

**Looking the Other Way:
A Critique of the Fair-Lending Enforcement System
and a Plan to Fix It ¹**

Stephen Ross
John Yinger

Stephen Ross is an Associate Professor of Economics, University of Connecticut. John Yinger is Trustee Professor of Public Administration and Economics, the Maxwell School, Syracuse University.

Introduction

In 2000, the homeownership rate in the United States reached 67.4 percent — an all-time high. The benefits of homeownership were not evenly spread across ethnic groups, however. In fact, the homeownership rate was 73.8 percent for non-Hispanic whites, 47.2 percent for blacks, and 45.5 percent for Hispanics (U.S. Census Bureau, 2001). See Figure 1. These homeownership gaps undoubtedly have many causes, but one of the key suspects is discrimination in mortgage lending. The vast majority of households cannot buy a house without a mortgage loan, and discriminatory barriers to obtaining a mortgage could have a dramatic impact on homeownership.

A hint about the possible role of discrimination in mortgage lending comes from data collected under the Home Mortgage Disclosure Act (HMDA), which records the ethnicity of the applicant and the disposition of the application for virtually all the mortgage applications filed in the United States. In 2000, black applicants were twice as likely as white applicants to be turned down for a loan, and Hispanic applicants were 41 percent more likely to be turned down (FFIEC, 2001b). These loan-approval disparities do not prove that blacks and Hispanics face discrimination in mortgage lending, because they do not account for possible differences in loan features or borrower creditworthiness across groups. Nevertheless, the differences are so dramatic that they focus attention on the possibility that this type of discrimination might exist.

The purpose of this policy brief is to explore the possibility that mortgage lending discrimination contributes to ethnic disparities in homeownership, to evaluate the current fair-lending enforcement system, and to propose reforms in that system to make it more effective in uncovering — and, ultimately, eliminating — mortgage lending discrimination.

What Is Discrimination in Mortgage Lending?

Discrimination in mortgage lending is prohibited by the Fair Housing Act of 1968 (FaHA) and the Equal Credit Opportunity Act of 1974 (ECOA).² According to ECOA, as amended,

It shall be unlawful for any creditor to discriminate against any applicant, with respect to any aspect of a credit transaction —

- (1) on the basis of race, color, religion, national origin, sex or marital status, or age (provided the applicant has the capacity to contract).³

This language reveals that a lender is allowed to consider an applicant's "capacity to contract" but is not allowed to define this capacity differently for people in a legally protected class.

ECOA also assigns fair-lending enforcement authority to the same federal financial institutions (namely the Office of Thrift Supervision, the Office of the Comptroller of the Currency, the Federal Reserve, and the Federal Deposit Insurance Corporation) that oversee other aspects of lender behavior. Each of these institutions regulates a different set of lenders. See FFIEC (2001a). According to ECOA, enforcement authority for mortgage bankers, which are non-depository lenders, is assigned to the Federal Trade Commission.⁴

FaHA also takes a strong stand against lending discrimination. This act gives enforcement power to the Departments of Housing and Urban Development and of Justice. In general, Justice is entitled to prosecute cases involving a "pattern and practice" of discrimination or an issue of national importance, whereas HUD is the main agency for dealing with discrimination complaints.⁵ The Federal Financial Institutions Examination Council (FFIEC), which consists of all the federal financial regulatory institutions, provides a guide to the fair-lending regulations of its members. This guide covers discrimination in many types of actions by lenders. For example, this guide says that it would be

discrimination for a lender to “Refuse to extend credit or use different standards in determining whether to extend credit” based on an applicant’s membership in a legally protected class (FFIEC, 1999, p. ii).

These civil rights laws also make a key distinction between different types of discrimination. As the FFIEC guide puts it:

The courts have recognized three methods of proof of lending discrimination under the ECOA and the FH Act:

- Overt evidence of disparate treatment
- Comparative evidence of disparate treatment
- Evidence of disparate impact.

The existence of illegal disparate treatment may be established either by statements revealing that a lender explicitly considered prohibited factors (**overt** evidence) or by differences in treatment that are not fully explained by legitimate nondiscriminatory factors (**comparative** evidence)....

When a lender applies a racially or otherwise neutral policy or practice equally to all credit applicants, but the policy or practice disproportionately excludes or burdens certain persons on a prohibited basis, the policy or practice is described as having a “disparate impact”(FFIEC, 1999, pp. ii-iv).

In this policy brief, the behaviors identified by the first two “methods of proof,” overt and comparative, will be called “disparate-treatment” discrimination, and careful attention will be paid to both disparate-treatment and disparate-impact discrimination. Indeed, recognizing that both type of discrimination exist is critical for evaluating — and reforming — the fair-lending enforcement system.

Finally, a case involving disparate-impact discrimination has three steps (FFIEC, 1999). The first step is to determine whether a practice has a disparate impact on a legally protected class of people. The second step is to determine whether the practice can be justified on the grounds of business necessity, and the third step is to determine whether there exists an alternative practice that

achieves the same business objectives without the same disparate impact. Disparate-impact discrimination in lending is said to exist if (a) an enforcement agency finds that a lending practice has a disparate impact on a protected group and either (b) the lender cannot show that the practice is justified on the grounds of business necessity or (c) the enforcement agency shows that this disparate impact can be avoided through the use of an alternative practice that achieves the same business objectives. Although these steps are well established in law, the precise legal requirements for building a “prima facie” case for discrimination, part (a) or for building a business necessity defense, part (b), are not yet clear (Mahoney, 1998).

Why Should We Care About Discrimination in Mortgage Lending?

Nobody claims to be in favor of mortgage lending discrimination, of course, but some people do not believe that we need to worry about it. Discrimination is a thing of the past, they say, and no lender could survive in today’s competitive market if it practiced discrimination. For three principal reasons, we believe that this position is incorrect and that this nation should still care about discrimination in mortgage lending.⁶ These reasons are the lack of change in the HMDA data, the results of a major study of mortgage lending discrimination, and the possibility of extensive disparate-impact discrimination.

Lack of Change in the HMDA Data

If discrimination were disappearing from mortgage markets, one would expect the loan-approval disparities in the HMDA data to be declining over time.⁷ This has not been the case. In fact, the black/white loan-denial ratio has fluctuated around 2.0 since 1995, with a high of 2.07 in 1998 and

a low of 1.92 in 1999.⁸ The current ratio, 2.0, is slightly higher than the 1995 ratio, 1.95. See Figure 2. The Hispanic/white denial ratio has fluctuated around the lower value of 1.5, but it exhibits a similar pattern over time, with a relatively high value in 1998 and a relatively low value in 1999. Its current value, 1.41, is slightly below its value in 1995, 1.43.

Evidence from the Boston Fed Study

An important study based on data from 1990, Munnell et al. (1996), found extensive evidence of mortgage lending discrimination. This study, which is known as the Boston Fed Study because its authors were researchers at the Boston Federal Reserve Bank, supplemented the HMDA data with extensive information on individual loan applications, including measures of the applicant's credit history. On the basis of these data, this study found that black and Hispanic applicants are 82 percent more likely to be turned down for a loan than are equivalent white applicants.⁹ This result provides strong evidence of discrimination.

This study's methodology has been criticized by many scholars. However, several careful examinations of its data and methods conclude that the study's main result cannot be explained by most of the issues raised by these critics, including omitted credit variables, data errors, and mis-specification of the estimating equation.¹⁰ See Carr and Megbolugbe (1993), Glennon and Stengel (1994), and Ross and Yinger (forthcoming).

One issue raised by several critics appears to have more bite. Specifically, the Boston Fed Study cannot rule out the possibility that underwriting criteria differ, for legitimate business reasons, across lenders and that the lenders selected by black and Hispanic applicants are not as well suited to their credit needs as are the lenders selected by whites (Glennon and Stengel, 1994, Stengel and

Glennon, 1999). In this context, “legitimate” variation in underwriting standards is defined as variation that arises because different lenders draw on different pools of applicants and therefore have different experiences about the impact of various credit characteristics on the probability that a borrower will default.

Any such legitimate variation should be associated with the characteristics of a lender’s loan portfolio, that is, with the characteristics of the loans a lender provides. By adding many characteristics of loan portfolios to the Boston Fed Study’s data set, Ross and Yinger (forthcoming) are able to test for this possibility. They find that underwriting standards do, indeed, vary across lenders based on portfolio characteristics, but that accounting for this has no impact in the estimated minority-white disparity in loan approval. Legitimate differences in underwriting standards cannot explain the Boston Fed Study’s main result, and one is left with the conclusion that this result is a sign of discrimination.

The Boston Fed Study is based on 1990 data and it has not been replicated.¹¹ As a result, there exists no direct evidence about the extent of discrimination in mortgage lending at the current time. Nevertheless, the Boston Fed Study provides the best available evidence and the HMDA data for the last several years provides no indication that discrimination is declining.¹²

The Potential Importance of Disparate-Impact Discrimination.

The third reason for concern is that disparate-impact discrimination in mortgage lending could be widespread, even if, as several scholars have argued, disparate-treatment discrimination is no longer a serious problem. The potential importance of disparate-impact discrimination is suggested by two principal arguments.

First, disparate-treatment discrimination can readily be transformed into disparate-impact discrimination. As clearly explained by Lundberg (1991), economic agents who want to practice disparate-treatment discrimination but who are prevented from doing so may be able to achieve virtually identical outcomes by using characteristics other than group membership to predict which group an applicant belongs to.¹³ This approach only works, of course, if there exist characteristics that are correlated with group membership. In the case of lending, this is clearly the case; on average, black and Hispanic loan applicants have poorer credit qualifications than do white applicants.

The possibilities for exploiting the correlation between credit characteristics and group membership are demonstrated by Buist et al. (1999) and by Blackburn and Vermilyea (2001), who show, using two different data sets, that the loan-approval decisions of lenders can be explained either by setting a lower approval rate for blacks and Hispanics than for whites with a common set of credit standards across lenders or by devising lender-specific underwriting standards that do not simply predict “the capacity to contract” but instead also predict group membership.¹⁴ The latter possibility is, of course, disparate-impact discrimination.

The second argument for concern about disparate-impact discrimination is that it can easily be built into a credit-scoring or other automated underwriting scheme, even one that appears to treat all groups equally.

This argument is important because of the recent growth in the uses of these schemes.¹⁵ Several private companies now provide credit scores, which are formulas that translate a loan applicant’s financial characteristics and credit history into a score designed to predict default on a loan. These formulas are based on a statistical analysis of the impact of applicant characteristics on loan performance, usually measured by loan default, for a sample of previous loans. More general

automated underwriting schemes bring in additional explanatory variables, such as the nature of the loan or of the property being purchased. For example, Fannie Mae and Freddie Mac, key institutions in the secondary mortgage market, have developed automated underwriting schemes for use by loan originators who want to sell mortgages to these institutions. In some cases, automated underwriting schemes are so complete that mortgage transactions based on them are conducted entirely over the internet.

As several scholars have pointed out, automated underwriting schemes make disparate-treatment discrimination more difficult because they provide a detailed formula linking applicant characteristics to loan decisions, without any consideration of an applicant's race, ethnicity, or gender. See Avery et al. (2000), Buist et al. (1999), Yezer (1995). Indeed, in the extreme case of loans provided over the internet, the lender may not ever observe the applicant and may therefore not be able to use different underwriting criteria for different groups. The growth in automated underwriting does not make disparate-treatment discrimination impossible, because most schemes leave some room for lender judgment, but it appears to lower the likelihood that this type of discrimination takes place.

These scholars also point out, however, that automated underwriting does not rule out the possibility of disparate-impact discrimination. Indeed, an apparently group-neutral procedure for developing an automated underwriting scheme can lead to disparate-impact discrimination whenever groups differ on credit characteristics that are unobserved by the lender, such as the probability that a relative will be able to provide financial assistance in the case of unemployment or some other negative income shock.

Suppose, for example, that an automated underwriting scheme is based on a statistical analysis that ignores group membership altogether, which appears to be the procedure behind existing

schemes.¹⁶ In this case, the estimated underwriting weights of observed credit characteristics capture not only the relationship between these characteristics and the probability of default, which is entirely legitimate, but also, to the extent that the observed credit characteristics are correlated with group membership, the role of average unobserved credit characteristics for each group, which lenders are not allowed to consider.¹⁷ The only way to avoid disparate-impact discrimination in this situation is to base the underwriting weights in the scheme on a statistical analysis that includes group membership variables but then to ignore the impact of these variables in making a loan-approval decision. By leaving group membership variables out of its statistical analysis, therefore, an automated underwriting scheme may appear to be group neutral but is, in fact, introducing disparate-impact discrimination.

More generally, it is possible to test whether one automated underwriting scheme represents a legitimate, that is, non-discriminatory improvement over another scheme by determining whether it improves the predictions of loan performance **within** each group (Ross and Yinger, forthcoming). Disparate-impact discrimination arises when a scheme selects either the variables used to rate an application or the weights placed on these variables so as to predict the group to which an applicant belongs. Improved prediction for the set of applicants from a single group, say, whites, obviously cannot be affected by provisions that predict group membership. As a result, switching to a scheme that is common across groups and that improves within-group predictions is non-discriminatory, whereas switching to a scheme that improves overall predictions only by doing a better job of identifying group membership by definition involves disparate-impact discrimination.

What Is the Fair-Lending Enforcement System?

As explained earlier, many federal institutions share responsibility for enforcing the ECOA and FaHA. The first line of enforcement at depository lenders comes from the financial regulatory agencies, which, as noted earlier, have jointly developed a set of enforcement procedures (FFIEC, 1999). These procedures, as implemented by the Federal Reserve, are described in Calem and Canner (1995). Alternative procedures developed by the Office of the Comptroller of the Currency are described by Stengel and Glennon (1999) and Courchane, Nebhut, and Nickerson (2000).

Traditional Enforcement Methods

Calem and Canner (1995) begin by describing what they call “the traditional fair-lending enforcement method.”

To help assess the consistency of underwriting decisions, examiners traditionally have applied a technique known as “comparative loan file review” or “matched-pair analysis.” Essentially, this procedure can be described as follows. The examiners begin by selecting a sample of applications. Next, they note on “Applicant Profile Worksheets” the key factors considered in the underwriting decision, and the disposition of each application. The examiners then evaluate the information on these spreadsheets to identify potential instances of disparate treatment of similarly qualified applicants (pp. 118-119).

Calem and Canner then discuss various problems with this approach. Our own evaluation, which is presented below, builds on this analysis. According to Calem and Canner,

The traditional matched-pair examination procedure suffers from two important limitations. First, it is difficult for examiners to find applicants that are perfect, or even close, matches; some differences in underlying financial or property related characteristics nearly always remain.

Such differences in creditworthiness make it difficult to identify cases of unequal treatment. Even if there exist close matches among an institution’s files, it may be difficult for an

examiner to find them through manual effort alone. Moreover, in some instances, there may not be many close matches among the pool of applicants.

The second difficulty with the traditional matched-pair approach is that even if some differences in treatment are detected, it is hard to determine whether these are isolated events that do not result from discrimination, or the result of a pattern or practice of discrimination. Differences in treatment observed for a particular “matched pair” could be a purely random outcome of the underwriting process (p. 119).

Another way to express these limitations is to say that it is difficult, if not impossible, to make judgments about the use of a multivariate procedure, such as loan underwriting, using one pair of observations. A multivariate procedure is one in which a decision is based on the weighted values of several different variables. In the case of underwriting, a comparison of one minority and one white application yields valid inferences about the treatment of that minority applicant only if those two applications are both comparable on all applicant, loan, and property characteristics and representative of other loans with those characteristics. This is an extremely demanding standard. Moreover, any procedure that does not meet the two above conditions could run into several problems not mentioned by Calem and Canner. For example, a case in which a minority applicant is expected to meet a higher standard could be mistaken for a case in which “comparable” minority and white applications are both approved.

The Use of Regression Procedures by Fair-lending Enforcement Agencies

Several fair-lending enforcement agencies have supplemented traditional enforcement procedures with regression analysis for individual large lenders. This approach has been used, for example, by the Justice Department (Siskind and Cupingood, 1996), the Office of Comptroller of the

Currency (Stengel and Glennon, 1999; Courchane, Nebhut, and Nickerson, 2000), and the Federal Reserve Board (Calem and Canner, 1995; Avery et al., 1997; Calem and Longhofer, 2000).

Calem and Canner (1995) explain that these procedures were developed at the Federal Reserve in an attempt to overcome the limitations of traditional enforcement techniques. The Federal Reserve's regression-based technique involves supplementing HMDA data for a sample of loan applications, both minority and white, submitted to a particular lender.

Once the data have been collected, the next step in the procedure is to estimate a loan-approval regression.

To gauge the effect of applicant race on the disposition of loan applications, examiners, in consultation with Reserve Bank economists, construct a statistical model of the lender's underwriting decisions. This model is developed on the basis of information gathered from the bank's written underwriting guidelines and from interviews with loan officers. Factors considered important to the decision of whether to approve an application are included as explanatory variables in the model of loan disposition (p. 121).

The next step involves interpreting the results of this regression.

If the results of the statistical analysis indicate that the race of the applicant is a statistically significant predictor of loan disposition, then this is viewed as an initial indication that a pattern or practice of discrimination may exist.

However, the statistical model is necessarily an abstraction that can only partially replicate the loan approval process. Each and every factor that might reasonably influence an underwriting decision cannot possibly be incorporated into a model. Therefore, the statistical results alone are not considered definitive. In order to more fully evaluate the discrimination issue, examiners select specific loan files for closer review (Calem and Canner, 1995, p. 123).

The loan files selected for further review are minority/white pairs consisting of "minority applicants who have been denied credit and who appear as well qualified as, or better than, white applicants who were approved" (Calem and Canner, 1995, p. 123). For these file pairs, which appear

to involve discrimination, the examiners try to identify a legitimate business explanation for the relatively unfavorable treatment of the minority applicant. If any such explanation is found, the file is not considered to be a case of discrimination. As Calem and Canner (1995, p. 124) put it: “examiners may find that factors omitted from the model may account for these decisions.” See also Calem and Longhofer (2000) and Stengel and Glennon (1999).

Although similar to those developed by the Federal Reserve, the OCC procedures described by Courchane, Nebhut, and Nickerson (2000) place more weight on the statistical analysis and less weight on the follow-up comparisons of loan files.

What Is Wrong with the Fair-Lending Enforcement System?

The new regression procedures used by several fair-lending enforcement agencies are valuable contributions to the fair-lending enforcement system. Most importantly, they recognize that building a prima facie case for discrimination requires a multivariate procedure. Even with these new procedures, however, this system retains two serious limitations: it misses many instances of disparate-treatment discrimination and it fails to look for disparate-impact discrimination at all.¹⁸

The Need to Obtain an Accurate Estimate of Disparate-Treatment Discrimination

Enforcement procedures to measure disparate-treatment discrimination should, of course, be as accurate as possible. According to the official interagency definition, discrimination in loan approval exists when, among other things, lenders “use different standards in determining whether to extend credit” to people in a legally protected class (FFIEC, 1999, p. ii). The underwriting standards to which this definition applies depend upon many applicant, loan, and property characteristics. These standards

cannot be directly observed but must instead be inferred from the actions taken by lenders through the use of a multivariate statistical procedure; in other words, an accurate enforcement procedure requires a multivariate analysis..

The new regression procedures used by fair-lending enforcement agencies represent a significant step in the right direction because they recognize this principle. Compared to traditional file reviews, in other words, these regressions lead to a process that is more likely to find discrimination when it exists and less likely to find discrimination when it does not exist. As they are currently designed, however, the file-review procedures used by the Federal Reserve appear to forget this principle and therefore have the potential to undermine the gains from using regressions. The problem here lies not with file reviews as such, but instead with the way information from file reviews is used by some enforcement agencies.

To be specific, information from post-regression file reviews can be used in two ways. The first way, which is the one built into the Federal Reserve procedure, is to search for “information that would legitimately account for the divergent credit decisions” (Calem and Canner, 1995, p. 124), that is, for benign explanations for cases in which minority applicants appear to have been treated less favorably than comparable whites.

Unfortunately, however, this approach runs into exactly the same problems as traditional file reviews, namely, that it may be difficult to identify comparable files and any two files identified as comparable may still differ in important ways. Calem and Canner admit this when they say that their new procedure “is very similar to the ‘matched-pair’ technique traditionally used by examiners” (p. 123). However, they go on to argue that the new approach is better because “the statistical model guides the identification of matched pairs for review” (p. 123). It is no doubt true that the quality of the

matches is improved through the use of the statistical model, but a model cannot eliminate the problem. Even if two matched files have identical values for “key underwriting variables,” they are bound to differ on some other characteristics, and it is not logically possible for a file review to determine the impact of these differences on the underwriting decision. In short, a file review cannot provide an alternative test for the hypothesis that discrimination exists.

The information from post-regression file reviews can also be used to improve the regression specification or to do tests for the robustness of the results. The OCC procedures in Courchane, Nebhut, and Nickerson (2000) appear to follow this approach. This second way of using the information is consistent with principle that underwriting discrimination cannot be identified without a multivariate procedure. Consider the examples provided by Calem and Canner (1995). If some applicants are unable to document all reported income, then regulators should re-estimate the regression with an “unable to document” variable. If underwriters make a distinction between revolving debt and installment debt in scoring late payments, then regulators should estimate a regression that incorporates this distinction. These revised regressions would make full use of the information in the file reviews without giving up the regression’s multivariate structure.

Another way to put this is that file reviews may be able to identify underwriting factors that were missed in an initial regression, but they cannot determine the weights placed on these underwriting factors. As explained earlier, these weights cannot be directly observed but instead must be inferred using multivariate statistics. It is not logically possible to determine whether a newly identified underwriting factor can explain a minority rejection without estimating the weight placed on this factor by the lender — and controlling for other factors.

The lesson from this analysis is that a formal test for disparate-treatment discrimination requires a multivariate underwriting model estimated with a carefully determined specification and carefully collected data. The specification of this model should reflect, as fully as possible, a lender's stated underwriting standards, and it should, to the extent possible, incorporate lessons learned from interviews or file reviews. Fair lending laws require lenders to use the same underwriting standards for all applicants, regardless of their group membership. Allowing lenders to evaluate applications on the basis of idiosyncratic factors and to place unobservable weights on these factors in making their underwriting decisions eviscerates these laws by making it impossible to determine whether common standards are applied to all applicants. Thus, fair lending laws cannot be enforced unless lenders are held to a standard of equal treatment based on an available and objective standard, namely, a multivariate analysis of the lender's loan-denial decisions.

Because a regression analysis inevitably involves judgements, a lender should, of course, be allowed to comment on a regression analysis that finds it practices disparate-treatment discrimination. In our view, a thoughtfully conducted loan-approval regression that finds a significantly higher loan-denial rate for minorities than for whites, controlling for credit characteristics, establishes a prima facie case for disparate-treatment discrimination and therefore shifts the burden of proof onto the lender. In this situation, the lender can escape the charge of disparate-treatment discrimination only if it can provide an alternative regression specification that is consistent with its expressed underwriting policies (and with principles of regression methodology) and that indicates no significant difference in loan approval between minority and white applicants.

The Need to Look for Disparate-Impact Discrimination

Both the traditional enforcement policies and the regression-based policies developed by several fair-lending enforcement agencies also have another major flaw: they are incapable of identifying most cases of disparate-impact discrimination. In fact, as stated in Avery et al. (1997), Stengel and Glennon (1999), and Courchane, Nebhut, and Nickerson (2000), the explicit purpose of the regression-based procedures is to identify disparate-treatment discrimination alone. As Avery et al. put it (p. 14):

In any statistical analysis of discrimination (parametric or nonparametric), the goal is to determine whether or not the treatment of an individual would have been different had the individual been of a different minority status.

This is a textbook definition of disparate-treatment discrimination and it completely ignores behavior that has a disparate-impact on member of a minority group.

Moreover, Courchane, Nebhut, and Nickerson (2000) explicitly reject regressions that pool information across lenders, such as those proposed in Section 10.3, on the grounds that pooling is a poor technique for isolating disparate-treatment discrimination:

Specifically, the finding of disparate treatment requires comparison of the treatment of all applicants at a particular institution, under the credit guidelines of that institution. While a merged set of data from several institutions might reveal an underlying problem, and analysis of public use HMDA data might indicate where problems might be more likely to exist, it is the examination of an institution under its own standards that can best indicate the extend and scope of the disparate treatment. (p. 2)

The problem we see here is not in the logic of this statement, which is fine, but in the choice of objective, which is not. The fair-lending enforcement agencies are responsible for identifying both disparate-

treatment and disparate-impact discrimination, and it makes no sense to rely exclusively on methods that, in effect, simply look the other way when confronted with the possibility of disparate-impact discrimination.

As shown by Ross and Yinger (forthcoming), disparate-impact discrimination can enter a loan-approval regression in two ways. First, it can show up in the estimated difference in loan approval between minority and white applicants, controlling for credit characteristics, if the regression specification does not exactly accurately reflect a lender's actual underwriting standards. Second, it can show up in the estimated weights for the credit characteristics, and therefore will not be recognized as discrimination in a loan-approval regression.

The first possibility needs to be considered because it helps to show why looking for disparate-impact discrimination is so important. Specifically, an investigator following the Federal Reserve procedures (or a lender responding to them) might be able to reduce apparent discrimination, as indicated by the estimated minority-white difference in loan approval, controlling for credit characteristics, by introducing a lender's idiosyncratic, but illegitimate underwriting standards into the specification of the regression. This step could shift the effect of disparate-impact discrimination from the estimated minority-white difference in loan approval to the estimated weights of individual credit characteristics, where it will not be observed. Thus, the search for the "correct" specification, that is, the specification most accurately portraying a lender's underwriting criteria, a search that is central to the logic of the Federal Reserve's regression procedure, can be seen as a way to ensure that disparate-impact discrimination is ignored.

The problem runs even deeper than this, however. As shown in such a compelling fashion by Buist et al. (1999) and Blackburn and Vermilyea (2001), lenders may be able to hide disparate-

treatment discrimination by transforming it into disparate-impact discrimination. In this case, the Federal Reserve's regression procedure could miss discrimination altogether, even when it is severe. Indeed, we believe it is inappropriate — if not irresponsible — for these agencies to use a procedure that violates the FFIEC guide by assuming that disparate-treatment discrimination is the only kind worth looking for.

How Can the Fair-Lending Enforcement System Be Improved?

In our judgment, the current fair-lending enforcement system is seriously inadequate because it is likely to miss some cases of discrimination in loan approval that take the form of disparate treatment and is incapable of identifying loan-approval discrimination that takes the form of disparate-impact.¹⁹

We propose three steps for eliminating these flaws.

First, the fair-lending enforcement agencies should come up with the resources needed to make certain that they are not missing a large share of existing disparate-treatment discrimination. Multivariate regressions should be employed by all these agencies; these methods based on virtually complete information; and loan file reviews should be treated as a method for improving, not overruling, regression analysis.

Second, these agencies should conduct loan-approval regressions based on applications submitted to a large sample of lenders. These regressions should recognize the complexity of underwriting standards and the possibility that these standards vary systematically across lenders based on their loan portfolios. This tool makes it possible to estimate the extent of discrimination by each lender in the sample, regardless of whether that discrimination takes the form of disparate impact or of disparate treatment. Moreover, because it is based on a large sample, this tool provides precise

estimates of the weights placed on a wide range of underwriting variables, yields an estimate of discrimination even for lenders that are too small for current regression procedures, and eliminates the arbitrary separation of lenders based on the agency that regulates them. In short, this tool provides the best possible lender-specific estimates of discrimination that are available without loan-performance information and is an ideal way to determine if there is a prima facie case for discrimination by any lender in the sample.

As before, a lender should of course be allowed to build a business-necessity defense. In this case, however, a lender cannot mount such a defense by adding its own idiosyncratic underwriting criteria to a loan-approval regression. Not only can a lender hide intentional discrimination by manipulating its underwriting weights, but, as shown earlier, these weights may reflect discrimination even if they are based on an apparently group-neutral analysis of loan performance. Instead, a lender cannot defend the underwriting weights it uses on business necessity grounds unless it can demonstrate that these weights do a better job of predicting loan performance (as measured, say, by loan default) than the weights implied by the enforcement agency's regression. Following the non-discrimination test developed earlier, this demonstration must apply **within** each ethnic group, not to all groups combined.

Third, the fair-lending enforcement agencies should implement a performance-based analysis of loan-approval decisions to supplement the first tool. This second enforcement tool requires an enforcement agency to estimate a model of the factors that determine loan default or some other measure of loan performance, which is the type of model on which an automated underwriting system is based. More specifically, this tool compares the minority composition of the applications that have the highest predicted loan performance based on this loan-performance model with the minority composition of the applications a lender actually approves.²⁰ Discrimination exists if significantly more

minority applications would be approved on the basis of the agency's predicted performance than are actually approved on the basis of the lender's underwriting standards.

This tool requires information on loan performance and on credit characteristics for a large sample of loans, which the fair-lending agencies have, so far, been reluctant to obtain, even though they have the power to do so. However, it does not require the investigator to know the formulas behind a lender's underwriting standards or credit scores, which may be considered proprietary. This tool, like the previous one, captures both disparate-impact and disparate-treatment discrimination but cannot tell them apart.

This tool would yield more precise answers about discrimination than the first tool, but it would obviously be more costly to implement. Loan performance is observed by the institution servicing a loan, which may not be the same as the institution that issued the loan. To examine discrimination in underwriting, therefore, regulators must develop procedures that link loan performance information with information about the issuing lender. These issues arise even for large lenders who originate and then continue to service many loans. After all, these lenders also sell some of their loans on the secondary market and the sample of loans they retain is not a random sample of the loans they originate.

To build a business-necessity defense in this case, a lender would have to show that his underwriting weights are derived from a loan-performance model that does a better job of predicting within-group loan performance than does the model estimated by regulators. If an enforcement agency has made a prima facie case for discrimination and the lender cannot supply an alternative loan-performance model that meets this non-discrimination test, then the third part of a disparate-impact case is automatically satisfied. Under these circumstances, the loan-performance model estimated by the

enforcement agency provides an alternative underwriting scheme that meets the lender's legitimate business objectives without any discrimination.

Although our second and third recommendations would require lenders to provide information from their loan files, they are designed, in part, to protect lenders from unwarranted charges of discriminatory behavior. Recall that we recommend stringent standards for establishing a prima facie case for disparate-impact discrimination, based on a multivariate procedure. Regulators should make it clear that the selection of a lender for further investigation does not imply that the regulator has already built a prima facie case for discrimination by that lender. Just as an income-tax audit does not imply that a taxpayer has cheated on his taxes, a lending investigation does not imply that a lender has practiced discrimination. Instead, a lender is charged with discrimination only if a statistical procedure finds a minority-white disparity after controlling for all legitimate underwriting variables. With these procedures, a lender who does not discriminate has nothing to worry about.

Despite their unique ability to collect the relevant data, the fair-lending enforcement agencies have decided not to provide the public with any credible evidence on the current extent of discrimination in mortgage underwriting. As in the case of fair-lending enforcement, they apparently favor looking the other way. Consequently, neither we nor anyone else knows how much of this discrimination still exists. According to the best available evidence, however, extensive underwriting discrimination existed in 1990 and there is no more recent evidence to show that this discrimination has gone away. Moreover, black and Hispanic households continue to have homeownership and loan-approval rates that are far below the rates attained by white households, even after controlling for income and other factors (Gyourko, Linneman, and Wachter, 1999).

Under these circumstances, this nation cannot begin to live up to the important principles embodied in its fair-lending laws without actively searching for mortgage discrimination in all its possible forms using the most accurate tools possible. The current fair-lending enforcement system does not even come close to meeting this standard.

It does not have to be this way. More comprehensive and accurate enforcement tools that build on a large body of scholarly research and are consistent with legal standards are readily available. We strongly urge the fair-lending enforcement agencies to make these tools a regular part of their enforcement activities. We also urge interested citizens, community groups, academics, lenders and other participants in the mortgage market, and public officials to work for improvements in the fair-lending enforcement system. Every American household should be able to enter the mortgage market feeling confident that it will not encounter discrimination.

Endnotes

¹This policy brief is based on Ross and Yinger (forthcoming).

²FaHA and ECOA also prohibit “redlining,” defined as unfavorable actions by a lender toward loans involving properties in neighborhoods where members of a protected class are located. Redlining is not considered in this policy brief.

³ U.S. Code Title 15, Chapter 41, Section 1691.

⁴Nondepository lenders obtain mortgage capital from investors in the secondary mortgage market, instead of from deposit. These investors want to receive their income in the form of mortgage interest payments. See the citations in note 14.

⁵For a more detailed discussion of the enforcement duties of these two agencies, see Schwemm (1994) or Yinger (1995).

⁶A fifth reason, which is too technical for full discussion in this policy brief, is that discrimination may be profitable, and therefore may not be eliminated by competition. For more on this view, see Ferguson and Peters (2000), Longhofer and Peters (1998), and Ross and Yinger (forthcoming).

⁷Avery et al. (1996) discuss several other HMDA results that are consistent with, but do not prove, the existence of discrimination.

⁸See FFIEC (2001b), which is the source of all the numbers in this paragraph. The pre-1995 HMDA data are not comparable to data for 1995 and later years. See Scheessle (1998).

⁹Munnell et al. also explored a wide range of alternative specifications for their estimating equation and found that their result was remarkably robust to these changes.

¹⁰Several scholars have also argued that Munnell et al. should have looked at loan defaults, not loan approvals. Ross and Yinger (forthcoming) examine this argument in detail and show that it is not correct.

¹¹The federal fair-lending agencies have the authority to collect the information needed to replicate this study but they have not done so. This lack of replication is itself a powerful indictment of these agencies. In our judgment, one of the principal responsibilities of any civil-rights enforcement agency is to educate the public on the magnitude of the problem.

¹²The trends in the HMDA data do not, of course, prove that discrimination remains at its 1990 level. In principle, a decline in discrimination since 1990 could have been accompanied by a deterioration in the relative creditworthiness of black and Hispanic applicants. We know of no evidence, however, that this type of deterioration has taken place.

¹³Because of this possibility, civil rights laws that only cover disparate-treatment discrimination have an enormous loophole.

¹⁴Both Buist et al. (1999) and Blackburn and Vermilyea (2001) interpret their results as evidence that one cannot tell whether lenders practice disparate-treatment discrimination, practice disparate-impact discrimination, or simply use different underwriting standards on legitimate business grounds. As explained earlier, however, Ross and Yinger (forthcoming) rule out the third possibility (using the same data as Buist et al.). Moreover, Blackburn and Vermilyea show that inter-group differences in loan approval are explained by across-lender differences in the definitions of underwriting variables, not in the weights placed on common underwriting variables. (For example, one lender might have special rules for mortgages with a loan-to-value ratio, LTV, above 0.90, whereas another might use an LTV cut-off of 0.95.) It seems unlikely that these idiosyncratic differences in definitions are justified by a link to performance data, which would be required for a business-necessity defense.

¹⁵The increased reliance on automated underwriting is related to several other trends, including a trend toward “unbundling” various mortgage services, the emergence of mortgage bankers, and the growth of the secondary mortgage market. See Follain and Zorn (1990), LaCour-Little (2000), Lea (1996), Ross and Yinger (forthcoming), and Van Order (2000).

¹⁶The actual statistical procedures are considered proprietary and are not released, but the available descriptions of the schemes never mention group membership variables.

¹⁷In technical terms, this is an example of omitted variable bias in a regression analysis. See Ross and Yinger (forthcoming).

¹⁸Several other weaknesses of the current enforcement system are discussed in Ross and Yinger (forthcoming).

¹⁹The fair-lending enforcement system could also do a better job preventing discrimination in lender actions other than loan approval, such as loan pricing. See Ross and Yinger (forthcoming).

²⁰If a lender has approved A applications, then this test compares the minority composition of approved loans with that of the A highest-ranking applications according to the enforcement agency's loan-performance model. See Ross and Yinger (forthcoming).

REFERENCES –Looking the Other Way

- Avery, Robert B., Raphael W. Bostic, Paul S. Calem, and Glenn B. Canner. 2000. "Credit Scoring: Statistical Issues and Evidence from Credit-Bureau Files," *Real Estate Economics*, 28 (Fall): 523-547.
- Avery, Robert B., Patricia E. Beeson, and Paul D. Calem. 1997. "Using HMDA Data as a Regulatory Screen for Fair Lending Compliance." *Journal of Financial Services Research*, 11 (April): 9-42.
- Avery, Robert B., Raphael W. Bostic, Paul S. Calem, and Glenn B. Canner. 1996. "Credit Risk, Credit Scoring, and the Performance of Home Mortgages." *Federal Reserve Bulletin*, 82 (July): 621-648.
- Blackburn, McKinley, and Todd Vermilyea. 2001. "Racial Discrimination in Home Purchase Mortgage Lending Among Large National Banks." Unpublished manuscript.
- Buist, Henry, Peter Linneman, and Isaac F. Megbolugbe. 1999. "Residential Lending Discrimination and Lender Compensation Policies," *Journal of the American Real Estate and Urban Economics Association*, 27 (Winter): 695-717.
- Calem, Paul S., and Glenn Canner. 1995. "Integrating Statistical Analysis into Fair Lending Exams: The Fed's Approach." In Anthony M. Yezer, editor, *Fair Lending Analysis: A Compendium of Essays on the Use of Statistics* (Washington, D.C.: American Bankers Association), pp. 117-126.

- Calem, Paul S., and Stanley D. Longhofer. 2000. "Anatomy of a Fair-Lending Exam; The Uses and Limitations of Statistics." Unpublished Manuscript, March.
- 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): 277-313.
- Courchane, Marsha, David Nebhut, and David Nickerson. 2000. "Lessons Learned: Application of Statistical Techniques to Fair Lending Compliance Examinations." Unpublished manuscript.
- Federal Financial Institutions Examination Council. 2001a. "Annual Report, 2000." Washington, D.C.: By the Author. Available through <http://www.ffiec.gov>
- Federal Financial Institutions Examination Council. 2001b. "FFIEC HMDA Aggregate National Reports." Washington, D.C.: By the Author. Available through <http://www.ffiec.gov>
- Federal Financial Institutions Examination Council. 1999. "Interagency Fair Lending Examination Procedures." Washington, D.C.: By the Author. Available through <http://www.ffiec.gov>
- Ferguson, Michael F., and Stephen R. Peters. 2000. "Is Lending Discrimination Always Costly?" *Journal of Real Estate Finance and Economics*, 21 (July): 23-44.
- Follain, James R., and Peter M. Zorn. 1990. "The Unbundling of Residential Mortgage Finance." *Journal of Housing Research*, 1(1): 63-89.

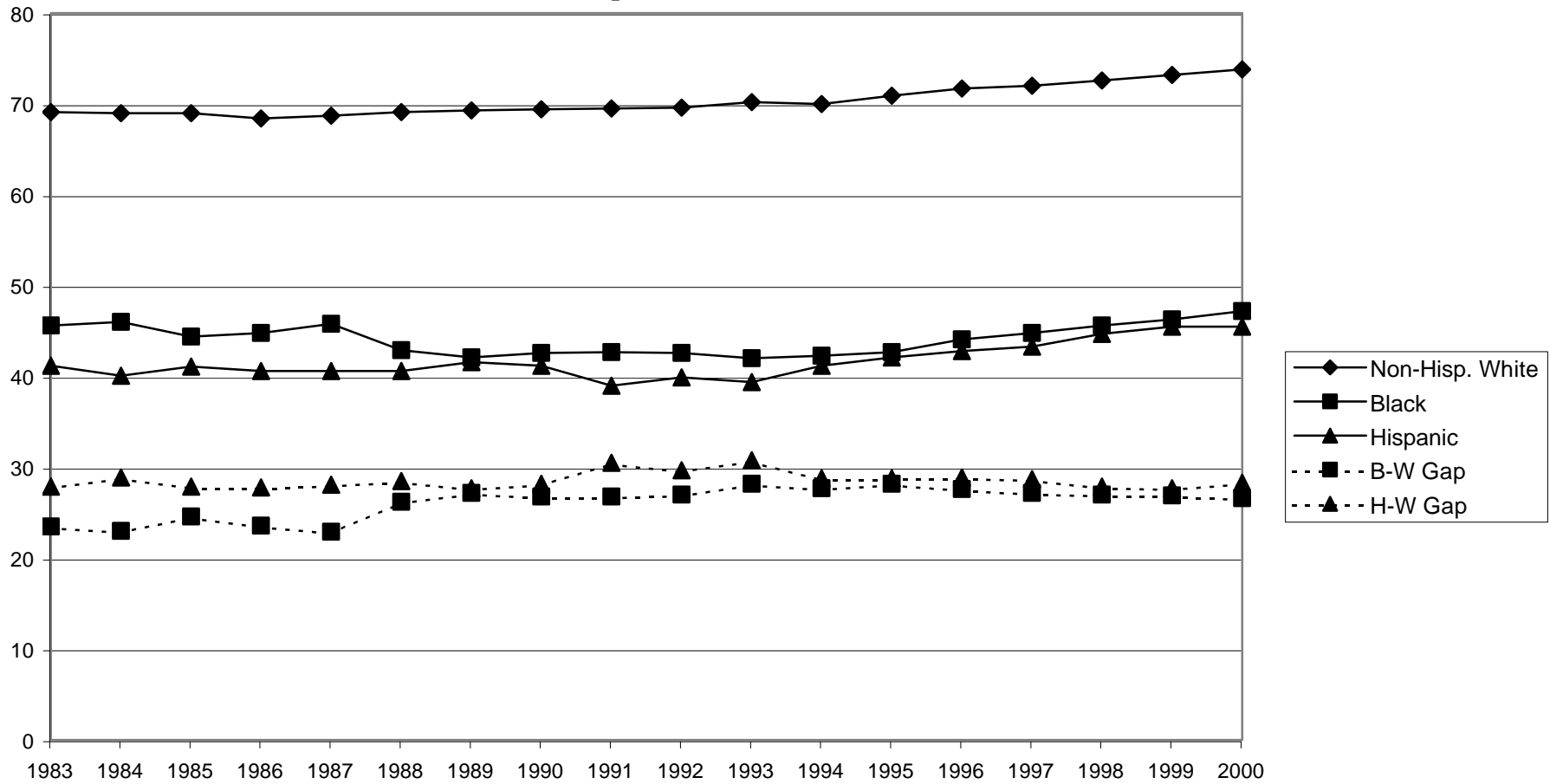
- Glennon, Dennis, and Mitchell Stengel, 1994 "An Evaluation of the Federal Reserve Bank of Boston's Study of Racial Discrimination in Mortgage Lending." Working Paper. Washington, DC: Office of the Comptroller of the Currency.
- Gyourko, Joseph, Peter Linneman, and Susan Wachter. 1999. "Analyzing the Relationships among Race, Wealth, and Home Ownership in America." *Journal of Housing Economics*, 8 (June): 63-89.
- LaCour-Little, Michael. 2000. "The Evolving Role of Technology in Mortgage Finance." *Journal of Housing Research* 11 (2): 173-205.
- Lea, Michael J. 1996. "Innovation and the Cost of Mortgage Credit: A Historical Perspective." *Housing Policy Debate*, 7(1): 147-174.
- Longhofer, Stanley D., and Stephen R. Peters. 1998. "Self Selection and Discrimination in Credit Markets." Unpublished manuscript. Cleveland: Research Department, Federal Reserve Bank of Cleveland, April.
- Lundberg, Shelly J. 1991. "The Enforcement of Equal Opportunity Laws under Imperfect Information: Affirmative Action and Alternatives." *The Quarterly Journal of Economics*, 106 (February): 309-326.
- Mahoney, Peter E. 1998. "The End(s) of Disparate Impact: Doctrinal Reconstruction, Fair Housing and Lending Law, and the Antidiscrimination Principle." *Emory Law Journal* 47 (Spring): 409-525.
- Munnell, Alicia H., et al. 1996. "Mortgage Lending in Boston: Interpreting HMDA Data." *American Economic Review*, 86 (March): 25-53.

- Ross, Stephen, and John Yinger. Forthcoming. *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair-Lending Enforcement*. Cambridge, MA: MIT Press.
- Schwemm, Robert G. 1994. *Housing Discrimination: Law and Litigation*. Deerfield, IL: Clark, Boardman, Callaghan.
- Scheessele, Randall M. 1998. "HMDA Coverage of the Mortgage Market." Housing Finance Working Papers #HF-007, Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.
- 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 John Goering and Ron Wienk, editors, *Mortgage Lending, Racial Discrimination, and Federal Policy* (Washington, DC: Urban Institute Press), pp. 451-468.
- Stengel, Mitchell , and Dennis Glennon. 1999. "Evaluating Statistical Models of Mortgage Lending Discrimination: A Bank-Specific Analysis," *Real Estate Economics*, 27 (Summer): 299-334.
- U. S. Census Bureau. 2001 "Housing Vacancies and Homeownership Annual Statistics: 2000, <http://www.census.gov/hhes/www/housing/hvs/annual00/ann00ind.html>.
- Van Order, Robert. 2000. "The U.S. Mortgage Market: A Model of Dueling Charters." *Journal of Housing Research*, 11(2): 233-256.

Yezer, Anthony M. 1995. "Editor's Notes." In Anthony M. Yezer, editor, *Fair Lending Analysis: A Compendium of Essays on the Use of Statistics* (Washington, DC, American Bankers Association), pp. iii- vi.

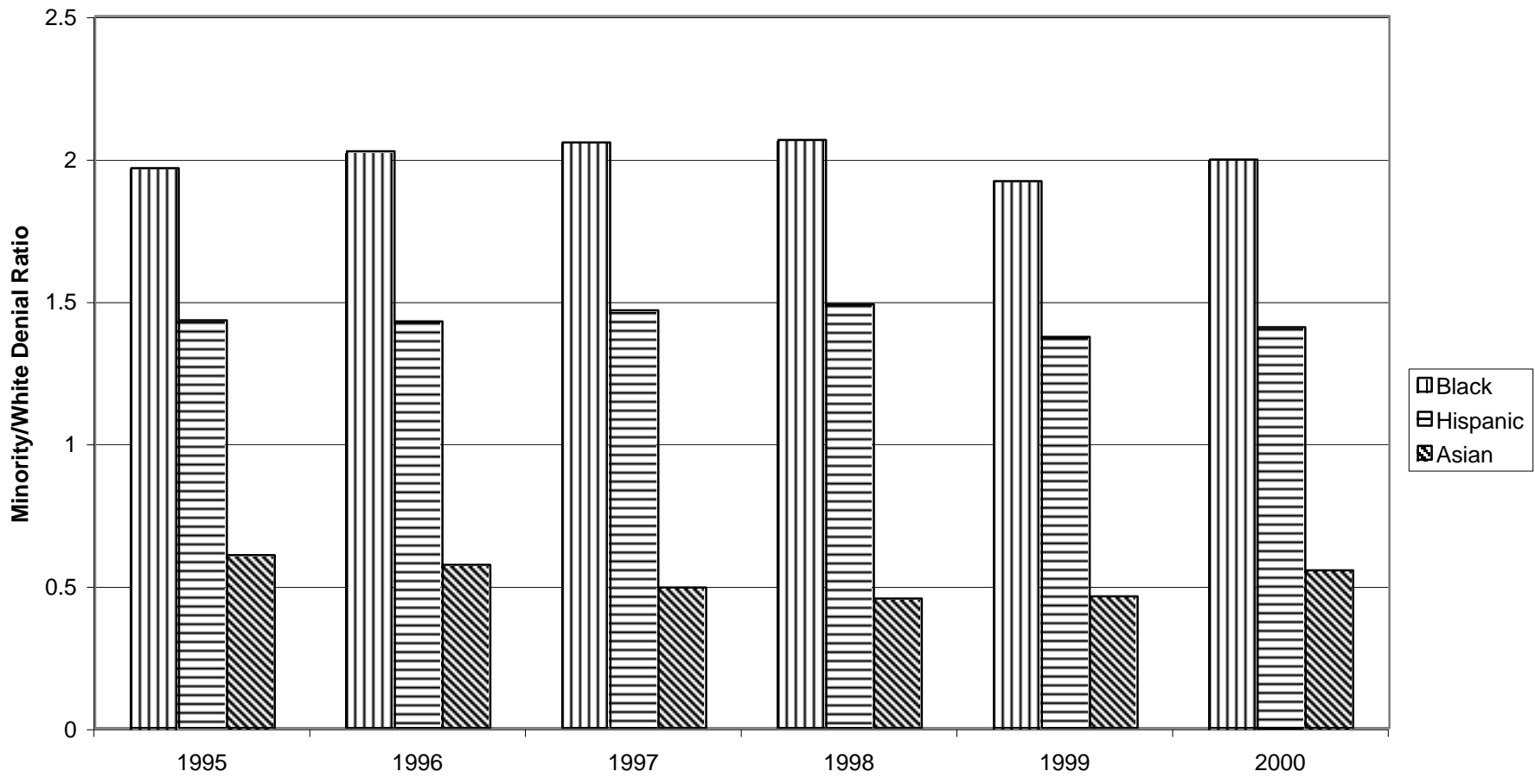
Yinger, John. 1995. *Closed Doors, Opportunities Lost: The Continuing Costs of Housing Discrimination*. New York: Russell Sage Foundation.

Figure 1
Homeownership Rates, 1983-2000



Sources: Before 1994, CEA (1998); 1994-2000, Census Bureau (2001)

Figure 2
Conventional Home-Purchase Loan-Denial Ratios by Year



Source: FFIEC (2001)