

Mortgage Lending Discrimination: A Review of Existing Evidence

Margery Austin Turner
and Felicity Skidmore,
Editors

Mortgage Lending Discrimination: A Review of Existing Evidence

**Margery Austin Turner
and Felicity Skidmore,
Editors**



**The Urban
Institute**

2100 M Street, N.W.

Washington, D.C. 20037

Phone: 202.833-7200

Fax: 202.429-0687

E-Mail: paffairs@ui.urban.org

<http://www.urban.org>

Copyright © June 1999. The Urban Institute. All rights reserved. Except for short quotes, no part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by information storage or retrieval system, without written permission from the Urban Institute.

The nonpartisan Urban Institute publishes studies, reports, and books on timely topics worthy of public consideration. The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders; they do not necessarily reflect the official policies or positions of the U.S. Department of Housing and Urban Development or organizations with whom the authors collaborated on this project.

Contents

1. Introduction, Summary, and Recommendations 1
Margery Austin Turner and Felicity Skidmore
 - Why It Is Difficult to Measure Lending Discrimination 3
 - Different Forms of Discrimination 3
 - Possible Reasons for Discrimination 4
 - Potential for Discrimination throughout the Mortgage Lending Process 5
 - Summary of the Evidence 5
 - Advertising and Outreach 7
 - Pre-Application Inquiries 8
 - The Loan Approval or Denial Decision 9
 - Loan Administration 13
 - Recommendations for Expanding the Knowledge Base 16
 - Launch Systematic Research on Office Locations, Outreach, and Referrals 16
 - Expand and Refine Paired Testing of Lenders 16
 - Conduct a Rigorous Statistical Analysis of Mortgage Approvals
 - Nationwide 19
 - Design and Conduct Research on Loan Terms and Conditions 20
 - Further Evaluate “Best Practices” for Remedyng Discrimination 21
2. New Evidence from Lender Testing: Discrimination at the Pre-Application Stage 23
Robin Smith and Michelle DeLair
 - The NFHA Tests 24
 - The Urban Institute Reanalysis Data Set 25
 - Results 26
 - Regional Differences 31
 - Lessons for Future Testing 31
 - Annex A: Mean Differences between White and Black Testers in Contact Length 34
 - Annex B: Information on Requested Loan Amounts by City 35
 - Annex C: Reanalysis Data Collection Form 36

| | | |
|----|---|-----|
| 3. | Does Discrimination in Mortgage Lending Exist? The Boston Fed Study and Its Critics | 43 |
| | <i>Stephen L. Ross and John Yinger</i> | |
| | Background | 43 |
| | Omitted Variables | 44 |
| | Evaluation | 47 |
| | Reanalysis | 47 |
| | Interpreting the “Meets Guidelines” Results | 50 |
| | Data Errors in the Explanatory Variables | 54 |
| | Evaluation | 55 |
| | Reanalysis | 56 |
| | Conclusions | 57 |
| | Misclassification in the Dependent Variable | 57 |
| | Evaluation | 59 |
| | Reanalysis | 61 |
| | Conclusions | 61 |
| | Incorrect Specification | 61 |
| | Evaluation: Reverse Regression | 64 |
| | Reanalysis: Reverse Regression | 64 |
| | Evaluation: Lender-Specific Underwriting Guidelines | 65 |
| | Conclusions | 68 |
| | Endogenous Explanatory Variables | 68 |
| | Evaluation | 70 |
| | Reanalysis | 72 |
| | Conclusions | 75 |
| | Conclusions Concerning the Boston Fed Study | 75 |
| 4. | Other Evidence of Discrimination: Recent Studies of Redlining and of Discrimination in Loan Approval and Loan Terms | 85 |
| | <i>Stephen L. Ross and John Yinger</i> | |
| | Loan Denial Studies Based on HMDA Data | 86 |
| | Loan Denial Studies Based on Applications to Individual Lenders | 88 |
| | Studies of Discrimination in Loan Terms | 93 |
| | Recent Studies of Redlining | 95 |
| | Recent Studies of Process-Based Redlining | 96 |
| | Recent Studies of Outcome-Based Redlining | 98 |
| | Conclusions | 99 |
| | Studies of the Causes of Discrimination | 99 |
| | Conclusions | 104 |
| 5. | The Default Approach to Studying Mortgage Discrimination: A Rebuttal | 107 |
| | <i>Stephen L. Ross and John Yinger</i> | |
| | Methodological Issues in the Default Approach to Studying Discrimination | 108 |
| | Unobserved Underwriting Variables | 108 |
| | Unobserved Borrower Characteristics | 110 |
| | Sample-Selection Bias and the Power of the Default Approach to Detect Discrimination | 111 |
| | Conventional versus Government-Insured Loans | 113 |

| | |
|--|-----|
| New Twists on the Default Approach | 114 |
| Simulations | 114 |
| A Default Test Based on Market Concentration | 117 |
| Conclusions | 120 |
| Technical Appendix to Chapters 3 through 5 | 129 |
| <i>Stephen L. Ross and John Yinger</i> | |
| Econometric Concepts | 129 |
| Chapter 3: Omitted Variables, Reanalysis | 131 |
| A Simple Model of Endogeneity in the Loan Denial Equation | 131 |
| A Complex Model of Endogeneity in “Meets Guidelines” | 132 |
| Chapter 3: Data Errors in the Explanatory Variables, Reanalysis | 133 |
| Chapter 3: Endogenous Explanatory Variables, Reanalysis | 133 |
| Chapter 5: New Twists, a Default Test Based on Market Concentration | 134 |
| 6. Inside a Lender: A Case Study of the Mortgage Application Process | 137 |
| <i>Kenneth Temkin, Diane K. Levy, and David Levine</i> | |
| Description of Case Study Lender | 138 |
| Lender’s Origination Process | 139 |
| Overview | 140 |
| Referral | 141 |
| Initial Application Interview | 142 |
| Underwriting | 144 |
| Post-Underwriting | 144 |
| Conclusions | 145 |
| The Lender’s HMDA Performance | 145 |
| Overall HMDA Performance | 146 |
| HMDA Performance Controlling for Applicant Income | 146 |
| HMDA Analysis for Government Loans | 148 |
| Conclusions | 149 |
| What the Lender Isn’t Doing and Should Do | 149 |
| Develop a Formal Mission Statement | 150 |
| Monitor Fair Lending Performance | 150 |
| Compensate Staff in a Manner Consistent with Fair Lending Objectives | 150 |
| Develop a Diverse Workforce | 151 |
| Provide Training in Fair Lending | 151 |
| Conduct Outreach to Minorities | 151 |
| Self-Monitor Fair Lending Performance | 152 |
| Designate an Employee to Receive Fair Lending Complaints | 152 |
| Implementation Strategies | 153 |
| Implementation Steps | 154 |
| Policy and Research Conclusions | 158 |
| References | 161 |
| About the Editors | 169 |
| About the Contributors | 171 |



THE URBAN
INSTITUTE

Chapter 1

Introduction, Summary, and Recommendations

MARGERY AUSTIN TURNER
FELICITY SKIDMORE

A major element of the American dream is a home of one's own in the neighborhood of one's choice. Owning a home is one of the primary ways of accumulating wealth in our society, a form of wealth acquisition that is especially protected in the U.S. tax code. Being a homeowner increases people's feelings of control over their lives and their sense of overall well-being. High rates of homeownership are believed to strengthen neighborhoods as well, by increasing residents' stake in the future of their communities.

Not all Americans, however, enjoy equal access to the benefits of homeownership. Federal law prohibits discrimination in the homebuying process, mandating that all would-be homebuyers must be treated equally by real estate agents, lenders, appraisers, and insurance brokers.¹ However, existing enforcement mechanisms may not be effective enough to guarantee equal treatment or equitable results. Indeed, research clearly shows that minorities still face substantial discrimination in the process of looking for a home to buy (or rent).²

Many people believe that minorities also face discrimination when they try to obtain a mortgage—a necessity for most Americans wanting to buy a home. There is no question that minorities are less likely than whites to obtain mortgage financing and that, if successful, they receive less generous loan amounts and terms. But whether these differences are the result of discrimination—rather than the inevitable result of objectively lower creditworthiness—is the subject of a raging debate. The problem is not that analysts or practitioners have ignored the question of discrimination in mortgage lending. Many research and investiga-

tive studies have addressed certain facets of it, using different data sets and analytic techniques to study various outcomes. The problem is that these studies have not produced a clear consensus on a set of conclusions.³

The purpose of this volume is to sort through the research evidence on mortgage lending discrimination, in order to provide policymakers with a comprehensive and comprehensible review of the current state of knowledge on lending discrimination and identify important questions that still need to be answered in order to recommend how best to further the goal of fair housing for all.⁴ Our review of the existing research evidence concludes that minority homebuyers in the United States do face discrimination from mortgage lending institutions. Although significant gaps remain in what we know, a substantial body of objective and credible statistical evidence strongly indicates that discrimination persists. Specifically, we find that:

- Discrimination in home mortgage lending takes two forms—differential treatment and disparate impact—and in many instances it is difficult, if not impossible, to disentangle the two.
- Despite individual instances of discrimination uncovered at every major stage in the mortgage lending process, almost no research has focused on the advertising, outreach, and referral stage or on the loan administration stage.
- Paired testing at the mortgage pre-application stage (conducted by the National Fair Housing Alliance) indicates that differential treatment discrimination occurs at significant levels in at least some cities. Minorities were less likely to receive information about loan products, they received less time and information from loan officers, and they were quoted higher interest rates in most of the cities where tests were conducted.
- Statistical analysis of data assembled by the Federal Reserve Bank of Boston finds large differences in loan denial rates between minority and white applicants, other things being equal. These differences have not been explained away by data errors, omitted variables, or other technical shortcomings. Although it is conceivable that these disparities are attributable to legitimate underwriting standards, the Boston Fed analysis establishes a strong presumption that discrimination exists, shifting the “burden of proof” to those who would argue that these differences are entirely due to legitimate underwriting criteria that reflect an applicant’s creditworthiness and therefore serve a business necessity.
- In-depth examination of the mortgage loan origination process from an individual lender’s perspective suggests that even among institutions with good intentions, minority customers may not be receiving equal treatment. Moreover, achieving significant reductions in lending discrimination may require systemic institutional reforms. If employees do not perceive the importance of change, and if reforms are not effectively integrated into the day-to-day operations of a business, they are unlikely to take root.

This introductory chapter begins with a brief review of the issues involved in measuring the incidence and severity of lending discrimination, including different ways in which discrimination can be defined and measured and the reasons why lenders might discriminate. This is followed by a brief summary of the evidence that highlights the new contributions of this volume. The chapter ends with



our recommendations for priority next steps in measuring mortgage discrimination and developing policies and practices to combat it more effectively.

Why It Is Difficult to Measure Lending Discrimination

Investigative activities by fair housing advocates and others have identified and successfully prosecuted individual cases of mortgage lending discrimination. However, analytic studies measuring the overall incidence of discrimination are subject to widely differing interpretations. The crux of the problem is that legal evidence of discrimination in specific cases is not the same as statistical measures of the overall level at which discrimination occurs. For analytic estimates of discrimination, researchers need to be confident that individual instances of discrimination are more than isolated occurrences, and that they add up to a consistent pattern that favors whites and outweighs in a statistical sense any corresponding pattern that favors minorities.

Two characteristics of mortgage financing make it especially difficult to reach definitive statistical estimates of discrimination. The first is that the home mortgage lending process is a complex series of stages. Discrimination could be occurring at any one or more of these, and it could take different forms at different stages. But until the stages themselves are clearly distinguished, and the incidence of discrimination measured at each, its overall incidence cannot be properly interpreted. The second is that deliberate discrimination by many institutions in American society in the past has left a legacy of economic inequality between whites and minorities that still exists today. This legacy includes racial and ethnic differences in characteristics that influence the creditworthiness of any mortgage applicant—income, accumulated wealth, property values in minority neighborhoods, and credit history. Much of the current debate about mortgage lending discrimination stems from disagreement about the extent to which differential success in obtaining a mortgage is due to credit-relevant factors that vary with race or ethnicity and how much is due to ongoing discrimination.

Different Forms of Discrimination

Discrimination in mortgage lending can take two different forms. It is important to understand the distinctions clearly, because the different forms of discrimination may require different measurement strategies, as well as different remedies. The fundamental distinction is between *differential treatment* and *disparate impact* discrimination.

Differential treatment discrimination occurs when equally qualified individuals are treated differently due to their race or ethnicity. In mortgage lending, differential treatment might mean that minority applicants are more likely than whites to be discouraged from applying for a loan, to have their loan application rejected, or to receive unfavorable loan terms—even after characteristics of the applicant, property, and loan request that affect creditworthiness are taken into account. A finding of differential treatment discrimination means



THE URBAN
INSTITUTE

that minorities receive less favorable treatment from a given lender than majority applicants with the same credit-related characteristics (as observable by the lender).

Disparate impact discrimination occurs when a lending policy, which may appear to be color blind⁵ in the way it treats mortgage loan applicants, disqualifies a larger share of minorities than whites but cannot be justified as a *business necessity*.⁶ A widely cited example is the policy of minimum mortgage loan amounts—setting a dollar limit below which a lending institution will not issue mortgages. More minorities than whites will be adversely affected by any given loan cutoff because—on average—minorities have lower incomes than whites and can only afford less costly houses. Policies such as minimum loan amounts, which disproportionately affect minorities, are illegal unless they serve an explicit business necessity. If these policies do not accurately reflect creditworthiness, or if they could be replaced by policies serving the same business purpose with a less disproportionate effect on minorities, then they are deemed under federal law to be discriminatory.

The point for public policy is that policies that are discriminatory in effect may have adverse consequences of equal or greater magnitude than practices that treat individuals differently on the basis of their race. Federal policy makes disparate impact discrimination illegal so that institutional policies do not simply perpetuate patterns of racial inequality, many of which are the consequence of past discrimination. In other words, achieving a world of truly fair lending will require remedies that go beyond color blindness.

Possible Reasons for Discrimination

The most straightforward explanation for why discrimination occurs is prejudice (often referred to by analysts as *taste-based* discrimination). If lenders—or their employees—are prejudiced against minorities, they consider them to be inherently inferior and prefer not to interact with them or have them as customers. The lending industry has long argued that it does not discriminate, because doing so would go against the very reason for being in business—maximizing profits. It is not the color of a customer’s skin that matters, according to an often-quoted statement of this viewpoint, but the color of his or her money.

This argument does not dispose of the discrimination issue, however. First, it is entirely possible for prejudice to persist among profit-motivated businesses, due to market imperfections, information barriers, and the large number of people who participate in a loan approval decision. In fact, suggestive, though not definitive, evidence that prejudice may indeed be a factor at work comes from one study in which black/white disparities in loan approval rates decline as minority representation in either a lender’s overall workforce or its management staff increases (Kim and Squires 1995).

Moreover, even if there is no taste-based discrimination in the industry, discrimination may in fact be in a mortgage lender’s perceived economic self-interest. Discrimination for this reason is referred to as *economic discrimination*, to distinguish it from discrimination due to prejudice. The key point here is that some factors that influence a lender’s expected rate of return may also be correlated with race or ethnicity. For example, minorities may be less likely



than whites to have affluent family members who can help them if they get into a financial bind, or they may be more likely to be laid off in the event of an economic downturn. If lenders think that race is a reliable proxy for factors they cannot easily observe that affect credit risk, they may have an economic incentive to discriminate against minorities. Thus, denying mortgage credit to a minority applicant on the basis of information about minorities on average—but not for the individual in question—may be economically rational. But it is still discrimination, and it is illegal.

Recent attention has focused on *cultural affinity* as another possible explanation for discrimination. This argument attributes discrimination to the lack of affinity among white loan officers for the culture of certain minority groups. Because they feel less comfortable with minority borrowers, or because they are not able to understand the way minorities communicate, loan officers may exert less effort to determine creditworthiness or to help minority borrowers meet underwriting criteria. The literature suggests several possible explanations for why this type of behavior might be occurring, but most turn out to be forms of either prejudice or economic discrimination. Another version of the cultural affinity argument is that blacks and whites tend to sort themselves by lender—black to black, white to white—and the resulting pattern of loan offerings is discriminatory to minorities. Indeed, there is some suggestive evidence that applicants may sort themselves by race in selecting lenders, but not that this form of “cultural affinity” results in differential loan denial rates (Longhofer 1996a; Hunter and Walker 1996; Black, Collins, and Cyree 1997; Bostic and Canner 1997).

Potential for Discrimination throughout the Mortgage Lending Process

Home mortgage lending is a complex process, composed of many different decision points and institutional policies. The potential for discrimination exists at any one or more points along the way. This makes research challenging, for several reasons. First, a finding of little or no discrimination at one stage in the process does not necessarily prove the absence of discrimination in the process as a whole. Moreover, discrimination may take different forms from one stage to the next, so that a single set of measurement techniques may not apply across the entire process. Finally, discrimination at one stage may influence the characteristics and requests of potential borrowers at a subsequent stage. For example, if a lender systematically steers minorities to apply for federally insured loans, while whites are encouraged to apply for conventional loans, analysis of the loan approval decision will be complicated by the fact that minorities and whites are requesting different types of loans, regardless of their qualifications.

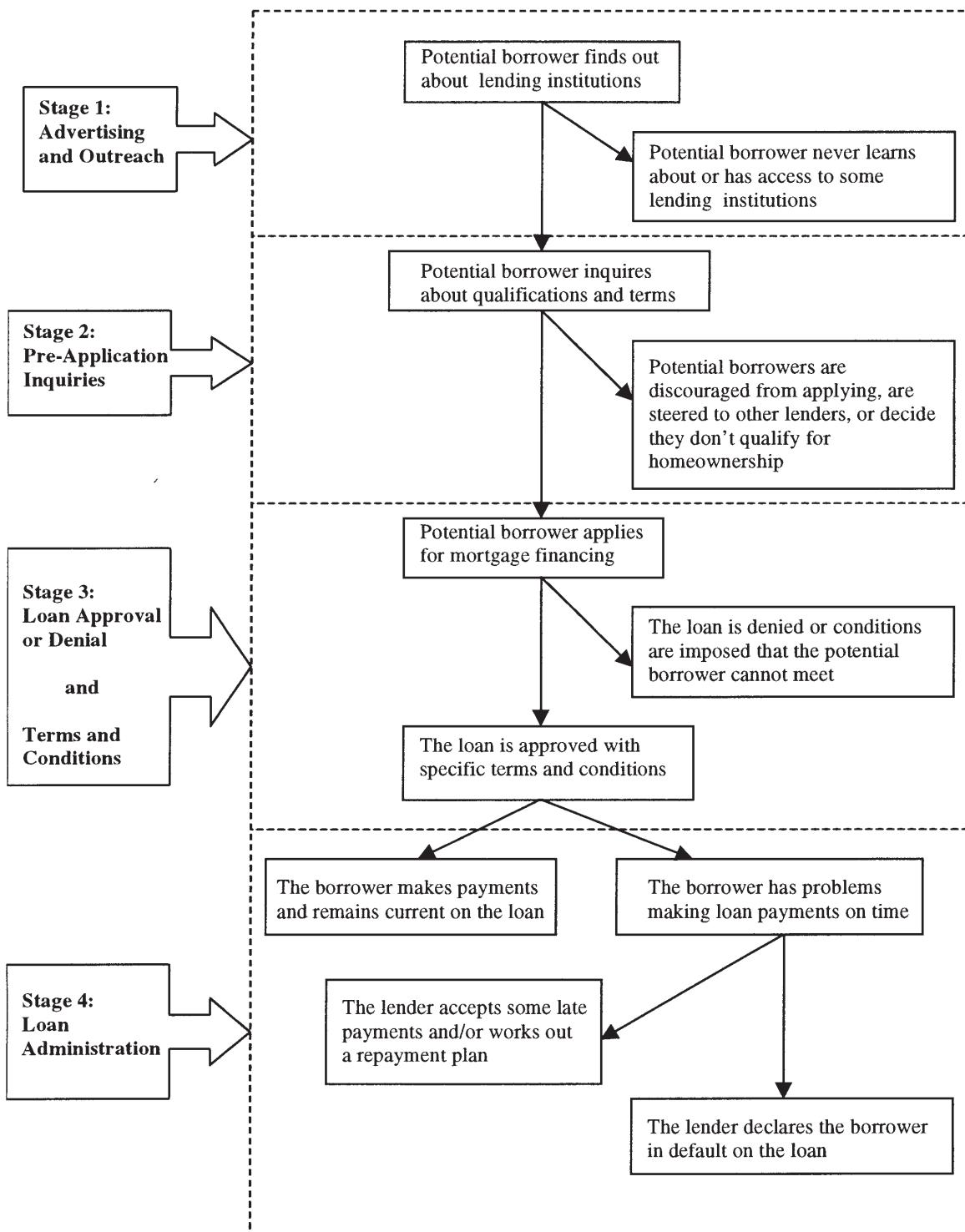
Summary of the Evidence

Although the loan approval/denial decision is what comes to mind when most people think about the mortgage lending process, the process starts considerably earlier than that, with the preliminary stages filtering out some would-be mortgage applicants before they even get to a loan officer (see exhibit 1 for an overview of key stages in the process).



THE URBAN
INSTITUTE

Exhibit 1. Key Stages in the Mortgage Lending Process



- The process actually begins with advertising and other outreach efforts—how potential mortgage applicants find out about lending institutions and loan alternatives. To some extent lenders use traditional means to advertise loans, such as newspapers and television, which are available on an equal basis to all who care to look. But they may also make special efforts to reach (or avoid) particular segments of the population.
- The second stage encompasses the information and encouragement people receive when they call or visit a lender's office to inquire about mortgage loan terms and conditions. Do minorities and whites receive different levels of services and assistance? Are they given different amounts or types of information? Are they told they may qualify for different types of loans? Do they receive different degrees of encouragement and help in understanding how to overcome barriers to applications and approval?
- The third stage in the process is the loan approval/denial decision. This stage involves submitting an application that includes a range of information required to determine the applicant's creditworthiness, confirmation (or not) of that information, the lender's up-or-down decision, and, if the loan is approved, which loan product it is.
- The final stage is loan administration. Even after a mortgage has been approved and issued, lenders can exercise considerable discretion about how to treat people who have missed one or more payments. They can accept penalties for several months, or they can start foreclosure proceedings.

Here we summarize briefly what is known about discrimination at each stage of the process, including new findings presented in this volume.

Advertising and Outreach

There is compelling legal evidence of discrimination in the placement of branch offices. This evidence comes from an investigation of the Decatur Federal Savings and Loan Association, which began in 1989 with a U.S. Justice Department investigation and ended with a consent decree signed by the two litigating parties in 1992. The investigation found that Decatur Federal had opened 43 branches in the Atlanta metropolitan area between its founding in 1927 and the late 1980s, only one of which was in a predominantly black neighborhood. During the same period, it closed two offices—the one originally opened in the predominantly black neighborhood and another one in a neighborhood that had become predominantly black.

Along with this history of branch closings, Decatur Federal explicitly applied different criteria for closing branches in black neighborhoods than in white neighborhoods. In addition, the Justice Department obtained evidence that Decatur Federal had explicitly excluded black census tracts from its market area, even though it was a large-volume lender able to compete throughout the Atlanta metropolitan area. Finally, a former Decatur Federal account executive told investigators that she was specifically instructed by the bank not to solicit loans south of Interstate 20, an area that included many of Atlanta's black neighborhoods (Ritter 1996; Siskin and Cupingood 1996).

How frequently does discrimination occur at the initial stage in the mortgage lending process? There is no research evidence, as yet, about the incidence of



THE URBAN
INSTITUTE

discrimination during the advertising and outreach stage of the mortgage lending process. This is an area where more research is clearly needed, which can build on the insights from the Decatur case as it defines and devises ways of measuring the incidence of these and similar practices across institutions and markets.

Pre-Application Inquiries

Existing knowledge about lender behavior at the pre-application stage comes primarily from paired testing (also known as fair lending audits) undertaken by fair housing advocacy agencies whose mission is to promote fair housing through a variety of channels, including litigation. Testing has been used widely for analytic as well as investigative studies of discrimination by landlords and real estate agents; however, only a few relatively small-scale investigative studies—primarily by the National Fair Housing Alliance (NFHA)—have been applied to mortgage lending.

The NFHA audits, funded by the U.S. Department of Housing and Urban Development's Fair Housing Initiatives Program (FHIP), were conducted by fair housing enforcement organizations using testers who posed as first-time homebuyers and refinancers at the pre-application stage. Testers, matched on ratios that relate a household's income and debts to the desired loan amount, visited lenders in person to inquire about the types and terms of loans for which they might qualify. After each visit, testers answered a set of closed-ended questions and wrote extensive narratives about their experiences. NFHA conducted tests in seven cities (Atlanta, Chicago, Dallas, Denver, Detroit, Oakland, and Richmond), with about two-thirds of the tests concentrated in Chicago and Oakland.

NFHA concluded that lenders often appeared to be less interested in giving information to black customers than to whites; urged black customers, but not whites, to go to another lender; and emphasized to black customers, but not whites, that application procedures would be long and complicated. According to these investigative audits, blacks were also more likely than equally qualified whites to be told that they did not qualify for a mortgage (before they had filed a formal application), and whites were more likely to be "coached" on how best to handle potentially problematic aspects of their credit profile (Smith and Cloud 1996).

Given their purpose, NFHA's tests were not designed to produce statistically valid measures of the incidence of discrimination across lenders or markets. However, NFHA provided the Urban Institute access to data from a large number of the tests it conducted, enabling researchers to construct and analyze a database of statistically tractable information (see chapter 2). It is important to keep in mind that findings from this analysis apply only to the specific sample of lenders tested by NFHA, which were selected in large part because they had already shown signs of potential discriminatory behavior. Nevertheless, these tests provide convincing evidence of significant differential treatment discrimination at the pre-application stage and highlight the need for further testing with a sample that is large and representative enough to allow for statistical estimation.

The most basic measure of service at the pre-application stage is whether a customer is seen by a lender and given information about specific loan prod-



ucts. In four of the five cities in the reanalysis data set, African American testers were more likely to be denied such information than white testers. In four of the five cities, lenders spent more time with white than with minority testers. And in three of the five cities, lenders provided whites with information about more possible loan products. What about the loan products themselves? The information available does not support detailed comparisons of the terms and conditions offered to whites and blacks, but it does indicate which testers were quoted a product with a 30-year term. Comparing the interest rates quoted for these 30-year mortgages reveals that African American testers were more likely to be quoted higher interest rates than their white counterparts in three of the five cities.

One notable feature of these paired test results is their regional variation. Although several treatment variables show the same general pattern across cities, differences between cities are substantial. This is particularly striking because the two cities with the most tests yield opposite results. In Chicago, all of the statistically significant findings were unfavorable to black testers. In Oakland, differences in treatment were rarely significant and, when they were, they often benefited African American testers. This contrast highlights the need to understand better the regional differences in mortgage lending practices and in the incidence and forms of lending discrimination.

The Loan Approval or Denial Decision

The decision about whether to accept or reject a mortgage loan application has been the subject of an impressive amount of sophisticated statistical analysis. The primary information used in these studies is a repository of data compiled as a consequence of the 1975 Home Mortgage Disclosure Act (HMDA). HMDA mandated the annual reporting of information, by mortgage lending institutions with at least \$10 million in assets, on the number and dollar amount of both home mortgage and home improvement loans, by census tract or county. Since passage of Section 1211 of the Financial Institutions Reform, Recovery, and Enforcement Act of 1989, HMDA data have also included the race, gender, and income of mortgage loan applicants.

HMDA data are routinely used to compare a lender's denial rates for minority and white loan applicants, as a measure of their loan performance with regard to minorities. But HMDA data alone cannot prove or disprove the existence of lending discrimination, because they do not provide enough information to control for all relevant differences between white and minority borrowers. Even though HMDA data now include borrowers' race and income, they do not include critical information on the wealth and debt levels of loan applicants, their credit histories, the characteristics of properties serving as collateral, the terms of loans for which applications were submitted, or the underwriting criteria used to determine eligibility. Herein lies a good part of the story behind the fierce analytical debate about what can and cannot be said about discrimination in mortgage lending.

The seminal study in the debate over discrimination at the loan approval stage is the so-called Boston Fed Study, undertaken by researchers at the Federal Reserve Bank of Boston, which was initially released in 1992 and published in



THE URBAN
INSTITUTE

final form in 1996 (Munnell et al. 1992, 1996). Other recent studies make valuable methodological and substantive contributions, but the lack of a data set comparable to the one collected for the Boston Fed Study casts a shadow over all of this other research and makes the results difficult to interpret.

The Boston Fed Study began with HMDA data for the Boston area and collected 38 additional variables for each application in the sample, covering the whole array of information needed to control for legitimate differences in applicant creditworthiness. That the Boston Fed sponsored the study and gained the cooperation of area lenders suggests that the lending community did not expect the study to find statistically compelling evidence of discrimination. But it did just that—concluding that race was indeed a statistically significant and fairly large influence in lending decisions, even when a mass of detailed information systematically related to the lending decision was controlled for in the statistical analysis.

The findings of the Boston Fed Study had an explosive effect on the mortgage lending discrimination debate, initially stimulating extensive soul searching by the industry, followed by a great deal of analytic scrutiny of both the study's data quality and its methodological approaches. The study findings have emerged remarkably intact in the face of most of this scrutiny. But certain complex analytical questions remain that some analysts conclude are enough to undermine the credibility of the original findings. Specifically:

- **Omitted Variables.** Key variables that affect the lending decision, and that are correlated with race or ethnicity, may have been omitted from the Boston Fed analysis. If so, the estimated impact of race on the approval decision may be overstated, because it partly reflects the impact of other, legitimate factors that vary with race and ethnicity.
- **Data Errors.** Mistakes in data entry or data coding may have distorted the Boston Fed analysis, possibly leading to overestimates of the importance of minority status. In addition, some loans in the Boston Fed data set may have been incorrectly classified as approved or disapproved.
- **Incorrect Specification.** When analysts “specify” a predictive equation, they have to make assumptions about how different factors interact to influence the approval or denial decision. If the Boston Fed Study’s equations were incorrectly specified, they might again overstate the importance of minority status.
- **Endogenous Explanatory Variables.** Some of the variables in the Boston Fed equations that are used to help explain or predict loan approval may in fact be decided at the same time the loan approval decision is made—or in conjunction with that decision. For example, many loan terms, such as the loan-to-value ratio, which changes if the applicant changes the down payment, are the result of negotiation or participation in a special loan program. If this is the case, single-equation models cannot disentangle the independent effects of minority status on loan outcomes.

Because of the importance of the Boston Fed Study, this volume presents a comprehensive review and reanalysis to assess these critiques (see chapter 3). In some cases, reanalysis shows that the critics are simply wrong; the problem they identify does not exist or the bias involved is empirically insignificant. In



several cases, however, we agree with the critics that a limitation in the Boston Fed Study could potentially lead to a serious overstatement of discrimination, and we have explored these cases in detail. Moreover, we find that the critical literature has raised several important issues concerning the interpretation of the Boston Fed Study's results. Our analysis leads to the following major conclusions:

- The large differences in loan denial rates between minority and white applicants found by the Boston Fed Study cannot be explained away by data errors, omitted variables, or interrelationships between factors that influence loan approval (endogeneity).
- The Boston Fed Study results do not definitively prove either the presence or the absence of differential treatment discrimination in loan approval, nor do they definitively prove either the presence or the absence of disparate impact discrimination.
- BUT, the Boston Fed Study results provide such strong evidence of differential denial rates (other things being equal) that they establish a presumption that discrimination exists, effectively shifting the “burden of proof” to lenders.

If a case as strong as the Boston Fed Study results were made in a courtroom setting, lenders could only escape the conclusion that they were discriminating if they could prove that their actions were based on “business necessity.” That is, they would have to prove they used underwriting guidelines with a clear connection to the return on loans, that they applied these guidelines equally to all groups, and that no equally profitable guidelines without differential effects on minorities were available. In our view, no scholar has come close to showing that the observed intergroup differences in loan approval in Boston can be justified in this way.

The best way to determine whether the observed minority-white differences in loan denial rates are the result of underwriting practices justified by business necessity would be to replicate the Boston Fed Study in other cities, with the addition of loan performance data. This approach would make it possible to determine which observed application characteristics are accurate predictors of loan returns and, therefore, which underwriting guidelines are legitimate. However, such a study still would not be able to distinguish between differential treatment and disparate impact discrimination. The best, and possibly the only, way to isolate differential treatment discrimination in loan approvals is with the paired testing methodology. Specifically, two applicants with the same credit histories and in need of the same type of loan would apply for a mortgage at the same lender. Differential treatment discrimination would exist if minority applicants were systematically treated less favorably in a large sample of cases.

Unfortunately, a paired testing study of loan approval faces major challenges. Perhaps the most important is that it would be difficult, and possibly illegal, to assign false credit characteristics to testers as a means of ensuring that teammates have identical loan qualifications. We do not believe that testing is a fraudulent activity, because testers would never actually close the loan transaction. But the courts have not yet ruled on this matter and any group that pushes paired testing into the loan approval stage of the mortgage process might face high legal bills, if not worse. It might be possible to conduct tests using



THE URBAN
INSTITUTE

people's actual credit characteristics, but this approach would be administratively difficult because testers would still have to be matched to have the same credit qualifications. As a result, a very large pool of potential testers would be required.

Using Defaults to Measure Discrimination in Loan Approvals. Some researchers have used information on differential default rates as a strategy for determining whether discrimination occurs at the loan approval stage. This approach is premised on the argument that lenders who discriminate against minority applicants do so by effectively raising their underwriting standards—rejecting minorities who meet the standard required of whites and only accepting minorities who meet a higher standard. If this is the case, minorities who receive loans will be less likely to default than whites. Therefore, the argument goes, if minority default rates are the same or higher than those of whites (other things being equal), lenders must not be discriminating. Analysis conducted for this project shows that this simple and intuitively appealing argument runs into severe methodological hurdles when used to measure discrimination in mortgage lending (see chapter 5). The difficulty of obtaining complete information on factors that influence default risk, as well as the covariance of many such factors with race, means that the default approach probably *understates* the incidence of discrimination at the loan approval stage.

A new specification of the default approach asks a new question in an effort to overcome these problems: Is the minority-majority default difference greater in locations where the lending industry is more concentrated, a situation that presumably gives the lender more leeway to discriminate? However, this new specification does not save the default approach because it depends on two mutually implausible assumptions: (a) that if lenders discriminate at all they do it more severely when market concentration is higher, and (b) that lenders do not alter any other aspect of their underwriting procedures in the presence of more concentration. Thus, we conclude that the default approach produces unreliable estimates of the incidence of discrimination in loan approvals.

Redlining. Discrimination against minority borrowers (both differential treatment and disparate impact) can take place at the neighborhood as well as the individual customer level. Discrimination based on location is often referred to as redlining because, historically, some lending institutions were found to have maps with red lines delineating neighborhoods within which they would not do business. Redlining is typically measured in two ways (see chapter 4). The first focuses on the case-by-case process of approving or denying loans. Redlining is said to occur when otherwise comparable loans are more likely to be denied for houses in minority neighborhoods than for houses in white neighborhoods, even though all credit-relevant characteristics of applicants, properties, and loans are the same. Studies of this kind of redlining face the same basic challenge as studies of discrimination against individual applicants, namely to find a data set with adequate information on loans and applicants, including applicant credit history. The only studies of redlining with such information turn out to be based on the Boston Fed Study's data. Two of these studies find no evidence of redlining, but a third, which accounts for the relationship between redlining and private mortgage insurance, finds redlining against low-income neighborhoods, which in Boston are largely black (Tootell 1996a; Hunter and Walker 1996; Ross and Tootell 1998).



The second approach to the measurement of redlining focuses on aggregate lending outcomes. In this context, redlining is said to occur when minority neighborhoods receive a smaller volume of mortgage loan funds than white neighborhoods that are comparable in all relevant respects. This approach has received more empirical attention than the individual-level approach. Most studies focus on outcomes by census tract, while one attempts to isolate the role of lenders (Schill and Wachter 1993; Phillips-Patrick and Rossi 1996). Many studies in this literature find evidence of redlining, but others do not, and no consensus has yet emerged on the extent of redlining or appropriate methods for measuring it.

Negotiating Loan Terms. At the loan approval stage, lenders do not simply decide whether to make a loan. They also set the terms of the loan, including the interest rate, loan fees, maturity, loan-to-value ratio, and loan type (conventional, adjustable rate, FHA, and so on). This is an important issue, because fair housing complaints often involve unfair terms and conditions for loans, and there is reason to believe that the lending industry may be in the process of shifting from “credit rationing”—where customers perceived to be high-risk are denied loans—toward “risk-based pricing”—where these same customers are simply charged a higher price for loans.

Chapter 4 of this volume reviews the existing empirical literature on this issue. One early analytic study found discrimination against blacks and Hispanics in interest rates and loan fees but not in loan maturities. Another also found discrimination against blacks in the setting of interest rates. Both studies used extensive statistical controls to isolate the effect of race and ethnicity from the effects of other factors. Two more recent studies examine discrimination in *overages*, defined as the excess of the final contractual interest rate over the lender’s official rate when it first commits to a loan. Both of these studies find cases in which the overages charged to black and Hispanic borrowers are higher than those charged white customers by a small but statistically significant amount.

With respect to type of loan, several research studies have examined the probability that a borrower will receive an FHA loan instead of a conventional loan. Both borrowers and lenders have an interest in this choice. FHA guidelines are relatively flexible and may qualify borrowers who do not meet conventional underwriting standards. This makes them attractive to both borrowers and lenders. But FHA loans may cost more than conventional loans and may also permit higher fees to the lenders. It is clear that minority borrowers, in fact, rely more heavily on FHA loans than do white borrowers. What the analytical literature shows is that, controlling for borrower, property, and loan characteristics, minorities are still more likely than whites to receive FHA loans. One plausible explanation is that minorities are steered in the FHA direction because of discrimination in the market for conventional mortgages.

Loan Administration

There is no systematic research evidence on potential discrimination in loan administration. However, anecdotal evidence—as shown, for example, on the investigative reporting TV show *60 Minutes*—suggests that at least some lenders take a harsher stance in foreclosure decisions against minority customers than

against whites. In extreme cases, some lenders may even increase their profits by making loans that encourage defaults, initiating foreclosure proceedings if any payment is late, and selling the property for a profit. This is clearly discriminatory behavior in itself. But if this practice occurs with any frequency, it also biases downward statistical estimates of discrimination in the initial mortgage lending decision, because it means that some lenders' acceptances of minority loans are made with the express intent to foreclose as soon as possible.

The Loan Approval/Denial Process from a Lender's Perspective. It is intriguing that the Boston-area mortgage lenders apparently believed that discrimination would not be found in the investigations of their practices. If they had, it is unlikely that they would have cooperated so fully with the Boston Fed survey. But the evidence reviewed here strongly suggests that their belief that they were not discriminating was false. Is it possible that lenders discriminate unknowingly? Can discrimination occur in the mortgage lending process even when people believe they are treating all applicants fairly? The answer to this question is vitally important in the quest for strategies to eliminate discrimination in home mortgage lending.

In an effort to shed new light on the issue, this project paid a field visit to a mortgage lending institution (see chapter 6). We conducted in-depth, structured interviews about the mortgage lending process to determine what role employees played in decisionmaking, whether they were aware of fair lending requirements, how they perceived fair lending issues, and how they were monitored by their company for fair lending compliance. After the visit, the impressions of our site visit team were compared with standard HMDA indicators of the lender's fair housing performance.

The lender we visited is a mortgage company, fully owned by a builder who develops housing for low- and moderate- as well as middle- and upper-income households. The lending institution has 31 employees and currently originates mortgages worth about \$70 million a year. Its loans are almost all for home purchases rather than refinancing, and it processes more minority applications than the average for its metropolitan area. The loan origination staff includes six loan counselors, who meet with prospective customers and take applications, and four loan processors, who collect the documentation needed to complete the applications. The branch manager of the company supervises both these groups. The company also has an underwriter who is responsible for assessing completed applications for government-insured loans (conventional loans are underwritten by an outside firm). The branch manager and the underwriter both report directly to the company's president.

Over the course of a two-day site visit, the research team scrutinized the process used to assess applications and was favorably impressed by the combination of a highly transparent review process, a strong commitment to qualifying marginal applicants, and the genuine belief by all staff that their process is color blind. The team's strong expectation was that the lender's HMDA data would show a relatively small denial disparity between white and minority applicants. However, that did not turn out to be the case.

The lender's denial rate for minorities is lower than average for its metropolitan area, indicating that it does a good job of qualifying marginal minority



applicants (and/or attracts minority applicants with above-average qualifications). But disparities between its denial rates for whites and for minorities are high, compared to metro-area averages. How can we reconcile these disparities with the lender's strong belief that its loan origination process contains absolutely no discriminatory treatment of minority borrowers? There are three possible explanations:

- A large share of the lender's minority loan applicants may actually be poor credit risks. It is possible that because the case study lender serves more minority customers than other area lenders, these customers may be less creditworthy—on average—than minority loan applicants in the metro area as a whole. If so, the case study lender's high denial disparities (relative to metrowide averages) may reflect the diversity of its customer base rather than the possibility of discrimination. However, this explanation seems inconsistent with the evidence that the case study lender approves a larger share of applications (from both minorities and whites) than the average for mortgage lenders metrowide.
- The case study lender may be applying underwriting standards that have a disparate impact on minority borrowers. In other words, minority customers may be denied at relatively high rates because some of the underwriting standards applied by the case study lender have a disproportionate effect on minorities and *do not* serve a clear business necessity. This explanation seems inconsistent with the fact that denial disparities between whites and minorities are significantly lower among other lenders in the metropolitan area.
- The lender's staff may be providing preferential treatment to white customers without realizing it. Our case study indicates that loan counselors work hard with customers to overcome problems in their applications. It is possible that the counselors are more at ease with white customers than with minorities, find it easier to communicate and sympathize, or feel more comfortable spending time with whites to solve credit problems. If this is the case, then minorities would be at a disadvantage, not because they were treated badly but because whites were treated better.

Given the information currently available, it is impossible to determine with certainty which of these explanations is correct. It is clear, however, that despite the commitment and good intentions of the case study lender, denial rates for minority loan applicants are unusually high, relative to denial rates for white customers. And these denial disparities appear to be out of line with comparable ratios for the metropolitan market as a whole.

Lending industry experts and fair housing advocates have identified a number of practices and procedures that lenders should implement to reduce the possibility of discrimination against minority applicants. Our case study reveals that the lender we visited has not fully implemented any of these fair lending best practices. Moreover, the research literature on organizational change contains clear lessons about “what it takes” to effectively change behavior within an institution. Thus, the case study illustrates how a lending institution might be discriminating against minorities despite its best intentions, and it reflects the challenges confronting lending institutions as they try to ensure full and fair service to both minority and white customers.



Recommendations for Expanding the Knowledge Base

The evidence just summarized, which is discussed at length in subsequent chapters of this volume, provides persuasive evidence that discrimination in home mortgage lending persists. Although we do not yet have reliable measures of the incidence of discrimination at each stage in the lending process, systematic monitoring and enforcement efforts are clearly justified by existing evidence that discrimination occurs at significant levels. But serious gaps remain in our collective knowledge about the incidence of discrimination, the forms it takes, and the circumstances in which it is most likely to occur. We recommend five key areas where more information and analysis can and should be assembled to inform both public policy and private action.

Launch Systematic Research on Office Locations, Outreach, and Referrals

Relatively little research has focused on the extent to which lenders may discriminate by avoiding or limiting contact with minority customers. Evidence from litigation suggests that some lending institutions locate their offices in predominantly white areas. It is also possible that some lenders target direct mail solicitations to white communities, or get their referrals primarily from real estate agents who serve white neighborhoods. If so, advertising and outreach practices steer minority and white borrowers to different lending institutions (which may offer unequal products and services). However, little is known about the extent of these practices or about their impact on potential homebuyers.

More basic research is needed to understand how white and minority borrowers identify potential lenders and whether practices such as office location, referrals, or advertising make a difference. If minority access to lending opportunities is significantly constrained by these practices, then best practice agreements and fair housing enforcement efforts can and should include strategies for reaching out to more minority customers. However, without better information about how homebuyers identify potential lenders, it is difficult to know what types of remedies make sense. For example, if most borrowers are referred to their mortgage lender by their real estate agent (as part of the home-buying process), then advertising or office locations may not matter much.

Understanding how borrowers identify potential lending institutions is also critical to the design of effective testing efforts. Paired testing, whether for research or enforcement purposes, generally attempts to replicate a typical encounter between a consumer (homebuyer) and a producer (mortgage lender). But we do not yet know enough to be sure what a typical encounter is. In the NFHA tests, individuals posing as first-time homebuyers walked into the offices of lending institutions to inquire about loan terms and conditions. However, this may not be a typical scenario, particularly if most homebuyers are referred to lenders by the real estate agent with whom they are searching for a house.

Expand and Refine Paired Testing of Lenders

Paired testing can and should be expanded at the mortgage pre-application stage. The testing conducted by NFHA demonstrates that paired testing is feasible and



that it uncovers instances of differential treatment that might otherwise go undetected. Because at least some lenders provide more information and assistance to white borrowers, minorities may be discouraged from submitting applications or may apply for loans with unfavorable terms. Discrimination at this stage cannot be detected through analysis of HMDA data or data drawn from lenders' application files. In fact, paired testing may be the only strategy for uncovering the incidence of discrimination at the pre-application stage. NFHA's testing (and our reanalysis of these test results) represents an important first step. But more work is needed to refine testing procedures and apply them to representative samples of lending institutions.

Paired testing can be effective for both research and enforcement purposes, although the procedures used for these two purposes are not identical. Research testing is designed to yield statistically reliable measures of the incidence (and severity) of differential treatment across a large number of transactions. Because all of the lender testing conducted to date was designed primarily for enforcement purposes, there are limits to what it can tell us in this regard. In order to learn more, the federal government should sponsor a paired testing effort whose primary goal is to quantify the incidence and severity of discrimination at the pre-application stage. Indeed, HUD is currently funding a pilot study that will develop several alternative paired testing methodologies and estimate levels of differential treatment at the pre-application stage for at least one market area.

Ultimately, such testing studies should be conducted in multiple markets, so that they can capture variation in levels and patterns of discrimination across sites. As discussed earlier, analysis of the NFHA test results suggests that there may be substantial differences between cities, and these differences need to be investigated more thoroughly. In addition, the lending institutions where tests are conducted should be selected systematically, to be representative of all lenders of a particular type or those serving a particular market. For example, tests might be conducted for a random sample of lending institutions with offices in a metropolitan area, for a sample of institutions over a certain size, or for a sample of those reporting a certain number of mortgage loans.

Test reporting forms should be as tightly structured as possible to permit objective comparisons of the treatment received by whites and minorities across a large number of tests. This may require advance research—or “scouting”—on the products offered and procedures followed by lending institutions in the study sites. Unless researchers and test supervisors know in advance how lending institutions treat potential borrowers prior to the formal application stage, what different loan products are called, and to whom potential borrowers might be referred, it is difficult for pairs of testers to make identical requests and to record accurately the treatment they receive. Moreover, testers should receive careful training and supervision to ensure that both members of each pair present the same attributes, qualifications, and financing needs and that both record their treatment fully and accurately.

Finally, more thought needs to be given to the specifics of lender testing scenarios. No single test pair can explore all possible requests that potential borrowers might make at the pre-application stage or all types of lending institutions in the market. The NFHA tests paired minorities and whites posing as relatively uninformed customers who were well qualified for the types of



financing about which they were inquiring. This scenario makes sense because it gives lenders the discretion to suggest different products, request different levels of information, or offer different amounts of assistance. However, other scenarios might capture different forms (and possibly different levels) of discrimination. For example, there is good reason to believe that marginally qualified whites receive more assistance and encouragement in correcting credit problems than do marginally qualified minorities. Thus, a study in which partners posed as marginally or poorly qualified borrowers might elicit different responses from lenders than a study in which testers pose as well-qualified applicants. The results of research testing could prove to be extremely sensitive to the specifics of the test scenario.

At the same time that work on research testing proceeds, fair lending enforcement testing should be refined and expanded. Pre-application testing is essential for finding out if lenders are discouraging minority borrowers from even applying, steering minorities to apply for particular loan products, or referring them to other types of lending institutions. Thus, this type of paired testing plays a critical role in the federal government's efforts to monitor fair lending compliance and to investigate complaints of discrimination. Fair housing organizations should be encouraged and supported in their efforts to conduct rigorous pre-application testing, both in response to complaints and to assess the extent to which differential treatment may be going undetected in the communities they serve. Moreover, lenders should be encouraged to conduct "self-testing," as a way to monitor the performance of their own operations. Experimentation with different testing scenarios should be encouraged to reflect different classes of potential borrowers, different segments of the lending industry, and different types of pre-application requests.

Testing should not be ruled out as a strategy for investigating and measuring discrimination beyond the pre-application stage. As discussed earlier in this report, paired testing appears to be the only research methodology that would disentangle differential treatment discrimination from disparate impact discrimination at the loan approval stage. Federal law makes it illegal to provide false information on a credit application,⁷ and many people believe that this precludes full application testing of mortgage lending institutions. However, some testing advocates argue that submitting false information as part of a paired test—when the tester does not actually intend to borrow money or incur any other financial obligation—does not violate this law. So it is possible that some organizations may be willing to incur the risk of conducting paired testing beyond the pre-application stage—or that the federal government could issue guidance that would allow and encourage greater use of testing. Moreover, it may be feasible to design a paired testing study using the actual income and credit characteristics of testers, although the challenge involved in recruiting equally matched testers would be substantial.

Some researchers also have argued for the use of *nonpaired* testing of mortgage lending decisions. This would involve finding a pool of actual candidates for mortgage loans. The applicants would then file genuine loan applications, and the progress that they made through the loan application and approval process would be monitored and documented. Analysis would focus on differential treatment of applicants from differing racial and ethnic backgrounds in



loan approvals and, in the case of approved loans, in the loan amount, interest rates, maturity, loan type, and collateral. Nonpaired testing could provide definitive estimates of the overall incidence of discrimination in loan approvals, but only paired testing can reliably distinguish differential treatment discrimination from disparate impact discrimination.

Conduct a Rigorous Statistical Analysis of Mortgage Approvals Nationwide

The Boston Fed Study should be replicated for more cities and enhanced to respond to the critical methodological issues discussed in this report. It constitutes the strongest and most complete analysis of discrimination at the loan approval stage. By assembling data on applicant characteristics and credit histories, it enabled researchers to estimate the extent to which minorities are more likely to be denied a mortgage loan, other things being equal. Despite the unprecedented scrutiny and criticism to which this study has been subjected, our reanalysis shows that it clearly disputes claims that blacks and whites receive equal treatment from the lending industry. However, this study is not able to distinguish differential treatment discrimination from disparate impact discrimination. And it cannot completely eliminate the possibility that high denial rates for minorities result from differences in their ability to meet legitimate underwriting criteria—criteria that meet the business necessity test. Moreover, the Boston Fed Study applies to only one urban area at one point in time. Comparable analysis for a representative sample of market areas is needed to assess the persistence of discrimination over time and across markets.

A multisite study of discrimination in loan approvals should build upon the intensive review and criticism generated by the Boston Fed Study. In particular, a national study should invest significant time and attention in the collection and verification of complete and accurate data on borrower characteristics, loan characteristics, property characteristics, and credit history to guard against omitted variables and data errors that may bias results.⁸ Because of widespread differences between whites and minorities in income, wealth, property values, and credit histories, analysis that fails to account fully for these factors may seriously overstate the extent of discrimination in mortgage loan approvals. Moreover, future analysis should explore alternative versions of a loan approval model and test extensively for possible interrelationships among explanatory variables to generate unbiased results.

In order to test the hypothesis that high rejection rates for minorities are entirely due to legitimate underwriting criteria, researchers need to assemble and analyze data on loan performance and defaults as well as information on loan applications and originations. As discussed earlier, evidence of higher default rates among minority borrowers than among whites *does not* prove the absence of discrimination at the loan approval stage. However, analysis of loan defaults does have an important role to play in the analysis of possible disparate impact discrimination. Specifically, underwriting policies and practices that disproportionately affect minorities even when they are even-handedly applied are discriminatory under the law if they do not serve a business necessity. Thus, if an underwriting criterion or requirement systematically disqualifies more minorities than whites, but does not reliably predict future loan performance, it



THE URBAN
INSTITUTE

is discriminatory. In fact, even if a criterion did predict future loan performance, it might be considered discriminatory if it could be replaced by an alternative criterion that had less of a disproportionate adverse effect on minorities. Data on underwriting criteria and loan terms, borrower and property characteristics, and long-term loan performance all need to be linked to support definitive analysis of disparate impacts in home mortgage lending.⁹

Finally, statistical analysis of discrimination in the loan approval process should attempt to distinguish discrimination based on the borrower's race or ethnicity from discrimination based on the racial or ethnic composition of the neighborhood in which a property is located. The existing empirical evidence on redlining (discrimination based on neighborhood composition) remains inconclusive. It may prove difficult to disentangle the effects of applicant race and neighborhood race, because most blacks currently live in black neighborhoods while most whites live in white neighborhoods. Nevertheless, the distinction is an important one from a policy perspective.

Design and Conduct Research on Loan Terms and Conditions

To date, relatively little statistical analysis has focused on the potential for discrimination in loan terms and conditions. Fair housing complaints often involve unfair terms and conditions for mortgage loans, and there are some indications that the lending industry is in the process of shifting from credit rationing to risk-based pricing. In other words, lenders may be more likely to charge higher interest rates and/or fees for customers they perceive to be risky, rather than denying them financing altogether. Thus, it will be increasingly important to understand how interest rates and fees are determined and to analyze the potential for either differential treatment or disparate impact discrimination in this area.

This issue is closely related to questions about credit scoring. Both risk-based pricing and credit-scoring schemes rely on data (or assumptions) about how the specific characteristics of borrowers relate to loan performance. More specifically, these schemes predict—or “score”—the risk associated with a particular borrower, based on past experience. Skeptics of risk-based pricing and credit scoring argue that the experience from which these predictive models are based may not be sufficiently diverse to reflect the favorable performance of loans to minorities and that the variables used in these models may put minorities at an unfair disadvantage. Moreover, none of these schemes has been evaluated by scholars. Thus, rigorous, objective analysis of the relationship between various borrower characteristics and loan performance is critically needed. Otherwise, these schemes may simply institutionalize disparate impact discrimination by imposing rules that put minorities at a disadvantage but that do not serve any business necessity.

In addition, researchers need to investigate systematically the uses of risk-based pricing and credit-scoring schemes, analyzing the criteria and procedures lenders use to determine interest rates and fees for individual borrowers. This type of research should be used to develop methods for analyzing the potential for either differential treatment or disparate impact discrimination. As several existing studies point out, it is not sufficient simply to compare the final



interest rates charged to different groups. Instead, analysis should compare final interest rates to the rates originally quoted when borrowers first inquired. And researchers should attempt to collect and analyze information on various loan fees, again exploring differences between “advertised” and “actual” fees.

Further Evaluate “Best Practices” for Remedyng Discrimination

In order to achieve significant reductions in mortgage lending discrimination, regulatory agencies must do a better job of identifying institutions that are discriminating. But, in addition, both regulators and lenders need to know what it takes to eliminate discriminatory practices. To the extent that discrimination is blatant and intentional, designing corrective remedies may be relatively straightforward. But much of the evidence summarized here suggests that lending institutions may be discriminating without realizing it—through policies and procedures that have a disparate impact on minority borrowers, through subtle differences in the level of encouragement and assistance provided to whites and minorities, or through unexamined assumptions about the types of products and terms for which minorities can qualify. Lending institutions may believe that their practices and decisions have been “color blind,” and the institutional changes they need to make to eliminate discrimination may not be obvious.

Fair lending advocates and industry experts have identified a set of strategies that lending institutions should implement in order to comply with anti-discrimination laws. Although these “best practices” appear logical and worthwhile, their effectiveness has not been systematically evaluated. Currently, there is a tendency to identify lending institutions as “high performers” if they are implementing a widely accepted set of best practices, not because they have eliminated unequal treatment of minorities. In other words, researchers need to compare fair lending performance for institutions with and without these best practices or for institutions implementing different remedial strategies. The goal of this research is to test the efficacy of various remedies and institutional reforms that lenders implement.

Finally, lending institutions need tools they can use to monitor and assess their own anti-discrimination efforts. The “stick” of litigation or regulatory action obviously creates an important incentive for lenders to care about the potential for discrimination in their policies and procedures. But lenders cannot take action if they do not realize that they are discriminating, and neither regulators nor fair housing groups have sufficient resources to investigate all lending institutions. Self-testing is one strategy lenders can and should use to monitor their performance and identify any problems that may exist. Research efforts that refine and promote practical methods for lenders to monitor and assess their own performance could help advance the cause of equal access to mortgage loans for minority homebuyers.



THE URBAN
INSTITUTE

Notes

1. The Fair Housing Act, 42 U.S.C.A. §3601 *et seq.*; the Equal Credit Opportunity Act, 15 U.S.C.A. §1691 *et seq.*; and the Civil Rights Act of 1866, 42 U.S.C.A. §§1981, 1982.
2. Federal law also prohibits discrimination in housing based on sex, family composition, religion, and disability. This volume focuses on the issue of racial and ethnic discrimination.
3. For a comprehensive discussion of the myriad and complex issues involved in legal and analytic investigations of mortgage lending discrimination, see Goering and Wienk (1996).
4. This report does not address potential discrimination by other important actors—such as real estate brokers, appraisers, insurers, and secondary loan institutions—who are not direct decisionmakers in the mortgage lending decision.
5. We use the term “color blind” in this volume to refer to policies and practices that appear to treat people equally regardless of their race or ethnicity.
6. The Policy Statement on Discrimination in Lending (*Federal Register* 1994) states that a business necessity must be manifest and may not be hypothetical or speculative. Factors that may be relevant to the justification could include cost and profitability.
7. See 18 U.S.C.S. §1014. Note that this would not bar lenders from conducting self-testing.
8. Although assembling such a database presents significant challenges, federal government regulators have sufficient leverage and resources to obtain the necessary information from lending institutions if they make it a priority.
9. For more information on the data and analysis required to test the business necessity of key underwriting standards, see Temkin et al. 1998.



THE URBAN
INSTITUTE

Chapter 2

New Evidence from Lender Testing: Discrimination at the Pre-Application Stage

ROBIN SMITH
MICHELLE DELAIR

In 1993, the U.S. Department of Housing and Urban Development (HUD) awarded funds through the Fair Housing Initiatives Program (FHIP) to the National Fair Housing Alliance (NFHA) for the paired testing of mortgage lenders in several cities across the United States. The tests were conducted by fair housing enforcement organizations using testers who posed as first-time homebuyers or refinancers at the pre-application stage. In most cases, testers visited lenders in person to inquire about the types and terms of loans for which they might qualify. These tests provide information on the actual treatment of white and minority applicants at the earliest phase of the loan application process and are the only test database available on the mortgage market that is large enough for statistical analysis.¹

At the pre-application stage of the mortgage loan process, prospective borrowers are collecting information to determine whether they can afford to purchase a home and what loan products may meet their needs. Tests conducted at this point provide information on the treatment of prospective borrowers, types of loan products discussed, and estimates of monthly payment and closing costs. In pre-application testing, a prospective borrower does not formally apply for the loan. No information is available on loan approval or the actual terms and conditions offered to these prospective borrowers later in the process.²

In 1998, NFHA gave the Urban Institute access to the mortgage lending tests it conducted under FHIP, and the Urban Institute constructed a database of selected information from the NFHA test report forms for statistical analysis. This database includes only discrete information items recorded in response to closed-ended questions. The more open-ended narrative material provided by testers is not included because it is not appropriate for a strictly statistical approach to the evidence.

Findings indicate that race-based differential treatment is occurring in some cities among the lenders that were tested. Although the data set does not provide enough information for estimating the incidence of discrimination in the broader market, it shows that testing can be used to measure statistically significant levels of differential treatment at specific stages of the lending process. Moreover, the wide variation we found by city suggests that regional market dynamics influence lender practices and that investigating local markets may be critical to the design and interpretation of lending tests.

The NFHA Tests

The NFHA FHIP-funded fair lending study was designed for enforcement purposes (both developing methodology for enforcement testing and gaining evidence for litigation). As such, it used test report forms and sampling techniques intended to further enforcement rather than research goals. Research testing seeks to quantify the incidence and forms of discrimination across a representative sample of firms and/or markets.³ To do this, design elements (such as sampling methods, test report forms, and the number of tests) are crafted to produce generalizable, statistical results. In contrast, testing done for enforcement is concerned less with a rigorous sampling methodology and more with identifying discriminatory lenders and generating legal evidence of discrimination. Both research and enforcement testing are based on the same core methodology and protocols. They differ primarily in the way results are recorded, analyzed, and used.⁴

NFHA conducted tests in seven cities (Atlanta, Chicago, Dallas, Denver, Detroit, Oakland, and Richmond), with roughly two-thirds of the tests concentrated in two of them (Chicago and Oakland).⁵ Some tests included as many as five test parts in order to distinguish differential treatment based on the race of the applicant and the racial composition of the neighborhood.

Tester pairs were matched on their front- and back-end ratios. These ratios compare a household's prospective mortgage loan payments and all debt payments to its income.⁶ NFHA matched testers on ratios rather than actual debt, income, and purchase amounts because it believed that minority testers inquiring about houses in minority neighborhoods at prices comparable to those prevailing in white neighborhoods would not be credible to lenders. This belief was based on the fact that homes in minority neighborhoods do not command comparable prices.⁷ NFHA felt that prospective borrowers with similar ratios should be treated in a similar manner by lenders and that a test scenario with inflated home prices for minority neighborhoods might compromise the whole testing effort.



The lenders tested in the NFHA study were selected as part of a purposive sample with three different selection groups. In the first two testing sites, Oakland and Chicago, lenders were selected through a structured process based on market share and HMDA performance relating to the share of loan applications from minority individuals and minority neighborhoods. The second selection group comprised follow-up tests conducted on lenders in Oakland and Chicago, where evidence of differential treatment was found during the initial test. The third and final group of tests were conducted in Atlanta, Dallas, Denver, Detroit, and Richmond. These lenders were selected based on differential treatment detected during testing of the same companies in Oakland and Chicago.

Since this sample of lenders was developed for enforcement purposes rather than according to a research-oriented, randomized sampling design, the resulting data set cannot be weighted to produce an estimate of differential treatment in the general market for the country or even for any of these cities in which the tests were performed. Thus, care must be taken to not generalize from the findings reported here.⁸ The Urban Institute's estimates of differential treatment apply only to the *specific* samples of lenders tested in the NFHA tests. Moreover, since the majority of tests were conducted on reputedly "bad lenders," the results should also be considered to reflect upper-bound estimates of discrimination in the aspects of the pre-application process analyzed here.

The Urban Institute Reanalysis Data Set

The data set constructed by the Urban Institute for reanalysis differs significantly from NFHA's original set of tests. First, NFHA's analysis assessed treatment across all test parts within a test.⁹ The Urban Institute, in contrast, concentrates on differential treatment between any two testers in a test but not between three or more test parts at the same time.¹⁰ For example, the original test might include a white tester in a white neighborhood, a minority tester in a minority neighborhood, and a minority tester in a white neighborhood. This is a three-part test. In this case, NFHA's analysis would have looked at the treatment each tester received in relation to the other two. The Urban Institute reanalysis focuses on differences between the treatment of white testers in white neighborhoods and minority testers in minority neighborhoods.

In addition, the Urban Institute reanalysis uses only about one-third of the items on NFHA's original form (a copy of the data extraction form is in annex C).¹¹ Items were selected based on their importance in the test analysis and their suitability for statistical analysis. The Urban Institute reanalysis, therefore, does not use information from the open-ended questions and the comprehensive narratives written by the testers on their particular experiences including instances of coaching, encouragement, and underwriting exceptions made by loan officers. This type of open-ended information is appropriate and powerful for case-by-case enforcement and is relied on heavily in presentations to legal professionals and juries making decisions about differential treatment in a judicial context. But narrative information is not well suited to research testing, which must rely on closed-end data elements that lend themselves to the stan-



THE URBAN
INSTITUTE

dardized comparisons based on large numbers of test pairs that are necessary to draw statistically valid conclusions.¹²

Because these portions of the NFHA test report form are not included in our reanalysis database, we did not construct any composite measure. We focused instead on individual treatment items (such as contact length and number of quotes) and looked for potential patterns and forms of differential treatment with respect to those items. A fundamental aim of the Urban Institute reanalysis, in other words, is to compare treatment across standard elements without a narrative in order to see if differential treatment can be determined from statistical analysis alone.¹³

The core of the Urban Institute reanalysis data set consists of 150 two-part tests pairing an African American tester buying a home in an African American neighborhood with a white tester buying in a white neighborhood. Although other scenarios were used in some tests, the results presented here focus on this test structure because it is the most common type found in the data set.¹⁴ Six cities are represented in the Urban Institute reanalysis data set.¹⁵ Two-thirds of the tests were done in two of these cities, 58 in Chicago and 54 in Oakland. The other test numbers in the Urban Institute reanalysis are Richmond (14 tests), Atlanta (12 tests), Denver (8 tests), and Detroit (4 tests).¹⁶

Forty-three lenders are represented in the reanalysis data set, with the number of tests per lender varying (see table 1). At one end of the spectrum, one-third of the tests were conducted at the same seven banks across different cities.¹⁷ This is not surprising, given that lenders were deliberately retested when differential treatment was suspected in a previous test of that lender. At the other end of the spectrum, 12 of the 43 lenders in the sample were tested only once.

Results

The most basic measure of service in pre-application testing is whether a tester is seen by a lender and given access to information on loan products. Testers attempted to get “a quote,” defined as information about a loan product with an

Table 1. Seven Lenders with the Most Tests by City

| | Atlanta | Chicago | Denver | Detroit | Oakland | Richmond | Total |
|------------------|---------|---------|--------|---------|---------|----------|-----------|
| Lender 1 | 1 | 2 | 0 | 2 | 3 | 3 | 11 |
| Lender 2 | 0 | 0 | 2 | 0 | 5 | 3 | 10 |
| Lender 3 | 0 | 3 | 0 | 1 | 4 | 0 | 8 |
| Lender 4 | 0 | 2 | 1 | 0 | 3 | 1 | 7 |
| Lender 5 | 0 | 1 | 2 | 0 | 0 | 3 | 6 |
| Lender 6 | 2 | 1 | 0 | 0 | 0 | 3 | 6 |
| Lender 7* | 0 | 0 | 0 | 0 | 6 | 0 | 6 |
| Total | 3 | 9 | 5 | 3 | 21 | 13 | 54 |

*This lender did not operate in other cities in the sample.



estimate of a monthly mortgage payment and closing costs.¹⁸ In most tests, both testers were given quotes. However, significant differences in treatment were detected in some cities (see table 2).

In four of the five cities in the data set, African American testers were more likely to be denied a quote than white testers; in two, these differences are statistically significant. In both Chicago and Atlanta, minority testers were significantly more likely to be denied a quote than their white counterparts.¹⁹ The denial of a quote includes both cases where a person is not allowed to speak to a loan officer (i.e., unless the person fills out an application or has a credit check) and cases where a tester speaks to a lender but is not given information on the basic question (“What type of loan do I qualify for and how much is it going to cost?”).

Besides getting in the door, how much time a tester is allowed with a loan officer may provide information on differential treatment. Presumably, lenders providing coaching and extra information will spend more time, on average, with a prospective applicant. However, the information gathered in the Urban Institute reanalysis does not reveal the content of the conversation a lender had with a tester, or whether more time means more chatting, coaching, or some other type of interaction.

In four of the five test cities, lenders spent more time with white testers than they did with their minority partners (see table 3). For Atlanta this difference is statistically significant.²⁰ In Atlanta, lenders spent an average of almost 30 more minutes with white than with minority testers. However, in Oakland, lenders spent more time with minority (47 minutes) than with white (41 minutes) testers.

Lenders did not only spend more time with whites. In most cities they also provided more quantifiable information to whites. Whereas table 2 looked at who received an estimate of monthly loan payments and who was denied such information, table 4 looks at who got more quotes in tests where both testers received quotes.

White testers received significantly more quotes than their minority partners in Atlanta, Chicago, and Denver—all cities where more time was also spent with whites. Lenders in Richmond, in contrast, were slightly more likely to give minority testers more quotes, although they did not spend more time with them.

It would be interesting to know more about the types of additional quotes testers received, because multiple quotes may represent either different products or different interest rates or points on the same product.²¹ It is also not clear

Table 2. Who Gets Quotes?

| | N | White/Not Black | Black/Not White | Both | Neither | t-statistic |
|------------------|-----------|-----------------|-----------------|------------|-----------|--------------|
| Chicago** | 58 | 16% | 3% | 76% | 5% | -1.82 |
| Oakland | 54 | 11% | 13% | 70% | 6% | 0 |
| Atlanta** | 12 | 25% | 0% | 75% | 0% | -2.35 |
| Denver | 8 | 13% | 0% | 87% | 0% | 0 |
| Richmond | 14 | 14% | 7% | 79% | 0% | -0.56 |

Using a one-tailed t-test because the direction of unfavorable treatment is unambiguous: *Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.



THE URBAN
INSTITUTE

Table 3. Contact Length Differences

| | N | White Longer | Black Longer | Equal | t-statistic |
|-------------------|-----------|--------------|--------------|------------|--------------|
| Chicago | 55 | 40% | 35% | 25% | 0.18 |
| Oakland* | 48 | 33% | 46% | 21% | -1.95 |
| Atlanta*** | 10 | 80% | 0% | 20% | 3.47 |
| Denver | 7 | 57% | 14% | 29% | 1.04 |
| Richmond | 13 | 54% | 46% | 0% | 0.50 |

Using a two-tailed *t*-test: * Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

Note: For this analysis, we used a range of ± 5 minutes to determine contact length differences given that the average contact time for both testers was around 45 minutes.

how the simple fact of more quotes should be interpreted. On the one hand, the number of quotes provides a proxy for the level of service, with more quotes possibly representing more information and more choice. On the other hand, more quotes could reflect other elements of a transaction, such as product steering. If both testers are told about conventional, fixed-rate loan products, but only one is also recommended to the FHA (Federal Housing Administration) program, is this favorable or unfavorable treatment? The second tester may benefit from additional information or may be steered to a less desirable program or product.

The Urban Institute reanalysis data set does include useful information on whether a lender discussed FHA (see table 5). Fewer than half of all testers were told about FHA loans, and there are dramatic differences by city in how often FHA was discussed with at least one of the testers. In Denver, someone in every test pair was told about FHA. In Oakland, no one was told about FHA in 88 percent of the tests.²² In Chicago, African American testers were significantly more likely than their white partners to be told about FHA.²³ In Atlanta, Denver, and Richmond, white testers were more likely to hear about FHA than minority testers, although the differences are not statistically significant.

These differences within and across cities speak to the different uses and value placed on FHA as a loan product. In some markets, FHA may be viewed as a less desirable loan product or reserved for more risky clients. In other markets, FHA may be viewed as advantageous because of features such as low down payment requirements. It is known that black borrowers are substantially

Table 4. Who Gets More Quotes?

| | N | White | Black | Equal | t-statistic |
|------------------|-----------|------------|------------|------------|-------------|
| Chicago** | 58 | 31% | 14% | 55% | 1.93 |
| Oakland | 54 | 24% | 24% | 52% | -0.17 |
| Atlanta** | 12 | 50% | 17% | 33% | 1.86 |
| Denver*** | 8 | 63% | 0% | 37% | 3.00 |
| Richmond | 14 | 28% | 36% | 36% | -0.25 |

Using a one-tailed *t*-test: *Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

Note: A one-tailed paired *t*-test is used for this analysis because the direction of unfavorable treatment is unambiguous.



Table 5. FHA Discussed

| | N | White/Not Black | Black/Not White | Both | Neither | t-statistic |
|-------------------|-----------|-----------------|-----------------|------------|------------|-------------|
| Chicago*** | 58 | 2% | 22% | 12% | 64% | 3.51 |
| Oakland | 54 | 4% | 6% | 2% | 88% | 0.44 |
| Atlanta | 12 | 33% | 17% | 25% | 25% | -0.80 |
| Denver | 8 | 25% | 0% | 75% | 0% | -1.53 |
| Richmond | 14 | 36% | 21% | 36% | 7% | -0.69 |

Using a two-tailed t-test: *Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

more likely to use the FHA program, but it is unclear why this disparity exists and what role lenders play in the process (Gabriel 1996). Although the results presented here are mixed, the Chicago results suggest that lender behavior may play a role in some cities.

The Urban Institute reanalysis captures information from the NFHA test report form on the questions lenders asked testers. Two questions of particular interest are about income and debts (see tables 6 and 7). Not asking for such information may indicate that a lender is not treating a person seriously as a potential applicant. If a lender does not have information on a prospective borrower's income and debts, in other words, it is hard to accept any quote that such a person receives as tailored specifically for that person because a lender needs that information to determine if the person qualifies for a loan.

Only in Chicago was the difference between the test partners statistically significant in how often both of these questions were asked. Lenders there were more likely to ask white testers about both income and debts. This is particularly interesting given how often minority testers were recommended to FHA in Chicago. In 21 percent of the Chicago tests, for example, lenders did not get income and debt information from minority testers but they did discuss FHA with them. This may imply that the lenders tested in Chicago steered minority testers to FHA based on race as opposed to income and debt calculations.

As discussed previously, the NFHA test report information on the type of loans discussed with testers is difficult to interpret without background information on lenders' products and programs. Although our ability to analyze the specific quotes received by testers is limited, some analysis is possible using

Table 6. Lender Requests for Income Amount

| | N | White/Not Black | Black/Not White | Both | Neither | t-statistic |
|-----------------|-----------|-----------------|-----------------|------------|------------|-------------|
| Chicago* | 57 | 19% | 7% | 58% | 16% | 1.85 |
| Oakland | 51 | 20% | 8% | 63% | 9% | 1.63 |
| Atlanta | 12 | 17% | 8% | 67% | 8% | 0.56 |
| Denver | 8 | 25% | 0% | 75% | 0% | 1.53 |
| Richmond | 13 | 0% | 15% | 85% | 0% | -1.48 |

Using a two-tailed t-test: *Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.



THE URBAN
INSTITUTE

Table 7. *Lender Requests for Debt Information*

| | N | White/Not Black | Black/Not White | Both | Neither | t-statistic |
|-------------------|-----------|-----------------|-----------------|------------|------------|-------------|
| Chicago*** | 58 | 26% | 5% | 57% | 12% | 3.02 |
| Oakland | 54 | 13% | 15% | 59% | 13% | -0.26 |
| Atlanta | 12 | 17% | 8% | 67% | 8% | 0.56 |
| Denver | 8 | 25% | 0% | 75% | 0% | 1.53 |
| Richmond | 13 | 8% | 8% | 84% | 0% | 0 |

Using a two-tailed *t*-test: *Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

data from 66 tests where both testers were quoted a product with a 30-year term.²⁴ While it is not known if the loan products quoted are identical, we assume that both tests refer to conventional, fixed-rate loan products with a 30-year term. Differences between testers in the monthly payment costs quoted in these 30-year loans sometimes favor whites and sometimes favor blacks.²⁵ The only statistically significant disparity—more expensive monthly payments quoted to African American testers in Chicago—favors whites (see table 8).

These mixed results on quoted loan terms reinforce the need to know more about the type of loan product and the specific details of a quote's individual parts (interest rate, points, tax rate, insurance assumptions). It could be argued that information and encouragement are more important at the pre-application stage than the financial details of the quotes.²⁶ Even so, providing a meaningful quote is part of the “encouragement” process, and this key piece of information may determine whether prospective borrowers think they can afford to purchase a home.

A baseline comparison of the interest rates quoted to each test partner in the tests with a 30-year term reveals that African American testers were more likely to receive a higher rate quote than their white partners in three of the five cities, a difference that is statistically significant in Atlanta (see table 9).

A complete analysis of interest rates would require more information on the timing and sequencing of test pairs because the frequent and legitimate changes in mortgage interest rates could lead to test partners being legitimately quoted different rates. However, as long as test management does not introduce bias into the process (for example, if all minority testers visited in a particular week

Table 8. *Total Monthly Payment Cost per \$10,000 of Loan Amount*

| | N | White Greater | Black Greater | Equal | t-statistic |
|-----------------|-----------|---------------|---------------|------------|--------------|
| Chicago* | 24 | 0% | 8% | 92% | -1.75 |
| Oakland | 25 | 16% | 8% | 76% | 1.58 |
| Atlanta | 6 | 0% | 67% | 33% | -0.99 |
| Denver | 5 | 20% | 20% | 60% | 0.10 |
| Richmond | 8 | 25% | 0% | 75% | 1.27 |

Using a two-tailed *t*-test: *Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.



Table 9. Interest Rate Differences

| | N | White Greater | Black Greater | Equal | t-statistic |
|------------------|----------|---------------|---------------|------------|--------------|
| Chicago | 30 | 33% | 40% | 27% | -1.37 |
| Oakland | 26 | 35% | 42% | 23% | -0.37 |
| Atlanta** | 7 | 0% | 71% | 29% | -3.27 |
| Denver | 5 | 40% | 40% | 2% | 0.95 |
| Richmond | 7 | 57% | 43% | 0% | 1.32 |

Using a two-tailed t-test: *Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

when higher rates are posted), such market fluctuations should cancel one another out across multiple test pairs.²⁷

Regional Differences

As noted, the unique sample developed for the NFHA tests, combined with the small number of tests available for each city, means that results should be viewed as suggestive, not definitive. This caution applies even more strongly when interpreting differences between cities. It is compelling that several treatment variables show the same general trend across cities (such as minorities being denied a quote), but it is equally interesting that there are substantial differences between cities. In particular, the two cities with the most tests, Chicago and Oakland, often yield opposite results. This suggests that the fair lending story at the pre-application stage may differ greatly across markets.²⁸

Chicago is the only site where differences in treatment are consistently statistically significant, and these results are always unfavorable to minority testers. African Americans in Chicago were significantly more likely to be denied a quote, receive fewer quotes, and be told about FHA. Further, the Chicago lenders tested by NFHA did not ask minorities for important information on income and debts as often as they did white testers, and the quotes minorities received were more expensive. The consistency of these findings is compelling.

However, this evidence of pre-application discrimination in Chicago is very different from the pattern of behavior found in Oakland. In Oakland, differences in treatment between white and minority testers were rarely significant and, more important, when disparities existed they often benefited the African American tester. This contrast reinforces the importance of understanding regional markets and analyzing findings by city.²⁹

Lessons for Future Testing

Since the NFHA tests were designed for enforcement purposes—both to refine methodology for enforcement testing and to produce direct evidence for litigation—the results reported here are not generalizable either nationally or locally.



THE URBAN
INSTITUTE

Moreover, the Urban Institute reanalysis does not include important treatment items such as coaching and exceptions, does not consider narrative data, and does not include full information on loan products—gaps that make it impossible to construct an overall measure of the incidence of differential treatment within our sample.

Nevertheless, the Urban Institute reanalysis does show that statistical analysis of selected items from mortgage lending test data can, and does, detect differential treatment. In particular, the findings suggest that for the lenders tested, African Americans in some cities are more likely to be denied a quote and to receive fewer loan alternatives than their white partners. These important findings suggest that more research testing is needed to fully understand and measure differential treatment at the pre-application stage of the mortgage process.

Such additional testing will be as challenging as it is important.³⁰ One of the primary challenges is to develop rigorous sampling rules geared to making the lender sample representative of the market place. City selection is also vitally important for any study that hopes to generate national incidence measures. The Urban Institute reanalysis finds mixed results across cities for specific behaviors, indicating that there may not, in fact, be a national mortgage market in the sense of having generally applied standards of behavior. Moreover, differences between the relative treatment of whites and minorities in different areas of the country may be large, in which case aggregate national incidence measures are likely to obscure important local patterns.

Local practices and standards may also influence how different products and actions are perceived and marketed in different cities, making it difficult to interpret certain outcomes as positive or negative. For example, how FHA is viewed and used may vary significantly by city and by lender. In this analysis, sampled Chicago lenders are directing blacks to FHA, but sampled lenders in Richmond are more likely to discuss FHA with their white clients. That the treatment is different is enough to raise concern, but not enough to make clear how to interpret the outcome. Moreover, once again, relying on an overall national incidence measure could mask key regional differences—showing no differential treatment when analyses of individual cities or lenders could demonstrate significant differences.

In addition to regional differences in the mortgage market, products and programs may vary among lenders within a city. This is part of the reason why so much information on the NFHA test forms was recorded in the narrative and in open-ended entries to collect the details of behavior that can vary tremendously. Structuring closed-ended items to accurately capture the product options in local markets requires a tremendous amount of advance information. Future testing efforts may benefit from “scouting” or “shopping” lenders before testing to obtain a better understanding of lender practices and product lines.³¹ This kind of information might save test resources in the long run, better prepare testers, and improve analysis by increasing the quality of information on loan products that is recorded by testers on test forms.

Scouting and shopping lenders before the full-scale tests can aid in developing forms and protocols that more accurately reflect the test experience.³² The pre-application stage of the mortgage process is about getting information. A tester is simulating a transaction where a would-be borrower approaches a



lender to find out if he or she can and should enter the market. To analyze the test appropriately, information on items that are not easily quantified, such as coaching and other forms of encouragement, is necessary. In order to be used in a statistical analysis, this information must be distilled and recorded in mutually exclusive, closed-ended elements. This requires an in-depth knowledge of the mortgage process and strict tester training and management.

Data collection forms can also build in summary variables that measure the amount of information provided. Summary variables can flag tests where a tester did not receive a quote or some other key piece of information. Summary variables can also indicate the number of quotes a lender offered on different products or the number of calculations prepared with different interest rates. In this way, more information will be available for comparisons of the type and quality of information each tester received.

The NFHA tests provide a good springboard for future testing efforts. The finding of differential treatment in the Urban Institute reanalysis sparks interest in the forms that discrimination may take at the pre-application stage and raises questions about how the circumstances of testers influence lender behavior. Future testing could employ different tester scenarios to determine both who is being discriminated against and what form the discrimination takes. For example, experts believe that differential treatment may often occur in situations where a lender, salesperson, or real estate professional has to “go the extra mile” to serve a marginally qualified person (Siegelman 1998).³³ This “race plus” hypothesis is an important area for investigation in the mortgage lending arena.

Further work could also explore the types of products discussed and recommended to testers to see if lenders are steering testers to different products by race. If more information is desired on differential treatment within specific elements in product quotes (interest rates, reserves, escrow amount, etc.), testers could be primed to ask for similar products as opposed to posing as customers who are fully ignorant of the market. In this way, comparisons could be made of the specific quotes given to testers as well as the help they received in making a product work for them.

Urban Institute reanalysis of NFHA’s enforcement testing data for a selected group of mortgage lenders shows that differential treatment is often present in some cities at the pre-application stage in statistically significant ways, such as being denied a quote and given fewer options on mortgage products. Future research could uncover more about other forms of differential treatment and the level of such treatment in the broader marketplace and could explore the circumstances of those likely to be discriminated against.



THE URBAN
INSTITUTE

Annex A

Mean Differences between White and Black Testers in Contact Length

| <i>Contact Length Differences</i> | | |
|-----------------------------------|-----------|---------------------------|
| | N | Mean Difference (minutes) |
| Chicago | 55 | 0.6 |
| Oakland | 48 | -6.4 |
| Atlanta | 10 | 27 |
| Denver | 7 | 13.7 |
| Richmond | 13 | 7.9 |



THE URBAN
INSTITUTE

Annex B

Information on Requested Loan Amounts by City

| <i>Requested Loan Amounts by City</i> | | |
|---------------------------------------|--------------------|-----------|
| | Median Loan Amount | |
| | Black | White |
| Chicago | \$93,100 | \$89,100 |
| Oakland | \$152,910 | \$152,910 |
| Atlanta | \$63,825 | \$109,153 |
| Denver | \$63,734 | \$91,105 |
| Richmond | \$75,030 | \$77,800 |



THE URBAN
INSTITUTE

Annex C

Reanalysis Data Collection Form

Data Collection Form Fair Lending Test Reports

Full Test # _____

| Site | Test # | Test Part | NIC | Type | Purp | R/NO | T/V |
|------|--------|-----------|-----|------|------|------|-----|
|------|--------|-----------|-----|------|------|------|-----|

Lender Name _____

Lender Address _____

Date of Contact ____ / ____ / ____

Type of Contact ____ (a = telephone, b = site visit)

Length of in-person contact ____ (enter in minutes)

Type of site visit ____ (a = appointment, b = approved walk-in, c = cold walk-in, d = other)

Wait time to be interviewed (in minutes) ____

Name of person who interviewed ____

Race/ethnicity of person who interviewed ____

Type of transaction for this test ____ (enter letter of response checked on form)

For the following items enter

A = asked, R = recorded, V = volunteered, N = not asked

Source of income _____

Amount of income _____

Debts/liabilities _____

Credit standing _____

Types of loans discussed and/or recommended by lender

(Check responses from test form)

| | Discussed | Recommended |
|--|-----------|-------------|
|--|-----------|-------------|

Fixed-rate mortgage _____

Adjustable rate mortgage _____

Balloon mortgage _____

FHA mortgage _____

VA mortgage _____

Community homebuyers program _____

First-time homebuyers program _____

Other _____

Other _____



Information on each loan recommended

(Check here if no loans recommended)

Recommendation #1 Recommendation #2

| | | |
|---------------------------------|---------|---------|
| Type of loan | _____ | _____ |
| Amount of loan | \$_____ | \$_____ |
| Loan term in years | _____ | _____ |
| Interest rate | _____ | _____ |
| Points | _____ | _____ |
| Monthly principal plus interest | \$_____ | \$_____ |
| Monthly taxes | \$_____ | \$_____ |
| Monthly homeowners insurance | \$_____ | \$_____ |
| Monthly PMI | \$_____ | \$_____ |
| Other monthly expense | \$_____ | \$_____ |
| Total monthly expense | \$_____ | \$_____ |
| Other _____ | _____ | _____ |

Information provided about closing and other costs

Type/term of loan these costs refer to

Interest rate for this loan

\$_____ Points (at ___ points)

\$_____ Homeowners insurance premium

\$_____ PMI premium

\$_____ Interest adjustment

\$_____ Monthly payment reserve

\$_____ Document preparation

\$_____ Title examination/title insurance

\$_____ Real estate tax escrow

\$_____ Attorneys' fees

\$_____ Down payment

\$_____ Appraisal

\$_____ Other _____

\$_____ Other _____

\$_____ Total closing/other costs

Closing costs provided in writing

a = good faith estimate, b = lending institution's form, c = computer printout, d = written on blank paper,
e = other, n = no answer checked)

Amount of application fee \$_____**How long did lender say application process would take (in days)?** _____**Status of written application** _____THE URBAN
INSTITUTE

Did test include evidence of phone or mail follow-up by lender after test?

- No
- Phone follow-up
- Mail follow-up
- Both phone and mail follow-up
- Other



THE URBAN
INSTITUTE

Notes

1. There has been a limited amount of lender testing by regulatory agencies and private fair housing organizations and self-testing by lenders. Most data sources either are not available for public review or are not large enough for statistical analysis. NFHA's test information presents a unique opportunity for independent statistical review of test data. The authors greatly appreciate NFHA's willingness and cooperation in giving us access to its data.
2. Most lending tests have stopped at the pre-application stage because federal law makes it illegal to knowingly provide false information on a credit application with intent to defraud. However, some testing advocates argue that submitting false information on a mortgage application as part of a paired test does not violate this law because the tester will not actually borrow money and therefore does not intend to defraud. The question has not yet been settled through litigation or regulatory rulings.
3. Testing is also referred to as auditing when used in a research context. This report uses the term *testing* for tests (or audits) conducted for either enforcement or research purposes.
4. For a more complete discussion, see Fix, Galster, and Struyk (1993).
5. The NFHA tests were conducted by private fair housing organizations working in conjunction with NFHA. NFHA developed all tester training manuals, protocols, and forms for the lender tests under FHIP and worked directly with a fair housing group in each city to recruit and train testers and locate test homes. Staff from NFHA and the local fair housing organizations jointly managed the tests. NFHA analyzed all tests.
6. The Urban Institute does not have detailed information on tester scenarios used in NFHA's tests, including on the actual income assigned to test partners within a pair. The FHIP test coordinator for NFHA, Cathy Cloud, reports that income was directly correlated to loan amount (see footnote below) and that, depending on the desired ratio, partners were matched as closely as possible in actual income.
7. The Urban Institute does not have information on the prices of homes used in the NFHA tests. However, the requested loan amounts were recorded on the test report form and are indicative of home price. Details on the median values of requested loan amounts by city are in annex B at the end of this chapter. For the two largest sites, requested loan values were the same (Oakland) or higher for minorities (Chicago).
8. Since results are not generalizable, no overall incidence measure is given across sites for the entire reanalysis sample.
9. The statistical test used for the reanalysis is the paired *t*-test, which looks at the relationship between pairs across tests. This statistical tool does not look at the relationship between three or more parts at one time.
10. To review NFHA's analysis of the full data set, refer to Smith and Cloud (1996).
11. The extraction form includes some data elements not analyzed here. This is because unavailable data reduced sample sizes in some cases, particularly when further divided by city, to the point where statistical analysis became impossible.
12. While narratives may be part of research test forms, in the research context they are used only as a secondary source of information for missing items or to corroborate items recorded elsewhere.
13. It is accepted in many areas of civil rights enforcement (housing, education, employment) that statistics showing serious racial disparities may indicate illegal differential treatment (Ritter 1996). However, the small amount of mortgage testing research available to date does not provide definitive evidence on how to collect and interpret statistical testing evidence appropriately in this field.
14. Ninety-five percent of the tests in the data set included a white tester linked with a home in a white neighborhood and an African American tester purchasing a home in an African American neighborhood. This test structure does not separate discrimination against individuals and neighborhoods, and subsequent analysis captures the combined differential treatment against both individuals and neighborhoods. Some tests included test parts with



THE URBAN
INSTITUTE

other tester scenarios such as a minority tester in a white neighborhood or a white tester in an integrated or changing neighborhood. Given the small sample sizes, the most common test type is presented in this analysis.

15. Dallas is not represented because no data were available on the relatively few tests conducted there. Most analysis also excludes Detroit because of the small number of tests done there.
16. While the reanalysis data set includes 150 tests, the small sample sizes in individual cities and missing data for some items (particularly because all types of activity did not occur in each test) limit our ability to analyze individual variables. Sample sizes are listed where appropriate.
17. Given the nature of this sample and the inability to generalize from the findings, the data set has not been adjusted to account for lender clustering.
18. The test report form captured a total monthly payment and individual parts of that total payment including monthly principal, interest, taxes, and insurance costs. A lender may or may not give a tester each of these items and a total. For this analysis, a test included a quote if it had an estimate of a total monthly mortgage payment OR an estimate that included at least the principal and interest portion of a total monthly payment (but no total recorded on the test form).
19. As noted earlier, the Urban Institute reanalysis data set did not include tester identifiers connected to each test. Therefore, analysis of the data set was conducted by pooling the tests without adjusting for tester clustering or homogeneity across pairs. For a fuller discussion of the merits of adjusting for tester effects, see Heckman and Siegelman (1993).
20. Mean differences in minutes between the white and black test pairs in each city are presented in annex A at the end of this chapter.
21. In this analysis, a tester is considered to have two quotes if the test form included quotes with different monthly payments. The type of loan for each of these payments is not known. The different monthly payments could reflect two calculations on the same loan product at different interest rates or completely different loan products (with a different mix of requirements, benefits, etc.). The level of “additional” information provided by the second quotes may not be equal. This analysis does not have enough information to judge the relative and comparable desirability and usability of the quotes received. It considers more information of any type to be favorable treatment.
22. In California, the prices of most of the homes made them too expensive to be eligible for FHA.
23. Advocates have long contended that FHA in Chicago has been an agent of neighborhood decline in minority communities; see Bradford (1979). In our analysis, the trend in Chicago for minorities to be told about FHA to a greater degree than whites was consistent regardless of the race of the neighborhood. When analyzing test pairs that included a white tester in a white neighborhood and a minority tester in a white neighborhood, the minority tester was told about FHA in 25 percent more cases. When whites with homes in integrated neighborhoods were paired with minorities in minority neighborhoods, the minority tester was still told about FHA in 57 percent more cases.
24. While this analysis is not definitive or generalizable, it does offer important insights for future testing on how the market works, the types of quotes possible, and how to construct a test instrument to capture loan information that can be statistically analyzed.
25. Monthly payment costs were standardized to be cost per \$10,000 of loan amount to account for differences in loan amounts across testers. Differences in monthly payment costs were determined using a range of ± 1 percent of the standardized monthly payment cost. Without this range, no test pair had exactly equal standardized monthly payments, but a good portion of the tests were within 1 percent of each other.
26. Again, the Urban Institute reanalysis does not have information on coaching or exceptions that might have been part of tests where similar product quotes were offered.
27. In more than two-thirds of the tests, minority testers preceded their white counterparts. However, the staggered nature of the tests led to white testers in different pairs conducting their tests on the same day or week as minority testers at different banks.



28. It would be interesting to analyze test results by lender across cities to see if differences are consistent within a given lender regardless of city. This is not possible with the Urban Institute reanalysis data set, because the already small sample sizes are reduced further by the different information collected in each test (for example, not all tests have a quote that can be compared).
29. NFHA believes that analysis by lender across cities is critical for enforcement purposes. However, for the Urban Institute reanalysis small sample sizes make such an analysis highly problematic.
30. For a fuller discussion, see Galster (1993). In 1998, HUD commissioned the Urban Institute to conduct a new paired testing study of potential lending discrimination at the pre-application stage. This study will explore and develop sampling methods, test scenarios, and testing protocols in order to advance paired testing methodologies for both enforcement and research.
31. NFHA found that in-depth knowledge gained during its scouting efforts was critical to successful testing and analysis.
32. Research testing differs from testing done for enforcement purposes and as such may have different test structures and test report forms. For example, in research testing, the selection of the lender and the issues being tested are determined by a research design, not by an actual complaint. Suggestions made here specifically target research testing.
33. NFHA notes that the most serious types of differential treatment it found in its testing program occurred in these types of tests.



THE URBAN
INSTITUTE

Chapter 3

Does Discrimination in Mortgage Lending Exist? The Boston Fed Study and Its Critics

STEPHEN L. ROSS
JOHN YINGER¹

The past decade has witnessed an explosion of scholarly interest in mortgage lending discrimination. The next three chapters in this report review the recent literature on this topic. This chapter focuses on the major study that has been at the core of the debate. Chapter 4 examines other recent research on discrimination in loan denial and on redlining in mortgage lending. Chapter 5 reviews the so-called default approach to studying mortgage lending discrimination. The focus is on articles published since 1996, because detailed literature reviews for earlier research are already available (Yinger 1996; Ladd 1998).

Background

The explosion in interest in mortgage discrimination was fueled by two events. First, starting in 1991, the Home Mortgage Disclosure Act (HMDA), as amended, resulted in the availability of data for the vast majority of mortgage loan applications in the country, complete with information on the outcome (approval or denial), the race and ethnicity of the applicant, and the location of the prop-

erty.² These data document wide disparities in loan approval across racial and ethnic groups—disparities that, in some cases, have hardly declined since 1991. In the case of conventional mortgages for home purchase, for example, the loan rejection rate for blacks was 2.07 times the rate for whites in 1991 and 2.05 times the rate for whites in 1997.³

Second, scholars at the Federal Reserve Bank of Boston supplemented the 1991 HMDA data on the greater Boston area with extensive information on each applicant's credit history and found that black and Hispanic applicants were 80 percent more likely than white applicants to be rejected for a loan after controlling for the characteristics of the loan, of the property and its neighborhood, and of the applicant (Munnell et al. 1996). This study, the so-called Boston Fed Study, has been subject to intense scrutiny and criticism since it was initially released in 1992, but it remains crucial to the literature because no comparable data set has yet been assembled.

Because of its centrality, the Boston Fed Study and criticisms of it are the subject of this chapter. We identify the major criticisms of this study and explore their validity using the public-use version of the data set that the study itself employed. We also examine the responses of the Boston Fed Study's authors to their critics.⁴ Because the broad issues in this literature are well known, we do not pause to present them here; instead, we build on the frameworks presented in Yinger (1996) and Ladd (1998).

Critics of the Boston Fed Study have identified many potential flaws in it. All of the potential flaws of which we are aware are reviewed in this section.⁵ To be specific, we investigate the claims that the Boston Fed Study overstates discrimination because of

- Omitted variables,
- Data errors in the explanatory variables,
- Misclassification in the dependent variable,
- Incorrect specification, and
- Endogenous explanatory variables.

Omitted Variables

Munnell et al. (1996) supplement the 1991 HMDA data for Boston with a survey of lenders that provides extensive information on an individual applicant's credit history, among other things. The resulting data set includes all the loan applications by blacks and Hispanics in the Boston area that year plus a large random sample of the loan applications by whites. It has by far the most complete set of information on loan applications ever assembled. These data allow the authors to estimate a model of loan denial that depends on an extensive list of variables associated with the probability and cost of default, including variables to measure an individual's credit history and dummy variables for the census tract in which the house is located and the lender to which the application is submitted. They also include a variable to indicate whether an applicant was black or Hispanic. The coefficient of this variable in their basic model indicates that the probability of denial is 8.2 percentage points higher for appli-



cants in one of these groups than it is for a white applicant, controlling for loan, property, and applicant characteristics. This result, which indicates that equivalent minority and white applicants are not treated equally in the mortgage market, is what all the excitement is about.

Munnell et al. (1996) also estimate a number of alternative models that add additional control variables, including loan characteristics, such as term, fixed rate, and cosigner, and individual characteristics, such as age, gender, and marital status. In addition, they replace the census tract dummies with explicit census tract characteristics, such as the percentage of units boarded up or vacant. The magnitude and significance of the minority-status coefficient is unaffected by these changes. Furthermore, the data set includes additional individual characteristics, such as years of job experience, education, and tenure in current job. The estimated minority-status coefficient is also unaffected by inclusion of these variables (Ross and Tootell 1998).

Many critics of the findings in Munnell et al. (1996), including Zandi (1993), Liebowitz (1993), Horne (1994), and Day and Liebowitz (1996), have argued that the study omits key explanatory variables. According to a well-known econometric theorem, the coefficient of one variable (in this case, minority status) will be biased if the estimating equation omits variables that are correlated with that variable and that help explain the dependent variable (loan denial). Moreover, if these omitted variables have a positive impact on loan denial (i.e., if higher values make loan denial more likely) and are positively correlated with minority status, then their omission will bias upward the coefficient of the minority status variable—or, to put it another way, will lead to an overstatement of discrimination. Because, on average, blacks and Hispanics have poorer credit qualifications, these critics conclude that the Boston Fed Study probably exaggerates discrimination because of these omitted variables. Examples of the omitted variables these authors discuss are “presence of cosigner,” “loan amount,” “dollar amount of gifts,” “home equity,” “lender toughness,” “whether data could not be verified” (henceforth called “unable to verify”), and “whether the applicant’s credit history meets loan policy guidelines for approval” (henceforth called “meets guidelines”).⁶

The available evidence reveals that most of these variables are not a source of bias in the Boston Fed Study’s equations. As noted by Browne and Tootell (1995), “cosigner” and “loan amount” are in fact present in the Boston Fed Study’s data set, and their inclusion in a loan denial model has no effect on the estimated minority-status coefficient. This data set also includes “whether a gift was used for the down payment” and “whether the applicant is a first-time homebuyer.” In addition, the “net worth” variable included in the Boston Fed equation includes “home equity.” The “net worth” variable is included in the Munnell et al. equations and is insignificant. Inclusion of the “gift” or “first-time homebuyer” variables has no effect on the estimated minority-status coefficient (see Browne and Tootell 1995 or Tootell 1996b). It seems unlikely that including the (unavailable) “actual amount of the gift” or “home equity” variables separately would have much influence on the minority-status coefficient, given that these related variables have no effect. In addition, Glennon and Stengel (1994) investigate many alternative sets of explanatory variables using the Boston Fed Study’s data. They find that the estimated minority-status coefficient is “remarkably” unaffected by their changes.



THE URBAN
INSTITUTE

The inclusion of the “unable to verify” and “meets guidelines” variables has a larger effect. With the Boston Fed Study’s data, the coefficients of the “unable to verify” and the “meets guidelines” dummy variables are statistically significant in a denial model. Moreover, according to Day and Liebowitz (1996), inclusion of “unable to verify” lowers the minority-status coefficient by 27 percent, and inclusion of both “unable to verify” and “meets guidelines” lowers the minority-status coefficient by 62 percent. Thus, including both variables substantially reduces the magnitude and significance of the minority-status coefficient; in fact, in some of Day and Liebowitz’s specifications, the minority-status coefficient is no longer statistically significant.

The effects of these variables on the minority-status coefficient are also examined by Carr and Megbolugbe (1993). They find that including “unable to verify” lowers the minority-status coefficient by 15 percent, and including both variables lowers the minority-status coefficient by 40 percent. However, even after including both variables, the minority-status coefficient is still statistically significant at the 1 percent level of confidence. The findings of Glennon and Stengel (1994), Browne and Tootell (1995), and Tootell (1996b) are similar to those of Carr and Megbolugbe.

Carr and Megbolugbe and Browne and Tootell argue that these variables, especially “meets guidelines,” should not be included in the denial equation. These variables were not recorded in the original loan file. Rather, they involve the after-the-fact judgment of the individual completing the HMDA data forms. The “unable to verify” variable could reflect the fact that lenders make extra efforts to verify the information of white applicants, and the “meets guidelines” variable could simply reflect the lender’s loan denial decision. In particular, Browne and Tootell contend that the “meets guidelines” question, which was answered a year after the loan approval decision, was interpreted by the respondents as whether “the sum total of applicant characteristics meet the institution’s credit guidelines for approval.” They note that some unsuccessful applications were coded as not meeting credit history guidelines even though those applicants had no credit problems. Browne and Tootell test their claim by estimating a model of the “meets guidelines” variable. As predicted by their claim, they find that loan terms, such as housing-expense-to-income ratio, debt-to-income ratio, and loan-to-value ratio, help explain the “meets guidelines” variable—a result that does not make sense if the “meets guidelines” variable only concerns the quality of an applicant’s credit history.

Horne (1994) and Day and Liebowitz (1996) counter that the “meets guidelines” variable is a suitable proxy for details of an individual’s credit history that were not collected for the Boston Fed Study. For example, Day and Liebowitz suggest that the “age of the credit problem” and the “size of the credit problem” should have been included in the analysis. They also suggest that standards may vary across lenders and claim that this is another justification for including the “meets guidelines” variable. Because credit history information often is provided by an outside agency, lenders also might determine whether an applicant meets their guidelines simply by comparing an external credit score to their internal standard. If so, then the “meets guidelines” variable can be considered exogenous and hence a legitimate explanatory variable.



Finally, Day and Liebowitz (1996) develop a “lender toughness” variable by matching the Boston Fed Study’s data to HMDA data. They add both the “lender toughness” and the “unable to verify” variables to the denial equation. Including both variables lowers the minority-status coefficient by 27 percent, which is the same as the effect of including only “unable to verify.” Therefore, “lender toughness” alone probably has little effect on the minority-status coefficient. Moreover, the basic equation in Munnell et al. (1996) already includes a set of lender dummy variables, which capture, of course, lender toughness and other fixed lender characteristics that affect loan denial. These variables lower the minority-status coefficient by about 20 percent—an effect that is already included in Munnell et al.’s basic estimate.⁷

Evaluation

The Boston Fed Study’s data set contains an extensive set of underwriting variables, including variables on credit history. Moreover, the magnitude and statistical significance of the minority-status coefficient in a loan denial equation are largely unaffected by the inclusion of a large array of risk, loan, borrower, unit, and neighborhood variables that are not included in the Boston Fed Study’s main equations. In our view, the Boston Fed Study’s equations contain a remarkably complete set of explanatory variables, and most claims concerning omitted-variable bias are implausible or have been shown to be incorrect. However, two variables omitted from the Boston Fed Study, namely “unable to verify” and “meets guidelines,” have been shown to have a substantial effect on the estimated minority-status coefficient. Browne and Tootell (1995) find that the minority-status coefficient is still large and statistically significant after including the variables, and they also argue that “meets guidelines” should not be included because it represents a subjective, after-the-fact opinion that could be influenced by a lender’s discriminatory behavior. Nevertheless, no consensus has yet emerged on the magnitude of the impact that these variables have on the minority-status coefficient or on the appropriate way to treat these two variables in a loan denial equation.

Reanalysis

The first question we examine in our reanalysis is the magnitude of the drop in the minority-status coefficient when “meets guidelines” and “unable to verify” are included in a denial equation. The Browne and Tootell (1995) and Tootell (1996b) conclusions on this point are not convincing, because they are based on a specification that is more parsimonious than many of the specifications investigated by Munnell et al. (1996) and others.

We reexamine the model with a more complete specification that includes additional loan, borrower, and neighborhood characteristics. We start with a base specification that does not include any credit history variables but does include an extensive list of control variables: housing-expense-to-income ratio; debt-to-income ratio; net worth; a proxy for the likelihood of unemployment; term; dummy variables for loan-to-value ratio (LTV) between 0.6 and 0.8, LTV between 0.8 and 0.95, and LTV above 0.95; fixed-rate mortgage; down payment



THE URBAN
INSTITUTE

includes gift; special program application; cosigner; age over 35; gender; marital status; self-employed; percent minority in tract above 30; median income in tract over \$39,111; multifamily unit; owner-occupied unit; private mortgage insurance application denied; and, of course, minority status (which means, as in Munnell et al., black or Hispanic).⁸ We then successively add the Boston Fed Study's credit history variables, the "unable to verify" variable, and the "meets guidelines" variable. The pseudo-R² or goodness-of-fit values for these four models are 0.454, 0.511, 0.568, and 0.611.⁹ Thus, each additional variable or set of variables has a substantial impact on the fit of the model. The minority-status coefficients (*t*-statistics in parentheses) for the four models are 0.537 (6.39), 0.356 (3.97), 0.327 (3.43), and 0.218 (2.047). When all of these variables are added, therefore, minority status is still significant, but only at the 5 percent level (two-tailed test) instead of the 1 percent level. Moreover, the effect of minority status on loan denial rates falls dramatically as these variables or sets of variables are included. Based on the characteristics of the minority applicants, the average effect of minority status on the probability of denial is 12.3, 7.7, 6.2, and 3.3 percentage points for the four models.¹⁰

We conclude that the two variables in question—"unable to verify" and "meets guidelines"—do, indeed, have a dramatic impact on the minority-status coefficient. As a result, the key question is whether it is appropriate to include these variables in the equation, include them as endogenous variables, or exclude them altogether. To answer this question, we first reestimate the model using a simultaneous equations technique that is appropriate for detecting a simple form of endogeneity when the dependent variable is binary (i.e., accept or reject).¹¹ This simple form arises when the unobservable factors in the equation to explain denial are correlated with the unobservable factors in an equation to explain one of the explanatory variables in the denial equation, that is, when some unobserved factors simultaneously influence both variables. The procedure we use allows us to calculate the correlation between unobservable factors across equations; a high value for this correlation indicates that endogeneity may be a serious problem.¹²

Consider first the case of "unable to verify," which may be endogenous because the probability of denial influences data verification efforts. We find that the across-equation correlation in unobserved factors is quite large (0.574 with a *t*-statistic of 0.94). This correlation could arise either because some underwriting variables are omitted from the specification or because "unable to verify" is influenced by the probability of denial. In either case, a loan denial equation that treats "unable to verify" as exogenous will yield biased results, but the simultaneous equations model we estimate will not. In this model, the estimated coefficient of "unable to verify" in the loan denial equation is 0.410 (with a *t*-statistic of 0.27), which is small and statistically insignificant. The coefficient of "unable to verify" without the endogeneity correction, 1.747 (with a *t*-statistic of 13.58), is obviously biased. Moreover, the estimated minority-status coefficient is basically unaffected by the inclusion of "unable to verify" when this variable is treated as endogenous. To be specific, it changes from 0.356 (with a *t*-statistic of 3.97) to 0.364 (3.68).¹³ The probability of loan denial is 7.4 percentage points higher for minorities than for whites with "unable to verify" treated as endogenous, which is close to the 7.7-percentage-point effect



based on a single-equation specification without the “unable to verify” variable.¹⁴ We conclude that, when properly treated as endogenous, the “unable to verify” variable has little or no impact on the minority-status coefficient, and we drop it from all further analysis.¹⁵

Now consider the “meets guidelines” variable, which may be endogenous because the actual underwriting outcome may influence a conclusion about whether an application meets a lender’s guidelines. As a point of reference, the minority-status coefficient in an equation that includes the “meets guidelines” variable but not the “unable to verify” variable is 0.245 (with a *t*-statistic of 2.43), and the effect of minority status on the probability of denial is 4.1 percentage points. Our simple simultaneous equations procedure estimates that the correlation between the unobservable factors in the loan denial and “meets guidelines” equations is -0.214 (with a *t*-statistic of 0.81). The estimated minority-status coefficient from a model in which “meets guidelines” is allowed both to be endogenous and to influence approval is 0.270 (with a *t*-statistic of 2.44), and the influence of minority status on the probability of denial is 6.5 percentage points. This “corrected” estimate of the impact of minority status on loan denial (6.5 percentage points) is bracketed by the single-equation estimate with the “meets guidelines” variable included (4.1 points) and the single-equation estimate with this variable excluded (7.7 points), but it obviously is closer to the latter. Thus, treating “meets guidelines” as endogenous eliminates most of its impact on the minority-status coefficient; correcting for the endogeneity of this variable makes a big difference.

This result may arise because this simultaneous equations model is too simple. As noted earlier, Browne and Tootell (1995) contend that the “meets guidelines” variable represents an after-the-fact judgment concerning the entire loan file made when someone filled out the Boston Fed Study’s survey. If this contention is correct, the issue is not simply whether unobservable factors are correlated across equations, but whether the loan denial decision itself influences the “meets guidelines” variable. Thus, we must ask whether denied applications are more likely than accepted applications to be coded as “does not meet guidelines,” all else equal. If so, then one would have to reject the Day and Liebowitz (1996) claim that “meets guidelines” is determined entirely by additional credit history details.

A full examination of these issues requires a complex model in which (a) both loan denial and the “meets guidelines” indicator depend on a loan officer’s unobserved opinion concerning whether the applicant meets the lender’s credit history standards and (b) the “meets guidelines” statement by another bank employee could be influenced by the loan denial decision. With this formulation, the “meets guidelines” variable itself cannot influence the denial decision because it is set at a later point in time. Instead, only one component of this variable can influence loan denial, namely, the component that reflects the initial loan officer’s evaluation of the applicant’s creditworthiness. This component is assumed to be the same for the loan officer who makes the loan denial decision and the bank official who later fills out the survey form. We develop and estimate such a model.¹⁶

The estimation results reveal that loan denial has a large influence on the “meets guidelines” variable; the coefficient is -2.040 (with a *t*-statistic of 2.48).



THE URBAN
INSTITUTE

This finding leads us to reject the view that lenders fill in the “meets guidelines” variable simply by comparing an applicant’s external credit score with some standard. Moreover, the fact that an application is coded as meeting the lender’s standards has a large influence on the likelihood of the loan being denied; the coefficient is -0.437 (with a t -statistic of 8.33). The estimated minority-status coefficient in the denial equation after controlling for the likelihood that the original loan officer felt that the application met the lender’s standards is 0.248 (2.60), and the effect of minority status on the probability of denial is 5.3 percentage points. Thus, the effect of minority status on the loan denial probability is still bracketed by the results from the two single-equation specifications, one excluding and the other including the “meets guidelines” variable. In other words, although “meets guidelines” is clearly influenced by the denial decision, it still has some impact on the minority-status coefficient even when this influence is taken into account.

Because we are ultimately interested in only the loan denial equation, not the “meets guidelines” equation, we also explore a simpler estimating procedure for use in our subsequent analysis. To be specific, we use the results of our complex estimating procedure to construct a variable, based solely on exogenous information, to measure the likelihood that an applicant meets a lender’s credit standards, as seen by the original loan officer. In other words, we “cleanse” the usual “meets guidelines” variable of any influence that flows from the loan denial decision. The introduction of this cleansed variable into a single-equation estimate of loan denial yields coefficient estimates that are very similar to those obtained with our more complicated simultaneous equations procedure. In addition, the effect of minority status on loan denial in the simplified procedure, 5.6 percentage points, is almost the same as the effect with the full model, 5.3 percentage points. In future models, therefore, we make use of the simpler procedure.

Interpreting the “Meets Guidelines” Results

This is not quite the end of the story, however, because the impact of the “meets guidelines” variable on the minority-status coefficient in the denial equation could have two different causes. The first possible cause is omitted variables. As noted earlier, Horne (1994) and Day and Liebowitz (1996) propose that the loan denial equation omits crucial credit history variables that are captured by the “meets guidelines” variable. Under this interpretation, the inclusion of “meets guidelines” eliminates an omitted-variable bias that would otherwise exist in the loan denial equation.

If this cause is at work, then there should be a strong negative correlation between the unobservable factors in the loan denial equation and in the “meets guidelines” equation. After all, the argument is that unobserved factors that show up in the “meets guidelines” rating are omitted from (and therefore obviously unobserved in) the loan denial equation. This prediction is not supported by the evidence. In our simultaneous equations procedure, the correlation between the unobserved factors in these two equations is only 0.032 (with a t -statistic of 0.070). Thus, we conclude that the issue here is not omitted variables.



The second possible cause of this impact is that the “meets guidelines” equation varies with minority status. This type of variation could have two different sources: variation in underwriting guidelines across lenders or bias in the evaluation of applications. Consider first the possibility that some lenders use underwriting guidelines that are particularly hard on applicants with characteristics that are relatively common among their minority applicants.¹⁷ The key issue in this case is whether these guidelines have an economic rationale, or, to use the legal terminology, whether they can be justified on the basis of “business necessity.” Guidelines that fail this test are said to involve disparate impact discrimination. Although litigation on this disparate impact standard has not yet occurred in the case of lending, the use of policies with a disparate impact in this sense is clearly illegal in the case of housing and employment.¹⁸

Now suppose that lenders have different underwriting guidelines, perhaps because of different experiences with past loans, and that all differences are fully justified on the basis of business necessity. Suppose, in other words, that there is no disparate impact discrimination. Even in this case, the “meets guidelines” equation could vary with minority status if blacks and Hispanics are not as successful as whites are in selecting a lender that meets their credit needs, perhaps because of poorer information. This “mismatch” would be picked up by the “meets guidelines” variable, so that including this variable in the loan denial equation might simply correct for the fact that different groups go to different lenders. The issue here is not that minorities simply go to “tougher” lenders or, indeed, to lenders with any particular characteristics. As just shown, the “meets guidelines” variable does not capture omitted variables in the loan denial equation.¹⁹ Instead, it reflects the possibility that black and Hispanic customers have poorer information than do white customers about the underwriting standards used by different lenders and therefore, on average, find a poorer match between their qualifications and the standards used by lenders to which they apply. In this case, white customers might be more likely than minority customers to be seen as meeting the lender’s standards, even though no individual lender used different standards for minorities than for whites. If so, adding the “meets guidelines” variable, treated endogenously, can be interpreted as a way to extract from the minority-status coefficient one component that reflects legitimate differences in underwriting guidelines, or, to put it another way, that is justified on the basis of business necessity.

The problem for clear interpretation of the “meets guidelines” results is that lenders with many minority applicants may tend to use underwriting standards that have a disparate impact on minorities, in the legal sense. As in the previous case, this possibility does not involve disparate treatment discrimination (defined as the application of different underwriting standards to applicants in different groups). However, it does involve disparate impact discrimination, so including the (endogenous) “meets guidelines” variable simply shows how much of the minority-status coefficient can be attributed to this one example of disparate impact discrimination.

The second possible source of intergroup differences in the “meets guidelines” variable is discrimination in determining whether an applicant meets a



THE URBAN
INSTITUTE

lender's underwriting standards. This discrimination could exist at two different points in the process. First, the loan officer who processes an application may use tougher standards for minorities, so that the bank official who later answers the "meets guidelines" question on the survey form, and who reads the comments entered by the loan officer, systematically gives lower ratings to minority applicants. Second, the bank official who fills out the survey form may simply be less likely to say that a minority applicant meets the lender's standards than does a white applicant—even when the two applications are otherwise identical. The first type of behavior is obviously more troubling, but they both have the same impact on the "meets guidelines" variable. Moreover, the two types of behavior could interact. If the loan officer filling out the survey form observes a case in which a minority applicant who actually meets the lender's guidelines is denied a loan by a discriminatory underwriter, then that officer may indicate that the applicant failed to meet the guidelines as a rationale for the underwriter's earlier decision.²⁰

The dilemma we face at this point is that we do not know which interpretation is correct. We have some evidence that "meets guidelines" affects the minority-status coefficient in the loan denial equation because the "meets guidelines" equation varies with minority status. But we do not know whether the impact of the "meets guidelines" variable reflects applicants' selection of lenders, which does not involve discrimination, or, instead, reflects either disparate impact discrimination by lenders or discrimination in determining whether applications meet a lender's underwriting guidelines. If the first possibility is true, then the procedure at the end of the preceding section is appropriate, and its conclusion, that the Boston Fed Study overstates discrimination by 37.5 percent ($(7.7 - 5.6)/5.6$), is compelling. If either of the last two possibilities is true, however, this procedure is not appropriate, because it implicitly assumes that differences in "meets guidelines" based on minority status have nothing to do with discrimination.

Under the last two possibilities, the correct procedure involves the introduction into the loan denial equation of a minority status-neutral variable indicating the likelihood that a candidate meets the lender's guidelines—that is, a variable that is not influenced by observed differences in this likelihood by minority status, after controlling for other observable factors. Once this step is taken, the "meets guidelines" variable no longer has a substantial impact on the minority-status coefficient. To be specific, the minority-status coefficient in an equation with a cleansed and minority status-neutral variable for "meets guidelines" is 0.355 (with a *t*-statistic of 3.81) and the impact of minority status on the probability of denial is 7.3 percentage points. This estimated impact is, of course, close to the estimated impact for a single equation specification that omits the "meets guidelines" variable altogether, namely, 7.7 points.

Although we know of no definitive method for deciding which of these possibilities is true, the available evidence suggests that the first possibility is not very likely. If minority households are denied more often, at a given level of creditworthiness, because they select lenders with relatively unfavorable underwriting guidelines, then the minority-white difference in loan approval should be larger among lenders that specialize in lending to minorities. This turns out not to be the case. Munnell et al. (1996) find that the minority-status coefficient is virtually the same for the sample of lenders who specialize in



dealing with minority applicants as for the sample of other lenders. Moreover, Browne and Tootell (1995) show that the minority-status coefficient is literally unaffected if one excludes two large minority lenders, who together account for half of the minority applications in the Boston Fed Study's sample.²¹ In addition, both Munnell et al. (1996) and Hunter and Walker (1996) find little evidence that individual underwriting variables receive different weights for minority and white applicants.²² In short, it seems likely, although it cannot yet be formally proved, that the decline in the minority-status coefficient that occurs when an endogenous "meets guidelines" variable is included in the analysis simply indicates that—following one of the last three possibilities listed above—the coefficient of the "meets guidelines" variable itself reflects some form of discrimination. Thus, this result does not imply that the Boston Fed Study overstates discrimination.

To account for the range of possibilities concerning the "meets guidelines" variable, we explore all other potential flaws in Munnell et al. (1996) using three models of the loan denial equation. Model 1 includes the "meets guidelines" variable in a single-equation estimation. Model 2 replaces the original "meets guidelines" variable with a cleansed (but not minority status-neutral) version of the variable. Model 3 excludes the "meets guidelines" variable from a single-equation estimation. Model 1 provides a lower bound measure of discrimination, because it ignores the possibility (supported by evidence presented here) that the "meets guidelines" variable is endogenous and that its inclusion leads to an underestimate of discrimination. Model 2 accounts for this possible endogeneity and provides the most accurate estimate of discrimination if the first possibility given earlier is correct—that is, if the observed across-group differences in the "meets guidelines" equation have nothing to do with discrimination. Model 3 provides the most accurate estimate of discrimination if either of the last two interpretations given earlier is correct—that is, if discrimination accounts for the across-group differences in the "meets guidelines" equation. The first row of table 1 reviews our estimates of the impact of discrimination on loan denial for the baseline version of these three models.

Table 1. *The Impact of Minority Status on Loan Denial for Various Models**

| | Model 1 | Model 2 | Model 3 |
|--|---------|---------|---------|
| 1. Basic model | 4.1 | 5.6 | 7.7 |
| 2. No condos or multifamily houses | 5.1 | 7.0 | 9.0 |
| 3. No applications with PMI denied or special program | 6.4 | 6.4 | 9.0 |
| 4. Use actual LTV, not LTV categories | 4.3 | 6.0 | 8.1 |
| 5. Use actual LTV and treat it as endogenous | 4.2 | 5.3 | 7.7 |
| 6. Treat LTV and housing-expense-to-income ratio as endogenous | 4.4 | 5.1 | 8.6 |
| 7. No applications for special programs | 5.7 | 7.2 | 9.5 |
| 8. Estimation conditional on choice of a special program | 6.6 | 8.0 | 11.1 |

*Impact in percentage terms

Notes: Model 1 includes the "Meets Guidelines" variable, with no correction for endogeneity. Model 2 includes the "Meets Guidelines" variable, with a correction for endogeneity. Model 3 excludes the "Meets Guidelines" variable.

All entries in this table are based on regression coefficients that are statistically significant at the 5 percent level (two-tailed t-test) or above.



THE URBAN
INSTITUTE

Data Errors in the Explanatory Variables

Liebowitz (1993), Horne (1994, 1997), and Day and Liebowitz (1996) present evidence of errors in loan terms and in other application characteristics in the Boston Fed Study's data. They claim that these errors lead to biased results and, in particular, to an overstatement of discrimination.

Horne (1994, 1997) focuses on a subsample of the Boston Fed Study's data. This subsample begins with a set of applications provided to the Federal Deposit Insurance Corporation (FDIC) to help it explore the underwriting policies of the lenders it oversees. This set was made up of all rejected loan applications in which the denial probability was below 50 percent—the so-called exceptions list. The subsample examined by Horne consisted of applications on the exception list filed with lenders that had at least one minority rejection on the exceptions list. Horne (1994) claims that he identified many coding errors in the Boston Fed Study's data in this subsample. Munnell et al. (1996) and Browne and Tootell (1995) obtained Horne's corrected data, reestimated the original Boston Fed Study equations using these data (along with the observations that Horne did not correct), and found no substantial effect on the minority-status coefficient.²³

Liebowitz (1993) and Day and Liebowitz (1996) analyze the public-use version of the Boston Fed Study's data set, which is the same set we use, and identify many observations that appear to have unreasonable values, such as a large, negative net worth; LTV above 0.80 without an application for private mortgage insurance (PMI); low or negative imputed interest rate; an annual income that does not match monthly incomes; or obligation ratios that are inconsistent with other variables. They conclude that these problems cast doubt on the integrity of the Boston Fed Study's data—and hence on the credibility of the results based on it.

In addition, Day and Liebowitz (1996) reexamine the Boston Fed Study's data after removing observations with questionable interest rates.²⁴ The magnitude of the minority-status coefficient drops from 0.0325 (with a *t*-statistic of 2.45) to 0.0293 (2.12) when observations with interest rates above 14 percent or below 5 percent are deleted. The coefficient drops to 0.0249 (1.69) when observations with interest rates above 12 or below 7 percent are deleted.²⁵

Munnell et al. (1996), Browne and Tootell (1995), and Tootell (1996b) argue that many of these problems are not errors at all. Some applicants actually have negative net worth, and the Boston Fed Study used the monthly incomes in the actual loan files, not the more problematic annual income in the HMDA data. Moreover, even though secondary market institutions generally require PMI when LTV is above 0.8, a lending institution may not impose this requirement if it does not intend to sell the loan on the secondary market. The Boston Fed Study's authors also explain that they contacted lenders to verify the accuracy of the information in many applications with unusual values and corrected many of the extreme values that appear in the public-use data before conducting their regressions. They also note that interest rates must be imputed from housing expense, loan term, and loan amount. This procedure does not work well for multifamily units, they say, and most of the applications with very low imputed interest rates are for loans on multifamily units. It also does not



work well for applications with very low LTVs, which are the applications that tend to have very high interest rates. In addition, they point out that the obligation ratios used in the Boston Fed Study were taken directly from each lender's worksheets in the loan files. Browne and Tootell also estimate several alternative models that omit observations with unusual values—with little or no impact on the minority-status coefficient.

Browne and Tootell (1995) also argue that the critics of the Boston Fed Study cannot eliminate the statistical significance of the minority-status coefficient without combining a large set of questionable steps. In fact, this coefficient becomes insignificant only if one eliminates a large number of observations because of potential data errors, includes both "unable to verify" and "meets guidelines," and takes a strong position concerning possible misclassification in the dependent variable. For example, Liebowitz (1993) claims that the minority-status coefficient is not significant in a sample of single-family homes with LTVs below 0.8 if six influential outliers are eliminated. However, Browne and Tootell show that this subsample contains only 14 minority denials to begin with (out of the original 200 in the Boston Fed Study's data), and yet Liebowitz must eliminate almost half of those observations in order to eliminate the impact of minority status.

These claims are supported by Carr and Megbolugbe (1993), who "clean" the Boston Fed Study's data by removing observations with very high LTVs, unreasonable interest rates, or inconsistencies between housing expenses and debt payments and by performing various other consistency checks. They also delete all observations with imputed interest rates either above 20 percent or below 3 percent. In all, Carr and Megbolugbe delete 1,045 observations out of 2,816. This procedure does not support the claim that data errors lead to an overstatement of discrimination; in fact, the estimated minority-status coefficient actually increases when these observations are deleted.

In addition, Carr and Megbolugbe (1993) point out that many of their so-called inconsistencies may not be data errors at all. About 800 of their deletions are for one of three reasons: the HMDA and Boston Fed Study's income variables are not consistent, the housing-expense-to-income ratio is not consistent with house expense and income considered separately, and the house price is greater than the loan amount plus liquid assets plus other money available. As noted earlier, differences between these income sources probably reflect problems in the HMDA data, not in the Boston Fed Study's data. Moreover, the housing-expense-to-income ratio often includes adjustments to account for the temporary nature of some income or the uncertain nature of rental income in multifamily houses. Finally, the house price may exceed apparent resources because the applicant plans to make a down payment with equity from a house he has not yet sold.

Evaluation

Overall, the key finding of the Boston Fed Study appears to be remarkably unaffected by corrections in the explanatory variables—both corrections that take the form of "cleaning" individual variables and those that take the form of dropping observations with extreme values on some variables. The only possible



THE URBAN
INSTITUTE

exception appears in Day and Liebowitz (1996), who find that the estimated minority-status coefficient declines by 24 percent when one drops observations with an imputed interest rate above 12 or below 7 percent. On the surface, at least, this data correction exercise appears to be independent of minority status, so it warrants further explanation. Moreover, the explanation provided by Browne and Tootell (1995) concerning problems with the imputation process for multifamily properties and for applications with low LTVs is not very satisfying. The estimated minority-status coefficient is not affected by eliminating applications for multifamily properties and with low LTVs.²⁶ If these applications are the ones with unreasonable interest rates, why does filtering out unreasonable interest rates affect the minority-status coefficient?

Reanalysis

To check the Day and Liebowitz (1996) claim that the minority-status coefficient is no longer significant when observations with extreme interest rates are dropped, we devised our own procedure for imputing interest rates. The formula for a mortgage is well known; it expresses the mortgage amount as a function of the monthly payment, interest rate, and term of the mortgage.²⁷ The interest rate is not in the Boston Fed Study's data set, but this formula can be used to calculate the interest rate using data on the mortgage amount, monthly payment, and term. The Boston Fed Study's data set includes the mortgage amount and term. It also includes monthly housing expense, but this variable includes property taxes and homeowner's insurance in addition to the monthly mortgage payment. This variable also includes condominium fees and, in the case of multifamily housing, is reduced by the expected rent from other units. One cannot estimate the interest rate, therefore, without making assumptions about these other components of the monthly housing expense variable.

Our procedure begins by dropping condominiums and multifamily units from the sample, as well as units with a negative imputed interest rate or with missing information on term, monthly housing expense, or loan amount. The resulting sample has 1,780 applications. With this sample, the effects of minority status on the probability of denial in our three models are 5.1, 7.0, and 9.0 percentage points. To cover a range of possibilities, we explore cases in which the sum of annual taxes and insurance expenses equals 2.0, 2.5, and 3.0 percent of the house price (20, 25, and 30 mils). Finally, we examine different possible interest rate cutoffs, in each case discarding observations with interest rates that are above one threshold or below another.

The results of these estimations reveal no clear pattern. When we use Day and Liebowitz's (1996) criteria of imputed interest rates between 7 and 12 percent, we delete between 20 and 40 percent of the sample, depending on the assumption about housing expense. Under the 20-mil assumption, the minority-status effects with our three models are 3.9, 6.7, and 8.7 percentage points, and the minority-status coefficient in model 1 is only significant at the 10 percent level. Under the 25-mil assumption, the minority-status effects are 1.9, 4.6, and 6.8 percentage points, and the minority-status coefficients in models 1 and 2 are both insignificant. Under the 30-mil assumption, however, the minority-status effects are 6.7, 8.0, and 10.2 points, and the minority-status coefficients



are highly significant in all three models. To make matters more confusing, when we use Day and Liebowitz's criteria of imputed interest rates between 5 and 14 percent, the largest minority-status effects (4.5, 6.7, and 8.6 points) occur under the 25-mil assumption—the mil rate that made the minority-status coefficient go away with the alternative interest rate cutoffs—and the minority-status coefficients are statistically significant in all three models.

Conclusions

The Boston Fed Study's data set is large and complicated and may contain some errors in the explanatory variables. However, the Boston Fed Study's authors made extensive efforts to check the data and to investigate the possible impacts of errors on their results. Several other scholars have also performed various tests for the possible impacts of data errors. All reasonable tests support the Boston Fed Study's conclusion: There is no reason to think that errors in the explanatory variables lead the Boston Fed Study to overstate the extent of discrimination.

Day and Liebowitz (1996) make the reasonable suggestion that one way to check the accuracy of the data in a mortgage lending study is to determine whether, using the standard mortgage formula, the payment, term, and mortgage amount imply a reasonable mortgage interest rate. However, this suggestion cannot be implemented in a compelling way with a data set, such as the one used for the Boston Fed Study, that does not include information on the mortgage payment. Because it is impossible to accurately impute interest rates without observing actual mortgage payments, any imputation procedure introduces measurement error, perhaps substantial error, into the interest rate variable. Any check based on such an imputation will identify many observations as having a data error when, in fact, the only error is in the imputation procedure. Thus, we do not find it surprising that the application of this procedure leads to confusing results.

More important, this type of procedure opens the door to another potentially serious source of bias in estimating the impact of minority status on loan denial. If minorities tend to live in areas with relatively high property taxes and insurance rates, for example, then assuming a low mil rate for tax and insurance costs may disproportionately—and inappropriately—filter out minority applications. Moreover, property taxes and insurance rates are undoubtedly positively correlated with loan denial, so this filtering process is correlated with the dependent variable. As is well known (see Greene 1993), any sample-selection process that is correlated with the dependent variable creates a sample-selection bias in the estimated coefficients. Thus, an interest-rate filter seems far more likely to create a sample-selection bias than to eliminate measurement error.

Misclassification in the Dependent Variable

Several authors have argued that the results in the Boston Fed Study are biased because the study's data misclassify the outcome of many applications (see Horne 1994 and 1997; Day and Liebowitz 1996). The Boston Fed Study's authors were themselves concerned about this issue and address several aspects



THE URBAN
INSTITUTE

of it (Munnell et al. 1996). Their basic model excludes applications that were withdrawn before the lender made a decision about them; they support this approach by showing that the factors determining withdrawals are very different from those determining loan denials. Moreover, they checked their final data set data carefully to remove “[r]efinancings, home improvement loans, and some business loans that institutions had mistakenly coded as mortgage originations in their original filings” (p. 31).

Based on the FDIC review, Horne (1994, 1997) argues that many of the apparently rejected loan applications on the exceptions list should not be considered “denied” applications. The FDIC reviewed 62 minority and 33 white loan application files. Out of these, Horne (1994) identifies 5 applications (3 minority and 2 white) that were actually approved; 8 applications (6 minority and 2 white) that were withdrawn; 6 minority applications to which the bank made a counteroffer that was turned down by the applicant; 5 minority applications in which the applicant applied for a special program loan and was overqualified; 1 minority application that was rejected because the VA would not approve the loan; and 1 minority application that was rejected because PMI was denied.

Day and Liebowitz (1996) found that the removal of these 26 applications (plus two others they identify as being in a bank-specific special program) from the Boston Fed Study’s sample lowers the minority-status coefficient by 39 percent, from 0.0531 (3.96) to 0.0325 (2.45). Even after this step, however, the estimated coefficient is still significant at the 1 percent level. Day and Liebowitz also observe that, since the FDIC reviewed only a small number of applications, additional file reviews would certainly yield additional errors, and all these errors might account for the significance of the minority-status variable in the Boston Fed Study.

Horne (1997) also examines the effect of observations with a potentially misclassified dependent variable using just the FDIC subsample of the Boston Fed Study’s data. First, Horne drops 111 observations because the application had one of the following characteristics: it was withdrawn, it involved a unit under construction, it involved refinancing, it was an investor application, it involved an applicant who was overqualified for a special program, or it had an LTV below 0.30. He also recoded four applications as approvals, arguing that they had been incorrectly coded as denials. The minority-status coefficient for the entire FDIC subsample is 1.12 and is highly significant statistically. After deleting and recoding observations, the minority-status coefficient falls to 0.67 and is still significant at the 1 percent level. Next, Horne dropped 61 applications because he believed that the outcome is ambiguous. For example, counteroffers were made by some lenders and turned down by applicants, or applicants were denied PMI. The FDIC file reviews also uncovered instances in which the lender appeared to be willing to provide credit but the transaction was precluded by outside factors, such as title problems or housing code violations. The exclusion of these applications and some modifications to the model specification result in a minority-status coefficient of 0.35, which is no longer statistically significant at even the 10 percent level.

The Boston Fed Study’s authors follow the HMDA reporting requirements and consider accepted counteroffers to be approvals and denied counteroffers to be rejections. They argue that an accepted counteroffer implies that the



lender provided credit based on the borrower's application package and preferences. A denied counteroffer implies that the lender was not willing to provide a loan at terms that were acceptable to the borrower. All counteroffers may not be equal. Some counteroffers may involve small changes to the terms of the loan, whereas others may involve dramatic changes. The data contain no information concerning the magnitude of the change proposed by lenders in their counteroffer, so, according to Munnell et al., the fact that the counteroffer was accepted is the best indicator that the change was minimal.

Horne (1994, 1997) proposes that all counteroffers be considered acceptances because the lender is willing to provide credit. However, one could also argue that all counteroffers should be considered rejections because the lender was not willing to provide credit based on the terms in the application package, which are in fact the terms that are included in the Boston Fed Study's data set. One possible way to deal with this difference of opinion would be to estimate a model based on three lender choices: approve, counteroffer, and deny.²⁸ In this type of model, a difference in either the likelihood or the nature of counteroffers based on minority status also constitutes differential treatment.

A study of an alternative sample yields some insight into the issues raised by Horne (1994, 1997). Glennon and Stengel (1995) examine applications from three different lenders. They find, on average, a significant impact of minority status on loan denial.²⁹ They identified all applications in which the applicant rejected the bank's counteroffer and all files in which the applicants were overqualified for special loan programs. The deletion of all of these applications did not affect their findings. Moreover, they reviewed all denied files at one lender and detected 41 withdrawals that had been missed previously. The deletion of these withdrawals did not affect the estimated minority-status coefficient for that bank.

An alternative approach is to see if the Boston Fed Study's results are driven by a few "influential" observations. If so, a few data errors or misclassifications might drive the results. Rodda and Wallace (1996) rank applications by their influence on the minority-status coefficient. Most of the highly influential applications are minority denials, and the elimination of the 23 most influential minority denials causes the minority-status coefficient to become insignificant. Rodda and Wallace conduct this analysis to determine which applications should be subject to file review—not to determine whether estimates of discrimination are flawed. Moreover, they point out that this finding is driven predominantly by sample size; there are nearly four times as many white applications as minority applications, and most applications are approved. A related approach that focuses on outliers in a broader sense is provided by Carr and Megbolugbe (1993), who calculate the influence of every observation on the minority-status coefficient and on all other coefficients. They exclude 27 applications that are highly influential on either measure and find that the minority-status coefficient does not change.

Evaluation

In the Boston Fed Study's data, as in any large, complex data set, it is reasonable to explore the accuracy and interpretation of the information, as Horne (1994,



THE URBAN
INSTITUTE

1997) and Day and Liebowitz (1996) have done. However, many of the cases they identify as “errors” are in fact cases that raise issues of interpretation about which reasonable people may disagree. For example, we find the interpretation of counteroffers by Munnell et al. (1996) to be entirely reasonable, although not the only interpretation possible. There is obviously room for more research on this topic. A model of three choices—accept, reject, and counteroffer, for example—might help to determine whether counteroffers are a relatively benign phenomenon, as some scholars claim, or are instead another type of lender behavior that involves differential treatment.

Moreover, the selection process used by the FDIC to identify files within the Boston Fed Study’s sample for review was designed to help bank examiners—not to shed light on a loan denial equation. For example, the FDIC did not review files at lenders that did not reject any minority applications. As a result, two-thirds of the files reviewed are minority rejections, despite the fact that more than half of the rejections in the Boston Fed sample are applications from whites. This feature of the FDIC exceptions list makes it inappropriate for use in a loan denial equation. The FDIC’s exclusive focus on rejected applications also is problematical for estimation purposes; after all, some, even many, denied or withdrawn applications could have been miscoded as approvals. Similarly, the Boston Fed Study’s data may contain underqualified minority applicants who were approved because they applied for a special program. Overall, Day and Liebowitz’s (1996) and Horne’s (1994, 1997) filtering of the data and the resulting conclusions must be rejected because they are not based on a random sample of the applications in the Boston Fed Study’s data. Indeed, alternative approaches to misclassification that are neutral with respect to minority status, such as the one in Glennon and Stengel (1995), find that misclassification has little or no impact on the relationship between minority status and loan denial.

Finally, studies such as Rodda and Wallace (1996) that identify and drop “influential” observations also do not shed much light on data errors or misclassification, although they might be useful for other purposes. In particular, it is not surprising that minority denials have the most influence on the minority-status coefficient in the Boston Fed Study; after all, this coefficient is supposed to determine whether applications with this outcome are treated differently from comparable white applications. Nor is it surprising that the list of influential observations did not include any white approvals, the effects of which are “watered down” by the presence of so many similar observations. As a result, this type of analysis sheds no light whatsoever on the credibility of the Boston Fed Study’s result. The study by Carr and Megbolugbe (1993) is more to the point, because it defines “influential” in a way that is relatively neutral with respect to minority status, but it also does not provide a compelling conceptual or methodological argument for dropping influential observations. We conclude that any procedure for identifying and dropping influential observations in a loan denial study, particularly one that defines influence by impact on the minority-status coefficient, is not appropriate for evaluating the role of errors or of misclassification in a study of mortgage lending discrimination.



Reanalysis

The public-use data from the Boston Fed Study contain some information to help shed light on the issues raised by Horne (1994, 1997) and by the FDIC file reviews. In particular, these data contain variables to indicate “whether an application for PMI was denied” and “whether a mortgage application was made for a special program.” We find that dropping all applications coded affirmative for “PMI denied” or “application for special program” actually increases the impact of minority status on loan denial. For our three models and the full sample, the effects of minority status are 4.1, 5.6, and 7.7 percentage points. When these applications are dropped, the remaining sample contains 2,336 applications, the effects of minority status are 6.4, 6.4, and 9.0 percentage points, and the estimated minority-status coefficient is highly significant in all three models (see table 1).

Conclusions

Although the available data do not permit a definitive conclusion concerning the impact of misclassification on the Boston Fed Study’s results, we find that these results are not affected by many of the misclassification problems discussed by Horne (1994, 1997) and Day and Liebowitz (1996). Moreover, even the strongest claims about bias due to misclassification are not compelling because they are based on a nonrandom sample of applications. There currently exists no evidence indicating that elimination from the data set of misclassified observations identified on a minority status–neutral basis has any substantial influence on the estimated minority-status coefficient.

Incorrect Specification

Anyone studying loan denials must decide how to “specify” the equation, that is, what algebraic form to use for estimating purposes. Different specifications may lead to different results. Munnell et al. (1996) were, of course, aware of this issue and investigated several different forms. For example, they conducted a test to determine whether the same model applied to applications for single-family houses, multifamily houses, and condominiums. On the basis of the affirmative test results, they pooled all these types of applications. They also investigated several interaction terms—terms that determine whether the impact of one explanatory variable depends on the value of another one—and could not find any such terms that affected the impact of minority status on loan denial.

Liebowitz (1994) and Day and Liebowitz (1996) suggest that the underwriting process may vary across the sample. For example, Liebowitz suggests that the sample should be split by type of unit or by LTV, with a separate analysis of each subsample. As noted above, however, Munnell et al. reject the hypothesis that the model differs by type of unit. Moreover, Browne and Tootell (1995) report that when separate equations are estimated for each type of unit, the



THE URBAN
INSTITUTE

minority-status coefficient actually increases in the single-family housing regression. The minority-status coefficient is insignificant in only the two- to four-family regression, which contains only 393 loan applications. They also find that the minority-status coefficient is still significant in what Liebowitz characterizes as the “core” sample, namely, single-family houses that do not involve an application for PMI, even though this subsample contains only 14 minority denials.

Buist, Linneman, and Megbolugbe (1997) separate the sample into applications that meet all standard underwriting criteria based on LTV, housing-expense-to-income ratio, and so on, and applications that fail to meet at least one of these criteria. They find that the minority-status coefficient is highly significant in the second subsample but is not significantly different from zero in the first. This result suggests that discrimination may arise in underwriters’ decisions concerning the interpretation of, or compensation for, an applicant’s failure to meet one or more underwriting criteria.

Glennon and Stengel (1994) suggest that the denial model should vary by minority status. They estimate separate models by minority status and use the Blinder-Oaxaca decomposition approach (a common approach in the literature on wage differentials) to estimate the effect of minority status on the likelihood of denial. They find that the results of the Boston Fed Study are the same with this alternative specification.

A related point is made by Hunter and Walker (1996), who suggest that individual elements of the underwriting model may vary by minority status.³⁰ They interact credit history, education, and obligation ratios with minority status. Only the obligation ratio interaction is statistically significant. They find that at low obligation ratios the minority-status coefficient is relatively small and only significant at the 10 percent level, but at high obligation ratios the minority-status coefficient is large and highly significant. A similar result is in Munnell et al. (1996).³¹ However, these results have no impact whatsoever on the average impact of minority status on loan denial, which is the focus of the Boston Fed Study and most of the other literature.

LaCour-Little (1996) estimates a model of loan demand and supply using a technique developed by Maddala and Trost (1982). If a loan is approved, the loan amount represents the minimum of the two unobserved variables: loan demand and loan supply. If the loan is denied, the loan amount is assumed to be the actual loan demand, and loan supply is presumably below this amount. LaCour-Little finds that minorities have a higher demand for credit than whites, but lenders supply less credit to minorities. In fact, the level of significance of the minority-status coefficient in LaCour-Little’s supply equation is comparable to the level of significance of the minority-status coefficient in Munnell et al. (1996). This finding is consistent with Maddala and Trost’s application of this technique to Schafer and Ladd’s (1981) data and confirms the view that lenders treat minority applicants less favorably than comparable whites.

LaCour-Little (1998) also applies the so-called reverse-regression approach to the Boston Fed Study’s data. In a reverse regression, a model of the variable of interest (in this case, loan denial) is estimated and is used to generate a quality index for each individual observation. The quality index is used as the dependent variable in a second-stage regression that controls for the original



variable of interest and the minority status and/or gender of the individual. This approach has been used in the wage discrimination literature (see Goldberger 1984). The approach is especially appropriate when the actual explanatory variables deviate from the ones that are called for theoretically, but it also may be useful when the explanatory variables suffer from measurement error. Typically, reverse regression is used when the variable of interest is continuous, as in the case of earnings, but LaCour-Little argues that the approach is also reasonable for a discrete dependent variable, such as loan denial. He predicts the probability of denial for all observations and uses this probability as the dependent variable in the second stage. Based on this analysis, LaCour-Little reverses the findings of the Boston Fed Study and concludes that minority applications are favored, on average, in the Boston Fed Study's sample.

Finally, Glennon and Stengel (1994, 1995) argue that specification changes are needed to account for the fact that different lenders may use different underwriting guidelines. According to the disparate-treatment standard, discrimination exists if a single lender uses different underwriting guidelines in evaluating white and minority applications, but it does not exist if minorities and whites with the same qualifications receive different treatment because they visit different sets of lenders with different guidelines. To separate these two possibilities, and therefore to isolate disparate treatment, Glennon and Stengel argue that the best approach is to estimate a separate loan denial regression for each lender. They cannot implement this approach with the public-use version of the Boston Fed Study's data, which does not identify lenders, but they can implement it using an entirely different data set for three large lenders. They find, for example, that the minority-status coefficient is not statistically significant in two of the three regressions for individual lenders.³² They conclude that the Boston Fed Study's results may reflect a specification error rather than market-wide discrimination.

Buist, Linneman, and Megbolugbe (1997) take this logic one step farther. For each lender in the Boston Fed Study's data, they identify a unique set of implicit underwriting criteria that differentiates between accepted and rejected applications. In other words, they identify the criteria that, if violated, always lead to a rejection. They then find that the minority-status coefficient in a loan denial equation is insignificant if a variable to identify applications that meet the lender's implicit standards is included. These authors are careful to explain that this finding cannot be used to reject the hypothesis that discrimination exists. Instead, it simply shows that there exists a set of lender-specific, minority status-neutral underwriting criteria that can explain observed intergroup differences in loan denial. Even if these criteria accurately describe how underwriters actually behave, they might have a disproportionate impact on minority applicants without any business justification and therefore might constitute discrimination by the disparate impact standard—an issue to which we will return.

The issue of variation in underwriting criteria across lenders is also raised by Horne (1997) and Black, Collins, and Cyree (1997), who argue that the minority-status coefficient in the loan denial equation reflects the behavior of lending institutions, perhaps even minority-owned ones, that specialize in lending to minorities. According to this argument, lenders who attract many minority applications have unusually high rates of minority denials, based in



THE URBAN
INSTITUTE

part on credit characteristics that are not observed in the Boston Fed Study's data or on unusual underwriting standards. For example, Horne finds that if two large minority-owned lenders are excluded from his subsample of the Boston Fed Study's data, the minority-status coefficient is no longer significant. However, Munnell et al. (1996) include lender dummies in their basic equation, which rules out fixed lender effects for minority-owned or any other lenders, and they show that the minority-status coefficient is virtually the same in a regression for lenders who have a high volume of loans to minorities as in a regression for lenders who do not.³³ Moreover, when Browne and Tootell (1995) drop minority-owned lenders and Tootell (1996b) drops the lenders identified by Horne, the minority-status coefficient is unaffected.

Evaluation: Reverse Regression

The only specification change that has any substantial effect on the Boston Fed Study's main result is the reverse-regression approach suggested and implemented by LaCour-Little (1996). However, we reject the reverse-regression approach for an equation with a discrete dependent variable. In the wage discrimination literature, the reverse regression asks the following question: Among workers who earn the same amount, do minority (or female) workers have higher average quality than white workers? This question is appropriate for examining wage differentials. If female workers make the same amount and yet are higher-quality workers, they must be facing discrimination. A similar question arises for the application denial decision. Among applicants with the same outcome (either denial or approval), are minority applications more creditworthy, on average? However, this question is not appropriate for a discrete outcome. Even if more marginal white applications are approved, the average quality of approved loans might be higher for white applications simply because the high-quality applications are predominantly from whites. Thus, the LaCour-Little conclusion that minorities are favored in loan decisions, on average, is unwarranted.

Reanalysis: Reverse Regression

To shed further light on this issue, we first obtain a measure of loan quality by estimating a loan denial model using only the applications from whites and using the resulting coefficients to predict the probability of acceptance for every application, regardless of minority status. A distribution of the resulting loan quality measure illustrates the problem with the reverse-regression approach (see table 2). Most of the very high quality applications are from whites and many of the very low quality applications are from minorities. There probably exist some applications that are of such high quality that they will be approved regardless of minority status, and there probably exist some applications that are of such low quality that they will be denied regardless of minority status. The existence of such applications should not influence a test for discrimination. As explained earlier, however, these applications undermine reverse regression, because the very high quality approved loans and the very low quality denied loans enter the calculation of average loan quality.



Evaluation: Lender-Specific Underwriting Guidelines

Glennon and Stengel (1995) and Buist, Linneman, and Megbolugbe (1997) argue that an analysis that pools lenders may not be able to isolate disparate treatment discrimination. Furthermore, the analysis by Buist et al. casts some doubt on the standard specification of a loan denial equation. In particular, if each lender uses a hard-and-fast rule that automatically determines acceptance or rejection, then the standard specification, which treats loan denial as probabilistic, is incorrect.

We disagree with Buist et al.'s second point, for two reasons. First, there is no evidence that lenders use hard-and-fast rules for loan denial. Decisions are based on judgments and interpretations and compensating factors that inevitably vary from case to case. Second, Buist et al.'s approach substantially limits the power of the analysis. The dummy variable they include, which indicates that an application fails the lender's implicit underwriting criteria, is set equal to 1 for 265 of the 410 denied applications in their sample. This approach eliminates the influence of these observations on any of the other estimated coefficients, including the minority-status coefficient. Including this variable lowers the minority status coefficient from 0.485 to 0.330 and renders it insignificant. But one cannot determine whether these results reflect a real behavioral effect or simply the decline in effective sample size.

Table 2. *The Distribution of Application Quality for White and Minority Applications*

| Application Quality* (1 = highest) | Number of White Applications | Number of Minority Applications | Minority Share of Applications (%) |
|---------------------------------------|---------------------------------|------------------------------------|---------------------------------------|
| 1 | 11 | 0 | 0.0 |
| 2 | 80 | 3 | 3.6 |
| 3 | 459 | 48 | 9.5 |
| 4 | 758 | 167 | 18.1 |
| 5 | 470 | 161 | 25.5 |
| 6 | 210 | 101 | 32.5 |
| 7 | 96 | 68 | 41.0 |
| 8 | 56 | 36 | 39.1 |
| 9 | 32 | 30 | 48.4 |
| 10 | 29 | 15 | 34.1 |
| 11 | 13 | 11 | 45.8 |
| 12 | 15 | 12 | 44.4 |
| 13 | 2 | 3 | 60.0 |
| 14 | 6 | 8 | 57.1 |
| 15 | 7 | 6 | 53.3 |
| 16 | 1 | 6 | 85.7 |
| 17 | 0 | 4 | 100.0 |
| 18 | 0 | 4 | 100.0 |
| 19 | 1 | 0 | 0.0 |
| 25 | 1 | 0 | 0.0 |

*This application quality index is the predicted propensity to be denied from an econometric model of loan denial, based on the public-use version of the Boston Fed Study's data and scaled so that it falls between 1 (highest quality) and 25 (lowest quality).



THE URBAN
INSTITUTE

We agree, however, with the conclusion that a regression analysis that pools all lenders may not be able to isolate disparate treatment discrimination. This puts us at odds with Munnell et al. (1996), who argue that they can isolate disparate treatment discrimination with their pooled approach. If lenders do, in fact, use different underwriting guidelines, then a pooled regression cannot separate cases in which individual lenders apply different guidelines to minority and white applicants from cases in which minority and white customers apply to lenders with different guidelines.

Even so, we do not think that regressions for individual lenders should replace pooled regressions as the primary method for studying discrimination, for two reasons. First, regressions for individual lenders inevitably involve fewer observations than pooled regressions. Given the often high correlation between applicant characteristics and minority status, a researcher may not be able to isolate the effect of minority status without pooling the data. Thus, for example, Glennon and Stengel's (1995) finding of an insignificant minority-status coefficient in separate regressions for two lenders should not be taken as convincing evidence that these lenders do not practice discrimination.³⁴

Second, as Buist, Linneman, and Megbolugbe (1997) make clear, scholars and policymakers are concerned about discrimination based on disparate impact as well as discrimination based on disparate treatment. To precisely isolate disparate impact discrimination, one would have to know the exact relationship between loan and applicant characteristics and expected loan profitability. This relationship might vary across lenders. With this knowledge, one could estimate the relationship between loan denial and expected loan profitability. Discrimination would exist by the disparate treatment standard if minority applications were more likely to be denied than white applications with the same expected loan profitability. Discrimination would exist by the disparate impact standard if the weight given to loan or applicant characteristics in a lender's actual underwriting standards differed from the impact of those variables on expected loan profitability.

Although no study has yet estimated the relationship between loan and applicant characteristics and loan profitability with a data set that includes applicant credit history, one could argue that lenders have an incentive to figure out this relationship and to incorporate it into their actual underwriting standards.³⁵ After all, lenders can make more money if they can accurately forecast the profitability of each loan application. Because the data do not exist to make this type of forecast with much precision, individual lenders will forecast with error, but lenders may get the relationship about right, on average. If so, then deviations from average standards could indicate "errors" (perhaps intentional) by individual lenders that may impose a disparate impact on minority applicants—and may therefore be discrimination by the disparate impact standard. In this case, a pooled regression, which controls for average actual underwriting standards (using loan and applicant characteristics), might capture both disparate treatment discrimination and disparate impact discrimination. Switching to separate regressions for individual lenders might provide a better estimate of disparate treatment discrimination (assuming that sample sizes are large enough), but only at the cost of ignoring disparate impact discrimination.



One cannot be confident that controlling for loan and applicant characteristics accurately captures disparate impact discrimination, however. First, lenders may not get the relationship between these variables and loan profitability right, on average. If that is the case, this approach hides disparate impact discrimination by some lenders and mistakenly finds disparate impact discrimination by others. Second, the relationship between loan and applicant characteristics and loan profitability might differ across lenders. In this case, deviations from the average relationship might be legitimate, that is, they might not involve disparate impact discrimination. At best, therefore, the standard, pooled regression approach provides a rough approximation for the sum of disparate impact and disparate treatment discrimination. No more accurate approach has yet appeared in the literature.

Another way to deal with possible variation in legitimate underwriting standards across lenders is to assume that the lenders to which minority and white customers apply have different underwriting standards but each set of lenders gets the underwriting model right, on average. In this case, the appropriate procedure is to estimate a separate model for minority and white applicants. As noted earlier, however, one cannot reject the hypothesis that the coefficients of the explanatory variables are the same for the minority and white applicants (Munnell et al. 1996), and splitting the sample in this way does not alter the estimated impact of minority status on loan denial (Glennon and Stengel 1995). Moreover, Horne's (1997) claim that the Boston Fed Study's results are driven by two minority-owned lenders is not compelling because it is based on a nonrandom subsample of the Boston Fed Study's data. Results from the whole sample indicate that these results do not depend on the behavior of a few lenders, minority-owned or otherwise. These findings provide further support for the use of a pooled model as a useful, if approximate, way to estimate the combined impact of disparate impact and disparate treatment discrimination.

As several scholars have emphasized (see, for example, Yinger 1996; Van Order and Zorn 1995), a more accurate estimate of overall discrimination would require a two-part analysis—one part to identify legitimate underwriting standards for each lender and the second part to determine discrimination by both the disparate impact and disparate treatment standards. Legitimate underwriting standards are those with an empirically verified link to loan profitability. Disparate impact discrimination would exist if a lending institution used standards that were not legitimate in this sense and if the differences between its actual standards and legitimate ones placed minority applicants at a disadvantage. Disparate treatment discrimination would exist if it used different standards for minority and white applicants. Two-part studies of this type are clearly needed.

The lending industry appears to be moving toward the use of so-called credit-scoring schemes, which involve relatively mechanical translations of loan and applicant characteristics into a measure of creditworthiness, and toward other types of automated underwriting. Although it may not be possible to completely eliminate human judgment in the application of these procedures, they are designed to make certain that, to the extent possible, all applicants are treated the same way. As pointed out by Lindsey (1995) and Buist, Linneman, and Megbolugbe (1997), among others, this implies that these



schemes help to eliminate disparate treatment discrimination. The problem, which Lindsey does not recognize but others do (including Yezer [1995] and Buist, Linneman, and Megbolugbe), is that credit-scoring or automated underwriting schemes also could institutionalize disparate impact discrimination. To the best of our knowledge, no scholar has published a formal analysis of the link between loan profitability and a full list of loan and applicant characteristics. Institutions that have produced credit-scoring schemes may have conducted such an analysis, but it has not yet been subjected to the scrutiny of scholars.³⁶ Without this type of analysis, it is impossible to determine whether any particular credit-scoring scheme does or does not involve disparate impact discrimination. Given the growing interest in credit scoring, shedding light on these issues should be one of the highest priorities of future research on mortgage discrimination.

Conclusions

Most of the specification changes discussed in the literature have little or no impact on the Boston Fed Study's main result. As in other cases, this result seems remarkably robust. There are, however, two possible exceptions: reverse regressions and the use of different regressions for each lender.

Our investigation of the reverse-regression approach indicates that it is not appropriate for analyzing discrete dependent variables. Thus, reverse regression does not provide a legitimate alternative to a loan denial regression, and one cannot legitimately reject the Boston Fed Study's findings on the basis of reverse-regression results.

Separate regressions for each lender might provide a clearer picture of disparate treatment discrimination if the number of applications from each individual lender is large. With only a few applications for each lender, however, this approach cannot provide any useful information. Moreover, even if this approach sheds light on disparate treatment discrimination, it does so at the cost of hiding disparate impact discrimination. The same issue arises in any credit-scoring scheme, which may minimize disparate treatment discrimination at the cost of promoting disparate impact discrimination. Pooling all lenders, as in the Boston Fed Study, provides the best approach currently available for measuring discrimination by both the disparate treatment and the disparate impact standards. However, this approach depends on several untested assumptions, and further research is needed on these important issues.

Endogenous Explanatory Variables

Rachlis and Yezer (1993) argue that single-equation models of loan denial are biased because many loan terms are actually the result of borrower choices. They highlight the endogeneity of LTV, arguing, in particular, that LTV is endogenous because negotiation with a lender will induce marginal loan applicants to increase their proposed down payment.³⁷ They show that this type of response will bias the coefficient of LTV downward and the coefficient of minority status upward. Yezer, Phillips, and Trost (1994) explore this issue further by construct-



ing a model in which the lender rejection decision depends on LTV and on an index of default likelihood, LTV depends on indices of approval and default likelihood, and default depends upon LTV. Using simulations, they demonstrate that if this structure is correct, a single-equation model of loan denial will overstate the minority-status coefficient. As evidence for the plausibility of their approach, they cite evidence from Schafer and Ladd (1981) that the effect of minority status on loan denial varies with number of times an application is reconsidered by the lender. During any reconsideration, they argue, loan terms may be altered and more negotiation may take place.

Browne and Tootell (1995) argue that Yezer, Phillips, and Trost (1994) have not established that negotiation is widespread. In addition, Tootell (1996b) argues that negotiation does not imply simultaneity. Endogeneity bias will not arise, he says, so long as any adjustments to loan terms have been finalized before the denial decision. To examine this issue, Brown and Tootell estimate a model in which LTV is treated as endogenous. As in most such procedures, they make use of new exogenous variables, called instruments, to "identify" the model, that is, to separate the true effect of the endogenous variable (LTV) on the dependent variable (loan denial) from the factors that affect both of these variables. The variables they use as instruments are the applicant's age minus years of education and the applicant's income.³⁸ They find that the minority-status coefficient in the loan denial equation is actually larger with this procedure than when LTV is treated as exogenous. Tootell (1996b) estimates a similar model with a somewhat different set of instruments and obtains a similar result. Yinger (1996) provides a conceptual basis for these findings. He points out that if white applicants receive additional aid in preparing their loan application or loan officers are more willing to negotiate with white applicants, then treating LTV as exogenous leads to an understatement of the minority-status coefficient in a loan denial equation—not an overstatement.

Another explanatory variable that may be endogenous is whether the application involves a special loan program. Phillips and Yezer (1995) investigate this possibility. Previous studies either control for whether an application is for a special program or drop all applications involving a special program. Phillips and Yezer observe that neither of these approaches yields consistent estimates if applying for a special program loan is, in fact, endogenous. They estimate a model in which the applicant first decides whether to apply for a special program. If this decision is negative, and the applicant applies for a regular loan, the lender then decides whether to accept that application. In technical terms, this model corrects for the "selection" of applications into a special loan program.³⁹ They observe that the applicant must decide whether to apply for a special program before the property is assessed, so that the assessed value cannot influence the special program decision. This observation identifies the exogenous variables they need for their simultaneous equations procedure; in particular, they include the ratio of the loan amount to the house price in the special program equation and the ratio of the loan to the assessed value in the loan denial equation. Using the Boston Fed Study's public-use sample, they find that the minority-status coefficient in this model is not statistically significant and is almost 30 percent lower than the minority-status coefficient in a model that simply omits the applications that involve a special loan program.



THE URBAN
INSTITUTE

One should be cautious in interpreting this result, however, for two reasons. First, their model also includes both the “unverified information” and “meets guidelines” variables, and the minority-status coefficient in the model that simply drops special program applications was only significant at the 10 percent level. Second, the estimated correlation between the error terms in the two equations was only -0.155 (with a t -statistic of 0.53), which provides only weak support for the claim that special program choice is endogenous.⁴⁰

In a study of redlining, Ross and Tootell (1998) treat yet another variable as endogenous, namely, whether the applicant obtained PMI.⁴¹ The first stage of their model explores the factors that determine the receipt of PMI, and the second stage is a loan denial equation in which one of the explanatory variables indicates whether the applicant obtained PMI. In estimating this model, Ross and Tootell drop all 77 cases in which an applicant applied for and was denied PMI. Dropping these observations greatly simplifies the model, because it avoids the necessity of estimating the PMI-denial decision, a step that Munnell et al. (1996) show has no impact on the minority-status coefficient in the loan denial equation. Ross and Tootell find little evidence to indicate that PMI is endogenous. The correlation between the unobservable factors in the PMI and denial equations is only -0.135 (with a t -statistic of 0.49), and the minority-status coefficient in their loan denial equation is not affected by allowing PMI to be endogenous.

Evaluation

The findings of Yezer, Phillips, and Trost (1994) are thought-provoking because they indicate that the Boston Fed Study’s results may be biased, even if the data are not flawed and no important underwriting variable is omitted from the equation. This bias arises because information about a lender’s treatment of an application affects the borrower’s behavior. In particular, LTV may decrease with the likelihood of loan rejection, even after controlling for all borrower characteristics that enter the underwriting problem, so LTV may be correlated with the unobservable factors in the loan denial equation. This type of correlation leads to biased coefficients, and the sign of the bias depends on the correlation between minority status and LTV. Empirically, minorities submit loan applications with higher LTVs than do otherwise-equivalent whites, which supports Yezer, Phillips, and Trost’s argument that single-equation loan denial models are biased toward finding discrimination. This does not prove, of course, that other biases are not also at work. For example, the possibility that whites receive more coaching or are treated better during negotiations, as suggested by Yinger, would lead to a bias in the other direction. The net impact of these, and perhaps other, biases on the estimated minority-status coefficient is, of course, an empirical matter.

Furthermore, we are not convinced by Tootell’s argument that the influence of the lender on LTV precedes the denial decision and so is not endogenous. An individual loan officer knows about his bank’s underwriting procedures and can probably give the applicants a good indication about whether their applications will be approved in any given form. This unobservable information about an application’s probability of denial may influence



both LTV and the eventual denial decision. Thus, a single-equation model of loan denial could suffer from endogeneity bias even if LTV is fixed at the time of the final loan denial decision. It may indeed be the negotiation that matters, not the timing.

Although it makes a valuable contribution, the model provided by Yezer, Phillips, and Trost (1994) understates the complexity of the problem. In fact, LTV is influenced by three sets of unobservable factors. In addition to the unobservable factors in the loan denial equation, there are unobservable factors in the LTV equation itself, and, because LTV depends on the likelihood of default as well as the likelihood of rejection, LTV is influenced by the unobservable determinants of default. Moreover, the likelihood of default may influence loan denial, in which case the unobservable factors in the default equation will be embedded in the unobservable factors in the loan denial equation. Only one of these relationships—and potential biases—is fully investigated by Yezer, Phillips, and Trost. They allow the unobservable factors in the default and loan denial equations to be correlated, a step that leads to a decrease in the minority-status coefficient in the loan denial equation. However, accounting for the other relationships among unobservable factors is likely to have the opposite impact on this minority-status coefficient and could even reverse the findings of Yezer, Phillips, and Trost. We conclude that their findings are far from definitive.

Although these studies do not convince us that endogeneity results in an overstatement of the minority-status coefficient, we also are not convinced that existing studies adequately deal with the endogeneity problem. A simultaneous equations procedure generally requires the use of variables, called instruments, that are highly correlated with the endogenous explanatory variable, in this case LTV, but not with the dependent variable, in this case loan denial, when the endogenous variable is also included in the regression. To put it more formally, good instruments must meet three criteria. First, they must make conceptual sense. Second, they should be significant in a regression to explain the endogenous explanatory variable, in this case LTV. Third, they should not have explanatory power in the regression of interest, in this case a regression to explain loan denial, when the potentially endogenous variable is also included.⁴² In our view, the last criterion requires a stronger standard than the usual significance test; if a variable is even marginally significant in the regression of interest, then it should not be used as an instrument.

Now consider the analysis by Brown and Tootell (1995). To account for the possible endogeneity of LTV, they use an applicant's income and age minus education as instruments. We do not find this approach compelling.⁴³ First, income is often significant in the loan denial equation, which violates the third of the above criteria. Age minus years of education is not available in the public-release data set, but the education variable in that data set is not statistically significant in the LTV model, which violates the second criterion. We conclude that they have not eliminated the bias from an endogenous LTV. The instruments proposed by Tootell (1996b) are no better. He uses liquid assets, marital status, gender, and years in this line of work but also includes years on the job (which is close to age minus education) and education. The new variables on this list all have explanatory power in some of the loan denial regressions in Munnell et al., which implies that they violate the third criterion for a good instrument.



THE URBAN
INSTITUTE

Reanalysis

This review reveals that LTV may indeed be endogenous in a loan denial equation and that estimates of the effect of minority status on loan denial may be biased if this endogeneity is not taken into account. On conceptual grounds, this bias could work in either direction, however, so the nature of the bias is ultimately an empirical question, which is not adequately answered by existing analysis.

To further investigate this issue, we estimate a simultaneous equations model with a new set of instruments, namely, house price plus the pairwise interactions among income, house price, and liquid assets. These instruments meet the first criterion stated above, that is, they make sense conceptually. The interaction between liquid assets and house price may indicate the ability of the household to make a down payment. Moreover, the interaction between income and liquid assets or income and house price may explain a household's desire to make a larger down payment and reduce monthly housing expenses. Finally, none of these interactions is likely to influence the loan denial decision, because the link between the underlying variables and default is already captured in LTV and in the housing-expense-to-income ratio.

These hypotheses are confirmed by the data in the sense that all of these variables are statistically significant in an equation to explain LTV. Thus, these variables appear to meet the second criterion for an appropriate instrument. In addition, each of these four variables has a small, insignificant coefficient in a loan denial equation that includes LTV, even at the 50 percent confidence level, which indicates that they meet the third criterion as well.⁴⁴

Having identified appropriate instruments, we then estimate the loan denial equation accounting for the endogeneity of LTV.⁴⁵ As a point of comparison, we first estimate the loan denial equation with the continuous version of LTV instead of the categories in our base case. The coefficients (*t*-statistics) of the minority-status variable for our three models are 0.260 (2.57), 0.272 (3.06), and 0.375 (4.19), and the impacts of minority status on loan denial are 4.3, 6.0, and 8.1 percentage points (see table 1). When LTV is treated as endogenous, the estimated minority-status coefficients are 0.253 (2.50), 0.233 (2.35), and 0.354 (3.62) and the minority-status effects are 4.2, 5.3, and 7.7⁴⁶ (again see table 1). Accounting for the endogeneity of LTV lowers the minority-status effect in every case, but the largest decline (in model 2) is only 0.7 percentage points, or less than 12 percent. These results support the view that single-equation estimates of the impact of minority status on loan denial can be biased upward but also indicate that this bias is likely to be small.⁴⁷

The loan-to-value ratio is endogenous because borrower or lender behavior may affect its numerator, the size of the loan. If loan size is endogenous, however, then so is the size of the mortgage payment, which is a key component of the housing-expense-to-income ratio.⁴⁸ In order to treat another explanatory variable as endogenous, an analysis generally must have at least one more instrument. In this case, we know that housing expense includes insurance and property taxes, as well as mortgage payments, and that the relationship between the loan amount and the monthly mortgage payment depends on the term of the mortgage.⁴⁹ Because insurance and taxes may vary by location, we include as instruments both a dummy variable for the central county in the Boston area and this dummy variable interacted with several factors that might influence



insurance or assessments—namely, whether the unit is to be owner-occupied, whether the unit is a multifamily unit, and three census tract characteristics. These characteristics are racial composition, tract income, and whether the percentage of units that are boarded up exceeds the metropolitan median value. It also seems reasonable to suppose that these instruments will affect loan denial only indirectly through the housing-expense-to-income ratio. Finally, the term of the mortgage is interacted with the age of the applicant on the assumption that older households may want a shorter mortgage. All of these instruments pass the second and third criteria listed earlier; that is, they are statistically significant in an equation to explain the housing-expense-to income ratio but have no explanatory power in the loan denial equation when the housing-expense-to-income ratio is also included.

With both LTV and the housing-expense-to-income ratio treated as endogenous, the minority-status coefficients in our three loan denial models are 0.265 (2.49), 0.218 (2.16), and 0.398 (4.38).⁵⁰ The corresponding impacts of minority status on loan denial are 4.4, 5.1, and 8.6 percentage points (see table 1). Hence, treating this additional variable as endogenous actually raises the impact of minority status on loan denial in models 1 and 3 and lowers it only slightly, from 5.3 to 5.1 percentage points, in model 2.

These results complicate our conclusions about the role of the endogeneity of the loan amount. If model 1 or model 3 is correct, then single-equation estimates actually understate the impact of minority status on loan denial to a small degree, but if model 2 is correct, then single-equation estimates overstate this impact by about 15 percent (row 6 relative to row 4 in table 1). Even in model 2, however, the impact of minority status on loan denial remains large and statistically significant when we account for this endogeneity. It is, of course, possible that a different set of instruments would lead to different results, but we believe that the instruments we have used clearly meet all the criteria for good instruments and that minority status has a large, significant impact on loan denial when any reasonable set of instruments from the Boston Fed Study's public-use data set is used.⁵¹

The bias identified by Yezer, Phillips, and Trost (1994) arises from negotiation, so one might also ask whether the minority-status coefficient is different when more negotiation takes place. Yezer, Phillips, and Trost suggest that there is more opportunity for negotiation when an application is considered more than once; if so, the bias in the minority-status coefficient should be larger for applications that are considered more often. To test this negotiation hypothesis, we first drop all applications that were considered four or more times, a sign of extensive negotiation, and reestimate the model with LTV treated as exogenous. This step, which results in a sample size of 2,717, changes the impact of minority status in our three loan denial models to 4.1, 6.7, and 8.4 percentage points. For models 2 and 3, these estimates are larger than our baseline estimates and therefore support the negotiation hypothesis, although the differences from the baseline estimates are not very large. As a further test, we drop all applications that were considered three or more times, resulting in a sample size of 2,301. According to the negotiation hypothesis, this step should lower the bias and hence lower the estimated impact of minority status compared with the first step. In this case, the impacts of minority status on loan



denial are 3.9, 5.7, and 8.7. These results provide further weak support for the negotiation hypothesis for models 1 and 2 but contradict it for model 3. Finally, we drop all applications considered two or more times, that is, all applications with evidence of negotiation. This step, which reduces the sample to 1,454, should lower the estimated impact of minority status still further. We find just the opposite, however, as the impacts of minority status on loan denial are now 7.1, 10.0, and 12.0. In short, we find no systematic relationship between minority-white differences in loan denial and the number of times an application is reviewed. We conclude that either negotiation is not the source of endogeneity in loan amount or else the number of times an application is reviewed is a poor measure of the extent to which negotiation takes place.

Following Phillips and Yezer's (1995) lead, we also estimate a simultaneous equations model in which an individual decides whether to apply to a special program. We use the same three specifications that we have used all along, except that we change one of the explanatory variables. Our previous models of loan denial include the ratio of loan amount to the minimum of assessed value and house price. Like Phillips and Yezer, we now include the ratio of loan amount to assessed value in the denial model and the ratio of loan amount to house price in our model of the special program application decision. As a point of comparison, we first estimate our three loan denial models after deleting all applications that involve a special program. The minority-status coefficients (*t*-statistics) from these models are 0.329 (2.93), 0.291 (2.90), and 0.418 (4.24), and the impacts of minority status on the probability of denial are 5.1, 6.5, and 8.7. With a correction for the selection of some applications into special programs, the minority-status coefficients in the three denial equations, which in this case apply just to loans that do not involve special programs, are 0.375 (3.70), 0.312 (3.64), and 0.439 (5.20), and the minority-status effects have increased substantially to 10.0, 9.7, and 13.7. These results differ from Phillips and Yezer's and indicate that, if anything, the failure to treat special programs as endogenous biases downward the impact of minority status on loan denial.

One possible explanation for the difference between our results and Phillips and Yezer's (1995) results is that they use a much more parsimonious specification and, in particular, omit a number of variables that are statistically significant in the denial model. For our three models, the estimated correlation between the unobservable factors in the special-program-choice equation and the unobservable factors in the loan denial equation is close to minus one.⁵² This result implies that the model is not well identified; that is, it cannot accurately account for the endogeneity of special program choice because we cannot distinguish between the unobserved factors in the two equations.⁵³ Phillips and Yezer do not run into this problem, because they drop several variables from the denial equation, thereby lowering the correlation between the unobservable factors in the two equations—and thereby introducing omitted-variable bias into their results. In other words, they solve the identification problem but cause another, more serious one. Given the difficult trade-off facing researchers in this case, we think future research should investigate other ways to identify the model.

As in the case of an endogenous LTV, variables for house price, liquid assets, and applicant income, along with their interactions, appear worth investigating



as possible determinants of the choice to apply for a special program. In fact, all six variables are statistically significant in a single-equation model for explaining the choice to apply for a special program, which indicates that these variables meet the second criterion listed above for a good instrument. These variables also are insignificant in a denial specification that controls for whether the applicant applied for a special program, which indicates that they also meet the third criterion.

On the basis of these results, we estimate three simultaneous equations models of loan denial and special program choice, with these additional variables included only in the special-program-choice equation. In addition, these models include the traditional LTV variable in the loan denial equation, and exclude the “meets guidelines” variable (or its instrument) from the special-program equation, because the borrower does not gain insight into whether he meets the lender’s underwriting standards until after he has submitted his application. This approach appears to do a better job of identifying the model. In particular, the estimated correlations between unobserved factors in the two equations are now all close to -0.5, instead of the almost perfect correlation with the previous approach.⁵⁴ Again for comparison, we find that the impacts of minority status on the probability of loan denial when special program loans are dropped are 5.7, 7.2, and 9.5 percentage points. With special program choice treated as endogenous, these impacts become 6.6, 8.9, and 11.1 points (see table 1). These results also support the conclusion that a failure to treat special program choice as endogenous biases the minority-status coefficient downward, but the bias is smaller in this case than in the previous one.

Conclusions

In conclusion, the key finding of the Boston Fed Study is unaffected when the model includes a complete accounting for the endogeneity of LTV—or of any other variable. In all of the models examined here, the impact of minority status on loan denial is statistically significant and close to the value in the Boston Fed Study’s equations. Moreover, we find that accounting for the endogeneity of special program choice actually boosts the impact of minority status on loan denial.

Conclusions Concerning the Boston Fed Study

Munnell et al. (1996), otherwise known as the Boston Fed Study, has been subjected to a phenomenal, and in our experience unprecedented, volume of criticism. Critics have argued that the study’s main finding, that minority applicants are more than 80 percent more likely to be turned down for a loan than are equivalent white applicants—an indication of discrimination in mortgage lending—is biased upward because the study (a) omitted variables; (b) used data with errors in the explanatory variables; (c) used a dependent variable that misclassifies loan outcomes; (d) used the wrong algebraic form, also known as misspecification; and (e) failed to account for endogeneity in several different explanatory variables.



THE URBAN
INSTITUTE

We have examined all of these arguments and, where possible, explored them with the public-use version of the Boston Fed Study's data. In some cases, we find that the critics are simply wrong: The problem they identify does not exist, or the bias involved is empirically insignificant. In several cases, however, we agree with the critics that a limitation in the Boston Fed Study could potentially lead to a serious overstatement of discrimination, and we have explored these cases in detail. Moreover, we find that the literature has raised several important issues concerning the interpretation of the Boston Fed Study's results.

This analysis leads us to five main conclusions.

First, we conclude that the large differences in loan denial between minority and white applicants identified by Munnell et al. cannot be explained by data errors, omitted variables, or the endogeneity of loan terms. No study has identified a reasonable procedure for dealing with any of these potential problems that eliminates the large positive impact of minority status on loan denial. One cannot, of course, prove that no bias exists in any particular equation, but one can examine all of the potential sources of bias identified by scholars for that case. Scholars have been unusually creative in identifying potential biases in the Boston Fed Study, but our analysis, based on the best data currently available, reveals that none of these potential biases can explain why the estimated minority-status coefficient in a loan denial equation is so large.

For example, some scholars have claimed that the "meets guidelines" variable should be included to correct for elements of applicants' credit histories that are omitted from the explanatory variables in the Boston Fed Study. If this variable does indeed capture such omitted elements, however, then the unobserved factors influencing "meets guidelines" will be correlated with the unobserved factors influencing loan denial. We show that this is not the case. It follows that the "meets guidelines" variable does not correct for omitted variables. In addition, we find that accounting for the endogeneity of various loan terms never results in a substantial reduction in the estimated minority-status coefficient, and in some plausible cases this step actually makes that coefficient larger.

Second, we conclude that no study has demonstrated either the presence or the absence of disparate treatment discrimination in loan approval, at least not in a large sample of lenders.⁵⁵ This conclusion puts us at odds both with the authors of the Boston Fed Study, who claim that they measure disparate treatment discrimination, and with several of their critics, who claim that there is no discrimination at all.

In our view, the Boston Fed Study's results measure disparate treatment discrimination only under the assumption that all lenders use the same underwriting guidelines. With this assumption, any group-based difference in treatment after controlling for underwriting variables implies that the guidelines are applied differentially across groups, which is, by definition, disparate treatment discrimination. Because virtually all lenders sell some of their loans in the secondary mortgage market, they have some incentive to use the underwriting guidelines that institutions in that market, such as Fannie Mae, have established. However, many loans are not sold in the secondary market, and the lending process often involves many individuals in the same lending institution, who may not all have the same incentives. Even on conceptual grounds,



therefore, the same-guidelines assumption is a strong one, and no existing empirical study can confirm (or deny) it.

Third, we conclude that no study has demonstrated either the presence or the absence of disparate impact discrimination in loan approval. The Boston Fed Study's results measure disparate impact discrimination only under the assumptions that (a) different lenders use different underwriting guidelines; (b) existing guidelines are accurately linked to loan profitability, on average; and (c) existing deviations from average guidelines cannot be justified on the basis of business necessity. These assumptions could be satisfied, for example, if underwriting guidelines vary across lenders solely for idiosyncratic reasons or if some lenders purposefully develop guidelines that have a disparate impact on minority applicants. However, no existing study sheds light on whether these assumptions are met.

Fourth, following the logic of civil rights legislation, the Boston Fed Study establishes the presumption that in 1990 lenders in Boston engaged in either disparate treatment discrimination, disparate impact discrimination, or both. This presumption can be rebutted only with evidence that the observed minority-white differences in loan approval can be entirely explained by profit-based differences in the underwriting guidelines used by the lenders to which minorities and whites applied. To use the legal term, the Boston Fed Study builds a *prima facie* case that discrimination exists. If such a case were made in a courtroom setting, the burden of proof would shift to lenders. To escape the conclusion that they are discriminating, lenders would have to prove that their actions were based on "business necessity," that is, that they used underwriting guidelines with a clear connection to the return on loans, that they applied these guidelines equally to all groups, and that no equally profitable guidelines without a disparate impact on minority applicants were available. Our conclusion builds on the spirit of this legal standard. In our view, the Boston Fed Study builds a strong *prima facie* case for discrimination, and no scholar has come close to showing that the observed intergroup differences in loan approval in Boston can be justified in business terms.

In fact, the available evidence, while far from conclusive, suggests that business necessity is unlikely to explain a large share of the observed minority-white difference in loan denial. In particular, legitimate differences in underwriting guidelines must be associated with real differences in lenders' experiences. They are therefore most likely to arise between lenders that specialize in groups of borrowers with different average creditworthiness. Thus, if differences across lenders in legitimate underwriting criteria have a major impact on the observed minority-white difference in loan denial, then allowing underwriting criteria to vary across lenders should dramatically lower the estimated minority-status coefficient. This turns out not to be the case. Munnell et al. (1995) can reject the hypothesis that the underwriting model is different for single-family houses, multifamily houses, or condominiums. Moreover, both Munnell et al. and Hunter and Walker (1996) find little evidence that individual underwriting variables receive different weights for minority and white applicants. In addition, Munnell et al. show that the minority status coefficient is virtually the same when separate regressions (and hence separate underwriting guidelines) are estimated for lenders that specialize in lending to minorities and



THE URBAN
INSTITUTE

for other lenders. Finally, Browne and Tootell (1995) show that the minority-status coefficient is literally unaffected if one excludes two minority lenders, which together account for half of the minority applications in the Boston Fed Study's sample.

As explained earlier, the "meets guidelines" variable might be related to the issue of business necessity. If we assume that minority households do a poorer job than white households in selecting lenders that meet their credit needs, then including the "meets guidelines" variable (and treating it as endogenous) can be interpreted as a way to account for legitimate differences in underwriting guidelines across the lenders visited by minorities and whites. In this case, we find that roughly 37.5 percent [(7.7 – 5.6)/5.6 from the first row of table 1] of the minority-white difference in loan denial is due to business necessity, not discrimination. However, this assumption is not consistent with the results in the previous paragraph. If minority households simply do a poorer job finding just the right lender, then, contrary to this evidence, the minority-white difference in loan approval should disappear for lenders that specialize in lending to minorities.

Fifth, we conclude that the best way to determine whether the observed minority-white differences in loan denial are the result of underwriting practices justified by business necessity would be to conduct a replication of the Boston Fed Study in other locations with the addition of loan performance data. This approach would make it possible to determine which observed application characteristics are accurate predictors of loan returns and therefore which underwriting guidelines are legitimate. Minority-white differences in loan denial that remain after accounting for legitimate underwriting guidelines are evidence of discrimination. Research along these lines is particularly important for policy purposes because credit-scoring and other automated underwriting schemes, which are becoming increasingly popular, have enormous potential to lessen disparate treatment discrimination while at the same time magnifying disparate impact discrimination.

Unfortunately, this approach would not be able to distinguish between disparate treatment and disparate impact discrimination. A combination of application data, including credit history, and performance data should make it possible to identify legitimate underwriting guidelines and even to determine if those guidelines vary by location or by some other variable. However, these data would contain only a few observations for each individual lender and therefore could not be used to identify each lender's actual underwriting guidelines. As a result, a researcher could not determine whether remaining minority-white differences in loan denial for an individual lender are due to that lender's use of different guidelines for minorities and whites (disparate treatment discrimination) or its use of illegitimate guidelines that place minority applicants at a disadvantage (disparate impact discrimination).

Sixth, we conclude that the best, and perhaps the only, way to measure disparate treatment discrimination is with audit methodology. In an audit, two applicants with the same credit histories and in need of the same type of loan would apply for a mortgage at the same lender. Disparate treatment discrimination exists if minority applicants are systematically treated less favorably in a large sample of audits. Audits of this type would shed no light on



disparate impact discrimination, because they would compare the treatment of identically qualified minority and white applicants at the same lender. Thus, observed differences in treatment could not be due to underwriting guidelines that illegitimately magnify differences in credit characteristics between minorities and whites, that is, to disparate impact discrimination.

Unfortunately, an audit study of loan approval faces many major practical challenges. Perhaps the most important is that it would be difficult, and might even be illegal, to assign false credit characteristics to auditors as a means of ensuring that audit teammates had identical loan qualifications. This step would be difficult because it would require the cooperation of the firms that maintain the credit records that lenders refer to. It might be illegal because laws prohibit false statements on credit applications with intent to defraud. We do not believe that auditing is a fraudulent activity. But the courts have not yet ruled on this matter, and any group that pushes audits into the loan approval stage of the mortgage process might face high legal bills, if not something worse. It might be possible to conduct audits using auditors' actual credit characteristics, but this approach would be administratively difficult because auditors would still have to be matched to have the same credit qualifications. As a result, a very large pool of potential auditors would be necessary.

By collecting, analyzing, and releasing their data, the authors of the Boston Fed Study have made an enormous contribution to the literature on lending discrimination, but their study is certainly not the last word on the subject. Can similar evidence of discrimination be found in urban areas other than Boston? Has the estimated level of discrimination declined? Do lenders engage in disparate-treatment discrimination, or disparate impact discrimination, or both? To what extent can observed minority-white differences in loan denial, controlling for applicants' credit histories, be explained by legitimate differences in underwriting guidelines across lenders instead of by discrimination? Given the potential importance of lending discrimination as a barrier to homeownership for minority households and the range of questions about lending discrimination that remain unanswered, further research on these questions is urgently needed.

Notes

1. Stephen L. Ross is an assistant professor of economics at the University of Connecticut and John Yinger is a professor of economics and public administration at the Maxwell School, Syracuse University. They are grateful to Anthony Yezer and Geoffrey Tootell for helpful comments. The views expressed in this and the following two chapters should not be attributed to anyone but the authors.
2. For a description of the lenders covered by the HMDA data, see Avery, Beeson, and Sniderman (1996).
3. The 1991 figures are presented in Yinger (1995), and the 1997 figures come from Federal Financial Institutions Examination Council (1998). By way of comparison, the Hispanic/white loan rejection ratio over this period declined from 1.74 to 1.47.
4. The original version of the Boston Fed Study was released as a working paper in 1992 and the final version was published in 1996. Because many of the criticisms focused on the original version and were published in 1994 or 1995, Munnell et al. (1996) includes considerable material responding to the critics. Subsets of the authors of the Boston Fed Study also have published additional responses. See Browne and Tootell (1995) and Tootell (1996b).



THE URBAN
INSTITUTE

5. A discussion of these potential flaws draws on several econometric theorems. A brief discussion of the key econometric concepts is provided in the technical appendix to chapters 3 through 5, which appears following chapter 5.
6. Another variable is considered by Hunter and Walker (1996), who argue that loan denial may depend on how “thick” an applicant’s file is, as measured by whether there are two or more credit checks in the file. This variable proves to be insignificant.
7. This effect was described to us in correspondence from Geoffrey Tootell. Note that our regressions do not include lender dummies because, to protect confidentiality, they are not included in the public-use data set. Because they omit these variables, our regressions overstate the minority-status coefficient by 20 percent. However, the public-use data set also does not, for the same reason, include census tract dummies, which raise the estimated minority-status coefficient. By coincidence, these two effects almost exactly offset each other, so our estimate of the minority/white denial gap using the methodology that is closest to the Boston Fed Study’s, 7.7 percentage points, is almost the same as the Boston Fed Study’s estimate, 8.2 percentage points.
8. This list is similar to the set of variables in the baseline estimation of Munnell et al. (1996), except that it substitutes census tract characteristics for tract dummies and excludes lender dummies. Munnell et al. could not reject the hypothesis that a separate coefficient for black applicants was the same as one for Hispanic applicants.
9. As discussed in the technical appendix following chapter 5, this is the percentage of the variance in the underlying latent variable that is explained by the model.
10. In a probit model (as used here) or the similar logit model (in Munnell et al. 1996), a percentage impact is determined by comparing the average predicted probability for all observations at two different values of the variable in question. In this case, we compare the average predicted probabilities of denial for minority applications with the minority-status coefficient set to zero and to its estimated value.
11. This technique is called a bivariate probit with recursion. Equations describing the model are presented in the technical appendix.
12. This correlation also might reflect omitted variables.
13. To make them comparable with the simultaneous equations procedures in the technical appendix that follows chapter 5, the single-equation results we present here and elsewhere are based on bivariate probit models, not the related logit models used by Munnell et al. However, logit and probit results for comparable equations are similar.
14. The reader may be puzzled by the fact that the minority-status coefficient is larger in the first specification but the percentage impact on loan denial is smaller. This apparent contradiction arises because predicted probabilities from a two-equation model reflect not only the estimated coefficients but also the estimated correlation between unobserved factors across equations.
15. An alternative response to these results would be to include “unable to verify” and treat it as endogenous. We implemented this alternative approach for many of the models discussed later in this chapter and found that the results are very similar to those using the simple approach of dropping this variable altogether. Consequently, we present only the results from the simpler approach.
16. The equations for this model, along with a discussion of econometric issues it raises, including the identification of the model, can be found in the technical appendix following chapter 5.
17. We return to the issue of disparate impact discrimination—and how to observe it—in the section titled “Endogenous Explanatory Variables,” which addresses issues of specification.
18. For further discussion of disparate impact discrimination, see Schwemmm (1992) and Yinger (1998).
19. We also pointed out earlier that a control variable for “lender toughness” has no impact on the minority-status coefficient and that the Boston Fed Study’s regressions already include lender dummies.



20. We are grateful to Geoffrey Tootell for suggesting this example to us.
21. This test is made in response to a claim by Horne (1997), who says that dropping these two lenders pushes the minority-status coefficient to zero in his subsample of Federal Deposit Insurance Corporation (FDIC) applications. However, when Browne and Tootell drop these two observations from the entire FDIC subsample, which is larger than the nonrandom subsample evaluated by Horne (discussed in the next section), they find that the minority-status coefficient only drops from 0.70 to 0.66.
22. The one exception is that Hunter and Walker find that the minority-status coefficient is larger at a high obligation ratio than at a low one.
23. There is some disagreement between Horne and the Boston Fed Study's authors about the extent to which they are making the same corrections. Munnell et al. (1996) cite an unpublished paper by Horne, which, after revisions, became Horne (1994). Tootell (1996b) cites another unpublished paper by Horne, which, after revisions, became Horne (1997). The last word, so far, is in Horne (1997).
24. They also remove observations on the exceptions list that were identified as problematical by Horne (1994).
25. Note that the baseline estimate presented here is lower than the estimates reported above because Day and Liebowitz also alter the data by removing the observations that Horne (1994) claims are misclassified. This issue is discussed in the next section.
26. Results dropping multifamily applications and dropping applications with low LTVs are presented later in this report.
27. This formula is presented in the technical appendix that follows chapter 5.
28. For one attempt to consider counteroffers, see Schafer and Ladd (1981).
29. Glennon and Stengel perform other regressions that qualify this finding. See the discussion in the section below titled "Incorrect Specification."
30. This argument is motivated by a theory about the causes of lending discrimination. We return to this topic in chapter 4.
31. Munnell et al. (1996) also find that self-employed minority applicants are less likely than other minority applicants to encounter discrimination. The impact of minority status on loan denial does not vary with any of their other explanatory variables.
32. Glennon and Stengel's regression specification differs across lenders in an attempt to capture differences in underwriting standards. When the same specification is used for all three lenders, the significance of the minority-status coefficient for one lender jumps from the (two-tailed) 8 percent level (which does not quite meet the conservative two-tailed standard for significance) to the 1 percent level. This study is discussed further in chapter 4.
33. The coefficient (*t*-statistic) is 0.98 (2.84) for lenders with many minority loans and 0.91 (2.22) for other lenders. See Munnell et al. (1996), table 5.
34. A similar conclusion applies to the Buist, Linneman, and Megbolugbe (1997) conclusion that there is no discrimination in the sample of loans that meet standard underwriting guidelines. Not only are there very few minority applications in this sample, but there are so few denials that no underwriting variable has a significant impact on loan denial.
35. We return to studies of loan performance in chapter 4.
36. Avery et al. (1998) find that applicants' credit scores are correlated with several locational factors, including the racial composition of a census tract. They point out that this result raises concerns about possible disparate impact discrimination but is far from conclusive.
37. The equations defining their model are presented in the technical appendix following chapter 5.
38. The technical appendix following chapter 5 presents a formal version of their approach.



THE URBAN
INSTITUTE

39. This model is estimated with conditional bivariate probit analysis, which controls for the selection of applicants into special programs.
40. Phillips and Yezer (1995) also estimate a model in which the first decision is the lender's loan denial decision and the second decision is whether the borrower accepts the lender's offer. This model, estimated with conditional probit, has a key technical flaw. The authors do not have any new exogenous variables to "identify" the model, so they identify it by making different assumptions about the functional forms of the two equations, but using the same set of explanatory variables. If these untested assumptions are not correct, their procedure yields biased estimates. Moreover, the estimated correlation between unobservable factors in the two equations is implausibly high, -0.997, and the authors indicate that many alternative specifications did not converge; that is, they could not find a solution for the underlying econometric formulas.
41. The PMI variable was also examined by Munnell et al. (1996). They found that the minority-status coefficient in the loan denial equation was not affected by dropping this variable or by dropping all observations that were denied PMI.
42. Although the validity of an instrument can be tested by including it in the regression of interest, an instrument is not included in the final form of this regression, but is instead used to "cleanse" the relevant explanatory variable of its endogenous component.
43. Rachlis and Yezer (1993) reject the use of instrumental variables in this case because any variable that explains LTV is likely to be used by lenders during the underwriting process, that is, to be correlated with denial. We agree that one must use caution in selecting instruments in this case, but we also believe it is possible to identify acceptable instruments.
44. To be specific, we estimated a loan denial model in which the LTV dummy variables are replaced with LTV itself and that includes income, house price, liquid assets, and all three pairwise interactions as explanatory variables. The coefficients of house price and of all three pairwise interactions are small and statistically insignificant (*t*-statistics of about 0.5 or lower). We also can reject the hypothesis that this set of variables is statistically significant in the loan denial equation. In particular, the χ^2 ratio for the appropriate likelihood-ratio test is below 1.3 for all three models in table 1. The 5 percent critical value for this test is 9.5.
45. Missing values for income, house price, and liquid assets required the deletion of 19 observations, resulting in a sample size of 2,912. In addition, we drop 16 observations in which the application's LTV is 1.5 or greater. We estimate LTV using ordinary least squares and use the predicted LTV as a regressor in a probit analysis of loan denial. For the specification with the "meets guidelines" variable when applications with extreme LTVs are not eliminated, the R^2 is 0.10 for the LTV models, but when the outliers are dropped the R^2 is 0.25. Moreover, the coefficient on the LTV variable is only 0.304 (*t*-statistic of 3.09) when the outliers are not eliminated, but 0.971 (3.97) when the outliers are dropped. Similar results arise for the other two specifications. In addition, making LTV endogenous requires us to alter the model in which "meets guidelines" is treated as endogenous, because LTV was treated as an (exogenous) instrument in that model. In particular, LTV is replaced as an instrument for the "meets guidelines" variable with the instruments used to identify LTV.
46. Our procedure is to obtain a predicted value of LTV using the exogenous instruments and then to include this predicted value in the probit regression for loan denial. The coefficients of actual LTV in the loan denial models are 0.804 (3.06), 0.744 (2.25), and 0.971 (3.97), whereas the coefficients of predicted LTV are 0.777 (1.01), 1.334 (1.86), and 1.114 (1.59). Thus, treating LTV as endogenous has little impact on its coefficient in the loan denial equation, except in model 2. These results for model 2 support the prediction by Yezer, Phillips, and Trost (1994) that LTV has a coefficient that is biased downward when it is treated as exogenous.
47. As an additional check on the validity of these instruments, we also estimated four new versions of model 2, that is, of the model that treats "meets standards" as endogenous. In each version we dropped one of the four instruments, leaving the other three. In every case the minority-status coefficient is larger than when all four instruments are used. Thus, our results are not driven by the use of too many instruments.
48. We are grateful to Anthony Yezer for pointing this out to us.



49. This relationship is presented in the technical appendix following chapter 5, in the section on errors in the explanatory variables.
50. The coefficients of LTV in the three models are 0.665 (0.49), 1.99 (1.51), and 0.602 (0.52), and the coefficients of the housing-expense-to-income ratio are 0.017 (0.54), 0.041 (1.32), and 0.011 (0.41). As in footnote 46, these results for model 2 support Yezer, Phillips, and Trost (1994).
51. Recall that Rachlis and Yezer (1993) doubt that enough acceptable instruments can be found to estimate a loan denial model with several endogenous variables. We disagree: Our variables meet all three criteria for a good instrument. However, one possible explanation for the lower impact of minority status in model 2 is that this model involves one more endogenous variable (“meets guidelines”) than the other models without any additional instruments; without sufficient instruments, a simultaneous equations model may not be able to sort out the impacts of various endogenous variables. Future loan denial studies should pay close attention to the collection and testing of potential instruments.
52. To be specific, these correlations (*t*-statistics) are -0.843 (7.70), -0.917 (13.59), and -0.911 (13.35).
53. When the ratio of loan to assessed value is replaced with the traditional loan-to-value ratio (based on the minimum of assessed value and house price), the estimated correlation moves even closer to 1, and two of the models do not converge, probably because the ratio of loan to house price is so similar to the traditional loan-to-value ratio. The only model to converge is the one that includes the “meets guidelines” variable. The results for that model are similar to those for the model that uses the ratio of loan amount to appraised value. These problems are similar to the problems experienced by Yezer in trying to estimate a bivariate model of loan denial and LTV—a model he identified through functional form assumptions, not through exogenous instruments. These problems did not arise in Yezer’s simultaneous equations model of loan denial and special program choice because he used far fewer explanatory variables than we do, leaving more left over in the “unobserved” factors. With many explanatory variables, the small differences between the ratio of loan amount to assessed value and the ratio of loan amount to house price do not provide enough new information to sort out the direct impact of special program choice on loan denial from the unobserved factors that influence both special program choice and loan denial.
54. To be specific, the correlations (*t*-statistics) for our three models are -0.462 (1.86), -0.467 (1.80), and -0.508 (2.10).
55. Disparate treatment discrimination has been established for a single lender. See, for example, Siskin and Cupingood (1996).



THE URBAN
INSTITUTE

Chapter 4

Other Evidence of Discrimination: Recent Studies of Redlining and of Discrimination in Loan Approval and Loan Terms

STEPHEN L. ROSS
JOHN YINGER

Over the past decade or so, dozens of scholars have examined redlining, which is defined as mortgage loan denial based solely on the location of the associated property, or discrimination in mortgage loan approval. A few have investigated discrimination in loan terms. But with the exception of the Boston Fed Study (Munnell et al. 1996) discussed in chapter 3, every published study on discrimination in loan denial faces one of two major obstacles, if not both.

The first obstacle is omitted-variable bias due to incomplete information on applicants' credit history. As explained earlier, a loan denial equation that does not include explanatory variables describing an applicant's credit history may produce biased results. Moreover, black and Hispanic applicants are likely to have poorer credit histories than do white applicants, on average, so the omission of credit history variables is likely to result in an overstatement of dis-

crimination and perhaps even in a finding of discrimination when none exists. The second obstacle is a limited sample. A few studies have obtained information on applicants' credit histories, but only for one, or at most a few, lenders. Thus, the results of these studies cannot be generalized to all the lenders in one region, let alone to the nation as a whole.

Because of these two limitations, there does not yet exist any study other than the Boston Fed Study that provides a credible estimate of discrimination in loan denial at the regional level. Similar limitations apply to the literatures on discrimination in loan terms and on redlining. Nevertheless, several recent studies into either discrimination or redlining raise important methodological issues. This chapter reviews these studies. The discussion is organized around five types of research: loan denial studies based on HMDA data, loan denial studies based on data for individual lenders, studies of discrimination in loan terms, studies that explore redlining, and studies that investigate the causes of discrimination.

Loan Denial Studies Based on HMDA Data

Because HMDA data cover the vast majority of mortgage loans in the country and are readily available, many scholars have used them to study lending discrimination, despite the fact that they do not contain any information on applicants' credit histories. The studies reviewed here are all aware of this limitation, and they either focus on results that are relatively unlikely to be subject to omitted-variable bias or develop methods to try to overcome such bias.

Avery, Beeson, and Sniderman (1996) explore intergroup differences in loan denial using the national HMDA data set for 1990 and 1991, which contains more than 4 million applications for conventional loans and almost as many applications for either refinancing or home improvement.¹ This study is distinguished by careful use of all the information available in the HMDA data and careful interpretation of the results. The explanatory variables include all borrower and loan characteristics available in the HMDA data, such as income, along with a dummy variable for each census tract and each lender.² The authors find large and statistically significant differences in loan denial across racial and ethnic groups but recognize that these results could be biased due to the omission of credit history variables.

They also find, however, that intergroup differences in loan denial rates are remarkably persistent across both geographic markets and loan products, results that are difficult to explain if there is no discrimination. For example, the black/white denial rate difference is similar for home-purchase mortgages and for both home-improvement loans and refinancing, despite the fact that applicants for the latter two types of loans have already been deemed credit-worthy and are, for a few years at least, default-free. Nevertheless, as the authors recognize, these comparisons are far from definitive, because they do not directly account for intergroup differences in credit history and other key underwriting variables, such as LTV.

Berkovec and Zorn (1996) match loans purchased by Freddie Mac to loans in the HMDA data in 1992 and 1993. More specifically, they compare the number of



loans actually purchased by Freddie Mac with the number of loans in the HMDA data that are reported as being sold to Freddie Mac. They find that the HMDA data contain only 70 percent of the loans actually purchased in 1992 and only 75 percent of the loans actually purchased in 1993. Moreover, these HMDA coverage rates are systematically higher in low-income neighborhoods. The authors conclude that these results have one of two causes. First, large lenders may be more careful in filling out HMDA reports and those lenders may be more active in low-income neighborhoods. Second, lenders in general may systematically underreport applications in high-income neighborhoods so as to exaggerate their relative effort in low-income neighborhoods and thereby enhance their ratings under the Community Reinvestment Act (CRA).

These findings raise concerns for the study of mortgage discrimination.³ If the HMDA data are subject to reporting biases, then those biases could affect the findings of a loan denial study. Suppose the first of the above explanations is correct. Then HMDA-based loan denial equations may still be viewed as representative of lenders with good reporting practices, but may not yield accurate results for other lenders. If the second explanation is correct, there could be an upward bias in the minority-status coefficient in a loan denial equation; in particular, this explanation implies that lenders underreport high-quality loans, for which, according to results presented in chapter 3, the minority/white loan denial ratio is relatively small. This argument is internally inconsistent, however, because it implies that lenders hide one fact about their record that might get them into trouble (loans in high-income neighborhoods) but fail to hide another fact that holds even more potential danger for them (a high minority/white denial ratio).

Berkovec and Zorn (1996) do not, however, provide proof that certain types of loans are underreported in the HMDA data. As they acknowledge, their study cannot distinguish between underreporting and misclassification of loans. In other words, some of the loans sold to Freddie Mac may not be identified because they are classified improperly in the HMDA data. In 1993, for example, for loans with an applicant income between 0 and 60 percent of the Metropolitan Statistical Area (MSA) median and in a tract with a minority percentage between 11 to 30, the number of Freddie Mac loans in the HMDA data was 27 percent greater than the actual number of loans purchased by Freddie Mac. This result could not arise unless either some loans are misclassified or lenders report some applications multiple times.

Meyers and Chan (1995) use the HMDA data to explore loan denial in New Jersey using a subsample that consists of all applications by blacks plus a 10 percent random sample of applications by whites. The key twist in this study is that it uses the HMDA variable indicating whether a loan was “denied due to poor credit” to “fill in” the credit history information missing in the HMDA data. In particular, the first stage of the study’s procedure is to estimate, for all denied loans, the relationship between “denied due to poor credit” and observable characteristics of the borrower, loan, and location. The estimation results are then used to calculate a predicted credit risk variable for all applications. In the second stage, this predicted credit risk variable is used as a proxy for the missing credit history information in a loan denial equation. This method is intriguing because it appears to control for an applicant’s credit history using the readily available HMDA data.



THE URBAN
INSTITUTE

On the basis of this procedure, Meyers and Chan (1995) conclude that the probability of loan denial is 12 percent higher for blacks than for whites with equal borrower and loan characteristics and equal predicted credit risk in equivalent neighborhoods. This difference is somewhat higher than the 8.2 percentage point gap in the Boston Fed Study.

The key issue raised by this study is whether its two-stage procedure is an adequate “patch” for the problem of missing credit information. At one level, this procedure is similar to the two-equation models of the “meets guidelines” variable discussed in chapter 3; after all, both “bad credit” and “meets guidelines” are after-the-fact evaluations of an applicant’s credit qualifications. However, the “meets guidelines” variable, which was collected for the Boston Fed Study, not HMDA, applies to all applications, not just to denied applications, and the two-equation models with “meets guidelines” are designed to account only for the credit history variables that are unobserved in the Boston Fed data set—not for all dimensions of credit history as in this case. Hence, this study raises two new methodological issues. The first issue is whether one can obtain an unbiased prediction of poor credit based on a sample of denied loans. We suspect that the answer is negative—at least without some form of sample-selection correction. As explained in chapter 3, any sample-selection procedure that is correlated with the dependent variable, in this case poor credit history, results in biased estimates. And there is an obvious correlation between poor credit history and loan denial.

The second issue is that the HMDA data do not contain much exogenous information to help predict poor credit, particularly exogenous information that is not already included in a loan denial procedure. As a result, one cannot be confident that the Meyers-Chan procedure adequately accounts for the simultaneous relationship between loan denial and a credit history evaluation.⁴ The impact of this problem on the estimated minority-status coefficient is difficult to predict, however. If it does not adequately account for the impact of a denial decision on a credit evaluation, the Meyers-Chan procedure will, according to the analysis presented earlier, actually result in an understatement of discrimination.⁵ But because it fails to directly control for any credit history variables, this procedure might also overstate discrimination. The net result is unclear. Until these issues are resolved, we think the results of this procedure should be interpreted with great caution.

Loan Denial Studies Based on Applications to Individual Lenders

The study by Siskin and Cupingood (1996) grew out of the U.S. Justice Department’s antidiscrimination case against Decatur Federal Savings and Loan, a large lender in Atlanta.⁶ In particular, the study examines 1,479 conventional, fixed-rate mortgage applications and 1,431 variable-rate mortgage applications processed by Decatur Federal in 1988 and 1989. The data are based on detailed file reviews, so the authors can estimate a loan denial equation with explanatory variables covering a wide range of loan, borrower, and credit history characteristics. This equation reveals that blacks are far more likely than equivalent whites to be turned down for a loan; this effect is highly significant statistically in both the fixed- and variable-rate samples.



The evidence of discrimination against blacks by Decatur Federal was not confined to these regression results. For example, Decatur Federal's market share in black census tracts was dramatically lower than its market share in white census tracts, even though Decatur Federal was a large-volume lender with an ability to compete throughout the Atlanta metropolitan area. Moreover, the Justice Department obtained evidence that Decatur Federal had explicitly excluded black census tracts from its market area. To be specific, Decatur Federal had a history of closing branches in black neighborhoods, and the criteria it stated as justification for closing a branch were not the same in black and white neighborhoods. Finally, Decatur Federal solicited referrals primarily from real estate agents whose business was concentrated in white neighborhoods. In fact, a former Decatur Federal account executive told investigators that she was specifically instructed by the bank not to solicit loans south of Interstate 20, an area that included many of Atlanta's black neighborhoods. (See Ritter 1996; Siskin and Cupingood 1996.)

Because the Siskin and Cupingood (1996) study is based on extremely detailed information about applicants and loans, its results are compelling. The main issue is that the results cannot be generalized to other lenders.⁷ Thus, this study provides a powerful illustration of the lengths to which some lenders will go to avoid lending to black households, but it does not reveal how common these practices are.

Glennon and Stengel (1995) evaluate three nationally chartered banks (banks A, B, and C) based on data collected by the Office of the Comptroller of the Currency (OCC). Random samples of applications for one- to four-family, conventional, nonpurchased, and home-purchase mortgage loan applications were drawn from the 1993 HMDA data for each of the three banks. The data contain applications from various minority groups, including Native Americans, Asians, blacks, and Hispanics. To preserve confidentiality, Glennon and Stengel sort the applications into minority groups I, II, and III, without indicating which group is which.

The first step Glennon and Stengel (1995) take is to estimate a loan denial model for each lender using a common set of explanatory variables, including variables to indicate the three minority groups. Because there are so few applications from some minority groups at some lenders, the authors are able to estimate coefficients for one of the minority groups for bank A, three of the minority groups for bank B, and two of the minority groups for bank C. Based on a general specification, they find statistically significant minority-status coefficients of 0.590 (with a significance level of 0.0062) and 0.723 (0.0029) for minority group I at banks A and B, respectively. The minority-status coefficient for group I at bank C is 0.309 (0.2970) and is not statistically significant. The only other significant minority-status coefficient is for minority group III; this estimated coefficient is 0.711 (0.001) at bank B.

The author's next step is to allow for the possibility that different lenders have different underwriting criteria by varying the estimating equation across lenders. For example, they replace LTV with the difference between LTV and a lender-specific threshold value based on the loan size, type, or program; this variable is set to zero if LTV is below the threshold or PMI is obtained. They also modify the debt-to-income ratio variable based on lender-specific criteria



and include a variable for whether the applicant has sufficient available liquid assets for closing. Based on these and other changes, the minority-status coefficient for group I at bank A is 0.473 (0.0791) and is not statistically significant. The minority-status coefficient for group I at bank B is 0.658 (0.0108) and for group III is 0.935 (0.0001), both of which are significant. Finally, they find that their lender-specific model has more explanatory power than their basic model and therefore reject the hypothesis that the underwriting guidelines are the same at all three lenders. They conclude that use of a common set of control variables across all lenders is inappropriate and, in the case of bank A and minority group I, leads to a false finding of discrimination.

As a follow-up to this analysis, the OCC conducted detailed file reviews and comparisons at bank B and failed to corroborate the existence of disparate treatment against any minority group. Glennon and Stengel (1995) argue that this apparent contradiction can be explained by two factors: First, several applications had unusual characteristics that were subject to special underwriting guidelines, and second, several data errors were found in the variables measuring credit history, debt ratios, and special program status. They suggest that bank B's underwriting model contained dozens of detailed limits and guidelines, some of which are only rarely applicable, and conclude that a statistical model cannot possibly capture every provision. They conclude that even their bank-specific model is insufficient and may itself have led to a false finding of discrimination.

Glennon and Stengel (1995) raise some important issues concerning the specification of a loan denial equation. One of these is their argument that a loan denial equation must reflect the fact that different lenders use different underwriting standards. We do not find this argument convincing, for three reasons. The first reason is that they focus exclusively on statistical significance. As pointed out in chapter 3, a lender-specific model has the disadvantage that it must work with smaller sample sizes, which make it harder to isolate discrimination when it does exist. Thus, the fact that the statistical significance of the group I coefficient at bank A becomes insignificant could reflect the small sample size, not a problem with the common model. In fact, the magnitude of the coefficient for this group is not much smaller in the lender-specific regression than in the common one, and we doubt if one could reject the hypothesis that the two coefficients are the same.

The second reason is that the authors do not actually test their hypothesis that a lender-specific model is needed. In shifting from a model that is the same for all three lenders to their lender-specific model, they not only introduce differences in explanatory variables across lenders but also change the form of several explanatory variables that are still the same across lenders. For example, they changed the LTV and debt-to-income ratios from continuous variables to discrete categories. They also added a variable to indicate whether the household had sufficient assets available for the closing. These changes do not require a lender-specific model of loan denial. As a result, the increase in explanatory power they observe in going from their common model to their lender-specific model could reflect these changes—not the fact that different variables are used for different lenders. Finally, they do not show that estimating separate regressions for each lender is preferable to a pooled regression, with common coeffi-



clients as well as common variables. The right procedure would be to develop the best possible model for all three lenders combined, then to compare that model with one in which the coefficients are allowed to differ across lenders, and finally to compare that model (and the second one) to a model in which some of the explanatory variables are allowed to differ across lenders as well. Because they have not used such a procedure, they have not shown that a lender-specific model is appropriate.

The third reason is that the authors do not show that differences in underwriting guidelines across lenders are legitimate, in the sense defined earlier—that they are connected with loan profitability. If these differences are not legitimate in this sense, then switching to a lender-specific model will, as shown in chapter 3, sharpen the focus on disparate treatment discrimination at the expense of hiding disparate impact discrimination. With no outside information on the relationship between underwriting guidelines and loan profitability, a pooled model is a more reasonable procedure, because it is likely to capture both types of discrimination.

Another important issue raised by Glennon and Stengel (1995) is that statistical analysis and traditional enforcement techniques (such as file reviews and comparisons) may yield contradictory results. Glennon and Stengel argue that loan files may differ on many idiosyncratic factors that are legitimately considered by underwriters, and that it is impossible to control for these factors because they affect so few applications. They conclude that the results of statistical analysis should not be taken as evidence of discrimination unless it can be confirmed by file reviews.

An alternative and, in our judgment, more compelling view appears in Browne and Tootell (1995), who argue that the existence of idiosyncratic factors makes file reviews less acceptable than statistical analysis. Because each file review takes a long time, any procedure that depends on file reviews is likely to be limited to a relatively small number of applications, which may make generalizing from file reviews difficult. Moreover, a file review cannot determine the weight that an underwriter puts on each idiosyncratic factor. If files differ on many such factors, it may be impossible for a file review to determine whether discrimination has taken place. Longhofer (1996b) makes a similar point: “Not surprisingly, paired file reviews rarely uncover any but the most egregious cases of illegal discrimination.” Indeed, the reliance on file reviews in the years before the Boston Fed Study implied that discrimination was almost never found. Fair housing enforcement officials identified some kind of credit problem in the file for virtually every denied minority application and, lacking any method for weighting the importance of each credit problem, concluded that differences in loan denial had nothing to do with race. Statistical analysis is the only way to prevent this type of unsupported conclusion, because it is the only objective way to determine the weights that lenders apply to each variable in a file.

Rosenblatt (1997) examines an HMDA-based sample of 12,725 applications taken in 39 states by a large New Jersey-based thrift between March 1989 and December 1990. His data contain many characteristics of the applicant and the loan but do not include any credit history information. This study treats loan underwriting as a two-stage process. In the first stage, prospective applicants are sorted into the conventional or the government-insured (FHA/VA) sector



THE URBAN
INSTITUTE

before they submit an application. The sorting process occurs because both the lender and the applicant want to avoid additional costs, such as appraisal costs, that arise once an application is submitted. In the second stage, lenders decide whether to approve applications.

This reasoning leads Rosenblatt (1997) to conclude that LTV should have little influence on the loan denial decision, because loans that are likely to be denied because of a high LTV usually are not submitted. He also argues that statistical discrimination, which occurs when lenders use minority status as a predictor of default or loss, will only occur at the first stage, as risky loans are steered to the FHA/VA sector, and that discrimination based on prejudice can occur in the second stage and will affect both sectors equally. Finally, he argues that education should reduce the likelihood of denial because more-educated applicants will be more confident of approval before they commit their resources to the process.

Rosenblatt (1997) then estimates a model of loan denial, with separate equations for the conventional and FHA/VA sectors. He finds that the coefficient of LTV is significant in both sectors but small in magnitude, thereby supporting his first hypothesis. He also finds that the minority-status coefficient is similar in magnitude and statistically significant in both sectors, as predicted by his second hypothesis. When separate variables are included for each minority group, however, the coefficient for blacks is considerably larger in the conventional sector (0.50) than in the FHA/VA sector (0.28), although it is highly significant in both.

A more complicated picture emerges when Rosenblatt (1997) examines a model of four possible outcomes: approval, withdrawal, denial based on financial considerations, and denial based on credit history.⁸ For conventional loans, the estimated marginal effect of being black instead of white is 2.56 (with a *t*-statistic of 1.04) for financial denials and 3.31 (2.01) for credit history denials. In the case of FHA and VA loans, these marginal effects are 0.75 (0.38) and 5.01 (5.34), respectively. He concludes that the racial difference in loan denial arises largely because of credit history factors that cannot be examined before the submission of an application. Finally, education is significant in all estimations, as predicted by Rosenblatt's third hypothesis.

Although it carries some intuitive appeal, the author's interpretation of these results is not compelling. First, the data set does not contain information on applicants' credit history, and many of the results may be biased because of this omission. In addition, the four-choice model implies that, contrary to the author's prediction, intergroup differences in loan denial are quite different in the conventional and FHA/VA sectors. In the conventional sector, being black appears to play a similar role in both financial and credit history denial. Although the race coefficient is not significant in the former case, this may simply reflect the fact that there are only 17 black denials for financial reasons among the 12,000 conventional applications. In the FHA/VA sector, however, the race coefficient for credit history denial is large and statistically significant, which would be expected if black applicants for FHA/VA mortgages had substantial credit history problems that are not reflected in the data, but the race coefficient for financial denial is small and insignificant, implying no discrimination in the FHA/VA sector. Finally, education may be significant because it



is correlated with omitted credit history variables. As reported in chapter 3, education is not statistically significant using the Boston Fed Study's data, which include credit history.

Ultimately, this study raises issues concerning pre-application interactions between potential borrowers and lenders that are similar to the issues raised by Yezer, Phillips, and Trost (1994) and provides some intriguing new evidence that there may be less discrimination in the FHA/VA sector than in the conventional sector. Without data on credit history and without a simultaneous equations framework, however, this study fails to shed much light on these issues.

Studies of Discrimination in Loan Terms

Few studies have examined discrimination in loan terms. The neglect of this topic is unfortunate for two reasons. First, the literature has not adequately addressed the complexity of mortgage pricing. Most existing studies focus on the interest rate as the price variable. But, in fact, the price of a mortgage also includes a wide variety of points, charges, and fees that can vary widely across customers. Second, many lending-industry analysts are now discussing a move away from a system that rations credit by turning down the least qualified borrowers to one that provides credit to almost everyone but sets prices that rise as a borrower's credit qualifications fall. If this type of shift does take hold, the opportunities for discrimination in loan terms will increase and policymakers will need to have a better understanding of this topic.

Yinger (1996) reviews early studies of discrimination in loan terms. In this section we review two recent papers on the subject. Both focus on overages, which are defined as the difference between the final interest rate on a particular mortgage and the rate set when the lender first committed to making the loan. One study focuses on the difference between final rates and lock-in rates, which are the rates lenders agree not to exceed. The other focuses on the difference between final rates and the minimum rate a loan officer is allowed to charge. Each lender decides on the minimum interest rate it will accept (for each type of loan in each time period) and then makes a commitment to a borrower not to exceed some lock-in rate when the loan is finalized. Two customers who enter a lender's office on the same day and receive a commitment for the same type of loan face the same policy concerning the minimum allowable rate. But they may not be offered the same lock-in rate and, for a variety of reasons these studies explore, the final rate could be higher or lower than the lender's "minimum."

Crawford and Rosenblatt (1997) examine final mortgage interest rate differences between whites, Asians, Hispanics, and blacks using information from one national home mortgage lender for 1988 and 1989. They begin with the observation that, near the beginning of the process that leads to a mortgage loan, a lender commits to a certain interest maximum interest rate, the lock-in rate. Interest rate differences across borrowers arise, either because a lender makes commitments that differ from the market rate on the commitment date or because the actual interest rate charged at the time of the loan is closed differs from the lock-in rate.⁹ If market rates drop in the period after the commitment is



THE URBAN
INSTITUTE

made but before the loan is finalized, the borrower may be able to reduce the contractual rate by threatening to switch to another lender. A lender also may be able to take advantage of an increase in market interest rates if some feature of the loan, such as the down payment, changes—an event that generally releases the lender from the commitment.

Based on this analysis, Crawford and Rosenblatt (1997) use a regression model to explain the difference between the actual interest rate on a loan and the market rate on the date of commitment—which they call the yield premium—as a function of group membership, loan terms, borrower characteristics, and whether market interest rates changed between the date of commitment and the date of finalization. The loan terms in their data set include LTV and loan amount; the borrower characteristics include whether the borrower is a first-time homebuyer, whether the loan is for a refinancing, and the borrower's years of education. In the case of conventional loans, they find no significant differences in the yield premium across groups. In the case of government-insured loans, however, they find that the yield premium is about 3 basis points higher for blacks and Hispanics than for whites—results that are highly significant statistically. They also find some evidence that this difference may arise because blacks and Hispanics have a harder time than whites in negotiating a new, lower rate when market interest rates fall below their lock-in rate.

Crawford and Rosenblatt (1997) downplay their results by saying that for the average government-insured loan, adding 3 basis points “adds about \$1.80 to the monthly payment.” Alternatively, one could say that this additional cost comes to roughly \$200 over the life of the loan. However, the average decline in interest rates below the lock-in rate was about 6 basis points. So one could also say that blacks and Hispanics were only half as successful as whites in negotiating lower rates when market rates declined after the lender’s commitment date. Moreover, the vast majority of the mortgages granted to blacks and Hispanics are government-insured (Gabriel 1996). So this result could signal widespread, if relatively modest, discrimination against black and Hispanic lenders.

Courchane and Nickerson (1997) report on an investigation of discrimination in loan pricing at three banks using similar concepts. This investigation was conducted by the Office of the Comptroller of the Currency (OCC). Courchane and Nickerson begin by pointing out that loan officers work from “loan pricing matrix sheets,” which indicate the minimum interest rate they are allowed to charge for each type of mortgage at any given time. They then define an overage as any excess of the actual interest rate charged above the interest rate on these sheets, measured by rate sheet points, after accounting for legitimate costs in the origination fee—to be precise, the final interest rate on the mortgage (in percentage points) minus the origination fee minus the rate sheet points.

Based on a detailed review of loan files, the OCC found evidence of discrimination in overage practices at one mortgage company. Blacks paid an overage of about 2 points on average, compared with a 1-point average overage for whites. Most of this black/white difference arose because of the actions of one black loan officer. But the OCC concluded that the lender contributed to the problem by offering lenders incentives to generate overages without providing guidelines or standards and by defining racially homogeneous territories.



At the second lender, the OCC conducted a regression analysis of the overage, measured in basis points, as a function of minority status (black or Hispanic), year issued, and several loan characteristics, including LTV, loan amount, and loan purpose.¹⁰ This regression indicated that blacks and Hispanics paid a significantly higher overage, but the minority-white difference was relatively small, only 0.176 basis points. This regression also uncovered higher overages for government-insured loans. Since blacks and Hispanics are far more likely than whites to receive these loans, this result could indicate a form of disparate impact discrimination; after all, legitimate differences in loan costs should be included in rate sheet points, not overages.

The OCC also conducted a two-part regression analysis of overage practices by the third lender. The first part looked into the probability that an overage would be charged. It found that both blacks and Hispanics were significantly more likely than whites to be charged overages.¹¹ The second part addressed the size of the overages using control variables similar to those used for the second lender. This regression indicated that, on average, the amount of overage charged was slightly (but significantly) lower for blacks and Hispanics than for whites.¹² This result is suspect, however, because one of the explanatory variables, namely, the interest rate on the loan, is clearly endogenous, and the estimated coefficients may be subject to endogeneity bias.¹³

According to Courchane and Nickerson (1997), extensive file reviews by the OCC indicated that this evidence of discrimination does not reflect “intentional behavior of the loan officer,” but instead reflects “changes in lock dates or close dates.”¹⁴ The evidence from the file reviews is not presented, however, so this claim is difficult to evaluate. Moreover, the disparities in overages uncovered by the OCC may constitute discrimination even if they are not intentional. Loan officers that base their decisions on unconscious stereotypes are still practicing disparate treatment discrimination. In addition, lenders that allow or encourage practices that result in higher or more frequent overages for minorities without any business justification are practicing disparate impact discrimination. More research is clearly needed to determine whether overage practices, and other practices that affect the cost of a mortgage, simply reflect legitimate business concerns or also include disparate treatment or disparate impact discrimination—as the results of this study, and the previous one, seem to imply.

Recent Studies of Redlining

Discrimination involves the differential treatment of an individual because of the group to which that individual belongs. Redlining is a form of discrimination based on location instead of group membership. There are two different definitions of redlining. The first definition, which focuses on the loan denial process, is that redlining exists when otherwise comparable loans are more likely to be denied when they apply to housing in a minority rather than a white neighborhood. Redlining by this definition is illegal according to the Equal Credit Opportunity Act of 1974. The second definition, which focuses on lending outcomes, is that redlining exists when minority neighborhoods receive a



THE URBAN
INSTITUTE

smaller flow of mortgage funds than comparable white neighborhoods. Redlining by this definition is illegal according to the Community Reinvestment Act (CRA) of 1977.

This section reviews recent studies of redlining by both the process-based and the outcome-based definitions and draws conclusions about the extent of redlining in mortgage markets today.

Recent Studies of Process-Based Redlining

The basic approach used by most studies of redlining using the process-based definition is to determine whether the probability that a loan application is denied is higher in minority neighborhoods than in white neighborhoods, all else equal. Thus, studies of redlining face the same key challenge as studies of discrimination, namely, to find a data set with adequate information on loans and applicants, including applicant credit history. Without this information, inferences about redlining, like inferences about discrimination, are likely to be subject to severe omitted-variable bias.

HMDA data, which do not contain information on applicant credit history, are therefore not adequate for isolating redlining. Indeed, HMDA data may be particularly unsuited for studying redlining, because it appears that lenders who are active in minority and low-income neighborhoods tend to attract applicants with relatively poor credit qualifications, based both on variables that are observed in the HMDA data and on variables that are not observed there. (See the discussion of Bostic and Canner 1997, later in this chapter.)

The first three studies of process-based redlining we review have access to information on applicants' credit histories, which implies that they are based on the only data set with such information, namely, the Boston Fed Study's data set. The final study is based on HMDA data combined with census tract data and data on house sales by tract.

Two articles, Tootell (1996a) and Hunter and Walker (1996), study redlining using the Boston Fed Study's data and a standard loan denial equation. Both authors add to this equation explanatory variables that describe the characteristics of the census tract in which the housing unit is located. These variables include the vacancy rate, the poverty rate, and the percentage of the population belonging to a minority group. None of these variables is statistically significant, and both studies conclude that there is no evidence of redlining in Boston.¹⁵ The results in Tootell are particularly compelling because he controls for the perceived risk to owners of home equity in a neighborhood, using variables that are not in the public-use version of the Boston Fed Study's data.¹⁶

Another study based on the Boston Fed Study's data set, Ross and Tootell (1998), examines a more complex model in which redlining is related to the market for private mortgage insurance (PMI). They examine redlining based both on the minority composition of a neighborhood and on the neighborhood's median income. They find evidence of redlining against low-income census tracts, defined as having a median income at least one standard deviation below that of the MSA, when the applicant did not apply for PMI. The coefficient estimate was 0.56 (with a *t*-statistic of 2.52).¹⁷ They also find some evidence that applications from low-income tracts are favored when the applicant applies



for PMI. In this case the coefficient was -1.96 (2.18). They find no evidence, however, that the probability of denial is higher in tracts with a minority percentage above 30 percent than in other tracts.

Ross and Tootell (1998) suggest that lenders may be meeting their CRA obligations to meet the credit needs of all members of the community from which a lender draws deposits at low risk by encouraging applicants from low-income tracts to apply for PMI. Because low-income tracts are seen as riskier, however, even after accounting for all the variables in the Boston Fed Study's data set, the authors find that lenders are more likely to deny applicants from those tracts when the applicant does not apply for PMI. Ross and Tootell also conclude that their test for redlining based on minority status has little power in the Boston area. In the Boston Fed Study's data set, the income and minority composition of tracts is highly correlated (a correlation coefficient of 0.7). Moreover, when the cut-off used to define a low-income tract is raised, the minority tract variable is statistically significant. In short, with this data set it is difficult to distinguish between income-based and minority status-based redlining.

Ross and Tootell obtain similar results for many different sets of explanatory variables and for several different models. In one alternative model they exclude cases in which the individual applied for PMI. In another they model the application for and receipt of PMI and allow this outcome to influence the lender's loan denial decision.¹⁸ In both cases, their main result, that lenders practice redlining against low-income neighborhoods, is upheld.

Finally, Ling and Wachter (1998) test a redlining hypothesis put forward by Lang and Nakamura (1993). The Lang-Nakamura hypothesis begins with the observation that lenders are uncertain about future developments in any particular neighborhood. Because they are risk averse, a greater degree of uncertainty about a neighborhood is associated, all else equal, with a higher probability of denying applications for loans to buy houses in that neighborhood. Lenders gain information by observing house sales. So, controlling for other things, the probability of loan denial should decline as the number of house sales goes up. This hypothesis should be of interest to policymakers because it implies that the flow of funds to some neighborhoods, particularly low-income neighborhoods where few house sales take place, may be restricted by a lack of information, which is the type of problem that markets cannot solve. If it is true, therefore, this hypothesis may serve as a justification for the CRA or other policies to offset redlining.

Ling and Wachter (1998) test this hypothesis by combining HMDA loan approval data for Dade County, Florida, where Miami is located, with census data on neighborhood (i.e., tract) characteristics and data on house sales and sales prices from the Florida Department of Revenue. The resulting data set has an extensive set of neighborhood variables (such as median income, median education, median house value, and percentage of housing units that are owner-occupied), along with a variable to test the Lang-Nakamura hypothesis (namely, the share of owner-occupied housing units that sold over a three-year period), and a related variable (namely, the percentage change in the price of housing). The HMDA data contain several applicant characteristics but do not, of course, indicate applicant credit history. Thus, the Ling and Wachter regressions, while thought-provoking, may be subject to severe omitted-variable bias.¹⁹



Ling and Wachter (1998) find that, as predicted by Lang and Nakamura, the probability of loan acceptance increases with the share of houses that sell. It also increases with the rate of increase in housing prices. Ling and Wachter also point out, however, that an increase in sales could signal an upward shift in the demand for housing in a neighborhood, so that their results are also consistent with the view that lenders see less risk in neighborhoods where housing demand is on the rise. This alternative hypothesis does not imply a need for governmental anti-redlining policies, so further research on this topic is clearly needed.

Recent Studies of Outcome-Based Redlining

A review of research on the outcome-based definition of redlining can be found in Schill and Wachter (1993). This section examines one recent study, Phillips-Patrick and Rossi (1996), not covered in that review. The authors focus on outcomes by tract, but, unlike others studying outcome-based redlining, they also attempt to isolate the role of lenders.

Phillips-Patrick and Rossi (1996) initially estimate a single-equation model of mortgage redlining using data on total loan originations by census tract. They begin with a simple equation in which loan originations are a function of racial composition, using a variable that indicates whether more than three-quarters of a tract's residents are black. They find that originations are significantly lower in largely black tracts. They then estimate a more complete single-equation model, in which the dependent variable is loan originations divided by salable housing units, and in which the explanatory variables include many tract characteristics that might influence underwriting risk. In this revised specification, the racial composition of the tract does not have a significant impact on originations. They conclude that great care must be taken in specifying a redlining equation.

Next, these authors estimate a simultaneous equations model of the demand and supply of mortgages. Their model is in the spirit of the method developed by Maddalla and Trost (1982) for applications data (see chapter 3). They measure demand for loans by the ratio of applications to salable units and supply by the ratio of loan originations to salable units. This approach allows them to separate lenders' role in approving loans from their role in influencing the number of applications (through advertising, branch location, and so on). In this model, they find that largely black tracts have a higher demand for originations after controlling for other tract characteristics. This higher demand masks racial differences in the supply of mortgages to largely black tracts, so that, with their simultaneous equations framework, they find that largely black tracts receive a significantly smaller supply of mortgages than other tracts, again controlling for other tract characteristics. These findings lead them to restate their cautions about interpreting any analysis of loan originations.

We agree with one conclusion of Phillips-Patrick and Rossi (1996), namely, that one should be cautious in interpreting studies of outcome-based redlining. This conclusion is not new, however, and is in fact emphasized in Schill and Wachter (1993)—and in many of the studies they review. What is new in the Phillips-Patrick and Rossi paper is the use of a simultaneous equations framework to study outcome-based redlining. The Phillips-Patrick and Rossi



finding that race-based redlining in the supply of loans appears when a simultaneous equations framework is used is an important contribution that is, in our view, downplayed by the authors. It suggests that some previous researchers may have failed to uncover race-based redlining because they did not account for such simultaneity. We hope that future work follows up on this suggestion.

Conclusions

Redlining has proven a difficult topic to study, both because it has two different definitions and because the underlying behavioral models are difficult to specify. As a result, no strong consensus has emerged in the literature. In the case of process-based redlining, the lack of data with adequate controls for borrowers' credit history makes inferences difficult. The few studies that have examined this topic using the Boston Fed Study's data also yield mixed results. Two studies find no evidence of redlining, but a third, which accounts for the relationship between redlining and private mortgage insurance, finds redlining against low-income neighborhoods, which are almost all largely black neighborhoods, at least in Boston. A fourth study, which may yield biased results because it does not have any credit history information, finds evidence that is consistent with the view that lack of house sales in some neighborhoods leads to lender uncertainty about future developments there and hence to redlining. In the case of outcome-based redlining, which has received the most attention, most but not all of the literature finds some sign of redlining, but there is no consensus on the appropriate methodology. More research on redlining is clearly needed.

Studies of the Causes of Discrimination

The principal objective of the Boston Fed Study and of all the other literature reviewed so far is to determine whether whites and minorities encounter discrimination in the mortgage market. If evidence for discrimination is found, it is also important to ask why this discrimination occurs. In other words, it is important to investigate its causes. This type of investigation provides not only a richer understanding of the underlying behavior but also valuable information to government officials who enforce fair lending laws. If discrimination against minority applicants is found to be particularly likely under some circumstances, for example, these officials can make better use of their scarce resources by concentrating their efforts on cases in which those circumstances arise.

The Boston Fed Study (Munnell et al. 1996) does not provide a formal analysis of the causes of mortgage discrimination. However, it does discuss one possible cause, called statistical discrimination, which is said to exist if lenders use minority status as a signal concerning unobserved credit characteristics. It seems possible, given their relatively disadvantaged socioeconomic outcomes, that black and Hispanic applicants are more likely than white applicants to be rated unfavorably on these unobserved characteristics—or at least that lenders perceive this to be the case. In either event, lenders will believe that minority applicants



THE URBAN
INSTITUTE

are more likely to default than are white applicants with the same observed credit characteristics and they have an economic incentive to discriminate against minority applicants. This behavior is illegal—a lender must base his or her decision on the observed credit characteristics of an applicant—but some lenders may respond to the economic incentive instead of to the requirements of the law. Munnell et al. (and, in a follow-up piece, Tootell 1996b) go to some lengths to dismiss this hypothesis on the grounds that there is not compelling evidence that blacks actually have higher default rates, controlling for observable characteristics. However, they do not test this hypothesis, or any other, directly.

In this section, we review a few loan denial studies that test hypotheses about the causes of discrimination. By way of preview, these studies barely scratch the surface of this important topic.

Another hypothesis about the cause of lending discrimination that, like statistical discrimination, has been widely discussed in recent years is the so-called “cultural affinity” hypothesis. This hypothesis, formally developed by Calomiris, Kahn, and Longdorfer (1994), is that the limited affinity of white loan officers to the culture of certain minority groups implies that these officers make less effort to determine the creditworthiness of minority than of white applicants; the resulting information disparity implies that minority clients are more likely to be rejected.

Hunter and Walker (1996) is offered as a test of this hypothesis. These authors interpret cultural affinity as a form of statistical discrimination. Loan officers must decide how much effort to make in finding additional information about an applicant and use group membership as a signal about the difficulty of obtaining this information. In the standard version of statistical discrimination, group membership is used as a signal for elements of creditworthiness that are unobservable or at least expensive to observe and that, in the loan officer’s estimation, differ across groups. For example, if blacks are believed to be less creditworthy by these unobserved indicators, on average, the loan officer has an economic incentive to use race as an indicator of creditworthiness. In the case of the cultural affinity hypothesis, however, there is no presumption that the two groups differ in their underlying creditworthiness. Instead, the issue is that the cost of collecting extra information is higher for minority applicants, presumably because the white loan officer is uncomfortable collecting this information.

This analysis leads to the following question: Why does the extra information collected for white applicants result in a more favorable outcome for whites, on average? We see three possible answers to this question.²⁰ The first is that what is really going on here is just standard statistical discrimination; the concept of “cultural affinity” adds nothing.

The second, due to Longhoffer (1996a), begins with the assumption that the credit characteristics lenders cannot observe have a larger variance for minority than for white applications, precisely because lack of cultural affinity with minority applicants makes it more difficult for lenders to find extra information or corroborating information. If lenders are risk-averse, this assumption implies that lenders will be more likely to turn down a minority than a white application even if the two applicants have the same observed creditworthiness. In other words,



lenders seek compensation for the added uncertainty in minority applications by holding those applications to a higher standard.

The third possible answer is that white loan officers are not trying to learn more about white applicants' creditworthiness, but are instead trying to build the best possible case for each white applicant, often by collecting additional *supporting* information. One could say that white loan officers do not provide this service to black or Hispanic applicants because they do not feel a cultural affinity with them. But this view is indistinguishable from the more traditional explanation that white loan officers are simply prejudiced against minority applicants. In this context, the term cultural affinity is nothing more than euphemism for a lack of prejudice. Thus, no one should think that a loan officer's behavior is either legal or in keeping with widely held values just because it is driven by a lack of cultural affinity for a particular group. In short, this version of the cultural affinity hypothesis boils down to the following: Loan officers are prejudiced against people in some groups and this prejudice induces them to withhold some services from applicants in those groups—that is, to discriminate against them.

The problem, of course, is that these three possible answers are quite different and no scholar has yet found a compelling way to distinguish among them.

Hunter and Walker (1996) begin their empirical investigation of cultural affinity with the Boston Fed Study's data set, then delete observations associated with a special program and observations identified as having a data error according to the procedure developed by Carr and Megbolugbe (1993). The resulting data set contains 1,516 white applications and 475 black or Hispanic applications. They use this data set to test "whether loan officers' decisions on white applicants depend less on formal information, such as credit history, financial obligations, and the like, than they do for minorities" (p. 60). This informal information is what the loan officer collects, because of his or her cultural affinity with white clients, to help make the best possible case for white clients. By relying on this extra, informal, largely positive information, the hypothesis goes, loan officers inevitably place less weight on formal information.

The main form of their test is to determine if two key variables—the obligation ratio and their credit history indicator—have a larger impact on loan denial for blacks and Hispanics than for whites. They find support for the first effect but not the second. The signs of both effects are in the expected direction. But the difference between the white and minority impacts is significant in the first case (*t*-statistic equals 2.87) but not in the second (1.14). As noted earlier, this obligation-ratio result can also be found in Munnell et al. (1996).

Hunter and Walker (1996) interpret this result as support for the cultural affinity hypothesis. We are not convinced. Even on the surface, the evidence is not very strong. If loan officers are willing to overrule the poor formal qualifications of whites on the basis of informal information, why does this effect show up in only a single coefficient? More important, the larger impact of the obligation ratio for minorities than for whites could arise because minorities and whites tend to go to different lenders with different underwriting guidelines (see chapter 3), not because individual lenders use different guidelines for minorities and whites.²¹



THE URBAN
INSTITUTE

Black, Collins, and Cyree (1997) test the cultural affinity hypothesis by focusing on the relationship between loan denial and the racial makeup of the people who own a lending institution. Their basic argument is that the cultural affinity problem, along with its impact on loan denial, arises in white-owned lending institutions but is unlikely to appear in a lending institution that is owned by blacks (or by people in some other minority group). A similar argument has been applied to discrimination by housing agents (Yinger 1986, 1995) and by car salesmen (Ayres and Siegelman 1995).

To explore this view, Black, Collins, and Cyree (1997) identify black-owned lending institutions in major metropolitan areas and matched or comparable white-owned lending institutions in the same locations. Then they obtain all applications to these lenders in the 1992–93 HMDA data. The resulting subsample contains 2,393 white applications and 925 black applications from 81 lenders, 32 of which were owned by blacks.

The first step in their analysis is to estimate the relationship between loan denial and the applicant and loan characteristics in the HMDA data for both black-owned and white-owned lending institutions. This estimation indicates large racial differences in loan denial for both black-owned and white-owned lending institutions. But the size and significance of the race coefficient is substantially larger for institutions owned by blacks. The second step is to estimate an enhanced equation that includes neighborhood characteristics and detailed financial characteristics of the individual lending institutions as explanatory variables. With this enhanced equation, the race coefficient is not significant for white lenders but it is still highly significant for black lenders.

These results contradict the cultural affinity hypothesis. It is not clear, however, what these results mean. One possibility is that they reflect lenders' trade-off between the gains from statistical discrimination and potential costs of being caught by federal regulators. If black-owned banks assume that their lending practices with regard to applicant minority status will receive little scrutiny, those banks may be inclined to practice statistical discrimination. Another possibility is that the results simply reflect limitations in the study's data. In particular, the relationship between minority status and loan denial for black-owned banks could arise because the applicants at black-owned banks are different from the applicants at white-owned banks on characteristics that are not recorded in HMDA data. See, for example, the discussion of Bostic and Canner (1997), immediately below.

The finding by Black, Collins, and Cyree (1997) that intergroup differences in loan denial are not significant for white-owned banks after one controls for bank financial characteristics is also provocative. Does the omission of bank characteristics in most studies result in an overstatement of discrimination? The answer to this question is clearly negative. First, this study investigates a small sample of white-owned banks selected to be comparable to black-owned banks. Moreover, bank financial characteristics appear to have little impact on loan denial. In fact, only one of the six financial characteristics in the study actually influenced the denial decisions of white-owned banks. Last, and most important, Munnell et al. (1996) already include a set of lender dummy variables, which, as noted in chapter 3, completely controls for financial or any other lender characteristics that do not vary across the loans made by a given lender.



Bostic and Canner (1997) also focus on the cultural affinity hypothesis, but they argue that this hypothesis applies to applicants as well as to lenders. This may help explain the Black, Collins, and Cyree (1997) results. As Bostic and Canner (p. 1) put it: "Minority applicants may feel more comfortable applying for mortgages at minority-owned banks, which could result in a relatively large volume of marginally qualified applicants at minority-owned banks. In such a case, minority-owned banks could have higher rejection rates than white-owned banks, even if only minority lenders exhibit cultural affinity or if lenders of both races applied the same underwriting standards."

Bostic and Canner (1997) begin by identifying minority-owned banks and comparable white-owned banks, following procedures similar to those in Black, Collins, and Cyree (1997). They then use 1994 and 1995 HMDA data to determine the share of applications at each bank that come from whites, blacks, and Asians. Their sample consists of 29 minority-owned and 52 white-owned banks in 1994, along with 32 minority-owned and 62 white-owned banks in 1995.

The basic form of their analysis is to explain the share of a lender's applications from a minority group (or from low-income neighborhoods) as a function of the financial characteristics of the bank and whether it was minority-owned. They find that the share of applications from blacks at black-owned banks was five times the share at white-owned banks, on average, in 1994 and eight times the share in 1995.²² Black-owned banks also had a much higher share of their applications from low-income neighborhoods.

Bostic and Canner (1997) then estimate a loan denial equation and reproduce the Black, Collins, and Cyree result. To account for the possibility that the applicants to black-owned lenders are less creditworthy, they then reestimate this equation using a subsample that consists of applications at black-owned banks along with comparable applications at white-owned banks. In this matched subsample, there is no evidence that the black/white denial gap is any higher at black-owned banks than at white-owned banks.

These results generally support the Bostic and Canner (1997) analysis. They indicate, in other words, that black-owned banks attract black applicants and applicants with relatively low credit qualifications, on average, and that failure to account for this effect leads to the false impression that black-owned banks are the only ones to discriminate against blacks. The results do not show, however, that blacks or applicants with low creditworthiness are attracted to black-owned banks because of cultural affinity. These applicants could, as Bostic and Canner say in their fourth footnote, end up at black-owned lenders because the marketing practices of those lenders differ from those of white-owned lenders. They point out, for example, that black-owned banks might be more likely than white-owned banks to work with black real estate brokers, who might, in turn, be more likely than other brokers to have black customers. Moreover, the results do not support the standard prediction from the cultural affinity hypothesis, namely, that the black/white denial ratio will be higher at white-owned than at black-owned banks.

Thus, the Bostic-Canner results support the notion that applicants sort themselves by race in selecting lenders, but they do not explain why this sorting occurs, and they do not find any evidence that cultural affinity affects lenders' loan denial decisions. However, their results should be interpreted



with care because they are based on HMDA data, with all its limitations. The authors cannot directly control for an applicant's credit history, for example, so they develop an ad hoc matching procedure with unknown properties. Nevertheless, we think the hypotheses in this paper are valuable and hope they will some day be pursued with a more complete data set.

Conclusions

The most remarkable feature of the literature reviewed here is that it does not contain a replication of the Boston Fed Study in another metropolitan area. Despite the facts that the Boston Fed Study was first released in 1992; that the Boston Fed Study is widely, but in our judgment incorrectly, viewed as seriously flawed; and that many large institutions, including lenders, financial regulators, and secondary mortgage market institutions, have a stake in knowing how much lending discrimination exists, no scholar has yet published a comparable, let alone an improved, mortgage lending study. As a result, none of the studies reviewed here provides an up-to-date estimate of the extent of lending discrimination.

Several recent studies of loan denial make valuable observations or methodological points, and several of them begin the important work of trying to determine the causes of lending discrimination. Nevertheless, the lack of a data set comparable to the one collected by the Boston Fed Study's authors casts a shadow over all this research and makes all the results difficult to interpret.

Notes

1. Another interesting use of HMDA data can be found in Avery, Beeson, and Calem (1996), who document a new HMDA-based procedure to identify lenders and loan applications that should be reviewed as part of a fair-lending enforcement program.
2. Avery, Beeson, and Sniderman (1996) use a linear specification for their equations, instead of the nonlinear logit or probit models used by most other scholars. This is a reasonable procedure, both because nonlinear models may be difficult to estimate with so many observations and because other scholars have found that linear models usually produce results that are similar to those of nonlinear models in problems of this type.
3. They also raise concerns for the use of HMDA data in enforcement of fair lending laws and the Community Reinvestment Act, because HMDA data may underestimate the differences in loan originations between low- and high-income neighborhoods.
4. In formal terms, Meyers and Chan use the total number of people living in a tract and the proportion of female-headed households in a tract as instruments in their simultaneous equations procedure. We do not believe these variables bring in very much exogenous information to identify the model. In fact, only one of the four coefficients (for two variables in two equations, one for blacks and one for whites) is statistically significant in explaining the likelihood of bad credit, which implies that these variables do not meet the second standard for an instrument given in chapter 3. The paper provides no information about whether they meet the first standard.
5. It also might underestimate discrimination because it runs separate bad credit regressions for blacks and whites. Hence, its predicted bad credit variable is not race-neutral and could build in discrimination in the evaluation of whether a person has bad credit.



6. See Ritter (1996) for a detailed discussion of this case.
7. The study also did not deal with the simultaneity issues raised in chapter 3, but, as in the case of the Boston Fed Study, it seems unlikely that they would alter the study's conclusions.
8. This model is estimated with multinomial logit analysis. This procedure is not ideal for the purpose because it requires the assumption that the trade-off between any two choices is not affected by the introduction of a third choice—an unlikely assumption in this case.
9. Crawford and Rosenblatt (1997) use a complex, and appropriate, definition of interest rate, or yield, that accounts for points and the prepayment option.
10. Unlike the Crawford and Rosenblatt (1997) regressions, the OCC regression for this lender does not control a change in the interest rate between the date of commitment and the date of closing. It also does not control for the market interest rate at the lock-in date, but instead control only for the year in which the mortgage was originated.
11. In addition, women were more likely than men to be charged overages.
12. The OCC regressions for the third lender control for an interest rate, but Courchane and Nickerson (1997) do not indicate whether the regressions use the actual interest rate for the individual loan or the average market rate, and they do not indicate whether the rate is measured at the lock-in date or the closing date. The regression coefficients may contain endogeneity bias if this variable was defined either as the individual loan rate or as the market rate at closing.
13. No clear definition of the interest rate used as an explanatory variable is given, but it appears from context to be the final interest rate on the loan, which includes, of course, the overage, if any. This problem can be avoided by following Crawford and Rosenblatt (1997), who use the market interest rate at the lock-in date, as well as the change in the market rate between lock-in and closing.
14. Courchane and Nickerson (1997) also argue that their results do not indicate disparate treatment discrimination, because loan officers are equally rewarded for all overages, regardless of the minority status of the borrower. We disagree. If a loan officer takes advantage of the lender's market power to charge a higher overage for groups with fewer options in the market, including minority groups, which is one of the mechanisms Courchane and Nickerson themselves identify, that is an example of applying different rules to different customers, which is the definition of disparate treatment discrimination.
15. These results are similar to earlier findings by Schafer and Ladd (1981).
16. To be more specific, Tootell (1996a) includes the rent-to-value ratio for rental property in each tract. This variable is highly significant. He argues that this variable has a relatively high value in neighborhoods where rental housing is an attractive option relative to owner-occupied housing, that is, when the risk to home equity is relatively high.
17. This coefficient and the following one are based on logit analysis. Ross and Tootell (1998) estimate several different models; these results are from the most complete one.
18. This simultaneous equations model, which was estimated with bivariate probit analysis, is discussed in chapter 3. As noted there, this model yields little evidence that the receipt of PMI is endogenous to the denial decision.
19. It should be noted, however, that Ling and Wachter's estimate that the impact of minority status (for blacks) on loan denial is 5.4 percentage points is similar to the estimate provided by Munnell et al. (1996), who can control for applicant credit history.
20. Hunter and Walker (1996, p. 67) present a third explanation, which does not, in our view, make sense. They point out that loan officers see many more white than minority applications and conclude that loan officers may therefore have more accurate predictions about the impact of less formal factors on outcomes for whites. As they say, "It is quite possible for factors other than credit history to matter for whites, possibly mitigating the impact of a weak credit history, while at the same time credit history may continue to play a dominant role in the accept/reject decision for minorities." This begs the question because it does not



explain why loan officers feel the need to treat black and white loan applicants any differently. Why can't their experience with white applicants be applied to black applicants?

21. Indeed, this effect could arise because of something even simpler, such as a nonlinearity in the relationship between the obligation ratio and loan denial.
22. Bostic and Canner also find that the share of Asian applications is higher at Asian-owned banks, all else equal, as is the share of white applications at white-owned banks.



Chapter 5

The Default Approach to Studying Mortgage Discrimination: A Rebuttal

STEPHEN L. ROSS
JOHN YINGER

Although analysts disagree about the best way to model many features of mortgage markets, one facet of the mortgage market elicits a clear consensus: Lenders care whether loans are repaid and, all else equal, are unlikely to approve a loan application with a relatively high probability of default. This simple premise leads to a number of straightforward conclusions. The sample of approved loans is of higher quality than the sample of rejected loans. The quality of a lender's portfolio depends upon the toughness of its underwriting criteria. Holding the composition of mortgage applications fixed, a lender that raises underwriting standards will raise the quality of its portfolio of approved loans. Similarly, if a lender raises underwriting standards for a specific group of applications, such as those from minorities or those for units in minority neighborhoods, the average quality of approved loans in this subsample will increase.

These conclusions have led to assertions that default rates or statistical analysis of defaults can be used to test for the existence of mortgage lending discrimination. Soon after the release of the Boston Fed study, a magazine column by Becker (1993) popularized the argument that if lenders discriminate against minorities by holding minorities to a higher underwriting standard than

whites, the average quality of approved mortgages will be higher for minority mortgages than white mortgages. Becker then claimed that the Boston Fed Study was “flawed” because it looked at loan denial rates instead of default rates. Perhaps because Becker’s earlier work (1971) is a seminal contribution to the economics of discrimination and because he is a Nobel Prize winner, this column has received a lot of attention. We are not convinced. The existence of an alternative method for studying discrimination hardly proves that the use of a loan denial equation is flawed. In fact, as we will see, the loan denial approach holds up quite well against the default approach.

In this chapter, we first review the key methodological issues involved in studying discrimination by looking at loan defaults. We then review the most recent literature on the subject.

Methodological Issues in the Default Approach to Studying Discrimination

Although it seems straightforward, the default approach raises many complex methodological issues. In this section, we discuss problems associated with unobserved underwriting variables, unobserved borrower characteristics, sample-selection bias and the power of the default approach to detect discrimination, and the use of data on government-insured loans to detect discrimination in the conventional-loan sector.

Unobserved Underwriting Variables

The most extreme form of the default approach to studying discrimination, which is illustrated in Becker’s (1993) column, claims that, in the presence of discrimination, the average default rate for minorities will be lower than the average default rate for whites. This version of the default approach makes no sense at all because the pool of minority mortgage applications is of lower quality than the pool of white mortgage applications. (See Peterson 1981; Galster 1993; Ferguson and Peters 1995; Tootell 1996b.) For example, minority applicants tend to average larger debt burdens, higher loan-to-value ratios, and poorer credit histories than white applications (based on the Boston Fed Study’s data). Even if lenders do not discriminate, therefore, the pool of approved minority applications will be of lower quality than the pool of approved white loans. This conclusion reflects, of course, a type of omitted-variable bias; intergroup comparisons of average default rates give biased results because they do not control for credit qualifications.

This point is illustrated by figure 1 at the end of this chapter. The horizontal axis represents the quality of loan applications, and the vertical axis represents the fraction (or density) of applications that have this quality. The distribution of minority and white loan applications are shown separately, with the minority distribution drawn so that minority loan applications are lower quality than white applications, on average. The quality cutoff below which loan applications are denied is labeled C. This cutoff is the same for whites and minorities, which



implies no mortgage lending discrimination based on minority status. The average quality of approved white applications, which is the mean of the white quality distribution above C, is substantially higher than the average quality of approved minority applications because so many high-quality white applicants applied for and received mortgages. These high-quality white applicants pull up the average quality of white mortgages and drive down the average white default rate. Discrimination against minorities lowers the average minority default rate, but might not lower it enough to offset the higher average quality of white loan recipients. Thus, a lower average default rate for whites cannot be interpreted as evidence that discrimination does not exist.

Van Order and Zorn (1995) agree with this point, but also point out that lower default rates for minorities can provide evidence that discrimination exists. Given the lower average quality of minority applications, minority mortgages could only experience lower default rates if minorities were held to a substantially higher underwriting standard. To put it another way, a lower default rate for minorities than for whites is sufficient but not necessary to show discrimination. In figure 2, the cutoff for minority applications is $D + C$, where D represents the increased underwriting standard for minorities. The average quality of minority mortgages can only exceed the average quality of white mortgages if D is very large, that is, if substantial discrimination exists in mortgage lending.

Van Order and Zorn (1995) examine default rates across the country. In the southeastern United States, they find that default rates are either unaffected by or fall with the share of black households in a census tract. They also observe that HMDA loan application rejection rates fall with black concentration in all regions, including the Southeast. They conclude that racial differences in default rates in the Southeast do provide evidence of discrimination, because the pool of minority applications is of lower quality than the pool of white applications and yet minority default rates are equal to or lower than white default rates. As noted earlier, Van Order and Zorn recognize that equal or higher default rates for minority mortgages do not provide evidence that lenders treat white and minority applicants equally and therefore do not contradict other findings of discrimination.

Berkovec, Canner, Gabriel, and Hannan (1994) attempt to avoid the problems inherent in examining average default rates, by using statistical analysis to examine whether the marginal minority applicant is treated the same as the marginal white applicant. They attempt to identify a marginal buyer by using regression techniques to control for many variables that lenders may consider during the underwriting process. Using a sample of FHA mortgages from 1987 through 1989, they find that minorities are more likely to default after controlling for the underwriting variables that are available in the data set of FHA mortgages. On the basis of this finding, they reject the hypothesis that minorities encounter discrimination.

Regression analysis controls for variables that are observed by the analyst and compares minority and white treatment based on unobserved factors, which may include borrower characteristics observed and used by the lender for underwriting but not observed by the analyst, henceforth called unobserved underwriting variables, along with lender differences in underwriting criteria and behavior. This exercise essentially replicates the comparison of average



THE URBAN
INSTITUTE

default rates, therefore, except that the influence of observed underwriting variables has been removed before the calculation of intergroup differences in default. Figure 3 is constructed in the same way as figures 1 and 2, except that the horizontal axis now represents the quality of loan applications based on unobserved underwriting variables and, as assumed by Berkovec et al. (1994), white and minority applications have the same average quality when only these unobserved underwriting variables are considered. In this case, even if minority loan applications are worse on observed underwriting variables, having higher underwriting standards for minorities results in a pool of approved minority mortgages that has higher-than-average quality based on unobserved underwriting variables than does the pool of white mortgages.

The assumption of equal unobserved loan quality is critical. If minority applications have lower quality on the basis of unobserved underwriting variables, figure 1 still applies, so long as the horizontal axis is relabeled “loan quality based on unobserved underwriting variables.” In this case, therefore, average quality can be lower for minority than for white loans even if lenders practice discrimination. In effect, the Berkovec et al. conclusion that there is no discrimination is conditional on a set of very strong assumptions, including no difference in unobserved underwriting variables for blacks and whites.

Note that the impact of unobserved underwriting variables on default is actually the flip side of the omitted-variable bias problem in an analysis of mortgage application denials. If minority applications are lower quality based on unobserved underwriting variables, a loan denial analysis will indicate a higher likelihood of denial for blacks even if lenders do not discriminate. With omitted underwriting variables, therefore, an analysis of loan denials is likely to overstate discrimination and an analysis of loan defaults is likely to underestimate discrimination. A key implication of this point is that any analysis based on loan performance or default must control for the same set of underwriting variables used in an unbiased loan denial equation. Otherwise, the loan performance or default analysis may suffer from an omitted-variable bias even if the loan denial equation does not. Unlike the Boston Fed Study, for example, Berkovec et al. (1994, 1998) do not control for applicants’ credit history. In the Boston Fed Study’s data, minority applicants have worse credit history than white applicants, and the credit history variables are highly significant in predicting denials. These results explicitly contradict Berkovec et al.’s key assumption and undermine their conclusion concerning discrimination.

Unobserved Borrower Characteristics

So far, we have only considered underwriting variables known to the lender but unobserved by the analyst. As discussed by Galster (1996), Yinger (1996), and Ross (1996b), however, many factors that determine actual mortgage quality and the likelihood of default are not observed by the lender; we call these variables unobserved borrower characteristics. For several reasons, it seems likely that on the basis of unobserved borrower characteristics alone, minority applicants are less qualified than whites. For example, most defaults occur because the borrower experiences an unexpected event, such as a layoff, and discrimination in the labor market may imply that minorities are more likely to experience



such negative events. Moreover, the legacy of past discrimination, which takes the form of lower skills and lower wealth for minorities, may imply that minority borrowers are less able to overcome negative events when they do occur. Thus, minority borrowers may be more likely to default than are white borrowers with identical applications.

The proponents of the default approach are correct in saying that holding minorities to a higher standard will decrease the likelihood that minority borrowers will default, but they are not correct in saying that this effect can be seen by comparing minority and white borrowers. The observed likelihood of default for minority borrowers needs to be compared with the likelihood of minority default that would have occurred without discrimination, not with the likelihood of default by white borrowers. This may be the most devastating critique of the default approach. Without knowing the minority-white difference in default likelihood after controlling for all lenders' underwriting variables, researchers cannot state a null hypothesis to use in a default-based test for discrimination.

Berkovec et al. (1996) counter that the default approach is only intended to test for discrimination caused by lender prejudice. According to this theory of discrimination, an application of the approach in Becker (1971), lenders hold minority applicants to higher standards because the lender must be compensated for his irrational prejudice against minorities by earning more profits when he lends to them. An alternative theory of discrimination is that lenders "rationally"¹ (but illegally) use minority status as a signal for unobserved borrower characteristics and therefore hold minority applicants to a higher standard because minority applicants are more likely to default, on average. Berkovec et al. (1996) insist that their approach does not attempt to capture this so-called statistical discrimination.

However, this counterargument does not address the key issue. Simply stating that one is testing for discrimination caused by one mechanism rather than another does not make the existing group-based differences in unobserved borrower characteristics—and in default—go away. Some lenders may practice statistical discrimination, others may discriminate due to their personal prejudice, and others may not discriminate at all. Any interpretation of default results must recognize all of these possibilities. One might think that the default approach tests whether the level of discrimination in the lending industry exceeds the level expected based on "rational" lenders, using minority status as a signal for the likelihood of default. In fact, however, the default approach cannot even provide a test of this limited type. Ross (1996a) shows that, even if lenders practice statistical discrimination to the extent dictated by their economic interest, minority borrowers will still be more likely than observationally equivalent white borrowers to default on loans.

Sample-Selection Bias and the Power of the Default Approach to Detect Discrimination

Several scholars have pointed out that the default approach cannot shed light on discrimination unless some underwriting variables are unobserved (see Galster 1996; Ross 1996b). To understand this argument, consider the following highly simplified version. First, suppose that loan applications are evaluated on



THE URBAN
INSTITUTE

the basis of “the number of late credit card payments.” Second, suppose for now that minority and white applicants have the same distribution of outcomes for this “late payments” variable. Third, suppose that lenders set a higher standard for minority than for white applicants, and in particular deny minority applications if more than two late payments are observed but deny white applications only if more than four late payments are observed. Fourth, suppose that this “late payments” variable is a good predictor of default; for example, people with two late payments (who will all be approved) are more likely to default than people with no late payments. Now if a researcher observes the “late payments” variable and includes it in an equation to explain defaults, he will find that the minority-status coefficient equals zero, even though lenders discriminate. In this case, the “late payments” variable fully describes the systematic component of default behavior and there is nothing left for a minority-status variable to explain. The higher hurdle that minority borrowers face in the loan approval decision pushes up their average creditworthiness, according to the “late payments” variable—pushes it high enough, indeed, so that minority borrowers have better creditworthiness on average, and hence experience fewer defaults, than do white borrowers.² But the impact of this higher hurdle is fully captured by the estimated coefficient of the “late payments” variable combined with the higher value of the “late payments” variable for minority than for white applicants.

Now consider what happens if the “late payments” variable is not observed by the analyst. In this case, the method of selecting the sample for a default analysis, namely, restricting it to people who actually receive a loan, introduces sample-selection bias. The people who receive a loan have fewer late payments, on average, than people who are denied a loan, and the number of late payments affects whether they will default. Thus, the sample-selection procedure is correlated with the outcome of interest, in this case default, which implies that the estimated coefficients will be biased. Ironically, however, this “bias” is what makes the default method work. When the default equation is estimated, now without the “late payments” variable, the minority-status coefficient will be negative because the selected sample of minority applicants has cleared a higher hurdle than the selected sample of white applicants. In other words, the sample-selection bias implies that the higher hurdle for minority applicants shows up in the minority-status coefficient.

In more formal terms, if a statistical analysis controls for all underwriting variables considered by lenders, the unobserved portion of the denial decision must be entirely due to random differences across banks and across loan officers. The underwriting variables in the analysis completely describe the quality of the mortgage from the lender’s perspective. And unobserved borrower characteristics that affect default are unrelated to lender differences in underwriting behavior. Therefore, the underwriting process has no influence on the likelihood of default after controlling for the observed underwriting variables. Under these circumstances, the default approach cannot detect discrimination regardless of the cause.

This result seems to imply that all one needs to do to ensure the success of the default approach is leave out a few variables. This is not the case, however, because the only variables that can legitimately be omitted are ones that



are not correlated with minority status. As pointed out earlier, the default approach yields biased results, and is likely to underestimate discrimination, if variables correlated with minority status are omitted from the estimated equation. In the example discussed earlier, suppose that the “late payments” variable is correlated with minority status and, in particular, that minority applicants have more late payments, on average, than do white applicants. Then, even with a higher hurdle for minority than for white applicants in the loan approval decision, minority borrowers—that is, minority applicants who made it through the hurdle—could still have more late payments, on average, than white borrowers. In this case, a default equation that omitted the “late payments” variable would yield a positive minority-status coefficient—a reflection of the omitted “late payments” variable and its correlation with minority status, not of reverse discrimination.

Conventional versus Government-Insured Loans

Finally, Berkovec et al. (1994) have been criticized for their use of government-insured, specifically FHA, loans (see Yinger 1996; Galster 1996). This issue is important because Berkovec et al. claim to shed light on discrimination in conventional loans even though their data consist of defaults on FHA loans. They build their case by observing that FHA loans cost the borrower more than conventional loans and arguing that mortgage applications will be sorted into three categories: the lowest-quality applications, which do not receive credit; the highest-quality applications, which receive credit in the conventional market; and the applications in between, which receive credit in the FHA market. If so, discrimination in the conventional sector would force higher-quality minority applicants into the FHA sector, which would result in a lower likelihood of default for approved minority mortgages in the FHA sector. The straightforward application of the default approach would focus on the possibility that only the highest-quality minorities are approved for conventional mortgages. Due to data constraints, Berkovec et al. examine the reverse, arguing that the pool of minorities who cannot receive a conventional mortgage, but qualify for an FHA loan, has a higher average quality than the pool of white applicants in the FHA sector.

This outcome is illustrated in figure 4, in which minority and white applications have the same unobserved quality distribution, separate cutoffs are specified for conventional and FHA mortgages, and lenders practice discrimination in the conventional sector. These conditions lead to the Berkovec et al. (1994) conclusion—higher loan quality for minorities in the FHA sector. However, the bias that arises when minorities and whites do not have the same unobserved quality distribution, which was illustrated in figure 1, still arises when the analysis focuses on FHA rather than conventional mortgages. Figure 5 illustrates this bias by allowing intergroup differences in application quality based on unobserved underwriting variables and removing discrimination in the conventional sector. Even though this example does not involve any intergroup differences in the underwriting guidelines used by lenders, it results in intergroup differences in default.



THE URBAN
INSTITUTE

Moreover, it is unlikely that borrowers are perfectly stratified across these three outcomes. Individuals have imperfect information about the underwriting behavior of lenders when they choose whether to apply for any loan, whether FHA or conventional. In addition, FHA loans have long been the primary source of credit for minority households (see Gabriel 1996), and low-quality minority applicants may be pulled into the FHA sector based on this history, just as high-quality minority applicants are pushed into it by discrimination. If so, the sorting assumptions on which the Berkovec et al. (1994) analysis depends do not hold. And a finding of higher defaults by minority FHA borrowers, holding constant observable borrower characteristics, need not imply lack of discrimination in the conventional sector—or, for that matter, in the FHA sector itself.

New Twists on the Default Approach

In the last few years, researchers have explored several new twists on the default approach. In this section, we examine simulation studies and a study that attempts to look for the impact of market concentration on discrimination.

Simulations

Ross (1998) observes that the loan denial approach and the default approach to studying discrimination are each valid in quite different situations. The default approach cannot detect discrimination unless some underwriting variables are unobserved by the researcher and omitted from the analysis. Otherwise, lender underwriting behavior is unrelated to unobserved borrower characteristics that influence default. In contrast, a loan denial analysis is valid only if it controls for all legitimate underwriting variables correlated with the minority status of the applicant. Most characteristics considered by lenders during the underwriting process are correlated with minority status, so a loan denial analysis requires controls for most or all underwriting variables.

Ross develops a test for the extent of omitted underwriting variables by estimating a joint model of loan denial and mortgage default. This joint model is not based on the ideal data set, which would involve applications, loan denial, and defaults for the same set of individuals. It is based, instead, on two separate data sets, the Boston Fed Study's public-use data on loan denial and the Berkovec et al. (1994) data on default. The Boston Fed Study's data set is used to estimate the loan denial model. The results of this estimation are combined with FHA default data, Berkovec et al.'s sample, in a second-stage analysis of loan default. This second-stage analysis corrects for the influence of the loan denial process on the likelihood of default and provides consistent estimates of the default process under the assumption that this process is the same for the two samples. The resulting estimates provide a relatively complete picture of the likelihood of default for mortgage applications—more complete than the picture in previous research.³

This analysis provides an estimate of the correlation between the unobserved determinants of default and the unobserved determinants of loan denial.



A correlation between these unobserved factors can exist only if some underwriting variables are unobserved and excluded from the analysis. The estimated correlation is 0.26 (with a *t*-statistic of 0.04) and 0.18 (0.05) using two different specifications. However, the FHA foreclosure sample does not include credit history variables. Ross estimates that the correlation between unobserved factors in the two equations with controls for credit history would be 0.12 (0.06) and 0.04 (0.06) for his two specifications.

These results do not support the view that the Boston Fed Study's data set omits important underwriting variables. Moreover, they imply that the default approach may not be valid, or at best may only provide a weak test for discrimination, once credit history variables are included in the default model. As discussed earlier, the default approach cannot be compared with a loan denial analysis unless it controls for the same variables, but the default analysis may not be able to detect any discrimination once those variables are included.

The key weakness of Ross (1998) is the use of the FHA foreclosure sample. Ross notes that his analysis depends on the assumption that the underwriting process for FHA applications is the same as, or at least similar to, the underwriting process for conventional mortgage applications. As stated above, the choice to apply for an FHA mortgage is a borrower decision. This choice should not influence the underwriting model, unless the decision reveals information about the borrower that is not observed during the underwriting process and explains default tendencies. An application to FHA is probably influenced by the likelihood of receiving a conventional mortgage, which is based on underwriting variables, and by the cost of a delay in receiving credit when a conventional application is denied, which is probably not correlated with default risk. Thus, if the model controls for most or all key underwriting variables, such as the ones in the Boston Fed Study's data, the approval model for FHA and conventional loans may be similar.

There is some evidence that FHA and conventional underwriting processes are, in fact, quite similar. Rosenblatt (1997) finds that the influence of loan-to-value ratio on loan denial does not vary between the FHA and conventional sectors of the market.⁴ In addition, Rosenblatt estimates a model in which an application may be denied for either financial considerations or credit history. The results for loan denials based on financial considerations do not vary between the two sectors. The results for loan denials based on credit history do vary between the two sectors. But Rosenblatt's sample does not contain information on credit history, so his estimates for this model may be severely biased. Avery and Beeson (1998) interact loan-to-income variables with whether the application is for an FHA mortgage and include these variables in a HMDA-based loan denial model. These interactions are insignificant. Although the interactions between income variables alone and FHA are significant, income does not provide a good basis for comparing approval models in the HMDA data, because it is correlated with key unobserved underwriting variables, such as credit history, loan-to-value ratio, and nonhousing-debt-to-income ratio. The use of FHA data is clearly not optimal for this analysis. Given the lack of good, publicly available default data, however, the Avery-Beeson analysis provides the best information currently available concerning the extent



THE URBAN
INSTITUTE

of omitted underwriting variables in default-based research on mortgage lending discrimination.

Ross (1997) simulates samples of loan applications in which information is present on both the loan denial decision and later default outcomes for approved mortgages. The simulated sample is based on the Boston Fed Study's public-use data, and the level of discrimination in the loan denial model is based on the minority-status coefficient from a loan denial model estimated with these data—a model that is similar to the basic version of the model 3 that we estimated in chapter 3 (that is, to the third entry, row 1, table 1, chapter 3). The parameters for the default model are taken from estimates based on FHA foreclosure data, namely, the Berkovec et al. (1994) sample. The advantage of a simulation approach is that it allows us to specify the “true” approval and default models, so that the estimates from a default model using only approved mortgages can be compared to “true” estimates.

As discussed earlier, discrimination in underwriting can only influence the minority-status coefficient in a default model if there exist unobserved underwriting variables that influence both loan denial and the likelihood of default. The presence of these variables is summarized by the correlation coefficient between unobservable factors in the two equations. Ross (1997) shows that the role of this correlation is nonlinear. Using a model in which intergroup differences in default do not actually exist, Ross estimates minority-status coefficients of 0.00, -0.07, -0.187, and -0.45, corresponding to correlations between unobservable factors in the two equations of 0.00, -0.25, -0.50, and -0.75. The negative values of these estimated coefficients supports Berkovec et al.'s (1994) assertion that minorities should default less frequently when lenders practice discrimination. These estimates also indicate, however, that the default approach may provide a weak test for discrimination when the correlation between unobservable factors is 0.25 or below. As noted earlier, the available evidence indicates that this correlation probably is below 0.25 for the Berkovec et al. sample.

Ross also examines the effect of omitting credit history variables from a default analysis. Again assuming that no intergroup difference in default exists after controlling for all underwriting variables, he estimates minority-status coefficients of 0.12, 0.04, -0.05, and -0.275 for correlations of 0.00, -0.25, -0.50, and -0.75. The positive values of the first two coefficients, along with the small values of the other two, indicate that the omission of credit history variables biases the results of the default approach. Recall that this analysis assumes that discrimination exists at roughly the level identified in the Boston Fed Study. Nevertheless, the default approach leads to the conclusion that discrimination does not exist—as indicated by a small negative or positive effect of minority status on default for minorities—if the correlation between unobservable factors is below -0.5.

These results cast serious doubt on the usefulness of the default approach as a test for discrimination. On the one hand, if credit history variables are omitted from the default equation, the default approach will probably fail to find discrimination when it exists, unless the correlation between unobservable factors in the loan denial and default equations exceeds 0.50 in absolute value. And at such higher correlations it probably understates the role of discrimina-



tion substantially. On the other hand, we have already shown that the default approach will probably not be very powerful if credit history variables are included in the default equation. In short, these simulations imply that the default approach yields misleading answers regardless of whether credit history variables are included or excluded.

A Default Test Based on Market Concentration

Berkovec et al. (1998) develop an alternative version of the default approach that may be insulated from the omitted-variable bias that arises because many determinants of default are unobserved by the lender or researcher and are correlated with minority status. Rather than test directly for differences in default based on minority status, these authors identify a proxy variable that may be related to the level of discrimination but not to minority status, and then test whether group-based differences in default are affected by this proxy variable. This is a clever attempt to eliminate omitted-variable bias. Instead of focusing on the coefficient of a minority-status variable, this version of the default approach focuses on the coefficient of an interaction between minority status and the proxy variable—a coefficient that, under certain assumptions, will not be subject to omitted-variable bias. To put it another way, this version of the default approach may allow the authors to take a step not possible with the standard version, namely, to specify a clear null hypothesis. This hypothesis is that without discrimination, minority status will not influence the relationship between the proxy variable and default, so the coefficient of the interaction term will be zero. A rejection of this hypothesis, they argue, is evidence of discrimination.

To be more specific, Berkovec et al. (1998) claim that lenders have more freedom to discriminate in markets where the lending industry is highly concentrated. Moreover, they argue that intergroup differences in default that persist after controlling for all underwriting variables are unlikely to be correlated with the degree to which lenders are concentrated. It follows that intergroup differences in default tendencies will not affect the estimated coefficient of the interaction between a borrower's minority status and market concentration. Using the FHA data in their earlier default study, they estimate a default model that includes this interaction variable. They find that minority borrowers are less likely to default in highly concentrated markets, which is consistent with higher levels of discrimination in more concentrated markets. But the estimated coefficient is small in magnitude relative to the minority-status coefficient itself and is not statistically significant. They conclude that there is no evidence of discrimination.

This is not the only way to interpret their results, however. In particular, one also might ask whether the impact of concentration on lenders' ability to discriminate is large relative to its impact on lenders' ability to ration credit more aggressively without regard to race. The latter impact is indicated by the coefficient of the concentration variable (not interacted with anything) in the Berkovec et al. (1998) default equations. As it turns out, in the three years examined by Berkovec et al., minority status increases the impact of market concentration on default by 100, 33, and 66 percent, respectively—which is a large impact even if it is not statistically significant.⁵



THE URBAN
INSTITUTE

Berkovec et al. (1998) provide a novel twist on the default approach, which helps insulate their hypothesis test from the problem of omitted-variable bias. However, this twist is built on two assumptions that are essentially contradictory, which makes it incapable of yielding credible results.

The first assumption, which is explicit, is that discrimination motivated by prejudice is stronger in locations where the lending industry is more concentrated—that is, when there are fewer lenders to compete against one another. We find this assumption to be plausible. It says, in effect, that lenders are in a better position to allow prejudice to affect their underwriting procedures when they face less competition and that, as a result, a lack of competition will be particularly hard on minority customers. However, this assumption has never been tested and is not tested by Berkovec et al. (1998). Given how little is known about the causes of discrimination, this assumption, or any other for that matter, should be treated with considerable skepticism. At best, therefore, this new test for lending discrimination must be seen as conditional on the validity of a so far untested assumption about discrimination.

The second assumption, which is implicit in the Berkovec et al. (1998) argument, is that underwriting standards are not affected by the degree of competition among lenders. Regardless of the degree of concentration, in other words, lenders must always set the same credit standard and place the same weight on each underwriting variable. This assumption is inconsistent with the first one. Even though a lack of competition causes lenders to be more aggressive in rationing credit to minorities (assumption 1), it does not cause them to be more aggressive in rationing credit on any other grounds (assumption 2).

Ironically, Berkovec et al. (1998) actually test their second assumption. As noted earlier, they include in their default equation the concentration variable itself, not interacted with anything. They find that the level of market concentration lowers the likelihood of default. They interpret this result as supporting the view that lenders ration credit more aggressively in more concentrated markets. But they fail to see that it undermines their interpretation of their results, or, to put it another way, that it alters their null hypothesis. In our terms, this result explicitly contradicts the second assumption.

So why is the second assumption necessary for the Berkovec et al. (1998) test to be valid? For the sake of illustration, suppose that this assumption is violated in the following way: a higher market concentration raises the quality cut-off for conventional mortgages but not for FHA mortgages. This effect can be represented by an increase in C_{Conv} in figure 5, which could arise because lenders set a higher standard using the same weights on all underwriting variables or using a higher weight on one or more individual underwriting variables. Thus, with an increase in market concentration, a large number of high-quality white mortgages, based on unobservable underwriting variables, are added to the FHA sector, compared with only a small quantity of high-quality minority mortgages. It follows that (a) an increase in market concentration raises intergroup differences in default, even if this increase has no impact on discrimination, and (b) this new version of the default approach is biased, in this case away from finding discrimination.

To put it another way, the Berkovec et al. (1998) approach cannot distinguish between two different reasons why the minority/white default ratio might change when market concentration increases. The first reason is that market



concentration facilitates discrimination. The second reason, which has nothing to do with discrimination, is that market concentration induces lenders to ration credit more aggressively, which changes the composition of successful applicants, in particular the average creditworthiness of minority borrowers relative to whites.

The second assumption is required even in a default model analysis of conventional mortgages instead of FHA mortgages. For the example in figure 5, an increase in the cutoff for conventional mortgages drops a large number of low-quality white mortgages but only a small number of low-quality minority mortgages. As before, the relationship between minority status and default increases with market concentration even if market concentration has no impact on discrimination. As a result, the coefficient estimated by Berkovec et al. (1998) will be biased toward finding no discrimination. Of course, figure 5 is only an example, and although a downward bias in this new version of the default approach is likely, it is not guaranteed. If higher concentration leads lenders to be less aggressive in rationing credit based on other factors, for example, the bias could work in the opposite direction. Nevertheless, this analysis shows that the Berkovec et al. revised default approach does not provide a good test for discrimination, because it yields a biased estimate of discrimination and there is no way to eliminate (or even measure) this bias.

The second assumption applies not only to the level of creditworthiness that applicants must meet but also to the weights placed on individual underwriting variables. In particular, the Berkovec et al. (1998) approach is valid only if market power affects the way lenders ration credit to minorities but does not affect credit rationing based on the loan-to-value ratio, the housing-expense-to-income ratio, the debt-to-income ratio, and possibly a host of other underwriting variables. Any change in the weights placed on these variables will alter the minority composition of successful applicants and, like a change in the level of creditworthiness, will alter the minority/white default ratio even with no discrimination.

Fundamentally, this problem is a form of omitted-variable problem. Just as omitted underwriting variables bias the minority-status coefficient in the standard default model, the omission of variables to capture the impact of market concentration on the use of underwriting variables biases the coefficient of the interaction between market concentration and minority status. In the case of observed underwriting variables, one could solve this problem by interacting all these variables with market concentration and including these interactions in the default equation. If all of these variables prove to be insignificant, then one could conclude that, as required by the second assumption, market concentration does not influence underwriting weights, at least for observed variables. If some of them are significant, however, then the second assumption is violated and it is necessary to include the new interaction terms as controls. Even though they could have done so, Berkovec et al. (1998) do not take this step.⁶ As a result, we cannot determine the magnitude of the bias in the interaction between minority status and market concentration that arises when these other interactions are omitted.

Including interactions between concentration and observed underwriting variables minimizes the problem. But it does not eliminate it, because concentration still might alter the weights lenders place on unobserved underwriting



THE URBAN
INSTITUTE

variables. To put it another way, once these other interaction variables are included, the required second assumption needs to refer only to omitted underwriting variables; that is, the role of unobserved underwriting variables for whites must not be affected by market concentration. As in the standard default approach, adding control variables does not alter the nature of the problem but it does narrow its scope. Thus, without a complete set of underwriting variables, this version of the default approach may yield biased results—even if it includes interactions between market concentration and all observed underwriting variables. This problem was first pointed out by Ross (1997) in connection with Berkovec et al. (1994). In that article, Ross argues that the omission of credit history variables may bias the Berkovec et al. approach away from finding discrimination.⁷

Thus, the conditions that must hold for this new version of the default approach to be valid are very stringent indeed. Not only must lenders respond to a lack of competition by increasing discrimination, but they also must refrain from responding to a lack of competition by altering their underwriting criteria in any other way. Berkovec et al. (1998) undermine their own argument by demonstrating one important way in which the second condition is violated, and further investigation of the problem could reveal additional violations. We conclude that Berkovec et al.'s (1994, 1998) revised default approach based on market concentration cannot be interpreted as a test of the hypothesis that prejudice-based discrimination exists in mortgage markets.

Conclusions

The so-called default approach to studying mortgage lending discrimination has received a great deal of attention in recent years. This approach builds on the simple, intuitively powerful idea that discrimination involves holding minority applicants to higher standards, so that loans given to minorities must perform better (that is, be less likely to default) than loans given to whites. As it turns out, however, using this simple idea runs into severe methodological obstacles.

The most fundamental problem is that it is impossible with existing data to steer the default approach between two obstacles. The first obstacle is the bias that arises when key underwriting variables are omitted from the analysis. In particular, the default approach yields results that are biased against finding discrimination unless it includes all variables that (a) influence default, (b) are observed by the lender at the time of loan approval, and (c) are correlated with minority status. No existing data set provides all this information.

The second obstacle is that the default approach has no power to detect discrimination when key underwriting variables are included in the analysis. To be specific, the default approach cannot detect discrimination even if it exists unless it omits some variables that (a) influence default, (b) are observed by the lender at the time of loan approval, and (c) are not correlated with minority status. A researcher obviously can determine which observed variables are correlated with minority status. However, even if all variables that influence default and are observed by the lender are available to the researcher, there is no



guarantee that some of these variables will be uncorrelated with minority status, which is a necessary condition for avoiding the second obstacle. In the more likely case that the researcher does not have access to all this information, there is no way to rule out the (likely) possibility that some of the omitted variables are correlated with minority status, which is a sufficient condition for omitted-variable bias to arise. Thus, it would take extraordinary circumstances, namely, complete data along with some variables that are not correlated with minority status, to overcome these two obstacles—circumstances that are not even close to being met for existing studies.

Even if a researcher could avoid these two fundamental sources of bias, however, he still might not be able to obtain unbiased estimates using the default approach. Perhaps the largest remaining problem is that characteristics of the borrower that are unobserved by the lender and by the researcher also can be a source of bias, generally downward bias, in an estimate of discrimination. Some scholars have argued that these characteristics give lenders a reason to practice statistical discrimination—that is, to use minority status as a signal of poor, unobservable credit characteristics—and that the default approach only looks for discrimination based on prejudice after statistical discrimination has already taken place. This is an enormous concession because it says that, at best, the default approach can only isolate discrimination that comes from one source and cannot provide a measure of overall discrimination. As it turns out, however, this concession does not go far enough: The default approach is still biased even in the presence of statistical discrimination.

Berkovec et al. (1998) claim to overcome these problems by asking a new question; instead of asking whether minority applicants are less likely to default, it asks whether the minority/white default difference is greater in locations where the lending industry is more concentrated, a situation that presumably gives lenders more leeway to discriminate. This new specification does not save the default approach because it depends on two virtually contradictory assumptions. In particular, it assumes that if lenders discriminate at all, they discriminate more severely when market concentration is higher but that lenders do not alter any other aspect of their underwriting procedures in the presence of more concentration. If the second assumption does not hold, then this new version of the default approach cannot distinguish changes in the minority/white default rate that are due to increased discrimination from changes that are due to shifts in the minority composition of successful applicants. Berkovec et al. provide evidence that contradicts the second assumption; specifically, they show that lenders ration credit more aggressively in more concentrated markets. We conclude that this revised version of the default approach cannot reveal whether prejudice-based discrimination exists in mortgage markets.

The most extreme proponents of the default approach claim that loan denial studies are flawed because they neglect to examine minority/white differences in loan default. Nothing could be farther from the truth. With appropriate data and careful attention to specification, the loan denial approach can yield a credible test of the hypothesis that discrimination exists in mortgage markets—regardless of whether alternative methodologies exist. Moreover, the default approach itself is so fraught with methodological difficulties that no scholar has yet found a legitimate way to use it as a test for discrimination.



THE URBAN
INSTITUTE

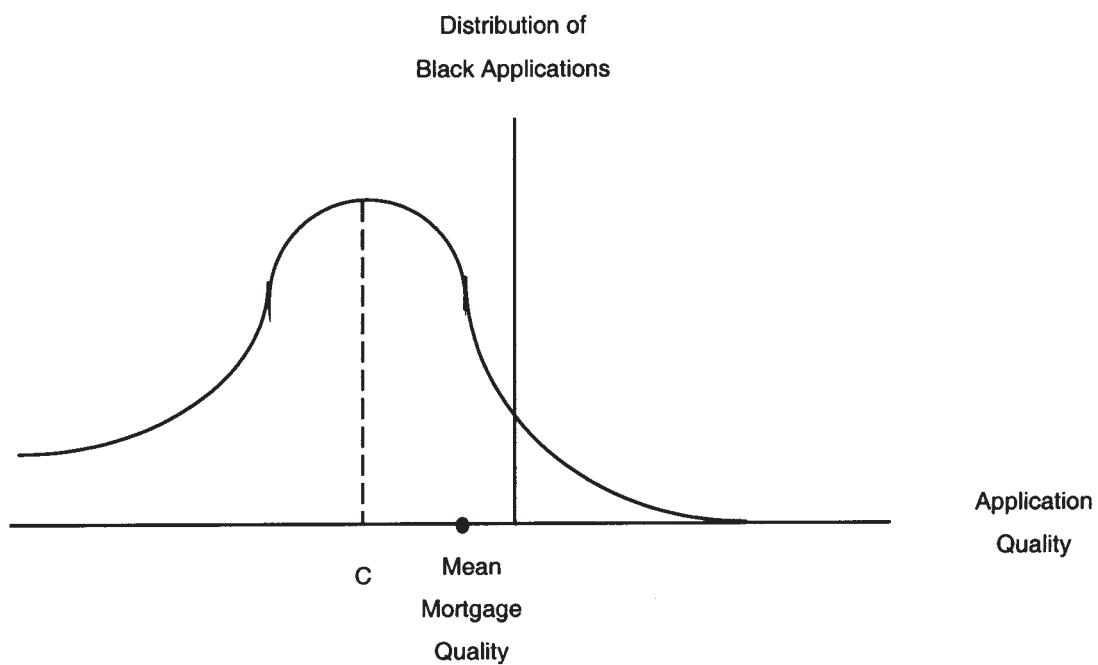
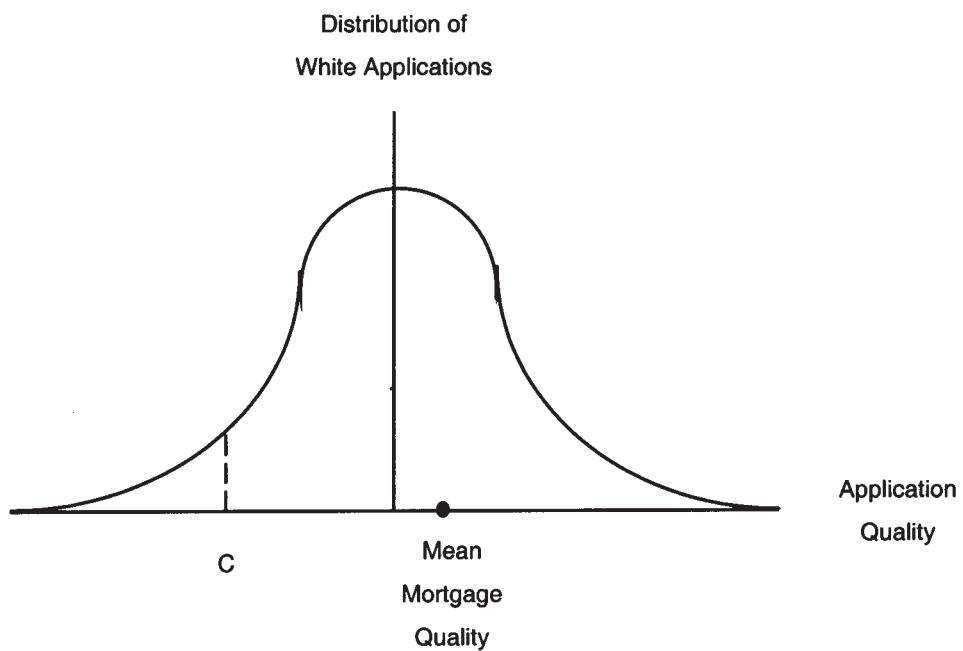
Notes

1. The term “rationally” is in quotation marks to indicate that lenders may be following an economic incentive to discriminate but they are still breaking the law—which may, of course, be irrational (not to mention wrong) given the associated penalties.
2. This outcome depends on the assumption that the “late payments” variable has the same distribution for minority and white applicants. Without this assumption, it could be true that there are so many white applicants with no late payments that the average white applicant has fewer late payments than the average minority applicant, even with the higher hurdle for minority applicants. We return to this issue below.
3. Ross uses a two-stage approach because he employs separate samples for estimating the denial and default models. Boyes, Hoffman, and Low (1989) estimate a default model for consumer credit that does not require a two-stage approach. Their approach cannot be applied to default on mortgages, because no existing sample contains both denied applications and default information on approved mortgages.
4. Further discussion of the Rosenblatt study can be found in chapter 4.
5. Comparing the coefficient of the interaction term with the coefficient of the minority-status variable also is suspect because, as shown earlier, the coefficient of the minority-status variable is biased upward by omitted variables. This bias is, of course, exactly the problem that Berkovec et al. are trying to avoid.
6. The step of interacting concentration with other underwriting variables was suggested to Berkovec et al. by Stephen Ross, but they have not pursued it as yet.
7. A detailed discussion of the potential bias can be found in the technical appendix following this chapter.



THE URBAN
INSTITUTE

Figure 1.



THE URBAN
INSTITUTE

Figure 2.

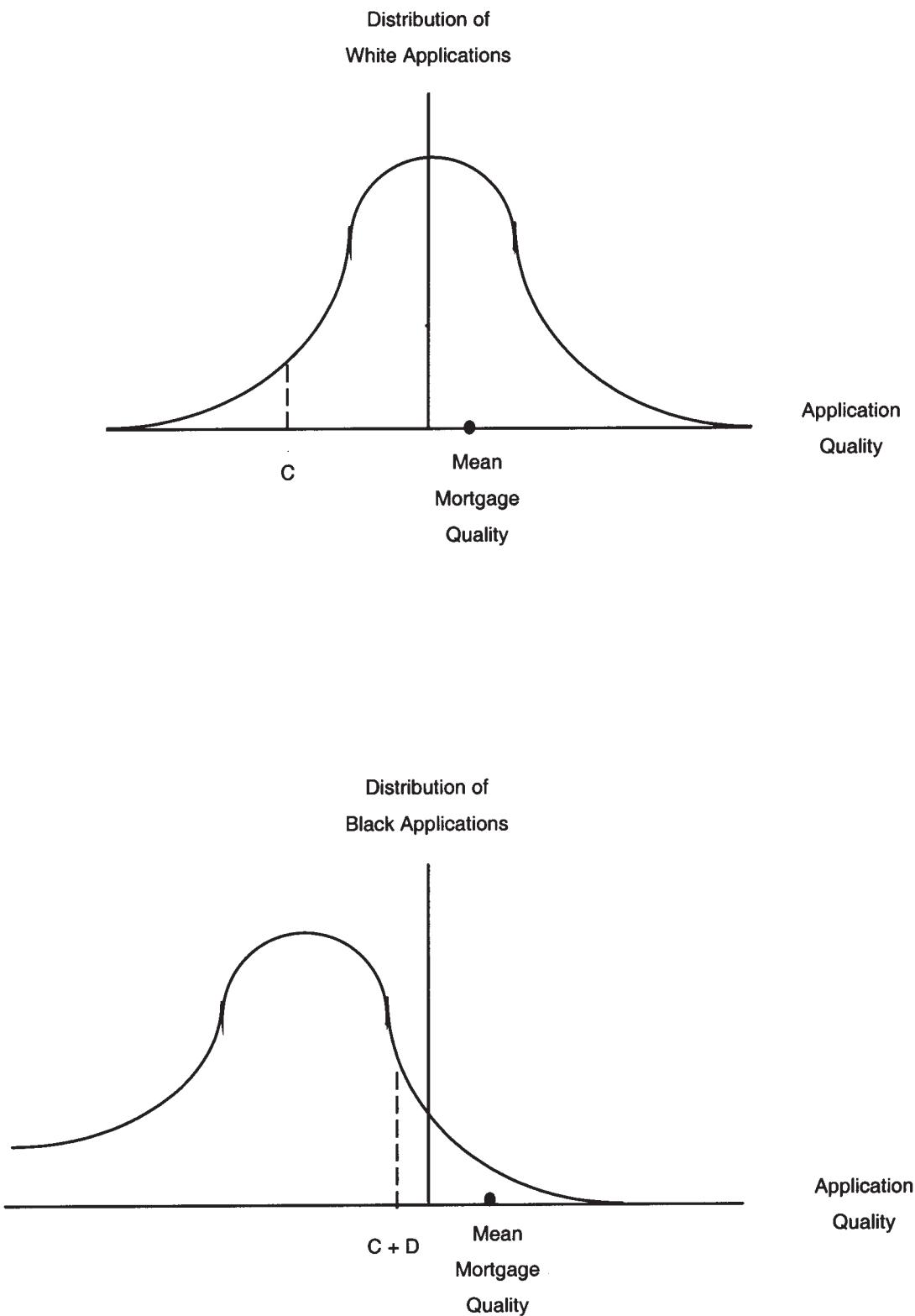
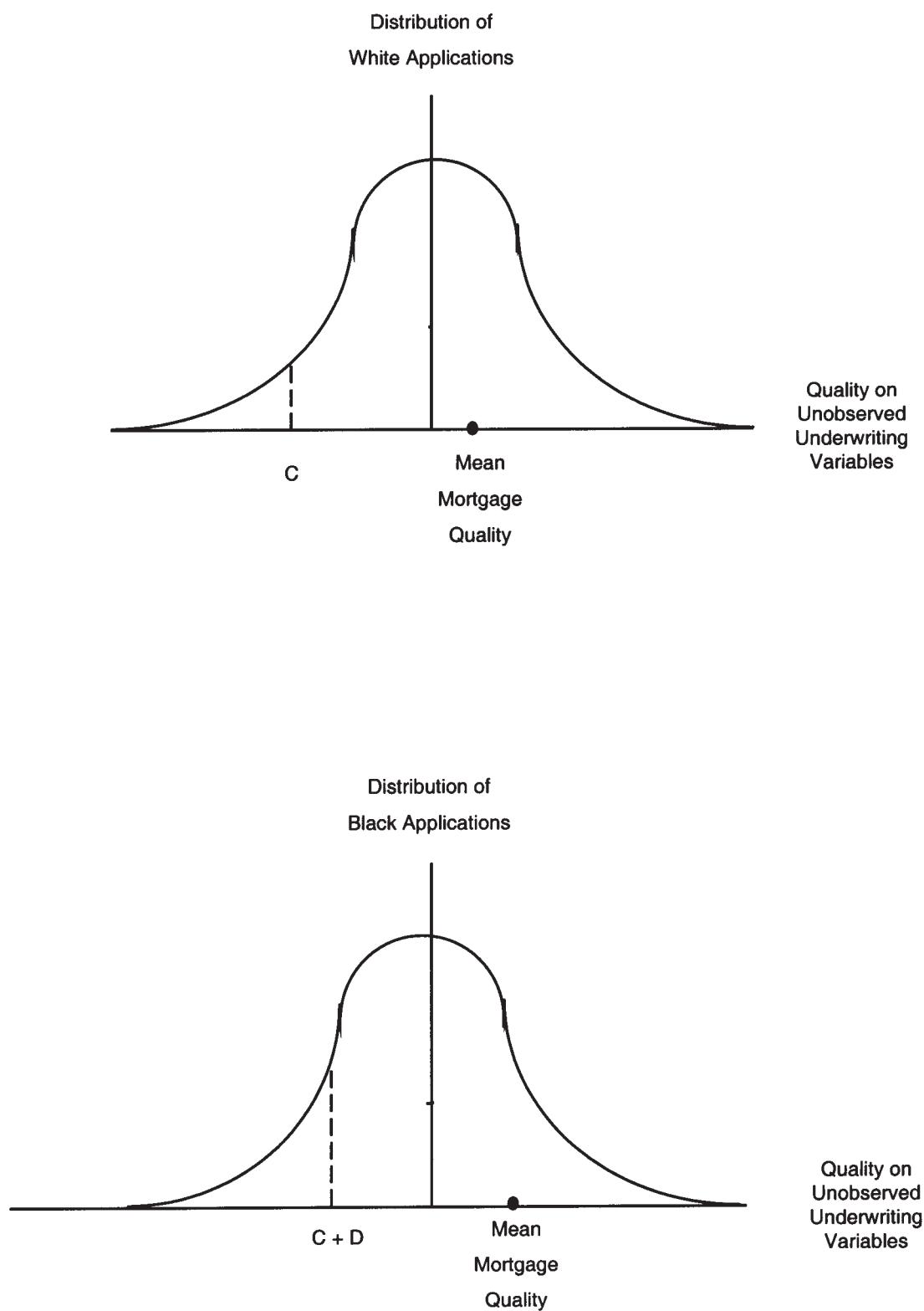


Figure 3.



THE URBAN
INSTITUTE

Figure 4.

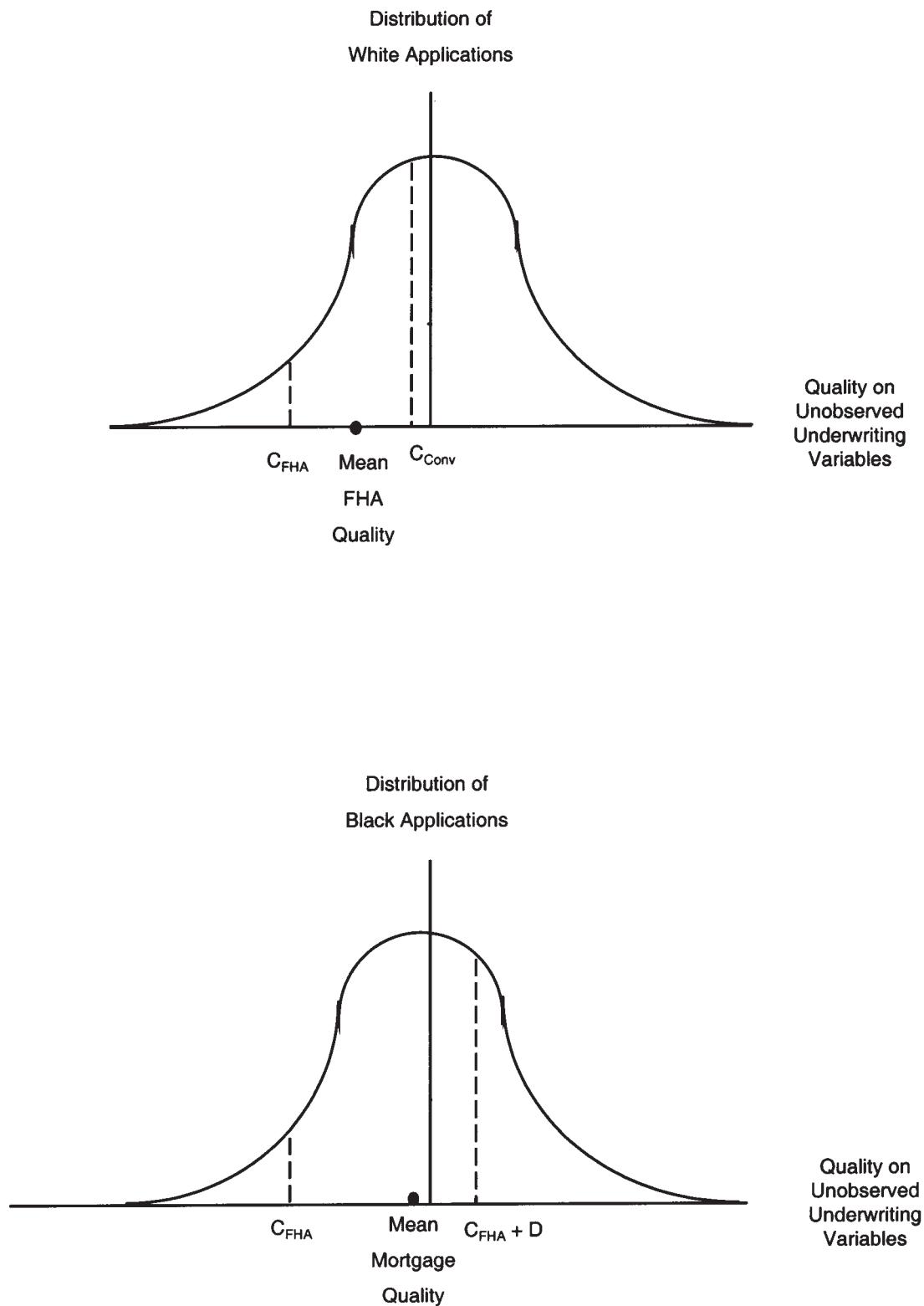
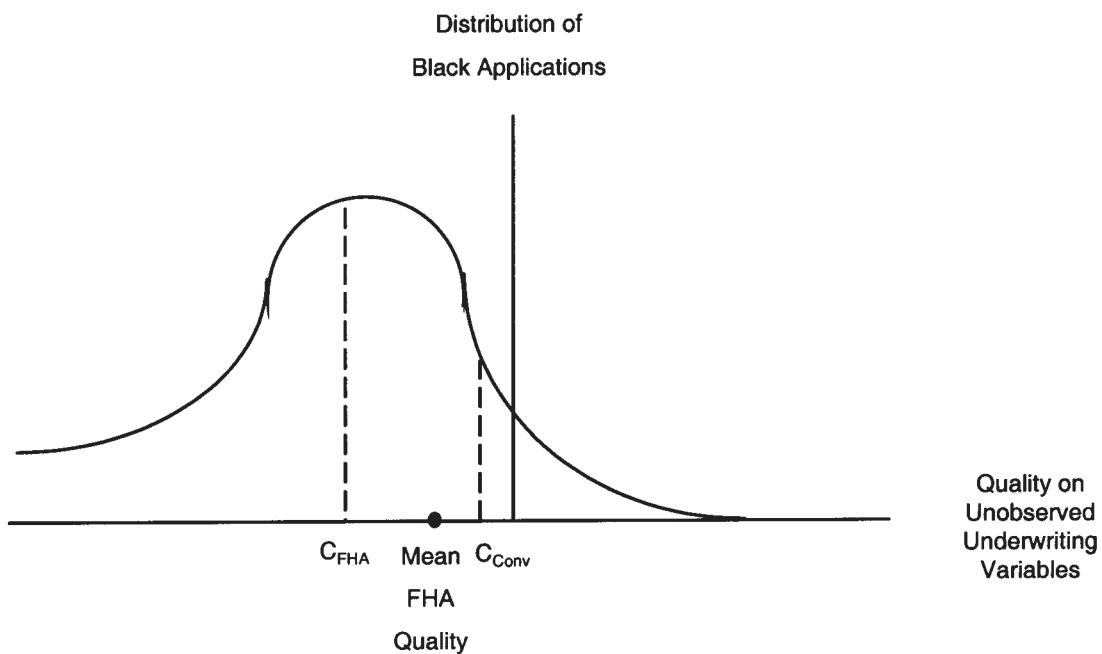
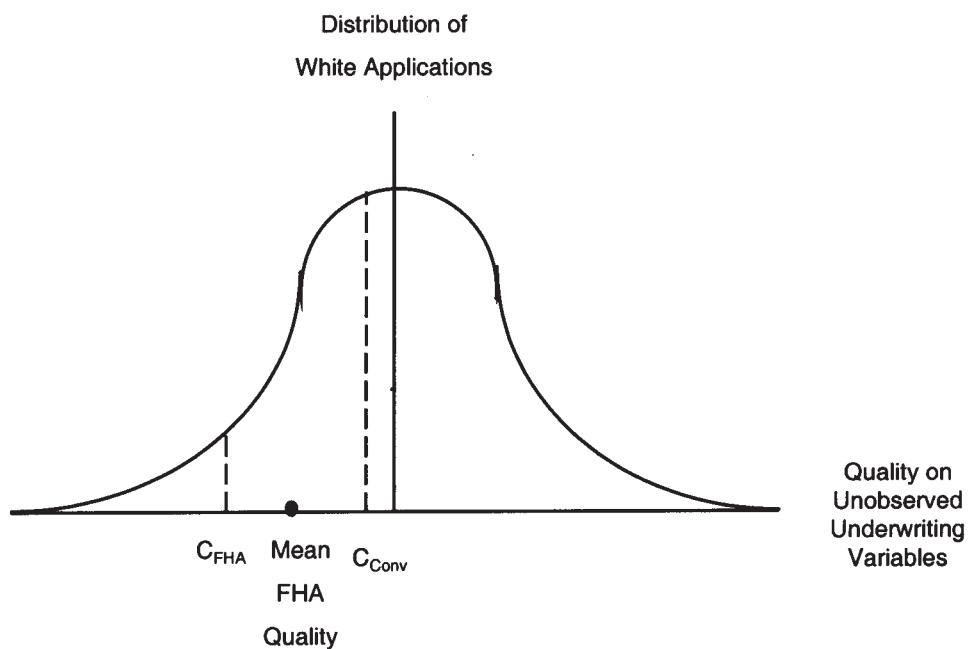


Figure 5.



THE URBAN
INSTITUTE

Technical Appendix to Chapters 3 through 5

STEPHEN L. ROSS
JOHN YINGER

This appendix contains equations and other technical material to further explain the estimation and analysis presented in chapters 3 through 5 of this report. The appendix begins with a brief review of some of the key econometric theorems to which the text refers. Thereafter, the appendix section headings match the relevant headings in the text chapters. For a far more complete and precise discussion, see an econometrics textbook (such as Greene 1993).

Econometric Concepts

The fundamental empirical tool used in the research reviewed here is regression analysis, which is designed to estimate a behavioral relationship based on a sample of observations. In a regression analysis, a dependent variable, say Y , is expressed as a function of a set of explanatory variables, say X , and a random error term, say e . In equation form, $Y = a + bX + e$, where a and b are the coefficients to be estimated. If there are many X variables, then b represents a set of coefficients, not a single one. Each b coefficient indicates the impact of one of the X variables on the dependent variable. In some cases, two different X variables, say X_1 and X_2 , are multiplied together (interacted) to form a new explanatory variable. An interaction term of this type indicates whether the

impact of X_1 on the dependent variable depends on the value of X_2 (and vice versa).

Standard regression techniques provide estimates of a and b for a sample of observations. The overall explanatory power of a regression is summarized by an R-squared, which indicates the percentage of the variance in the dependent variable that is explained by the model. When Y is a yes-or-no variable, such as whether a loan was accepted or denied, the expression $(a + bX)$ can be interpreted as the probability that an event (such as loan denial) will occur.¹ In this case, the well-known logit or probit models are often used to estimate a and b . These models focus on an underlying latent variable under the assumption that the dependent variable takes on a value of one if the latent variable exceeds a certain value. A standard R-squared does not apply to these models, but researchers can use pseudo R-squared, which indicates the share of the variance in the latent variable explained by the model.

Scholars are interested in both the magnitude and the statistical significance of the estimated values of b . An estimate that is large indicates that the associated X variable has a large impact on the dependent variable. An estimate that is statistically significant indicates that the estimated value is very unlikely to have arisen by chance, so a scholar can be confident that the impact of the X variable on the dependent variable is real. Statistical significance is measured with a t-statistic, which is the ratio of the coefficient estimate to its standard error. The conventional indicator of statistical significance for a reasonably large sample is a t -statistic above 1.96, which indicates that the probability that the estimated coefficient simply reflects random factors is less than 5 percent. This is a so-called two-tailed test, which is appropriate if the estimated coefficient could, in principle, be either positive or negative. If the estimated coefficient can be only positive (or only negative), then a one-tailed test is appropriate. In this case, statistical significance at the 5 percent level is indicated by a t -statistic above 1.65.

Regression analysis is based on several assumptions. When those assumptions are satisfied, the estimated values of a and b can be shown to be accurate in the sense that they are close to the true behavioral parameters. In some cases, a violation of one or more of these assumptions will result in biased estimates, that is, in estimated values of a and b that differ systematically from their true values. Many of the methodological problems in this report involve issues of bias.

One key source of bias is an omitted variable. If the true behavioral relationship involves a variable, but this variable is omitted from the regression analysis, then the estimated coefficients of variables that are included in the analysis may be biased. This bias arises because one of the basic regression assumptions is violated, namely, the assumption that all explanatory variables are uncorrelated with unobserved factors, that is, with the error term. In general, the bias depends on the true coefficient of the omitted variable and the correlation between the omitted variable and the included variable. Intuitively, some of the impact of an omitted variable on the dependent variable may be incorrectly assigned to one or more of the included variables. The solution to this problem is to obtain data on all key variables and to include them in the regression; otherwise one cannot rule out the possibility of biased results.



The basic regression model also assumes that all variables are exogenous, which means that they are determined by factors that are not part of the behavior under investigation. Estimated coefficients can be biased if some of the explanatory variables are endogenous. Endogeneity can arise either if some unobserved variable affects both an explanatory variable and the dependent variable, or if the dependent variable has a causal impact on the explanatory variable. In either case, the regression coefficients may not simply reflect the direct impact of an explanatory variable on the dependent variable, but may also pick up indirect effects or effects flowing from the dependent variable. As in the case of omitted variables, the problem here is that, in violation of the basic assumptions, the explanatory variables are correlated with unobserved factors. This type of problem can be solved with a simultaneous equations procedure, in which the explanatory variable is “cleansed” of its endogenous component. The cleansed variable is then used in the regression and the estimated coefficient reflects only the desired direct impact of the explanatory variable on the dependent variable.

In general, a simultaneous equations procedure requires the use of instruments, which are variables that do appear in the basic regression themselves but are related to the endogenous explanatory variable. As indicated in the text, a good instrument should (a) make conceptual sense, (b) be significant in a regression to explain the endogenous explanatory variable, and (c) not have explanatory power in the regression of interest (when the potentially endogenous variable is also included). In principle, one can estimate a simultaneous equations model so long as one has one instrument for each endogenous explanatory variable. In practice, however, a high correlation among instruments may make it difficult to sort out the effects of more than one endogenous variable and better results can often be obtained with more instruments than endogenous variables.

In a study of mortgage lending discrimination, instruments that meet these tests are difficult to find, because all the variables observed by the researcher are also observed by the underwriter and could, in principle, influence the denial decision. In practice, however, lenders collect a huge array of information during the application procedure. To keep the process manageable, they simplify this information using broad indicators of the likelihood of default, such as a credit score or LTV. Our search for instruments, therefore, focuses on variables that may not be directly considered by lenders because they are summarized in a broader indicator that lenders are known to use.

Chapter 3: Omitted Variables, Reanalysis

A Simple Model of Endogeneity in the Loan Denial Equation

This section presents a model to account for endogeneity in “unable to verify.” The same model can be used to account for endogeneity in “meets guidelines.” Let y_v be a binary variable equal to 1 when the loan officer is “unable to verify” certain information. The value of this variable is determined by a latent vari-



THE URBAN
INSTITUTE

able, marked with an asterisk, which is, in turn, influenced by a set of exogenous variables, labeled X :

$$y_v = \begin{cases} 1 & \text{if } y_v^* \geq 0 \\ 0 & \text{if } y_v^* < 0 \end{cases}, \text{ where } y_v^* = \beta_v X + \varepsilon_v. \quad (1)$$

In addition, let y_d be a binary variable equal to 1 when a loan application is denied. The value of this variable is determined by a latent variable, also marked with an asterisk, which is, in turn, influenced both by X and by y_v :

$$y_d = \begin{cases} 1 & \text{if } y_d^* \geq 0 \\ 0 & \text{if } y_d^* < 0 \end{cases}, \text{ where } y_d^* = \beta_d X + \gamma y_v + \varepsilon_d. \quad (2)$$

We estimate this model using a bivariate probit with recursion. Note that the influence of “unable to verify” or “meets guidelines” on the latent variable for denial is identified without exclusion restrictions, because these variables are not continuous. See Amemiya (1974) and Sickles and Schmidt (1978).

A Complex Model of Endogeneity in “Meets Guidelines”

Let y_{m1} be a binary variable to indicate whether the original loan officer believes a loan application meets the lender’s guidelines. It is driven by a latent variable, marked with an asterisk, that is a function of a set of exogenous variables, X (a different set of variables from the one in the previous model). The actual binary variable is not observed, but we can write:

$$y_{m1}^* = \beta_{m1} X + \varepsilon_{m1}. \quad (3)$$

Now let y_{m2} be a binary variable that equals 1 when the bank official filling out the HMDA form indicates that an applicant meets the lender’s guidelines. It is driven by a latent variable, marked with an asterisk, and is influenced by the same factors that influence the original loan officer’s “meets standards” decision and by the original loan officer’s loan denial decision:

$$y_{m2} = \begin{cases} 1 & \text{if } y_{m2}^* \geq 0 \\ 0 & \text{if } y_{m2}^* < 0 \end{cases}, \text{ where } y_{m2}^* = \gamma_1 y_{m1} + \gamma_2 y_d + \varepsilon_{m2}. \quad (4)$$

Finally, let us modify the model of loan denial given earlier by introducing into it the original loan officer’s latent variable for “meets guidelines”:

$$y_d = \begin{cases} 1 & \text{if } y_d^* \geq 0 \\ 0 & \text{if } y_d^* < 0 \end{cases}, \text{ where } y_d^* = \beta_d X + \gamma_3 y_{m1}^* + \varepsilon_d. \quad (5)$$

We cannot estimate (3) because the latent variable for y_{m1} is not observed, but we can substitute (1) into (2) and (3). The result:

$$y_{m2} = \begin{cases} 1 & \text{if } y_{m2}^* \geq 0 \\ 0 & \text{if } y_{m2}^* < 0 \end{cases}, \text{ where } y_{m2}^* = (\gamma_1 \beta_{m1}) X + \gamma_2 y_d + \varepsilon_{m2} + \gamma_1 \varepsilon_{m1}. \quad (6)$$

$$y_d = \begin{cases} 1 & \text{if } y_d^* \geq 0 \\ 0 & \text{if } y_d^* < 0 \end{cases}, \text{ where } y_d^* = (\beta_d + \gamma_3 \beta_{m1}) X + \varepsilon_d + \gamma_3 \varepsilon_{m1}. \quad (7)$$

We estimate (4) and (5) using a bivariate probit model with recursion. In estimating these equations, the parameter γ_1 is not identified and is initialized



to one. In addition, the variances of the error terms are not identified, and the variance of the reduced-form error terms in equations (4) and (5) are also both initialized to one.

As an econometric matter, the relationship between the latent “meets guidelines” variable for the actual loan officer and loan denial can be identified only if a variable exists that influences “meets guidelines” but does not have any direct influence on loan denial. The standard criteria for determining whether a variable meets this test is that it must be significant in explaining the first dependent variable (“meets guidelines”) but insignificant in explaining the second dependent variable (loan denial) when the first dependent variable is included in the specification. The consumer credit history variable appears to meet this test. Past mortgage payment history is probably closely related to future payment, and defaults that appear in an applicant’s record undoubtedly result in a stigma that alters the entire underwriting process. However, to keep their process manageable, lenders may not refer to individual credit history variables once they have been incorporated into the decisions about whether an applicant meets the lender’s guidelines. As it turns out, the consumer credit history variable is highly significant in explaining “meets guidelines” but has no explanatory power (as indicated by a t -statistic of 0.15) for the denial variable when “meets guidelines” is included in the equation. Once consumer credit history has been used to determine whether the application meets the lender’s credit history standards, consumer credit history has no additional influence on the underwriting decision.

Chapter 3: Data Errors in the Explanatory Variables, Reanalysis

The present value, A_{PV} , of a \$1 annuity stream that lasts T periods with an interest rate i can be written as follows:

$$A_{PV} = \frac{1}{i} \left(1 - \frac{1}{(1+i)^T} \right). \quad (8)$$

If all the other terms are known, (8) can be used to solve for the interest rate.

Chapter 3: Endogenous Explanatory Variables, Reanalysis

The model proposed by Rachlis and Yezer (1993) is as follows. Let y_l stand for LTV. Then, using the notation defined earlier,

$$y_l = \beta_l X + \gamma_l y_d^* + \varepsilon_l \quad (9)$$

and

$$y_d = \begin{cases} 1 & \text{if } y_d^* \geq 0 \\ 0 & \text{if } y_d^* < 0 \end{cases}, \text{ where } y_d^* = \beta_d X + \gamma_d y_l + \varepsilon_d. \quad (10)$$



THE URBAN
INSTITUTE

Substituting (10) into (9) results in

$$y_l = \frac{\beta_l + \beta_d \gamma_l}{1 - \gamma_d} X + \frac{\varepsilon_l + \gamma_l \varepsilon_d}{1 - \gamma_d} = \beta_l X + \hat{\varepsilon}_l. \quad (11)$$

Hence, a regression of LTV on a set of exogenous variables can be interpreted as a reduced form of equation (9). We employ a reduced form of this type to obtain an “instrument” for LTV, that is, to correct for the possible endogeneity of LTV in a loan denial equation.

Chapter 5: New Twists, a Default Test Based on Market Concentration

As noted in the text, market concentration may affect credit rationing on many underwriting variables including credit history, which is omitted from Berkovec et al.’s (1998) analysis. Lenders in highly concentrated markets will hold applicants to higher standards on credit history variables. So just as omitted credit history variables bias the race coefficient in the default model, the relationship between credit rationing and market concentration biases the coefficient on the interaction between race and market concentration.

The path of the bias is quite complex, but it can be untangled (see Ross 1997). First, we consider the bias of a default analysis based on a sample of conventional mortgages. The omission of credit history variables biases the race coefficient upward in a default model. More stringent credit rationing in highly concentrated markets should weaken the relationship between credit history and default in the sample of approved mortgages for highly concentrated markets. Therefore, the bias in the race coefficient should be smaller in more concentrated markets, which would create a negative relationship between market concentration and the size of the race coefficient. This bias is consistent with Berkovec et al.’s (1998) negative estimate for the coefficient on the interaction between market structure and race, and biases the default approach toward finding discrimination. Since they did not find statistically significant evidence of discrimination, this bias would not be a problem for their results if the results were based on conventional mortgages.

For a sample of FHA mortgages, the problem is even more complex. First, the effect of omitting credit history variables on the relationship between race and defaults for FHA mortgages is unknown. The omission of credit history variables lowers the quality of the distribution of minority applications on unobserved underwriting variables. At first glance, a leftward shift in the distribution of black applications relative to white applications will increase racial differences in default, because many low-quality FHA mortgages are falling below the cutoff and out of the FHA sample. However, this assumes that the FHA and conventional cutoffs are fixed, but these cutoffs are for application quality based on unobserved underwriting variables. The nature of this quality measure changes when credit history variables are omitted. Most white applications are for loans in the conventional sector, and most of those applications are approved. So most white applications have acceptable credit history, other



things equal, and omitting credit history variables may actually shift the FHA and conventional cutoffs to the left relative to the white distribution of mortgage applications. Therefore, omission of credit history variables shifts both the cutoffs and the distribution of minority applications to the left, and it is uncertain whether omitting credit history increases or decreases racial differences in default for an FHA sample. Moreover, the effect of market concentration on the relationship between credit history and default is unknown. For conventional loans, more aggressive underwriting filters out lower-quality applications, decreases the variance in the sample, and reduces the strength of the relationship between credit history and default. However, more aggressive underwriting by conventional lenders shifts more mortgages into the FHA sector, which may increase the variance in the FHA sample and increase the relationship between credit history and default. Thus, no conclusions can be made concerning the direction of the omitted-variable bias when the default approach is pursued with a sample of FHA mortgages.

Note

1. Yes-or-no variables are known as binary, or dummy, variables.



THE URBAN
INSTITUTE

Chapter 6

Inside a Lender: A Case Study of the Mortgage Application Process

KENNETH TEMKIN
DIANE K. LEVY
DAVID LEVINE

Although fair lending laws mandate that all loan applicants receive equal treatment, all of the evidence reveals wide disparities in origination outcomes between white and minority loan applicants. Some of these differences are attributable to income and wealth differences between minorities and whites. Rigorous statistical analysis, however, continues to find loan denial disparities between minority and white loan applicants, even when differences in applicant creditworthiness and loan characteristics have been controlled for, and even when lenders appear to believe that no disparate treatment exists.

This case study examines the loan application process of one lender in detail, to shed light on the relationship between a lender's organizational practices and staff perceptions and its loan outcomes as reflected in its HMDA scores.¹ While the results of a case study of one lender have no statistical generalizability, the case study approach is valuable because it "allows an investigation to retain the holistic and meaningful characteristics of real-life events—such as...managerial processes" (Yin 1989, p.14).

The research team conducted interviews with seven employees over a two-day period and conducted follow-up interviews with the lender's president and two loan counselors. The initial interviews, following the discussion guides

presented in annex A (which appears at the end of chapter 2), were conducted to determine if employees were aware of fair lending requirements, had received fair lending training, and had had their performance monitored for compliance with fair lending requirements. In addition, each employee was asked to describe his or her role in the lender's loan origination process. Interviewees were assured anonymity; therefore, neither the lender nor any employee is named.

The case study is followed by a discussion of specific managerial practices that will affect fair lending performance. We also discuss the challenges associated with instituting such policies.

Description of Case Study Lender

The lender analyzed in this case study is a mortgage company, fully owned by a builder who develops housing for low- and moderate-, middle- and upper-income households. Founded in 1991, the company has grown from 2 to 31 employees and currently originates roughly 1,000 mortgages per year worth about \$70 million. Nearly all mortgages originated by the company are for a home purchase, rather than to refinance an existing mortgage. The lender operates in a large city that has substantial numbers of black and Hispanic residents. It processes more minority applications, as a proportion of its total volume, than the average for its metropolitan statistical area (MSA).

About three-quarters of the loans originated by the company are underwritten according to government (Federal Housing Administration [FHA], Veterans Administration [VA], and Farmers Home Administration [FmHA]) guidelines that have more flexible underwriting standards than conventional mortgages. As a result, the lender is able to qualify applicants with less-than-perfect credit and fewer resources for a down payment than associated with conventional mortgage standards.

The lender does not service any of the mortgages it originates. Its conventional loans are sold to the company that underwrites the loan. Its government loans are sold to one of two financial organizations. The first requires a four-month recourse period, during which the lender is responsible for the loan balance in the event of a borrower's default. The second requires a one-month recourse period. Because of the recourse terms of its loan sales, the lender has an incentive to apply conservative underwriting guidelines.

The lender submits all FHA-insured loan files for post-close audits. FHA audits a sample of loan files submitted to ensure that underwriters comply with its standards. In the event of a questionable underwriting judgment, FHA contacts the lender and informs the company about its concern. Lenders who continue to make loans with features that are unacceptable to FHA underwriters risk sanctions, including being dropped from the FHA program.

The lender employs six loan counselors, who work in one of three offices and who are responsible for meeting prospective customers and taking applications. Unlike many mortgage companies, loan counselors do not take applications in the field. The loan counselors all report to the branch manager of the company, who also supervises a team of four loan processors responsible for



collecting the documentation needed to complete a loan application. The company employs an underwriter who is responsible for determining the creditworthiness of all mortgage applicants. The underwriter and the branch manager report directly to the president. The company also has staff who work on closings and quality control. However, these staff are not involved in the origination decision process.

Four of the six loan counselors are white and two are Hispanic. The company had a black loan counselor, but she recently left to relocate to another part of the state. The president, underwriter, and processors are all white. The president was employed by another mortgage company before she moved to her present position. Her previous employer went bankrupt, and she was hired for her current position in 1991. Most of the people interviewed by the research team had worked for the president's previous employer, and found out about job openings through conversations with her. One Hispanic loan counselor, however, was hired after two rounds of interviews following his response to a local newspaper advertisement asking explicitly for a Spanish speaker.

The president estimates that 85 percent of the customers served by the company are referred by the builder's sales representatives after potential home buyers complete sales contracts. The remaining 15 percent are referred by real estate agents familiar with the company. Home purchase applications from blacks account for 25.2 percent of the lender's total home purchase application volume, almost three times the MSA figure of 8.9 percent. Hispanics account for 17.6 percent of the lender's purchase mortgage applications, higher than the MSA figure of 13.3 percent.

As mentioned earlier, roughly three-quarters (73.6 percent) of mortgages originated by the company are FHA, VA, or FmHA loans, compared with 16.3 percent of all mortgages originated in the MSA. Blacks account for slightly more than 30 percent of the lender's government loan applications, Hispanics for another 21.4 percent. These proportions are higher than for the MSA as a whole, where blacks account for 13.4 percent and Hispanics 17 percent of government loan applications.

The lender's applicants are disproportionately middle-income. Almost 40 percent of them have incomes that fall between 80 and 120 percent of the MSA median, compared with 22.9 percent of all loan applicants in the MSA. Only 28.9 percent have incomes 120 percent above the MSA median, compared with 39.6 percent of all applicants in the MSA. And 31.2 percent have incomes less than 80 percent of the MSA median, compared with 37.3 percent of all applicants in the MSA.

Lender's Origination Process

The lender's origination process is designed, according to respondents, to qualify as many applicants as possible, irrespective of race or ethnicity. Most applicants are referred to the lender with a contract on a house built by the owner of the mortgage company. The whole purpose of the company, according to one employee, is to get people into homes. As a result, the lender does not conduct prequalification assessments. Every customer completes a hard-copy loan



THE URBAN
INSTITUTE

application, and all the information from it is entered into an electronic version of the application form located on the lender's computer system.

Overview

All respondents said they have a very strong commitment to treat every customer fairly, based on their personal conviction that discrimination is wrong and must not be tolerated. The lender does not provide any specific fair lending

training, however, and has only a one-paragraph discussion of fair lending in its procedures manual. Respondents also said that it does not make business sense to turn away potential business based on an applicant's race. Nevertheless, the company has been subject to discrimination claims by minority customers who were denied loans. Staff said these claims were baseless, and the company has never been found liable.

"We originated a mortgage to a lady that had three jobs, two child support payments, and SSI for her nephew; so we had six different sources of income. Anybody else would have looked to only one source. We were able to tie in all three jobs, and we used the child support. She got a \$101,500 loan. She had been to another mortgage company and been denied. She works at the Wal-Mart down the street and gives me a big hug every time she sees me."

—A loan counselor

In order to accomplish its mission, the lender's origination process, detailed below and outlined in figure 1, includes multiple reviews so that no employee can make a unilateral decision about a particular application. The lender uses a "team" approach, whereby a loan counselor, a processor, the branch manager, and the underwriter or president use as much creativity as possible to qualify applicants. The status of every loan application is discussed at the weekly staff meeting attended by loan counselors, processors, and the branch manager. One purpose of these meetings is to have staff brainstorm about strategies that can be used to qualify marginal applicants.

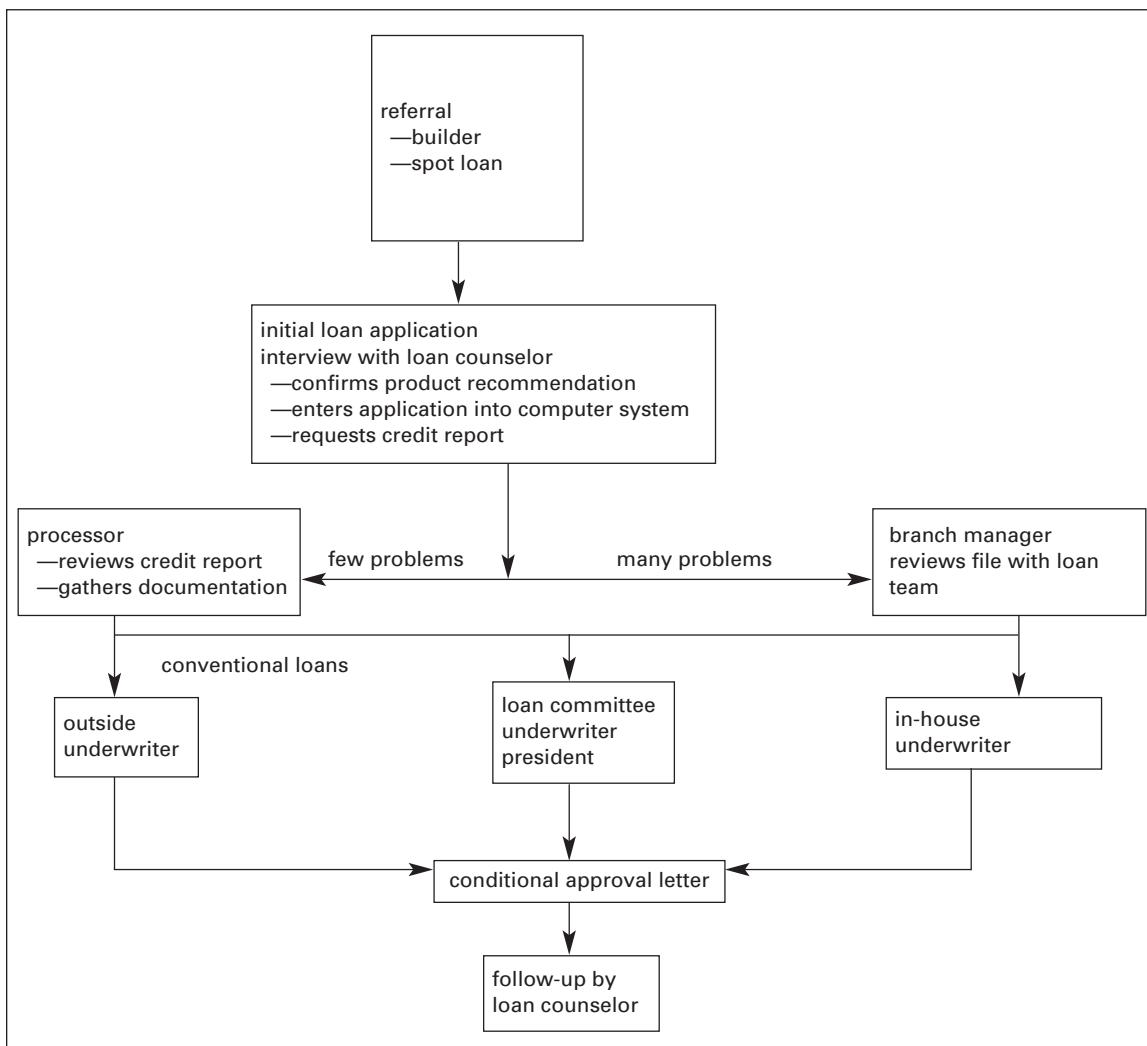
The lender's origination process never results in an outright denial. Rather, every applicant receives a conditional approval, with a mortgage originated once the specified conditions are met. These conditions are based on the perceived underlying risk associated with the potential borrower and are tailored to meet the needs of that customer. An applicant who meets all the guidelines will receive a mortgage subject only to receipt of an appraisal report. This is a relatively straightforward and rapid process. Borrowers who fail a greater number of underwriting guidelines may have to pay past due debts, or lower their overall monthly financial obligations. A more complex conditional approval does not preclude the applicant from receiving a mortgage from the lender. Indeed, the lender sometimes originates mortgages to applicants one year after the application was initially processed.

Some applicants decide they will be unable to meet the conditions set forth by the lender and tell the lender to withdraw their application. In these cases the lender sends the applicant an adverse action letter, and the loan application is classified as a denial.

All of the lender's staff interviewed by the research team expressed great pride in their ability to work with borrowers, even with borrowers whose loan



Figure 1. Lender's origination process



applications have multiple problems. Indeed, many staff members said the company originates loans to many customers who would not receive mortgages from other companies where staff are not as dedicated to working with marginal applicants.

Referral

Since most of the lender's customers are referred by a sales representative of the builder that owns the company, loan applicants typically have already signed a contract for a house before they contact the lender. After signing the contract, customers from a particular subdivision are referred to a particular loan counselor, with each loan counselor servicing about 10 subdivisions. These customers are encouraged to use the mortgage company, and are given the loan counselor's business card, which contains contact information as well as the documentation needed to complete a loan application. In addition, customers



THE URBAN
INSTITUTE

are offered a discount on closing costs if they choose to use the lender. According to one respondent, the sales representative says to the customer, “Why don’t you give [loan counselor’s name] a call. Here’s [his/her] card. They can help you with the financing.”

The builder has three types of subdivisions: entry-level homes priced between \$70,000 and \$90,000; trade-up homes between \$90,000 and \$120,000; and luxury homes that start at \$150,000. Loan counselors are assigned a mix of subdivisions to ensure that each loan counselor serves a variety of applicants. This mix is important, because a portion of the loan counselor’s compensation is based on the dollar volume of mortgages originated. Loan counselors receive a base pay plus a commission of 10 basis points once the mortgage closes.

“We had a woman; she had been denied by two other mortgage companies. I looked at her credit report. She seemed to have \$17,000 in collections. It took nine months to sort through it all. Some of the information was wrong. She got a mortgage from us.”

—A loan counselor

Most customers make initial contact with the company via a telephone call to the loan counselor, whose name they have been given. Because many of the lender’s customers have signed a contract for a specific house, they are highly motivated to provide as much information as possible in the initial

interview. One respondent said, “Our customers have seen the house, or walked through a model. They have this picture in their mind already.” The loan counselor tells the customer to bring the documentation described on the business card to the initial loan application interview. The two then agree on a mutually convenient time for that interview.

Initial Application Interview

The lender has a corporate headquarters and two branch offices. A small number of applications are completed via mail or over the telephone. Almost all applicants complete the initial application interview at either the corporate headquarters or the main branch because the other branch is staffed by a loan counselor only one morning a week. Both the corporate headquarters and the main branch are located far from the city center, off major roads in commercial areas approximately 30 miles apart. Because the city is quite spread out, most area businesses are not located in the central business district.

Every customer completes a hard-copy loan application, which asks the customer about his/her income, employment history, existing debts, and other relevant information. This information is entered by the loan counselor into an electronic version of the form, which has a field for each item on the hard-copy loan application. In addition, customers must indicate their race, which is entered on a separate field that is visible only to readers who scroll down several screens. This information is never used in the origination decision process, according to the lender’s staff. It is required for disclosure purposes,² however, and most of the staff refer to it as “disclosure information.” The loan counselors ask for the disclosure information at the end of the initial application interview. One loan counselor said she turns the computer screen toward the customer so that the customer can type in the appropriate category. She said she does this



because she does not want to guess which category is appropriate, and most customers type in their own choice.

The initial loan application takes about two hours to complete. In most cases, the customer's contract already indicates whether the borrower should receive a government or conventional loan. Since government loans allow for higher loan-to-value (LTV) ratios, the builder's sales representative recommends government loans for customers who do not have sufficient funds for a conventional down payment. According to the lender's president, conventional loans are most suitable for middle-income applicants with good credit, and most of the lender's customers have problematic credit and few resources for a down payment. While the loan counselors review the suggestion made by the sales representative, and can make a different decision, respondents said it was very unusual for an FHA-eligible customer to receive a conventional mortgage.

At the end of the initial application interview, the loan counselor explains the next steps in the process. At this point the applicant also has to provide \$75 to pay for a credit report and is told that the loan counselor will be in touch once the credit report is received. All four loan counselors interviewed by the research team said they never forecast outcomes with the customer. According to one counselor, "You don't want to get anybody's hopes up. You also don't want to be too discouraging."

Once the customer leaves, the loan counselor adds comments to the computer version of the loan application. None of the respondents said it was acceptable to add subjective feelings about an applicant. Instead, they said, comments are factual and relate the applicant's employment history, credit history, income, and whether the loan is government or conventional. Because these comments are on the electronic version of the application, they are accessible to everyone in the company and meant to provide information to the underwriter and branch manager about any financial issues that warrant attention.

After completing the comments, the loan counselor sends the file, with the completed hard-copy application, to either a processor or the branch manager. The processors receive relatively problem-free applications. They then secure the documentation needed to complete the file before submitting it to the underwriter. A completed loan application file has information collected during the application interview, a full credit report, an appraisal, and documentation of the customer's employment and financial statements.

The branch manager receives the applications that have a high front-end (house-expense-to-income) and/or back-end (total-monthly-debt-to-income) ratio or information customers have volunteered about past credit problems. She works with the processor to develop a plan to handle each application referred to her. She then forwards the applications to the underwriter after the questions have been answered. Applications that fail more than one underwriting guideline are sent by the branch manager directly to the president of the company. These applications are said to have gone to "loan committee," which means they will be reviewed by the underwriter and the president.

Completed conventional mortgage applications are sent to outside underwriters. The company uses three, but most of its conventional application busi-



THE URBAN
INSTITUTE

ness is sent to a mortgage insurance company. Completed government mortgage applications that are sent to the loan committee by the branch manager are assessed by the committee. Applications sent directly to the underwriter may also be referred to the loan committee.

Underwriting

The underwriter does not use an automated underwriting system or credit scores to measure a borrower's creditworthiness. Rather, she judges the application by evaluating all relevant information in the file. According to her own responses to us, she applies the underwriting guidelines in as flexible a manner as possible and starts with a review of the credit report. In addition, she judges whether the applicant's income and current rental payment performance offset instances of derogatory credit.

The underwriter prefers to approve applications where the front-end ratio is less than 31 percent. She also likes to see applicants with back-end ratios below 43 percent, although she will approve applications with a back-end ratio of 46 percent so long as the applicant has a good credit history. In addition, the underwriter looks to see if the applicant's income is increasing over time, and will calculate a second set of ratios to include payments for overtime work.

The underwriter said the applicant's race never enters into her decisions and she could not imagine not being fair to different people. She does not receive any information about the race of the people who actually receive loans from the company, and could only provide a guess as to the percentage of minorities who receive them. In addition, the underwriter expressed great pride in the company's ability to qualify marginal applicants. She told us that it can be quite moving when new homeowners come into the office to make their first payment.

The underwriter also shared with us some of the letters submitted by applicants to explain past instances of derogatory credit. She said these letters indicated the level of credit problems she had to evaluate in her underwriting decisions. She reads every credit explanation letter to see if derogatory credit episodes resulted from a one-time illness or other family crisis, or represent a pattern. She read a few letters aloud that detailed stories of family illnesses and job losses and that contained many spelling and grammatical errors. The research team noticed that some of the letters were typewritten and asked the underwriter: Who sends handwritten credit letters? In a stage whisper, she said, "Mainly the minorities." She also said that poorly written credit letters did not invalidate an applicant's reasons for a derogatory credit episode.

Post-Underwriting

The underwriter's review of an application never results in an outright denial or an unconditional approval. All conditional approvals are transmitted by letters reviewed and signed by the president. Therefore, no customers receive conditions before the underwriter, the president, and often the branch manager have reviewed their files. Thus, no single person in the company unilaterally can reject a loan application or impose a condition without another employee's review.



For each conditional approval, the loan counselor calls once the conditions are finalized, to tell the customer to expect the president's letter in the mail with the conditional approval, and to explain the steps needed to meet each of the conditions that will be contained in the letter. In addition to the detailed information about the steps needed to secure approval, the conditional approval letter contains a timetable to complete these steps. The loan counselors are expected to assist customers with meeting conditions.

Conclusions

The loan origination process used by the lender ensures that all applications receive a careful review by at least two staff members. The lender's staff take great pride in their commitment to qualify borrowers who seem, at first, to be unacceptable credit risks. This commitment, according to the lender's staff, is provided to both minority and white applicants. Every staff member interviewed said the entire process was "race-blind." None of the staff could think of a reason why minorities would be treated differently from whites. Although a customer's application contains information about his/her race, the underwriter

"We're currently closing an application from a lady in a homeless shelter. I'm working with her. She had judgments and a repossession and we got her approved. She has worked in the same job for the past 15 years. She takes care of her grandson. She has no transportation, but her application is workable."

—A loan counselor

and the president, who actually evaluate applications, say they are unaware of an applicant's race. The point of the application process, according to everyone interviewed at the company, is to get customers into homes, and staff seemed bewildered by the suggestion that race could be a factor in the lender's origination decisions. Finally, the lender's presi-

dent expressed a high level of enthusiasm for this research project, was eager to participate and answer questions about fair lending, and scheduled interviews for the research team.

Over the course of the site visit, the research team scrutinized the process used by the lender to assess applications. The combination of a highly transparent review process and a seemingly genuine commitment to fair lending suggested to us that minorities were *not* receiving differential treatment from anybody on the lender's staff. We also were impressed with the high level of personal satisfaction the lender's staff received from working in the mortgage finance industry. Specifically, many staff members expressed a sense of pride in helping people who would probably be denied mortgages from other companies achieve their dream of homeownership. We left with the expectation that the lender's HMDA data would show little denial disparity between white and minority applicants.

The Lender's HMDA Performance

We were wrong. The HMDA data indicate that the case study company denies minority loan applicants at a rate lower than other MSA lenders. This suggests



THE URBAN
INSTITUTE

that the lender is doing a better job than other area lenders in qualifying marginal minority applicants and is consistent with the statements of all interviewees that they work hard to qualify problematic applicants irrespective of race. But the lender's HMDA data also reveal large disparities in denial rates for minority versus white loan applicants. Disparities that are wider than MSA averages persist even after controlling for the applicant's income and loan product type.³ As noted in previous chapters of the report, HMDA data alone cannot prove either differential treatment or disparate impact discrimination. When corrected for applicant income and loan product, however, they certainly raise questions.

Overall HMDA Performance

The lender's overall HMDA performance for a representative recent year,⁴ as summarized in table 1, indicates that the lender's denial rate for black and Hispanic applicants is below MSA averages. This finding is consistent with many respondents' statements that the lender makes an extra effort to qualify marginal applicants irrespective of race or ethnicity.

The lender's denial disparity index (DDI),⁵ however, which is the ratio of minority-to-white denial rates, is much higher than the MSA average. This is due to the lender's lower-than-average denial rate of white applicants. Applications submitted to the lender by black applicants are rejected 2.73 times as often as those submitted by white applicants. Applications from Hispanics are rejected 2.5 times as often as those from whites. The area DDIs for blacks and Hispanics are 1.58 and 1.36, respectively.

HMDA Performance Controlling for Applicant Income

The company's high black and Hispanic DDIs do not necessarily mean the lender is treating minorities differently from whites. They may reflect differences in the ability of minority applicants to meet underwriting guidelines. HMDA data do not include all of the information used by an underwriter in making origination decisions. The data, however, do provide information about an applicant's income. Since income correlates with creditworthiness (Avery et al. 1998) and wealth (Simmons 1997), analyzing denial rates and DDIs by applicant income at least partially controls for the quality of an individual's application.

Table 2 presents denial rates for the case study lender and the MSA, controlling for income. In interpreting this and other tables that categorize appli-

Table 1. Comparison of Denial Rates by Race and Ethnicity

| | Lender Denial Rates | | | MSA Denial Rates | | |
|------------------------------|---------------------|----------|-------|------------------|----------|-------|
| | Black | Hispanic | White | Black | Hispanic | White |
| Denial Rate | 28.0% | 25.6% | 10.3% | 38.9% | 33.4% | 24.6% |
| Denial Disparity Index (DDI) | 2.73 | 2.50 | n/a | 1.58 | 1.36 | n/a |

Source: HMDA.

Note: n/a = not available.



Table 2. Comparison of Denial Rates by Race and Ethnicity by Applicant Income

| | Lender Denial Rates | | | MSA Denial Rates | | |
|------------------|---------------------|----------|-------|------------------|----------|-------|
| | Black | Hispanic | White | Black | Hispanic | White |
| Low | 56.3% | 53.6% | 71.4% | 55.3% | 42.6% | 54.1% |
| Moderate | 28.6% | 21.3% | 10.6% | 42.3% | 32.7% | 39.6% |
| Middle | 27.4% | 22.7% | 8.8% | 32.0% | 31.5% | 26.0% |
| Upper | 22.2% | 15.6% | 9.5% | 28.7% | 22.0% | 12.0% |
| Weighted Average | 28.3% | 25.3% | 10.2% | 39.3% | 33.4% | 24.6% |

Source: HMDA

Note: Low-income applicants have incomes less than 50 percent of the MSA median; moderate-income applicants have incomes between 50 and 80 percent of MSA median; middle-income applicants have incomes between 80 and 120 percent of MSA median; and upper-income applicants have incomes above 120 percent of MSA median.

cants by income, it is important to remember that nearly three-quarters (70 percent) of all black and Hispanic applicants processed by the lender are in the moderate- and middle-income groups.

The lender's denial rates for minorities are below MSA averages for moderate- and middle-income applicants. Only low-income Hispanic applicants are denied by the lender at a higher rate than similar applicants in the MSA, but the lender only processed 15 such applications. The lender also denies a much smaller proportion of moderate-, middle-, and upper-income white applicants compared with MSA averages. Low-income white applicants are more likely to be denied by the lender compared with MSA lenders, but this figure is based on only 7 applications.

However, the DDI data show that the lender's relative denial rates for minorities versus whites, controlling for income differences, tend to favor whites more than is true for the area as a whole. The lender's DDIs for moderate- and middle-income black and Hispanic applicants are higher than MSA averages. The lender's DDIs for low- and upper-income black and Hispanic applicants are lower than MSA averages. As already noted, data for these income groups need to be interpreted with caution because they represent only small numbers of applicants.

Table 3. Comparison of DDIs by Race and Ethnicity by Applicant Income

| | Lender's DDIs | | | MSA DDIs | | |
|------------------|---------------|----------|-------|----------|----------|-------|
| | Black | Hispanic | White | Black | Hispanic | White |
| Low | 0.79 | 0.75 | n/a | 1.02 | 0.79 | n/a |
| Moderate | 2.70** | 2.02** | n/a | 1.07 | 0.83 | n/a |
| Middle | 3.11** | 2.58** | n/a | 1.23 | 1.21 | n/a |
| Upper | 2.34 | 1.65 | n/a | 2.40 | 1.85 | n/a |
| Weighted Average | 2.67 | 1.96 | | 1.41 | 1.06 | |

Source: HMDA

**Significant difference from MSA DDI at the .01 level

Note: Low-income applicants have incomes less than 50 percent of the MSA median; moderate-income applicants have incomes between 50 and 80 percent of MSA median; middle-income applicants have incomes between 80 and 120 percent of MSA median; and upper-income applicants have incomes above 120 percent of MSA median.



THE URBAN
INSTITUTE

Table 4. Comparison of Denial Rates by Race and Ethnicity by Applicant Income for Government Loans

| | Lender Denial Rates | | | MSA Denial Rates | | |
|------------------|---------------------|----------|-------|------------------|----------|-------|
| | Black | Hispanic | White | Black | Hispanic | White |
| Low | 60.0% | 53.6% | 75.0% | 42.4% | 21.8% | 24.0% |
| Moderate | 27.1% | 22.4% | 10.8% | 24.0% | 15.1% | 12.3% |
| Middle | 24.2% | 23.6% | 10.6% | 20.9% | 14.2% | 9.6% |
| Upper | 27.3% | 15.8% | 8.5% | 19.3% | 14.5% | 9.4% |
| Overall | 27.9% | 27.5% | 11.2% | 24.1% | 16.1% | 10.9% |
| Weighted Average | 28.1% | 27.2% | 10.9% | 24.9% | 16.4% | 11.2% |

Source: HMDA

Note: Low-income applicants have incomes less than 50 percent of the MSA median; moderate-income applicants have incomes between 50 and 80 percent of MSA median; middle-income applicants have incomes between 80 and 120 percent of MSA median; and upper-income applicants have incomes above 120 percent of MSA median.

HMDA Analysis for Government Loans

As discussed earlier, most of the lender's originations are government loans because so many customers served by the company combine less-than-spotless credit with few resources for a down payment. In addition, the lender's underwriter only examines government loans. These factors indicate that government loans better reflect the company's treatment of loan applications because they are the only loans processed entirely by the company; therefore, tables 4 and 5 show denial rates and DDIs for government loan applications controlling for applicant income.

The picture is even clearer. The lender's denial rates for minority government loan applicants at all income levels are higher than MSA rates for such borrowers (see table 4). This pattern contrasts with the lender's white denial rates, which are lower than MSA levels for moderate- and upper-income and virtually the same for middle-income white government loan applicants. The lender's DDIs for black and Hispanic applicants (except for those with low incomes) are also higher than MSA averages (see table 5). This pattern reflects

Table 5. Comparison of DDIs by Race and Ethnicity by Applicant Income for Government Loans

| | Lender's DDIs | | | MSA DDIs | | |
|------------------|---------------|----------|-------|----------|----------|-------|
| | Black | Hispanic | White | Black | Hispanic | White |
| Low | 0.80 | 0.71 | n/a | 1.77 | 0.91 | n/a |
| Moderate | 2.50** | 2.06** | n/a | 1.86 | 1.17 | n/a |
| Middle | 2.28 | 2.22** | n/a | 2.18 | 1.49 | n/a |
| Upper | 3.19 | 1.85 | n/a | 2.05 | 1.54 | n/a |
| Overall | 2.50** | 2.46** | n/a | 2.22 | 1.48 | n/a |
| Weighted Average | 2.43 | 1.86 | n/a | 2.28 | 1.77 | n/a |

Source: HMDA

**Significant difference with MSA DDI at the .01 level

Note: Low-income applicants have incomes less than 50 percent of the MSA median; moderate-income applicants have incomes between 50 and 80 percent of MSA median; middle-income applicants have incomes between 80 and 120 percent of MSA median; and upper-income applicants have incomes above 120 percent of MSA median.



the lender's higher minority and lower white denial rates compared with other lenders in the MSA. It is difficult to reconcile these results with the lender's belief that the origination process contains absolutely no differential treatment of minority borrowers.

Conclusions

The lender's higher denial disparities surprised the research team members who were unaware of the lender's HMDA performance before the site visit and, on the basis of site visit observations, expected the lender to have a very low denial disparity.

There are three potential explanations for the differences identified with the HMDA data, only the third of which implies that minority borrowers receive differential treatment. First, the DDI values may result from a higher number of minority applicants, relative to whites, failing to meet underwriting guidelines. As discussed earlier, the lender is fully owned by a builder. It may be that the builder's sales staff refer as many applicants as possible in order to increase the potential for a sale. As a result, the lender may be evaluating a large number of applicants who would not be processed by lenders not owned by a builder. To the extent that minority applicants have problematic applications, the DDIs may reflect racial differences in the creditworthiness of different applicants, rather than differential treatment of minority applicants.

A second possibility is that the HMDA-generated DDI figures are imprecise measures of discrimination, and may represent a "false positive." However, because the company has not instituted any of the management practices identified by Listokin and Wyly (1998) that may promote fair lending, discussed in the next section, a third possibility is that the case study represents a "false negative." The company's staff may assume they are carrying out a race-blind application process, but there is no training and monitoring, and staff are unaware of differential outcomes. Good intentions among staff without good monitoring and feedback may not lead to good (nondiscriminatory) outcomes. It may be, for example, that the lender's staff are unwittingly providing more assistance to marginal white applicants, despite a high level of assistance to minority customers. Such a result would be consistent with the Boston Fed Study's findings described in chapter 3 of this report.⁶

What the Lender Isn't Doing and Should Do

Of the managerial practices and procedures identified by fair lending experts as reducing the possibility of differential treatment of minority loan applications, the lender has implemented none completely and two only nominally. The lender's management does not see the need for specific fair lending training or monitoring, believing that the company's origination process contains sufficient checks to eliminate the potential for differential treatment. Without proper monitoring, however, differential treatment cannot be ruled out.

There is a growing literature about managerial and organizational procedures that help companies comply with fair lending laws.⁷ These strategies pre-



THE URBAN
INSTITUTE

scribe specific training, monitoring, compensation, and managerial strategies to lenders who want to guarantee all customers equal treatment (Listokin and Wyly 1998). This section contrasts the recommended activities with the current practices of the lender.

Develop a Formal Mission Statement

Recommended Activities

Include goal of fostering minority lending in institutions' overall mission statement, lending policy statement, or similar defining document.

Lender's Activities

The lender posts a description of fair lending laws in its office. In addition, the lender has a one-paragraph fair lending statement in its procedures manual. This statement was developed in 1994 in response to a HUD letter defining fair lending. The policy statement says, "All...employees are to be knowledgeable of and comply with existing fair lending related laws and requirements." Employees are also asked to keep current on fair lending requirements and discuss any concerns with management.

Monitor Fair Lending Performance

Recommended Activities

Involve senior-level management in developing, implementing, and monitoring goal of minority lending.

Lender's Activities

Company staff, from the president on down, all said that any race-specific monitoring would compromise the lender's race-blind mortgage application process.

Compensate Staff in a Manner Consistent with Fair Lending Objectives

Recommended Activities

Provide compensation practices that reward, or at least do not indirectly penalize, employees working to foster minority lending.

Lender's Activities

The loan counselors and processors receive base salaries along with a small commission based on the dollar volume of loans they help to close. The other employees receive a base salary and a profit-sharing bonus. These compensation policies do not discourage employees from working on loans that are small and/or take time to process, but they are not explicitly linked to employees' fair lending performance.



Develop a Diverse Workforce

Recommended Activities

Practice staff recruitment and promotion to foster minority lending through the racial diversity of employees.

Lender's Activities

At the time of the research team's site visit, none of the lender's loan origination staff were black, although two of the six loan counselors were Hispanic. All of the staff with whom we met except one heard about their current job through a personal contact. The exception was one Hispanic loan counselor, who was hired after responding to an advertised solicitation for a Spanish-speaking loan counselor. He said he was interested in working for the lender because its compensation was in the form of a straight salary, rather than commissions.

Provide Training in Fair Lending

Recommended Activities

Provide staff training in multicultural interactions generally and fair lending practices specifically.

Lender's Activities

The lender does not provide multicultural training or any specific training in fair lending practices. The lender uses "on-the-job" training for its employees. Newly hired loan counselors, with little previous experience, work in the processing department first, and then observe loan application interviews. After this stage, newly hired loan counselors conduct initial application interviews jointly with the branch manager. When considered ready, they are allowed to take applications without direct supervision. New employees also receive a procedures manual for the company, which contains the fair lending statement discussed above. However, most of the staff interviewed by the research team had not read the statement, although they were able to define fair lending.

Employees hired within the last year received general customer service training conducted by a consulting company hired by the lender's owner. This training took place in three half-day sessions. Attendees received training about treating customers courteously and putting people at their ease. However, as one employee put it, "The training didn't say 'treat black people like this, white people like this, and Hispanics like this.'"

Conduct Outreach to Minorities

Recommended Activities

Work with third parties committed to fostering minority lending.



THE URBAN
INSTITUTE

Lender's Activities

Almost all of the lender's customers are referred by a sales representative of the builder who owns the lender. The research team does not know whether the builder is making formal outreach efforts to minorities.

Self-Monitor Fair Lending Performance

Recommended Activities

Systematically test for fairness, so that at all stages of the lending process minorities are treated equally.

Lender's Activities

The lender does not monitor its fair lending performance. All of the employees said they never receive any reports or information about how the lender treats minorities. The lender's staff saw no reason to track outcomes by race and said such information would take away from its race-blind origination process.

Designate an Employee to Receive Fair Lending Complaints

Recommended Activities

Appoint an ombudsman or comparable official to receive complaints from customers.

Lender's Activities

All customer complaints, including those alleging discrimination, are directed to the lender's president. The president is responsible for follow-up and resolution of complaints. In addition, loan recipients are asked to complete a customer satisfaction survey after the closing that asks about their experience with the process. The lender has a sample of these surveys in a loose-leaf binder for review in the lobby. Only customers who receive mortgages provide feedback to the company, so it is not surprising that these surveys contain highly favorable comments.

The lender has not instituted a number of strategies recommended to promote fair lending. Two of the most serious omissions by the lender are: (1) a complete lack of internal tracking and monitoring by the lender of its fair lending performance; and (2) no specific fair lending training for staff. While respondents said they treat all customers in the same manner, none of the respondents said they receive information that verifies that they are, in fact, conducting business in a race-blind manner. In addition, the president said the company's monitoring system is incapable of identifying instances in which staff inadvertently discriminate against minority applicants, although overtly negative comments entered into the company's electronic application system would generate attention. In a follow-up interview, the president told us that she had given more thought to the question about inadvertent discrimination, and had come to the conclusion that the lender's internal controls would not detect such behavior.



She hastened to add, however, that nobody employed by the lender would inadvertently discriminate, because everybody who works for the lender is committed to fair lending.

The lender has received complaints about discriminatory treatment, and is currently under investigation in connection with a suit filed by a minority applicant who was denied a mortgage. The president said she occasionally receives a phone call from a customer who claims he or she was denied due to race. The president said she reviews the loan application files to see why the customer's application was denied. In all cases, she said, the loan applications contained serious problems, such as a poor credit or income history.⁸ She said it upset her when allegations of discrimination were made because the company qualifies so many marginal applicants and has no reason to discriminate. Two other employees said that one of the discrimination suits was filed by a black customer who was working with a black loan counselor. This story was related to the research team to indicate the weakness of discrimination suits filed against the company.

The lender does not provide any specific fair lending training, although most employees could define fair lending. Employees said fair lending meant that "You treat everybody equally." The lender's staff said fair lending training was not necessary because discrimination is wrong, and they would never do it.

Implementation Strategies

The research team asked representatives of CLC Compliance Technologies, Inc. (a consulting firm that provides technical assistance to lenders seeking to improve fair lending performance), to identify specific strategies the lender should implement to assess whether or not the denial discrepancies result from differential treatment.⁹

CLC Compliance Technologies staff have several recommendations for the lender. First, the lender's quality assurance staff should expand its activities to include fair lending enforcement. Currently, the lender's quality assurance staff ensure that loan applications contain accurate documentation. CLC Compliance Technologies staff said they should also review origination decisions for similarly qualified applicants of different races. A staff member should be assigned to identify a number of loan applications that are similar in income, credit history, front- and back-end ratios, and so forth, but differ by race. He/she should then compare the number and types of conditions transmitted to marginal white, black, and Hispanic applicants and look for systematic differences. If differences exist and sample size is large enough, it is worthwhile assessing whether particular staff members have systematic disparities in treatment of similar files.

To the extent that borrowers are fairly randomly assigned to loan counselors, the lender can identify loan counselors with potential problems by computing DDIs by loan counselor for each subdivision where he or she has dealt with a substantial number of loans. A loan counselor with consistently high DDIs should then be given further scrutiny to determine the sources of the persistent racial gaps.

Second, the lender should institute fair lending training for its staff. This training would focus attention on the subtle ways in which white employees



THE URBAN
INSTITUTE

can discriminate against minority applicants. Annex B (which appears at the end of this chapter) outlines the topics covered in a fair lending training seminar offered by CLC Compliance Technologies. This training seminar includes discussions of scenarios where differential treatment inadvertently occurs and proposes actions to avoid such behavior. For example, seminar participants analyze the following scenario in which two applicants call one lender. In the first case, the teleservice representative greets the applicant in the usual way, then mentions that he lives in Forsythe County, Georgia, where the teleservice representative grew up. The representative proceeds to take the phone application while striking up a great conversation about people they both know. After the representative collects all the information, the applicant asks, "Well, how do I look? Will I get the loan?" The representative responds, "You need to get a cosigner so you can show more income. Or, you could reduce your monthly expense by \$150 a month." The applicant responds, "Great! My brother lives with me, he can cosign." The loan was approved. In the second case, a similarly qualified applicant from central city Atlanta calls, is greeted in the usual way, has the information taken, and has the application denied without even being notified.¹⁰ Although employees may not believe such a scenario represents outright discrimination, it does underscore for them the subtle nature of differential treatment that may occur in the loan origination process.

Third, the company should develop quantitative fair lending objectives. Previous CLC Compliance Technologies customers have committed to reducing minority denial rates, and the consultants recommend that the company create similar measurable goals. Once these goals are developed, the lender should use internal data, as well as HMDA data, to assess its progress in meeting these goals. Such a system would provide the lender with the means to measure its fair lending performance relative to its stated quantitative goals.

Finally, the lender should hire white, black, and Hispanic "mystery shoppers" to see if the lender's origination process is, in fact, race-blind. These mystery shoppers would go to different subdivisions and pose as interested buyers. The lender would be able to use the results of these audits to see if minority customers are receiving differential treatment by the builder's sales representative or the lender's loan counselor.

According to the consulting firm, none of these recommended strategies is very complex or costly for the lender to implement, given its size and origination volume. Moreover, the results of these activities would allow the lender to detect the presence of differential treatment and take effective measures to reduce the possibility it will occur in the future.

Implementation Steps

A number of steps would be needed for a lender to implement changes such as these. For each recommended change, we briefly identify advantages and likely obstacles.

Identify whether a problem exists. Before starting any change process, management and front-line employees must understand that a problem exists.

There is a "chicken and egg" problem concerning the data-gathering needed to identify whether a lender discriminates. To the extent a lender does not



believe it has a problem, it has no incentive to gather more data or perform more analysis on existing data. In addition, organizations face disincentives to gather data because any findings might be subpoenaed or leaked to the press. (Subpoenas are most likely not a problem if a lender's law firm carries out the analysis, because then the research may be protected by lawyer-client privilege. This alternative does raise costs.) At the same time, without detailed data it is difficult to identify discrimination.

Fortunately, HMDA data that present race-specific denial rates by income can be used for a first look; this approach has low costs and the data are based on public information. The first set of suggestions noted above, looking at differential treatment of very similar files, and sending out "mystery shoppers" that differ by race, can indicate or rule out race as a factor in explaining differential DDIs. Even if no problem is identified, the lender's large HMDA data discrepancy makes it prudent for management to institutionalize some ongoing monitoring. Only then can the lender be sure discrimination does not arise, reduce the risks of lawsuits, and be prepared to counter bad publicity based on HMDA data.

Create the case for change. If the more detailed analyses do indicate a problem with (presumably inadvertent) discrimination, management must take the next step of making a business case for change.

The need for a business case is not part of the typical "best practice" advice on fair lending (e.g., Listokin and Wyly 1998, Vartanian et al. 1995). However, fair lending (or other) changes proposed without a business case are likely to be viewed as something alien to good business and "tacked on" to an otherwise profit-maximizing concern. That is, many employees and managers may feel that fair lending initiatives are implemented for "feel-good" reasons. They will expect the initiatives to hurt their measured bottom-line performance and not to last very long. Such employees and managers are likely merely to go through the motions of compliance. Even if the changes are tied to compensation, they will be perceived as not likely to remain after the sponsoring executive moves on.

The business case must clearly link the objective data on differential treatment to three issues—law, ethics, and profits—and must emphasize that discrimination is illegal and is perceived as unethical by management. It must also make the point that discrimination increases the lender's potential to be sued or face (deserved) bad publicity. Finally, it must highlight that discrimination can lead the lender to deny loans to qualified applicants and thus lose business.

The business case must then identify actions to address the problems identified. A sensible starting point is to go through a list of fair lending practices (e.g., the list in Listokin and Wyly 1998, summarized above) and identify policies that are both cost-effective and tied to the problems identified in the self-analysis. The proposals discussed above for the case study lender to increase fair lending training and to create fair lending goals are examples of such changes.

Ideally, the business case would be based on objective data describing the situation and clearly explaining the discrimination findings. The key is not to place blame but to create a shared understanding of the problem. Unfortunately, a lender that circulates its own analysis indicating likely discrimination increases the likelihood that it will be sued. Thus, the lender may need to



THE URBAN
INSTITUTE

describe the situation on paper in terms that are both vague and less severe than would be ideal, other things equal.

At a minimum, circulating the HMDA data internally should alert employees to the possibility and the appearance of discrimination. Unfortunately, because employees will be able to identify the weaknesses of HMDA data and many employees will discount “feel-good” messages that diversity will increase profits, the business case will probably be weaker than if it could lay out the true scope of the problem.

Create an integrated fair lending plan. A recurring theme in the literature on organizational change is the necessity of creating a coherent strategy, whereby the various management policies of the organization reinforce each other. (See, e.g., Milgrom and Roberts 1995 for theoretical exposition, and the literature review in Ichniowski et al. 1996 for empirical evidence concerning one set of human resource practices.) The standard fair lending best practice list follows this precept by including a coherent list of practices to support fair lending. Within the firm the emphasis includes: management leadership; training that both increases awareness of problems and gives skills to address them; pay systems that reinforce fair behaviors; and information systems that provide feedback on fair lending outcomes to both employees and their managers. Policies that interact with the environment of the enterprise also support the fair lending goal: recruitment of a diverse workforce; partnerships with organizations that can assist in achieving fair lending goals; and learning from customers (e.g., with an impartial complaint system).

Two additional sets of policies that are not in the Listokin and Wyly checklist of good practices also appear in the standard literature on organizational change. First, many successful change efforts involve front-line employees in identifying and solving problems. In many organizational change efforts, front-line employees, such as loan counselors, have insights into why the data turn out the way they do. Employees often have insights into successful strategies for change as well (Levine 1995). Finally, any change effort will involve changes in the behavior of front-line staff. They are more likely to implement the changes if they perceive that their views were incorporated into the change effort.

Second, successful change efforts often include means for organizational learning across subunits performing similar tasks. Unlike the case study above, the majority of mortgages are originated by very large lenders or mortgage companies, not a local lender. Such organizations face the challenge of implementing good practices in many different workplaces, while still retaining flexibility to learn and to adapt to local conditions. Best practice companies will be those that identify means to give all employees the skills, technology, organizational structures (e.g., cross-functional teams), and incentives both to share their good ideas, and to learn and adapt good ideas from elsewhere in the organization (Gilbert and Levine 1998). In the fair lending setting, organizations should identify subunits with impressive and consistent success in lending to underrepresented groups, and benchmark their practices for use in other parts of the organization.

Implement new pay systems. As with any change effort, the move to promote fair lending will face many barriers. For example, if loans to minorities are disproportionately for smaller amounts and require more assistance than loans to white applicants, minority loan applicants may not receive the same treat-



ment as whites for that reason alone. Consistent with profit maximization, pay systems at many lenders reward the dollar volume of loans. Such a practice aligns incentives with a lender's costs and benefits of loans. A fair lending initiative that weakens the relationship between total lending amounts and compensation will face problems in the short run from employees who were compensated highly based on a high volume of loans processed and whose expected compensation will be cut if the fair lending initiative is implemented.

Some lenders put only a subset of their loan officers on salary or on lower-stakes commissions. This avoids the demotivating effects of cuts in expected compensation. But it ensures that low-income loans are ghettoized into a subset of lender employees. Moreover, because a pay system such as salary or paying per loan (rather than compensation based on loan size) does not align employee incentives with profit maximization, it may present particular challenges to implement.

Change decisionmaking. Fair lending programs often involve “second look” and other additional reviews for marginal mortgage applications, especially those of minorities. These programs can be quite important if, as some recent research suggests, marginal minority borrowers receive less attention than marginal whites (Yinger 1995). At the same time, multiple reviews of loans that front-line employees feel should be denied can reduce front-line employees’ perceived power. This can increase their resistance to change. Permitting front-line employees more flexibility in finding creative ways to show ability to repay a loan can increase front-line employees’ decisionmaking power. In this case, however, the reduced power of those who established the (formerly) rigid guidelines can lead to resistance to change. More generally, employees may feel that “fair lending” is mainly just more oversight by uninformed outsiders.

Individual-specific monitoring of denial rates by race and incentives for lending to underrepresented groups can lead to resentment among front-line employees, particularly if they feel pressure to meet a (reverse) discriminatory quota of loans. This resentment can be mitigated by technology that provides employees with feedback on their fair lending performance and the skills and other tools needed to identify the sources of any problems and implement effective solutions. In this case, fair lending initiatives can be integrated with a general effort to increase front-line empowerment. That fair lending goals and monitoring can be either oppressive or empowering is an instance of a more general capability of technology to be used in ways either that automate decisions or that create tools for front-line employees to use (Zuboff 1984).

Implement fair lending and diversity training. The fair lending program should be tied to the business plan of the lending institution. Thus, training should be presented in terms of satisfying the needs of a diverse clientele and complying with existing fair lending laws. Each initiative is more likely to succeed if it is tied to making good loans than if it is solely motivated by legal or “do-gooder” perspectives. At the same time, programs that are not tied to profitability goals typically do not flourish in organizations. To the extent that a lack of understanding of other cultures is a barrier to judging marginal minority applicants accurately, the cost of serving these populations is actually higher. Thus, it is appropriate for lenders to pay for fair lending training of employees as well as homebuyer counseling



THE URBAN
INSTITUTE

and education for applicants who may not be familiar with the mortgage lending process (Longhofer 1995).

Recruit and promote a diverse workforce. A lender can enhance its fair lending performance by practicing staff recruitment and promotion to create a racially diverse workforce. It may be that seeing an all-white office may make minority borrowers feel unwelcome. (Kim and Squires [1995] provide some evidence of lower minority denial rates at banks with more minority employees.) At the same time, recruiting people based on the color of their skin, even if it makes customers more comfortable, is not always allowed by law. That is, customer discrimination is not an excuse for an employer to discriminate on the basis of race or ethnicity in hiring or promotions. Particularly for language minorities, having bilingual staff can be helpful. Fortunately for diversity efforts, hiring based partly on language skills is legal. Interestingly, employees at the lender in this case study claimed language was not a barrier even for the many borrowers born in other countries. Non-English speakers almost always came to a branch with a bilingual friend or relative.

Policy and Research Conclusions

The employees of the lender analyzed in this case study strongly believe that they participate in a race-blind origination process. Moreover, the research team, unaware of the lender's HMDA data at the time of the field visit, also believed the lender was acting in a race-blind manner based on its discussions with employees and review of the lender's origination process. The HMDA data, however, reveal that origination outcomes are different for whites, blacks, and Hispanics—differences that are not eliminated by controlling for the applicant's income or type of loan. This outcome suggests that the lender may not be acting in a race-blind manner, although other explanations are possible.

Although the lender denies only a relatively small proportion of minority applications, it denies an even smaller proportion of white applications. This may result from the lender's staff making greater efforts to qualify marginal white applicants compared with marginal black and Hispanic applicants. Because the lender does not monitor its fair lending performance, it is impossible to rule out the possibility that staff are inadvertently treating whites differently from minority customers.

These results indicate that lenders cannot simply assume they are using a race-blind origination process. It may be that all of the racial differences in origination outcome result from lower income, poorer credit, and other factors among minority applicants that predict loan payment performance. If so, none of the racial differences constitutes differential treatment. However, the lender's lack of a monitoring system precludes ruling out the possibility of differential treatment. The only way the case study lender could distinguish between outcomes that result from valid underwriting standards and those from differential treatment is to conduct a side-by-side review of origination outcomes for similarly qualified minority and white applicants. Such a review should be supplemented with more extensive fair lending training for staff and an enhanced internal monitoring and control system that tracks outcomes and reports information about racial differences in origination outcomes.



It cannot be assumed that fair lending can only be enhanced through voluntary lender self-assessments. As discussed earlier, institutional inertia will prevent many lenders from undertaking strategies that promote fair lending. In addition, some lenders have no idea that they may be treating minorities differently from whites. Therefore, regulators must ensure that lenders comply with fair lending laws and provide data to lenders in a form that assists lenders in monitoring fair lending performance.

At a minimum, lenders should be given HMDA data that show the company's denial rates for different types of borrowers as well as racial denial rate discrepancies. Lenders should also receive the same information for other lenders in the MSA in order to provide a sense of their relative performance. The data analysis software used in this case study, HMDAware®, could be provided to lenders, perhaps on the internet, in order to allow for analyses of HMDA data, and so provide an opportunity for all lenders to start evaluating their fair lending performance.

These data, however, only provide information about the possibility that a lender is not treating minorities the same as white applicants. As discussed earlier, implementing organizational change is difficult and will be a challenge for most companies. Therefore, lenders may require substantial technical assistance from HUD, as well as other organizations and agencies, to identify and implement new managerial processes.

Notes

1. The lender was one of eight companies invited to participate in the study. Lenders were chosen because they had processed a minimum of 100 minority applications in a recent year. Such lenders were then categorized as having a high or low Denial Disparity Index, a measure used by CLC Compliance Technologies that measures the extent to which minority applicants are more likely than whites to be denied a loan. Each lender was contacted in writing and by telephone by a member of the research team. Some lenders gave no reason for their decision not to participate. Other lenders said they could not dedicate staff time to participate in discussions with the research team. Future research projects based on in-depth case studies of lenders will have to be designed cognizant of the potential for poor cooperation among potential respondents.
2. The lender reports origination outcome data for HMDA. Since the lender is a mortgage company, however, it is not subject to CRA regulations.
3. Other important underwriting factors, such as credit history, are not included in this analysis, however.
4. We omit the actual year in order to protect the lender's anonymity.
5. The Denial Disparity Index is a measure of loan outcomes developed by CLC Compliance Technologies and used with permission.
6. Some attribute such behavior to a "cultural affinity" between white lending staff and white customers (Hunter and Walker 1996). As a result, white loan officers and underwriters are more likely to help marginal white applicants qualify for a mortgage (Lindsey 1997). Another body of literature, however, posits that affinity can develop between people of different races because such connections are not solely based on race (Harrison 1991).
7. Indeed, Listokin and Wyly (1998, p. 23) identified 12 such studies, issued by organizations ranging from the Neighborhood Reinvestment Coalition to the Federal Reserve Bank of Boston to HUD. They argue "[t]here is now a broad body of knowledge in the lending indus-



THE URBAN
INSTITUTE

try and among regulators regarding strategies that are successful in expanding the availability of mortgage credit to traditionally underserved markets.”

8. The president, however, does not look for applications with similar characteristics filed by whites to see if such applications were approved. Therefore, the president’s review does not rule out the possibility that white marginal applicants receive different treatment.
9. All of the HMDA analyses presented in this case study were conducted using CLC Compliance Technologies’ HMDAware® software package. In addition, CLC Compliance Technologies staff assisted the research team in developing discussion guides used during the site visit, reviewed the research team’s field notes (with identifying information removed), and were briefed by the research team after the site visits were completed.
10. CLC Compliance Technologies.



References

- Ambrose, Brent, William Hughes, and Patrick Simmons. 1995. "Policy Issues Concerning Racial and Ethnic Differences in Home Loan Rejection Rates." *Journal of Housing Research* 6 (1):115–136.
- Amemiya, Takeshi. 1974. "Multivariate Regression and Simultaneous Equations When the Dependent Variables Are Truncated Normal." *Econometrica* 42: 999–1012.
- Arrow, Kenneth. 1973. "The Theory of Discrimination." In *Discrimination in Labor Markets*, edited by Orley Ashenfelter and Albert Rees (Princeton: Princeton University Press), pp. 3–42.
- Avery, Robert B., and Patricia E. Beeson. 1998. "Neighborhood, Race, and the Mortgage Market: Evidence from HMDA Data." Paper presented at the annual AEA/NEA meetings. Chicago, January.
- Avery, Robert B., Patricia E. Beeson, and Paul D. Calem. 1996. "Using HMDA Data as a Regulatory Screen for Fair Lending Compliance." *Journal of Financial Services Research* 11 (February/April): 9–42.
- Avery, Robert B., Patricia E. Beeson, and Mark S. Sniderman. 1996. "Accounting for Racial Differences in Housing Credit Markets." In *Mortgage Lending, Racial Discrimination, and Federal Policy*, edited by John Goering and Ron Wienk (Washington: Urban Institute Press), pp. 75–142.
- Avery, Robert B., Raphael W. Bostic, Paul S. Calem, and Glenn B. Canner. 1998. *Credit Scoring: Issues and Evidence from Credit Bureau Files*. Washington: Board of Governors of the Federal Reserve System.
- Ayres, Ian, and Peter Siegelman. 1995. "Race and Gender Discrimination in Bargaining for a New Car." *American Economic Review* 85 (June): 304–321.
- Becker, Gary S. 1993. "The Evidence Against Banks Doesn't Prove Bias." *Business Week*, April 19, p. 18.
- _____. 1971. *The Economics of Discrimination*, 2d ed. Chicago: University of Chicago Press.

- Berkovec, James A., Glenn B. Canner, Stuart A. Gabriel, and Timothy H. Hannan. 1998. Discrimination, Competition, and Loan Performance in FHA Mortgage Lending. *Review of Economics and Statistics* 80: 241–250.
- _____. 1996. “Response to Critiques of ‘Mortgage Discrimination and FHA Loan Performance’.” *Cityscape: A Journal of Policy Research* 2 (1): 49–54.
- _____. 1994. “Race, Redlining, and Residential Mortgage Loan Performance.” *Journal of Real Estate Finance and Economics* 9 (3): 263–294.
- Berkovec, James A., and Peter Zorn. 1996. “How Complete Is HMDA? HMDA Coverage of Freddie Mac Purchases.” *Journal of Real Estate Research* 11: 39–56.
- Black, Harold A. 1995. “HMDA Data and Regulatory Inquiries Regarding Discrimination.” In *Fair Lending Analysis: A Compendium of Essays on the Use of Statistics*, edited by Anthony M. Yezer (Washington: American Bankers Association), pp. 147–154.
- Black, Harold A., M. Cary Collins, Ken B. Cyree. 1997. “Do Black-Owned Banks Discriminate against Black Borrowers?” *Journal of Financial Services Research* 11 (February/April): 189–204.
- Bostic, Raphael W., and Glenn B. Canner. 1997. “Do Minority-Owned Banks Treat Minorities Better? An Empirical Test of the Cultural Affinity Hypothesis.” Unpublished manuscript, December.
- Boyes, William J., Dennis L. Hoffman, and Stuart A. Low. 1989. “An Econometric Analysis of the Bank Credit Scoring Problem.” *Journal of Econometrics* 40: 3–14.
- Bradford, Calvin. 1979. “Financing Home Ownership: The Federal Role in Neighborhood Decline.” *Urban Affairs Quarterly* 14: 313.
- Brimelow, Peter. 1993. “Racism at Work?” *National Review*, April 12, p. 42.
- Brimelow, Peter, and Leslie Spencer. 1993. “The Hidden Clue.” *Forbes*, January 4, p. 48.
- Browne, Lynn E., and Gregory M. Tootell. 1995. “Mortgage Lending in Boston: A Response to the Critics.” *New England Economic Review* (September/October): 53–78.
- Buist, Henry, Peter Linneman, and Isaac F. Megbolugbe. 1997. “Residential Lending Discrimination and Lender Compensation Policies.” Unpublished manuscript.
- CLC Compliance Technologies, Inc. *CLC Compliance Technologies Fair Lending Training Program outline*. [<http://www.compliancetech.com/floutln.htm>].
- Calomiris, Charles W., Charles M. Kahn, and Stanley Longhoter. 1994. “Housing Finance Intervention and Private Incentives: Helping Minorities and the Poor.” *Journal of Money and Banking* 26 (3) August: 634–74.
- Canner, Glenn B., and Wayne Passmore. 1994. “Residential Lending to Low-Income and Minority Families: Evidence from the 1992 HMDA Data.” *Federal Reserve Bulletin* (February): 79–108.



- Carr, James H., and Isaac F. Megbolugbe. 1993. "The Federal Reserve Bank of Boston Study on Mortgage Lending Revisited." *Journal of Housing Research* 4 (2): 277–314.
- Courchane, Marsha, and David Nickerson. 1997. "Discrimination Resulting from Overage Practices." *Journal of Financial Services Research* 11 (1997): 133–152.
- Crawford, Gordon W., and Eric Rosenblatt. 1997. "Differences in the Cost of Mortgage Credit: Implications for Discrimination." Unpublished manuscript.
- Day, T., and Stanley J. Liebowitz. 1996. "Mortgages, Minorities, and HMDA." Paper presented at the Federal Reserve Bank of Chicago, April.
- Federal Financial Institutions Examination Council. 1998. "Press Release, August 6, 1998." [<http://www.ffiec.gov/pr080698.htm>].
- Ferguson, Michael F., and Stephen R. Peters. 1995. "What Constitutes Evidence of Discrimination in Lending?" *Journal of Finance* 50: 739–748.
- Fix, Michael, George Galster, and Raymond Struyk. 1993. "An Overview of Auditing for Discrimination." In *Clear and Convincing Evidence*, edited by Michael Fix and Raymond Struyk (Washington: Urban Institute Press), pp. 1–68.
- Gabriel, Stuart A. 1996. "The Role of FHA in the Provision of Credit to Minorities." In *Mortgage Lending, Racial Discrimination, and Federal Policy*, edited by John Goering and Ron Wienk (Washington: Urban Institute Press), pp. 183–250.
- Galster, George. 1993. "Use of Testers in Investigating Discrimination in Mortgage Lending and Insurance." In *Clear and Convincing Evidence*, edited by Michael Fix and Raymond Struyk (Washington: Urban Institute Press), pp. 287–334.
- Galster, George C. 1996. "Comparing Loan Performance between Races as a Test for Discrimination." *Cityscape: A Journal of Policy Development and Research* 2 (2): 33–40.
- _____. 1993. "The Facts of Lending Discrimination Cannot Be Argued Away by Examining Default Rates." *Housing Policy Debate* 4 (1): 141–146.
- Gilbert, April, and David I. Levine. 1998. "Knowledge Transfer: Managerial Practices Underlying One Piece of the Learning Organization." COHRE briefing paper, University of California at Berkeley [<http://socrates.berkeley.edu/~iir/cohre/knowledge.html>].
- Glennon, Dennis, and Mitchell Stengel. 1995. "Evaluating Statistical Models of Mortgage Lending Discrimination: A Bank-Specific Analysis." Economic and Policy Analysis Working Paper 95-3. Washington: Office of the Comptroller of the Currency.
- Glennon, Dennis, and Mitchell Stengel. 1994. "An Evaluation of the Federal Reserve Bank of Boston's Study of Racial Discrimination in Mortgage Lending." Economic and Policy Analysis Working Paper 94-2. Washington: Office of the Comptroller of the Currency.
- Goering, John M., ed. 1996. "Race and Default in Credit Markets: A Colloquy." *Cityscape: A Journal of Policy Development and Research* 2 (1).



THE URBAN
INSTITUTE

- Goering, John M., and Ron Wienk, eds. 1996. *Mortgage Lending, Racial Discrimination, and Federal Policy*. Washington: Urban Institute Press.
- Goldberger, Arthur D. 1984. "Reverse Regression and Salary Discrimination." *Journal of Human Resources* 18 (3): 294–318.
- Greene, William. 1993. *Econometric Analysis*, 2d ed. New York: Macmillan.
- Harrison, Faye. 1991. "Ethnography as Politics." In *Decolonializing Anthropology: Moving Further toward an Anthropology of Liberation*, edited by Faye Harrison (Washington: Association of Black Anthropologists), pp. 88–109.
- Heckman, James H. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator of Such Models." *Annals of Economic and Social Measurement* 5: 475–92.
- Heckman, James, and Peter Siegelman. 1993. "The Urban Institute Audit Studies: Their Methods and Findings." In *Clear and Convincing Evidence*, edited by Michael Fix and Raymond Struyk (Washington: Urban Institute Press), pp. 187–258.
- Hendershott, Patrick, W. LaFayette, and David Haurin. 1997. "Debt Usage and Mortgage Choice: The FHA-Conventional Decision." *Journal of Urban Economics* 41: 202–217.
- Herzog, J.P., and J.S. Earley. 1970. *Home Mortgage Delinquency and Foreclosure*. New York: National Bureau or Economic Research.
- Horne, David K. 1997. "Mortgage Lending, Race, and Model Specification." *Journal of Financial Services Research* 11 (February/April): 43–68.
- _____. 1994. "Evaluating the Role of Race in Mortgage Lending." *FDIC Banking Review* (Spring/Summer): 1–15.
- Hunter, William C., and Mary Beth Walker. 1996. "The Cultural Affinity Hypothesis and Mortgage Lending Decisions." *Journal of Real Estate Finance and Economics* 13 (July): 57–70.
- Ichniowski, Casey, et al. 1996. "What Works at Work: A Critical Review." *Industrial Relations* 35, (3): 299–333.
- Judy, Nancy Ness. 1996. "Fair Lending: Telling Your Story through the Media." *ABA Bank Compliance* 17 (11): 14–18.
- Kim, Sunwoong, and Gregory Squires. 1995. "Lender Characteristics and Racial Disparities in Mortgage Lending." *Journal of Housing Research* 6 (1): 99–113.
- Kohn, Earnest, Cycl E. Foster, Bernard Kaye, and Nancy J. Terris. 1992. "Are Mortgage Lending Policies Discriminatory? A Study of 10 Savings Banks." New York State Banking Department, March.
- LaCour-Little, Michael. 1998. "Application of Reverse Regression to Boston Federal Reserve Data Refutes Claims of Discrimination." *Journal of Real Estate Research* 11: 1–12.

- _____. 1996. "Race Differences in Demand and Supply of Mortgage Credit." Unpublished manuscript.
- Ladd, Helen F. 1998. "Evidence on Discrimination in Credit Markets." *Journal of Economic Perspectives* 12 (Spring): 41–62.
- Lang, William W., and Leonard I. Nakamura. 1993. "A Model of Redlining." *Journal of Urban Economics* 33 (March): 223–234.
- Laitila, Thomas. 1993. "A Pseudo-R² Measure for Limited and Qualitative Dependent Variable Models." *Journal of Econometrics* 53 (3): 341–356
- Levine, David I. 1995. *Reinventing the Workplace: How Business and Employees Can Both Win*. Washington: Brookings Institution.
- Liebowitz, Stanley. 1993. "A Study That Deserves No Credit." *Wall Street Journal*, September 1, p. A14.
- Lindsey, Debby. 1997. "The Mortgage Industry's Attitude on Race, Culture, and Discrimination in Lending," Paper presented at the annual meeting of the Eastern Economic Association, Arlington, Va., April.
- Lindsey, Lawrence B. 1995. "Foreword." In *Fair Lending Analysis: A Compendium of Essays on the Use of Statistics*, edited by Anthony M. Yezer (Washington: American Bankers Association), pp. ix–xiv.
- Ling, David C., and Susan M. Wachter. 1998. "Information Externalities and Home Mortgage Underwriting." *Journal of Urban Economics* 44 (November): 317–332.
- Listokin, David, and Elvin Wyly. 1998. *Successful Lending Industry Strategies, Volume 1: Successful Strategies* (New Brunswick, N.J.: Center for Urban Policy Research).
- Longhofer, Stanley D. 1996a. "Cultural Affinity and Mortgage Discrimination." *Economic Review* 32 (3): 12–24.
- _____. 1996b. "Discrimination in Mortgage Lending: What Have We Learned?" *Economic Commentary* (Federal Reserve Bank of Cleveland), August 15.
- _____. 1995. "Rooting Out Discrimination in Home Mortgage Lending." *Economic Commentary* (Federal Reserve Bank of Cleveland): 1–4.
- Maddala, George S., and Robert P. Trost. 1982. "On Measuring Discrimination in Loan Markets." *Housing Finance Review* 1: 245–268.
- Meyers, Samuel L., Jr., and Tsze Chan. 1995. "Racial Discrimination in Housing Markets: Accounting for Credit Risk." *Social Science Quarterly* 76 (September): 543–561.
- Milgrom, Paul, and John Roberts. 1995. "Complementarities and Fit: Strategy, Structure, and Organizational Change in Manufacturing." *Journal of Accounting and Economics* 19 (2, 3): 179–208.



THE URBAN
INSTITUTE

- Munnell, Alicia H., Geoffrey M.B. Tootell, Lynn E. Browne, and James McEneaney. 1996. "Mortgage Lending in Boston: Interpreting HMDA Data." *American Economic Review* 86 (March): 25–53.
- Peterson, Richard L. 1981. "An Investigation of Sex Discrimination in Commercial Banks' Direct Consumer Lending." *Bell Journal of Economics* 12 (2): 547–561.
- Phillips, Robert F., and Anthony M. Yezer. 1995. "Self Selection and the Measurement of Bias and Risk in Mortgage Lending: Can You Price the Mortgage If You Don't Know the Process?" Paper presented at the American Real Estate and Urban Economics Association meetings, Washington, January.
- Phillips-Patrick, Fred J., and Clifford V. Rossi. 1996. "Statistical Evidence of Mortgage Redlining? A Cautionary Tale." *Journal of Real Estate Research* 11: 13-24.
- Policy Statement on Discrimination in Lending. 1994. *Federal Register*, Vol. 59, no. 73. April 15, 1994: 18266–18274.
- Quercia, Roberto G., and Michael A. Stegman. 1992. "Residential Mortgage Default: A Review of the Literature." *Journal of Housing Research* 3: 341–380.
- Rachlis, Mitchell B., and Anthony M. Yezer. 1993. "Serious Flaws in Statistical Tests for Discrimination in Mortgage Markets." *Journal of Housing Research* 4 (2): 315–336.
- Ritter, Richard. 1996. "The Decatur Federal Case: A Summary Report." In *Mortgage Lending, Racial Discrimination, and Federal Policy*, edited by John Goering and Ron Wienk (Washington: Urban Institute Press), pp. 445–450.
- Rodda, David, and James E. Wallace. 1996. "Fair Lending Management: Using Influence Statistics to Identify Critical Mortgage Loan Applications." In *Mortgage Lending, Racial Discrimination, and Federal Policy*, edited by John Goering and Ron Wienk (Washington: Urban Institute Press), pp. 531–560.
- Rosenblatt, Eric. 1997. "A Reconsideration of Discrimination in Mortgage Underwriting with Data from a National Mortgage Bank." *Journal of Financial Services* 11: 109–132.
- Ross, Stephen L. 1998. "Mortgage Lending, Sample Selection, and Default." Working paper, University of Connecticut.
- Ross, Stephen L., and Geoffrey M.B. Tootell. 1998. "Redlining, the Community Reinvestment Act, and Private Mortgage Insurance." Unpublished manuscript.
- _____. 1997. "Mortgage Lending Discrimination and Racial Differences in Loan Default: A Simulation Approach." *Journal of Housing Research* 8 (2): 277–297.
- _____. 1996a. "Mortgage Lending Discrimination and Racial Differences in Loan Default." *Journal of Housing Research* 7 (1): 117–126.
- _____. 1996b. "Flaws in the Use of Loan Defaults to Test for Mortgage Lending Discrimination." *Cityscape: A Journal of Policy Development and Research* 2 (1): 41–48.
- Schafer, Robert, and Helen F. Ladd. 1981. *Discrimination in Mortgage Lending*. Cambridge: MIT Press.



- Schill, Michael H., and Susan M. Wachter. 1993. "A Tale of Two Cities: Racial and Ethnic Geographic Disparities in Home Mortgage Lending in Boston and Philadelphia." *Journal of Housing Research* 4 (2): 245–275.
- Schwemm, Robert G. 1992. *Housing Discrimination: Law and Litigation*. Deerfield, Ill.: Clark, Boardman, Callaghan.
- Sickles, Robin C., and Peter Schmidt, 1978. "Simultaneous Equation Models with Truncated Dependent Variables: A Simultaneous Tobit Model." *Journal of Economics and Business* 31: 11–21.
- Siegelman, Peter. 1998. "Racial Discrimination in 'Everyday' Commercial Transactions: What Do We Know, What Do We Need to Know, and How Can We Find Out?" Paper presented at the Urban Institute Conference on Testing for Racial and Ethnic Discrimination in American Economic Life. Washington, March.
- Simmons, Pat. 1997. *Housing Statistics of the United States*. Lanham, Md.: Bernan Press.
- Siskin, Bernard R., and Leonard A. Cupingood. 1996. "Use of Statistical Models to Provide Statistical Evidence of Discrimination in the Treatment of Mortgage Loan Applicants: A Study of One Lending Institution." In *Mortgage Lending, Racial Discrimination, and Federal Policy*, edited by John Goering and Ron Wienk (Washington: Urban Institute Press), pp. 451–468
- Smith, Shanna, and Cathy Cloud. 1996. "The Role of Private, Nonprofit Fair Housing Enforcement Organizations in Lending Testing." In *Mortgage Lending, Racial Discrimination, and Federal Policy*, edited by John Goering and Ron Wienk (Washington: Urban Institute Press), pp. 589–610.
- Temkin, Kenneth, et al. 1998. *A Study of GSEs' Single Family Underwriting Guidelines*. Report prepared for U.S. Department of Housing and Urban Development (Washington: Urban Institute), October.
- Tootell, Geoffrey M.B. 1996a. "Redlining in Boston: Do Mortgage Lenders Discriminate against Neighborhoods?" *Quarterly Journal of Economics* 111: 1049–1079.
- Tootell, Geoffrey M.B. 1996b. "Turning a Critical Eye on the Critics." In *Mortgage Lending, Racial Discrimination, and Federal Policy*, edited by John Goering and Ron Wienk (Washington: Urban Institute Press), pp. 143–182.
- Turner, Margery Austin, Michael Fix, and Raymond Struyk. 1991. *Opportunities Denied, Opportunities Diminished*. Washington: Urban Institute Press.
- _____. 1996. "Turning a Critical Eye on the Critics." In *Mortgage Lending, Racial Discrimination, and Federal Policy*, edited by John Goering and Ron Wienk (Washington: Urban Institute Press), pp. 143–182.
- Van Order, Robert, and Peter Zorn. 1995. "Testing for Discrimination: Combining Default and Rejection Data." In *Fair Lending Analysis: A Compendium of Essays on the Use of Statistics*, edited by Anthony M. Yezer (Washington: American Bankers Association), pp. 105–112.



THE URBAN
INSTITUTE

- Vartanian, Thomas, Robert Ledig, Alisa Babitz, William Browning, and James Pitzer. 1995. *The Fair Lending Guide*. Little Falls N.J.: Glasser Legal Works.
- Williams, Alex O., William Beranek, and James J. Kenkel. 1974. "Default Risk in Urban Mortgages: A Pittsburgh Prototype Analysis." *AREUEA Journal* 2: 101–112.
- Yatchew, Adonis, and Zvi Griliches. 1985. "Specification Error in Probit Models." *Review of Economics and Statistics* 67 (1): 134–139.
- Yezer, Anthony M. 1995. "Editor's Notes." In *Fair Lending Analysis: A Compendium of Essays on the Use of Statistics*, edited by Anthony M. Yezer (Washington: American Bankers Association), pp. iii–vi.
- Yezer, Anthony M., R. Phillips, and R. 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* 9: 197–216.
- Yin, Robert. 1989. *Case Study Research: Design and Methods*. Newbury Park, Calif.: Sage Publications.
- Yinger, John. 1998. "Evidence on Discrimination in Consumer Markets." *Journal of Economic Perspectives* 12 (Spring): 23–40.
- _____. 1996. "Discrimination in Mortgage Lending: A Literature Review." In *Mortgage Lending, Racial Discrimination, and Federal Policy*, edited by John Goering and Ron Wienk (Washington: Urban Institute Press), pp. 29–74.
- _____. 1995. *Closed Doors, Opportunities Lost: The Continuing Costs of Housing Discrimination*. New York: Russell Sage Foundation.
- _____. 1986. "Measuring Discrimination with Fair Housing Audits: Caught in the Act." *American Economic Review* 76 (December): 881–893.
- Zandi, M. 1993. "Boston Fed's Study Was Deeply Flawed." *American Banker*, August 19, p. 13.
- Zuboff, Shoshana. 1984. *In the Age of the Smart Machine*. New York: Basic Books.



About the Editors

Margery Austin Turner directs the Metropolitan Housing and Communities Policy Center at the Urban Institute, where her research focuses on spatial and racial dimensions of anti-poverty policies within urban regions. Ms. Turner served as deputy assistant secretary for research at the Department of Housing and Urban Development from 1993 through 1996, where she focused HUD's research agenda on the problems of racial discrimination, concentrated poverty, and economic opportunity in America's metropolitan areas. She co-authored a national housing discrimination study, which used paired testing to determine the incidence of discrimination against African-American and Hispanic home-seekers, and applied the paired testing methodology to employment, producing the first direct statistical evidence of discrimination against young African-American jobseekers. Ms. Turner also directed research on racial and ethnic steering and conducted an exploratory study of differential real estate marketing practices in minority and white neighborhoods.

Felicity Skidmore is senior research editor for the Urban Institute, and has edited and produced a number of books on social policy reform. She served as the director of the Urban Institute Press from 1988 to 1998, supervising the writing, editing, and production of books, reports, and briefs for the Institute. She also coedited *Rethinking Employment Policy* with D. Lee Bawden (Urban Institute Press, 1989). Before joining the Urban Institute, Ms. Skidmore worked as a writing and manuscript consultant for a number of social research organizations, including Abt Associates, Mathematica Policy Research, and the National Institute of Child Health and Human Development.

About the Contributors

Michelle DeLair is a research assistant with the Urban Institute's Metropolitan Housing and Community Policy Center. She is currently assisting in studies on fair lending analysis and welfare geography analysis. Prior to joining the Urban Institute, Ms. DeLair served as an economic and financial analyst in the housing group at Price Waterhouse LLP, where she worked on econometric and financial cost models for Ginnie Mae, Fannie Mae, and Freddie Mac.

David Levine is an associate professor at the Haas School Business at the University of California, Berkeley. He is also editor of the journal *Industrial Relations*, associate director of the Institute of *Industrial Relations*, and director of Research at U.C. Berkeley's Center for Organization and Human Resource Effectiveness. His research examines how management policies, such as training and quality control, contribute to high-skill/high-performance workplaces and on the role and design of public policies in promoting such workplaces.

Diane Levy, a research associate at the Urban Institute's Metropolitan Housing and Community Policy Center, has over five years of experience in housing and community-related research and a background in social anthropology and city planning. Since joining the Urban Institute in 1998, she has built upon past experiences in housing and community economic development research by studying public housing desegregation efforts and fair lending practices among mortgage lenders.

Steven Ross is an assistant professor at the University of Connecticut and is considered an expert on the use of defaults in testing for discrimination in mortgage markets. His other research interests include discrimination in hous-

ing markets, urban labor markets, urban public finance, and general equilibrium models of urban economies.

Robin Smith, a research associate in the Urban Institute's Metropolitan Housing and Communities Center, has studied federal housing programs, fair housing, neighborhood revitalization, community building, employment and training, and child welfare. She is currently involved in several of the center's research projects for the U.S. Department of Housing and Urban Development (HUD), including a baseline assessment of the impact of HUD's major public housing desegregation cases and an evaluation of the Neighborhood Networks program.

Kenneth Temkin, a research associate with the Urban Institute's Metropolitan Housing and Communities Center, has written on segregation in public housing, fair lending practices, Fannie Mae's and Freddie Mac's underwriting guidelines, neighborhood change, social capital and other housing market-related issues. His current research interests include fair lending and the impacts of community development activities on lower income urban neighborhoods.

John Yinger is professor of economics and public administration and associate director of the Center for Policy Research at the Maxwell School, Syracuse University. He is a specialist in racial and ethnic discrimination in housing, state and local public finance, and urban economics. His most recent book, *Closed Doors, Opportunities Lost: The Continuing Costs of Housing Discrimination*, won the Myers Center Award for the Study of Human Rights in North America.





The Urban Institute

2100 M Street, N.W.
Washington, D.C. 20037

Phone: 202.833.7200

Fax: 202.429.0687

E-Mail: paffairs@ui.urban.org

<http://www.urban.org>