home Mortgage Disclosure Act (HMDA), although often the only data available for developing random samples for use in the regression analysis, suffers from serious shortcomings. Despite uniform regulatory guidelines, banks may use different reporting criteria, such as the distinction between denials and withdrawals. In addition, the extremely limited nature of the data may conceal unusual characteristics among subgroups of the applicant population (bank employees, CRA loan programs, OREO properties, or private banking relationships). The HMDA data also do not identify changes in underwriting policy guidelines. Expanded discussions with bank personnel prior to drawing the sample may prevent some of these difficulties; however, the problems posed by the fact that these programs or changes in guidelines may lead to differences in treatment or impact (especially if the normal underwriting standards are modified or suspended) would still remain.

And lastly, the question of how to reconcile differences between the conclusions of the regression model and those of a non-statistical, judgmental examination at the same bank must be refined further. A statistical model can identify only cases of possible discrimination. The results of a statistical analysis should be verified by a systematic review of the files in which discriminatory behavior is suspected and of other, comparable files, since unusual or idiosyncratic reasons that were not captured in the statistical model may legitimately explain the observed differences in outcomes. Additional work is warranted in developing alternative methods of identifying files for second-round reviews and of matching them for comparative file review.

The use of statistical modeling techniques in the fair lending area holds promise, but until these questions are answered, we believe that their application in the bank supervisory process is premature.

Appendix

Artificial regression

The simplest artificial regression procedure, called the Gauss-Newton Regression, is based on a first-order Taylor series approximation of the (non-linear) regression function $F(x'\beta)$ around an arbitrary parameter vector β^* (Davidson and MacKinnon, 1993; MacKinnon, 1992). To illustrate, for the regression model

$$y = g(x'\beta) + u$$

the first-order Taylor series expansion around β^* is given by

$$y = g(x'\beta^*) + g'(x'\beta^*)\, x\, (\beta - \beta^*) + \text{higher-order terms} + u.$$

Letting $b = (\beta - \beta^*)$ and $e =$ higher-order terms $+\ u$, and re-arranging terms, we get

$$y - g(x'\beta^*) = [g'(x'\beta^*)\, x]\, b + e$$

where b is the artificial regression coefficient. A model properly specified using $g(x'\beta^*)$ would gain no additional information from the expression $[g'(x'\beta^*)\, x]$. The estimated value of b, under the hypothesis of a properly specified model, would be zero.

If the errors are heteroscedastic, the artificial regression equation can be corrected by dividing both sides of the equations by the standard errors of the regression. A statistically valid test of the hypothesis that $b = 0$ can be preformed on the transformed data.

Using artificial regressions in non-nested hypothesis testing

A non-nested hypothesis test is used to formally test alternative model specifications when the specification proposed under the alternative hypothesis (H_1) cannot be written as a restriction of the specification given under the null (H_0). This condition exists when the competing model includes (excludes) variables excluded from (included in) the original model. For example, the bank-specific model specification is not "nested" within the market-level model. More explicitly, we have two competing methods of specifying the rejection equation as represented by the regressions:

$$H_0: \quad E(y|\,\Omega) = F_m(\mathbf{x}_m, \beta_m)$$
$$H_1: \quad E(y|\,\Omega) = F_s(\mathbf{x}_s, \beta_s),$$

where y is 1 if applicant is denied, 0 otherwise; $E(y|\,\Omega) = \text{Prob}(y=1|\,\Omega)$; Ω is an information set; $F_i(.)$ is the logistic cumulative distribution function; $\mathbf{x}_i$ ($i= m,s$) denotes the vector of borrowers' characteristics for the market-level (m) and bank-specific (s) models, respectively; and β_i denote the vector of estimated parameters for each model. Conventional methods of testing H_0 against the alternative H_1 are no longer appropriate.

The artificial regression analysis provides a relatively simple framework under which to test the non-nested hypothesis. Davidson and MacKinnon (1984, 1993) developed an artificial regression for limited dependent variable models that is similar to a P test procedure developed for standard regression analysis.[28] By embedding the models under the null and alternative hypothesis into a more general "compound" model of the form:

$$H_c:\ y = (1-\alpha)\, F_1(x_m\beta_m) + \alpha\, F_2(x_s\beta_s), \qquad (A.1)$$

we can test the separate specifications against the compound (H_c) specification. If $\alpha = 0$, A.1 collapses to $y = F_1(x_m\beta_m)$; and if $\alpha = 1$, A.1 collapses to $y = F_2(x_s\beta_s)$. Identifying all the parameters in A.1 (β_s, β_m, and α), however, is problematic given the degree of partial overlapping of explanatory variables used in each specification. Davidson and MacKinnon suggest replacing the β_s with any consistent estimate of their values, and constructing a Gauss-Newton Regression (artificial regression) model of the form:

$$\hat{V}_1(x_m\beta_m^*)^{-1/2}(y-F_1(x_m\beta_m^*)) = \hat{V}_1(x_m\beta_m^*)^{-1/2}\, f_1(x_m\beta_m^*)\ x_m\, b$$
$$+\ a\ \hat{V}_1(x_m\beta_m^*)^{-1/2}[F_2(x_s\beta_s^*) - F_1(x_m\beta_m^*)] + e \qquad (A.2)$$

where y, $F_i(.)$, $\mathbf{x}_i$, and β_i, are defined above; $f_i(.) = \delta F(.)/\delta x$; β_i* is a consistent estimator of β_i; $V_1(\mathbf{x}_m\beta_m)$ is the estimated variance of $F_1(\mathbf{x}_m\beta_m)$ (e.g., $V_1(\mathbf{x}_m\beta_m) = F_1(\mathbf{x}_m\beta_m)\,(1 - F_1(\mathbf{x}_m\beta_m))$; *a* and *b* are vectors of artificial regression coefficients; and e is an error term.[29] A test of the hypothesis that *a* is zero (such as a conventional t-test) is a test of the hypothesis that α is zero in A.1.

The artificial regression is constructed to test the hypothesis that $\mathbf{x}_s$ is the appropriate specification of the rejection equation against the null hypothesis that $\mathbf{x}_m$ is correct. Since the

---

[28] See Davidson and MacKinnon (1993), Greene (1993), MacKinnon (1992), Godfrey (1987), and McAleer (1987) for more detailed discussions of the non-nested hypothesis testing approach first proposed by Cox (1961, 1962) and developed by Atkinson (1969, 1970), Pesaran (1979), and Pesaran and Deaton (1978).

[29] The error terms of limited dependent variable models are non-normal and heteroscedastic. Since y takes on only the values 0 and 1, y is like a Bernoulli trial that has a variance p(1-p). The variance of the limited dependent variable models can be shown then to be $V(x\beta) = F(x\beta)\,(1-F(x\beta))$. Thus, both sides of the artificial regression equation are multiplied by the inverse of the standard error to correct for heteroscedasticity. This correction allows us to perform a valid t-test for the significance of the artificial regression coefficients *a* and *b*.

order of the equations is arbitrary, a second test is performed with the order of the equations switched. Thus, for each test of the non-nested hypothesis, two tests must be performed to determine which (if either) specification better fits the data.

Intuitively, the artificial regression is used to test the ability of the regressors in $\mathbf{x}_s$ to explain y after adjusting for the explanatory power of the set of regressors in $\mathbf{x}_m$. That is, the coefficient *a* reflects the *additional* variation in y explained by the regressors in $\mathbf{x}_s$. Thus, if *a* is significantly different from zero, it suggests that $F_2(\mathbf{x}_s \beta_s)$ performs better than $F_1(\mathbf{x}_m, \beta_m)$.

References

Atkinson, A.C. 1969. "A Test for Discriminating Between Models," *Biometrika*, vol. 56, pp. 337-47.

__________. 1970. "A Method of Discriminating Between Models," *Journal of the Royal Statistical Society*, series B, vol. 32, pp. 323-53.

Avery, Robert B. and Thomas M. Buynak. 1981. "Mortgage Redlining: Some New Evidence." *Economic Review*, Federal Reserve Bank of Cleveland. Summer, pp. 18-32.

Benston, George J. and Dan Horsky. 1991. "The Relationship Between the Demand and Supply of Home Financing and Neighborhood Characteristics: An Empirical Study of Mortgage Redlining." *Journal of Financial Services Research*, vol. 5, pp. 235-260.

Bradbury, Katharine L., Karl E. Case, and Constance R. Dunham. 1989. "Geographic Patterns of Mortgage Lending in Boston, 1982-1987." *New England Economic Review*, vol. 3, pp. 3-30.

Carr, James H. and Isaac F. Megbolugbe. 1993. "The Federal Reserve Bank of Boston Study on Mortgage Lending Revisited." *Journal of Housing Research*, vol. 4, pp. 277-313.

Cox, D.R. 1961. "Test of Separate Families of Hypothesis," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 105-23.

_________. 1962. "Further Results on Tests of Separate Families of Hypothesis," *Journal of the Royal Statistical Society*, series B, vol. 24, pp. 406-24.

Davidson, R. and J.G. MacKinnon. 1984. "Convenient Specification Tests for Logit and Probit Models," *Journal of Econometrics*, vol. 25, pp. 241-262.

_________. 1993. *Estimation and Inference in Econometrics*, New York: Oxford University Press.

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

Galster, George C. 1992. "Research in Discrimination in Housing and Mortgage Markets: Assessment and Future Directions." *Housing Policy Debate*, vol. 3(2), pp. 639-683.

Glennon, Dennis, and Mitchell Stengel. 1994. "An Evaluation of the Federal Reserve Bank of Boston's Study of Racial Discrimination in Mortgage Lending." Office of the Comptroller of the Currency, Economic & Policy Analysis Working Paper 94-2. April.

Godfrey, L. 1988. *Misspecification Tests in Econometrics*, Cambridge, University Press.

Greene, W. H. 1993. *Econometric Analysis*, New York, Macmillan Publishing Co..

Horne, David K. 1994. "Evaluating the Role of Race in Mortgage Lending." *FDIC Banking Review*, vol. 7, pp. 1-15.

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

Department of Housing and Urban Development (HUD), Department of Justice, Office of the Comptroller of the Currency, Office of Thrift Supervision, Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation, Federal Housing Finance Board, Federal Trade Commission, National Credit Union Administration, and Office of Federal Housing Enterprise Oversight. 1994. "Policy Statement on Discrimination in Lending." March 8.

Jaeger, R. M. 1984. *Sampling in Education and the Social Sciences*. New York: Longman, Inc.

Department of Justice. 1993. "Statement of James P. Turner, Acting Assistant Attorney General, Civil Rights Division, before the Committee on Banking, Housing, and Urban Affairs, United States Senate, Concerning Mortgage Lending Discrimination." February 24.

Lawton, Rachel. 1993. "Pre-Application Mortgage Lending Testing Program: Lender Testing by a Local Agency." U.S. Department of Housing and Urban Development, Home Mortgage Lending & Discrimination: Research and Enforcement Conference. May 18-19.

Leeds, Barry. 1993. "Testing for Discrimination during the Pre-Application and Post-Application Phases of Mortgage Lending." U.S. Department of Housing and Urban Development, Home Mortgage Lending & Discrimination: Research and Enforcement Conference. May 18-19.

MacKinnon, J.G. 1992. "Model Specification Tests and Artificial Regression," *Journal of Economic Literature*, vol. 30, pp. 102-146.

McAleer, M.J. 1987. "Specification Tests for Separate Models: A Survey," in *Specification Analysis in the Linear Model*, eds. M.L. King and D.E.A. Giles, London, Routlegde and Kegan Paul. pp. 146-96.

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.

Office of the Comptroller of the Currency. 1993. "Interim Procedures for Examining for Racial and Ethnic Discrimination in Residential Lending." Examining Bulletin EB 93-3. April 30.

Pesaran, M.H. 1974. "On the General Problem of Model Selection," *Review of Economic Studies*, vol. 41, pp. 153-71.

__________ and A.S. Deaton. 1978. "Testing Non-Nested Nonlinear Regression Models," *Econometrica*, vol. 46, pp. 677-94.

Rachlis, Mitchell B. and Anthony M.J. Yezer. 1993. "Serious Flaws in Statistical Tests for Discrimination in Mortgage Markets." *Journal of Housing Research*, vol. 4, pp. 315-336.

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.

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.

Smith, Shanna L. and Cathy Cloud. 1993. "The Role of Private, Non-Profit Fair Housing Enforcement Organizations in Lending Testing." U.S. Department of Housing and Urban Development, Home Mortgage Lending & Discrimination: Research and Enforcement Conference. May 18-19.

Snedecor, George W. and William G. Cochran. 1989. *Statistical Methods*. Ames, Iowa: Iowa State Press.

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