Mortgage Lending in Boston—A Response to the Critics

Lynn Elaine Browne and Geoffrey M.B. Tootell

Senior Vice President and Director of Research, and Assistant Vice President and Economist, respectively, Federal Reserve Bank of Boston. *The authors thank Faith Kasirye-Nsereko for her indefatigable assistance on this and related projects and their colleagues at the Boston Fed for helpful comments.*

Three years ago, the Federal Reserve Bank of Boston released an examination of racial patterns in mortgage denial rates in the Boston area (Munnell, Browne, McEneaney, and Tootell 1992 (MBMT)). The study was motivated by newly available data on mortgage applicants, showing that black and Hispanic applicants were two to three times as likely to be turned down for mortgages as white applicants. The study gathered additional data on applicants' debt burdens, credit histories, and other financial characteristics to see whether economic factors explained the racial difference in denial rates. Although the additional information did explain much of the difference, after taking account of economic factors the applicant's race still affected the probability of getting a mortgage.

The study was promptly given "landmark" status by some in the press, and in some respects the designation is warranted. The data collection was a major undertaking; thus, despite many calls for studies of racial lending patterns in other cities or sets of institutions, only one somewhat similar work had appeared as of mid-1995 (Stengel and Glennon 1995). It has also been influential. The study alerted both the mortgage industry and its regulators to the possibility of discrimination in mortgage lending. It has stimulated many financial institutions to re-examine their lending practices and has caused the federal supervisory authorities to change their examination procedures pertaining to fair lending. It has spurred efforts by the major secondary market agencies both to ensure that lenders do not interpret their credit guidelines excessively strictly and to reassess the appropriateness of some of these guidelines. The study may have provided some of the impetus to revise the Community Reinvestment Act regulations and it probably reinforced the Department of Justice's efforts to pursue fair lending more vigorously.

Given the attention the study has received, criticism is to be expected. Some of the criticism has been scholarly. Some has been strident, with one critic even hinting the study was "consciously fraudulent" (Roberts 1993). Much of the criticism seems to reflect a view that discrimination simply cannot occur in lending; much, especially some of the most vociferous, appears driven by concerns over policy directions that the study might inspire. These concerns have taken on new life in the past year in response to the Justice Department's more aggressive stance towards redlining and fair lending violations.

Thus, it seems appropriate to respond to the major criticisms of the study, showing why the study is sound and why its finding that discrimination and economic factors both contributed to the racial disparities in mortgage denials in Boston is solid. At the same time, it should be noted that the study itself did not advocate any specific remedial policies, simply concluding that "a serious problem exists in the market for mortgage loans" such that "lenders, community groups, and regulators must work together to ensure that minorities are treated fairly."

Primary responsibility for addressing the problem of discrimination in mortgage lending lies with the industry.

Subsequently, in testimony before the Senate Banking Committee, Richard Syron, then President of the Federal Reserve Bank of Boston, made clear that primary responsibility for addressing the problem of discrimination in mortgage lending lies with the industry, stating, "the most critical step is for mortgage lenders to acknowledge at least the possibility that the results of their lending process are discriminatory. As long as lenders sincerely believe their procedures are beyond reproach, efforts to get them to change will have limited success.... Lenders' reactions to the study suggest that they are now questioning what they always took for granted. They are starting to recognize that simply having a policy that prohibits discrimination does not prevent discrimination" (Syron 1993). Fostering this self-questioning was a major accomplishment of the Boston Fed study and it would be most unfortunate if it were reversed.

Although criticisms of the Boston Fed's findings in the media have been numerous, most of these repeat the arguments of three sources (Brimelow and Spencer 1993; Liebowitz 1993; and Horne 1994a and 1994b). In a separate category stands the more technical criticism by Yezer and various co-authors (for example, Yezer, Phillips, and Trost 1994) that negotiations between borrowers and lenders preclude finding discrimination. The appendix summarizes the issues raised by each of these and other major critics and provides point-by-point rebuttals.

The criticisms can be grouped into five categories:

1. Default rates—If discrimination exists, the average default rate of minority borrowers will be below that of white borrowers, whereas data in the Boston Fed study suggest that minority and white default rates are similar.

2. Omitted or missing variables—Variables have been omitted from the analysis that might explain the influence of race on the mortgage decision.

3. Misspecification—A different specification of the mortgage decision process might lead to a conclusion that discrimination is not occurring. The argument that the mortgage decision process is a negotiation is a specification issue.

4. Goodness of fit—The Boston Fed's regression analysis does not explain mortgage outcomes very well.

5. Data errors—The data collected from the lending institutions and used in the Boston Fed's analysis contain errors.

Each of these criticisms is addressed below. It will be shown that comparisons of average default rates tell little about the existence of discrimination if the distribution of default probabilities is different for minority applicants than for white applicants. Most of the allegedly missing variables were included in the regressions presented in the study's appendix or were well proxied by included variables. Alternative specifications do not affect the influence of race on the outcome, unless the sample is split so as to eliminate most of the minority rejections from the analysis. The regression explains denial outcomes well. Most "data errors" are not errors at all; and such errors as do exist do not affect the study's results. Before responding to these criticisms in detail, however, the study and its findings are summarized.

Recap of Boston Fed Study

The Boston Fed's study was undertaken in response to the release of Home Mortgage Disclosure Act (HMDA) data for 1990 that showed black and Hispanic applicants for home purchase mortgages being turned down much more frequently than white applicants. This was true in almost all major metropolitan areas. In Boston, approximately 30 percent of black and Hispanic mortgage applicants were rejected, compared to 11 percent of white applicants.

This was new information. Although community and minority groups had previously complained about the small number of mortgages made in minority areas, the available information covered only approved loans and told nothing about the characteristics of the applicants. Thus, it was not possible to distinguish the role of the lending industry from that of buyers, sellers, realtors, and other actors in the housing and mortgage markets. In 1989, however, the Home Mortgage Disclosure Act was amended to require information on the disposition of all mortgage applications according to the applicant's race, gender and income.

The implications of the disparities in denial rates were hotly debated, with some people seeing them as proof of discrimination and others arguing that they could be explained by differences in applicants' loan-to-value ratios, obligation ratios, credit histories, and other economic characteristics. In an effort to clarify the importance of these economic factors, the Federal Reserve Bank of Boston, with the support of the Federal Reserve Board and other federal supervisory agencies, undertook to gather the missing information and perform the necessary analysis for the Boston metropolitan area.

The project was a major effort. The 131 financial institutions that had been the most active mortgage lenders in the Boston metropolitan area were asked to provide additional information on 38 financial, credit history, and employment variables for all their black and Hispanic mortgage applicants and a random sample of white applicants.[1] The final sample consisted of more than 700 black and Hispanic ("minority") applicants for conventional home purchase mortgages and 2300 white applicants.

It should be noted that the Boston Fed researchers did not have direct access to the lenders' files. The Boston Fed was not the primary regulator of the institutions, and thus researchers were precluded from seeing the files. To ensure accuracy, the Boston Fed ran the data through various computer checks and screened the information visually. Institutions were asked to verify that unusual-looking variables corresponded to the information in their loan files. The institutions also had an incentive to be accurate, having been told that the information could be turned over to their primary regulators.

The choice of the variables to be collected was based on numerous conversations with underwriters, examiners, and others familiar with the mortgage lending process. While media accounts of industry explanations for the racial disparities in denials focused on a relatively small number of variables, the Boston Fed study tried to include everything that might possibly be relevant to the mortgage decision. The information collected from the financial institutions was then combined with the institutions' final HMDA submissions and data on neighborhood characteristics from the 1990 Census.

Until the release of HMDA data for 1990, it was not possible to distinguish the lending industry's contribution to racial patterns in mortgage originations from that of buyers, sellers, realtors, and other actors in the housing and mortgage markets.

The resulting data set contained more than 60 variables, although some of the information from the HMDA submissions, such as the purchaser of the mortgage and the date the application was submitted, were not considered pertinent to the analysis of mortgage denials. This data set was made available to the research community in 1993. A limited number of variables were deleted from this public research data set in order to prevent individual mortgage applicants and lending institutions from being identified.[2] Regressions run using these data are virtually identical to those using the original data set.

The data were analyzed using a logit regression, in which the probability of being denied a mortgage loan was a function of obligation ratios, credit history, measures of wealth, and a variety of other economic

[1] The Boston Fed did not, as some critics have asserted, collect information on 6.6 million mortgage applications (see Macey 1994). The 6.6 million figure appears to refer to mortgage applications nationwide.

[2] Among the more noteworthy changes were the deletion of the lender identifier, the census tract number, and information relating to the applicant's occupation.

characteristics, as well as the applicant's race. Many specifications were examined, of which a sample appears in the Boston Fed study (MBMT).

Although logit regressions are well-suited to modelling discrete outcomes, in this case whether an application was approved or denied, the output of the regression is an estimate of the *probability* that an application will be denied. Probabilities are continuous. Both critics and fans of the study have resisted the probability concept, wanting to interpret a 49 percent probability of denial as a surefire prediction of approval and a 51 percent probability as clear evidence the application should be turned down. Additionally, because the actual outcomes are discrete, whether a loan was denied or not, and the estimates are probabilities, there is no simple measure of how well the regression explains the variation in outcomes comparable to the r-squared that is traditionally used to measure a regression's "goodness of fit."

The Boston Fed's analysis confirmed that credit histories, loan-to-value ratios, and other factors cited by lenders as influencing the mortgage decision did indeed explain much of the gap between minority and white denial rates. As the study stated, "Including the additional information on applicant and property characteristics reduces the disparity between minority and white denials from the originally reported ratio of 2.7 to 1 to roughly 1.6 to 1" (MBMT, p. 2). Nevertheless, after taking into account obligation ratios, loan-to-value ratios, credit histories, and other factors affecting the loan decision, black and Hispanic mortgage applicants were still more likely to be turned down than white applicants. Specifically, given white applicants' financial, credit history, employment, and neighborhood characteristics, minority applicants would experience a 17 percent denial rate compared to the white applicants' denial rate of 11 percent.

The study also provided some insight into why this outcome might occur. Most applicants, white as well as minority, are not "perfect." They exceed some secondary market guideline for obligation or loan-to-value ratios or for credit history, or they possess some characteristic, such as self-employment or purchase of a two- to four-family home, that requires additional documentation. Thus, approving a mortgage involves considerable judgment on the part of the lender. The decision is not a mechanical process in which loan originators unthinkingly apply guidelines set by the secondary market or their institution.

Discretion is desirable. Residential mortgages are generally seen as very safe investments, implying that applicants need not be perfect to be creditworthy. But discretion opens the door to the possibility of discrimination. In addition, the relative scarcity of perfect applications means that discovering discrimination through file-by-file reviews is very difficult. Almost always, some blemish is present that could be cited as justification for denial. A search might reveal approved applications with the same flaws, but they probably will not be the same in all respects. And even if they are, the tricky issue of probabilities remains. The denial of a single minority applicant while a similarly situated white applicant was approved could be a chance outcome. Only by looking at large numbers of applications can patterns be discerned.

Fostering self-questioning by the mortgage lending industry was a major accomplishment of the Boston Fed study and it would be most unfortunate if it were reversed.

The Boston Fed's findings about the role of judgment in the mortgage decision and the difficulties of trying to identify discrimination through file-by-file reviews are important in and of themselves. In combination with the finding that race affected lending decisions in Boston in 1990, they mean that discrimination is possible and that lenders cannot pass off responsibility for their loan outcomes to the secondary market or take comfort in past favorable exam results. Moreover, while it may be more palatable to think that discrimination arises from subtle differences in the exercise of judgment rather than through overt policies, ensuring fair treatment may actually prove more difficult in the former situation. Training, changes to hiring and promotion practices, self-monitoring, and other steps may be required to ensure that all borrowers are treated equitably. A new policy statement, alone, is unlikely to do the job.

Default Rates

Of all the criticisms of the Boston study, the one that resurfaces with greatest persistence is the claim by Brimelow and Spencer (1993) that data in the

Boston Fed study showing similar foreclosure rates in white and minority neighborhoods disprove the finding of discrimination. Their argument was given stature by Becker (1993) and continues to be repeated in media commentaries; but as pointed out by Galster (1993), Carr and Megbolugbe (1994), and Tootell (1993), and acknowledged by Berkovec, Canner, Gabriel, and Hannan (BCGH 1994b), themselves proponents of default analysis, average default rates tell almost nothing about discrimination. Moreover, the foreclosure rates presented in the Boston study are not the appropriate data for an analysis of racial default experience.

Most advocates of default analysis believe it provides a way of sidestepping the problem of determining what factors the lender considered when deciding whether to approve or deny a mortgage application. Instead of worrying about the relevance of loan-to-value ratios, obligation ratios, or credit history, one simply looks at default rates. If minorities are being treated unfairly, the argument runs, lenders are passing up profitable loans to minorities with low default probabilities while making less profitable loans to whites with a greater likelihood of default. Accordingly, the average default rate for successful minority borrowers will be lower than for whites if discrimination is occurring—and the absence of a lower minority default rate can be taken as evidence that lenders are not discriminating.

Appealing as this reasoning may seem, it depends critically on three assumptions. If any one of the three fails to hold, no inferences about discrimination can be made from average default rates. All three assumptions are open to question on general principles, and the third was demonstrably incorrect for Boston-area borrowers.

Assumption one is that the loan originators know with a high degree of precision what determines defaults.[3] But little hard information exists on what causes defaults. Because home mortgages are seen as very low-risk, not much effort has been expended on monitoring loan performance. Most lenders have not been tracking the determinants of defaults. Nor has the secondary market; a study of the FHA experience was recently completed, but that broke new ground (BCGH 1994b).

One indication of the paucity of data is that the only information cited by critics in support of their argument is a table in the Boston Fed study showing foreclosure rates in City of Boston neighborhoods. These data on foreclosure rates by racial composition of the neighborhood were intended as a measure of neighborhood risk and cannot reveal much about racial default probabilities. The foreclosure figures are not limited to owner-occupied, one- to four-family properties; they include a number of instances of multiple properties owned by the same individual. Nor is the race of the property owners known. Whites may be property owners in minority neighborhoods and blacks and Hispanics may own property in white areas. In particular, the geographic distribution of the homes being purchased by the black and Hispanic mortgage applicants in the Boston study differed quite substantially from the distribution of the black and Hispanic population, with over half of the minority applicants planning to purchase in predominantly white areas.[4]

Assumption two is that discrimination in mortgage markets consists of requiring minority applicants to be more creditworthy than white applicants. In other words, the maximum probability of default that lenders will accept is lower for minority applicants than for white, resulting in a pool of approved minority applicants that, on average, has a lower probability of default than the white pool. One can, however, postulate forms of discrimination in which minority rejections are not concentrated among more marginal applicants and which need not, therefore, result in a lower average probability of default. Discrimination could be random, for example, if it arose because some white loan officers simply disliked blacks and Hispanics. Even reactions to different styles of speech or dress and misunderstandings of cultural differences in communication, such as the significance of looking people in the eye, could lead to rejections of minority applicants across the entire spectrum of default probabilities.[5]

[3] For profit-maximizers, the relevant consideration is really the expected profitability of the loan and not simply the likelihood of default. Profitability also depends upon the cost of making the loan and, if the loan is kept in portfolio, the probability of prepayment.

[4] A further concern is the use of foreclosure rates as the measure of default probability. Borrowers default and lenders foreclose; and while foreclosures may be more closely related than defaults to the lender's primary objective, profitability, the use of foreclosures rather than an indicator of borrowers' failure to pay introduces complicating issues such as resale opportunities and the lender's foreclosure policies. If the housing market is healthy, borrowers who fall into default will sell their properties themselves rather than experience foreclosure. This can be seen in the Boston foreclosure data. Foreclosures were very infrequent until 1990, when the housing market softened.

[5] In this context, it is worth noting that the Boston Fed's analysis implies that mortgage applicants who were not approved because of their race represent less than 10 percent of all minority applicants. A figure of this magnitude could plausibly be explained by random acts of prejudice.

Table 1
Effect of Distribution of Default Probabilities on Average Default Rates—An Example

Default Category	Average Default Probability	White: Applicants (percent)	White: Approved Borrowers (percent)	Minorities with Equal Default Probabilities: Applicants (percent)	Minorities with Equal Default Probabilities: Approved Borrowers (percent)	Minorities with Higher Default Probabilities: Applicants (percent)	Minorities with Higher Default Probabilities: Approved Borrowers (percent)
0–1	.5	20.0	22.2	20.0	26.7	5.0	7.1
1–2	1.5	25.0	27.8	25.0	33.3	15.0	21.4
2–3	2.5	15.0	16.7	15.0	20.0	30.0	42.9
3–4	3.5	15.0	16.7	15.0	20.0	20.0	28.6
4–5	4.5	15.0	16.7	15.0	0	15.0	0
>5		10.0	0	10.0	0	15.0	0
Total		100.0	100.0	100.0	100.0	100.0	100.0
% Denied		10.0		25.0		30.0	
Average Default Rate			2.3		1.8		2.4

Note: This is a hypothetical example. Highlighted numbers indicate denials.

Assumption three is that the distribution of default probabilities is the same for white applicants and for minority applicants who meet the white default standard. Only if the distributions are the same will requiring minorities to meet a more stringent standard necessarily result in accepted minority applicants having a lower average default rate than successful white applicants; and only then can the absence of lower default rates be taken as evidence that discrimination is not occurring. This assumption is not valid for the mortgage applicants examined in the Boston study; the distributions are not the same.

The importance of this assumption is illustrated in Table 1, which is based loosely on an example in Galster (1993). The numbers in the table are hypothetical. Column (3) shows the distribution of white mortgage applicants according to their expected probabilities of default. Thus, 25 percent of white applicants have an expected probability of default between 1 and 2 percent, with an average probability of default of 1.5 percent. If lenders are willing to approve applications with default probabilities up to 5 percent, the distribution of approved white borrowers will be that shown in column (4); and if lenders have accurately assessed the default probabilities, the average default rate will be 2.3 percent.

Columns (5) and (6) show what happens if the distribution of creditworthiness is the same for minority and white applicants and lenders discriminate against minority applicants by requiring a default probability of 4 percent or less rather than the more lenient 5 percent cutoff used for white applicants. The result is that 25 percent of minority applicants are rejected, compared to 10 percent of white applicants, and the average default rate is 1.8 percent—lower than the white average. Thus, in this case, the average default rate does reveal the existence of discrimination.

In columns (7) and (8), however, the distributions of default probabilities are not the same for minority and white applicants. There are fewer minority applicants with very low default probabilities, more in the 2 to 4 percent range. Minority applicants are still subjected to a tougher standard than white applicants, with minority applicants whose default probabilities are 4 to 5 percent being rejected while similarly risky white applicants are approved. Now, however, even though discrimination is occurring, the average default minority rate is 2.4 percent, higher than the white rate. Thus, in this case, a comparison of average default rates would lead to the false conclusion that no discrimination is taking place.

The distributions of characteristics thought to indicate default probabilities are not the same for the minority and white applicants examined in the Boston Fed's study. As shown in Tootell (1993), the situation is more like that shown in columns (7) and (8), with minority applicants tending to be concentrated in the lower ranges of the acceptable creditworthiness spectrum (higher default probabilities). This is also the message of Table 2, which presents key economic characteristics for the approved minority and white applicants in the Boston Fed's data set. Even though

Table 2
Key Characteristics of Approved Mortgage Applicants in Boston Fed Study

Variable	White	Black/ Hispanic
Ability to Support Loan		
Housing Expense/Income (percent)[a]	26.0	26.0
Total Debt Payments/Income (percent)[a]	33.0	34.0
Net Wealth ($)[a]	93,000	39,000
Monthly Income ($)[a]	4,666	3,333
Liquid Assets ($)[a]	38,000	19,000
Risk of Default		
Percent with Poor Credit History[a]	14.6	23.4
Probability of Unemployment	3.2	3.2
Percent Self-Employment	12.0	7.5
Potential Default Loss		
Loan/Appraised Value (percent)[a]	77.3	85.0
Rent/Value in Tract (percent)	4.6	7.3
Percent Applied for Private Mortgage Insurance	21.6	42.2
Percent Denied Private Mortgage Insurance[c]	.7	1.3
Loan Characteristics		
Percent Purchasing Two- to Four-Family Homes	7.7	24.8
Percent Fixed-Rate Loans	68.6	60.6
Percent 30-Year Loans	85.9	91.1
Percent in Special Loan Programs	12.6	40.6
Personal Characteristics		
Age[a]	34.0	36.0
Percent Married	63.0	53.7
Percent with Dependents	37.6	52.6

[a]Median value.
[b]Poor credit defined as having more than two late mortgage payments or delinquent consumer credit histories (more than 60 days past due) or bankruptcies or other public record defaults.
[c]Base is those applying for private mortgage insurance.
Source: MBMT (1992), Table 4.

minority applicants experienced discrimination and were denied more frequently than white applicants with the same characteristics, a larger fraction of white applicants with strong economic characteristics meant that the median values for most attributes were less favorable for the pool of accepted minority applicants than for the pool of accepted whites.[6]

[6] Some researchers are attempting to address the problem of differing applicant characteristics by comparing white and minority default experience, holding constant obligation ratios and other expected indicators of default experience. Thus, they face the same challenge as the Boston study—trying to include all the factors that could affect default probabilities. And even the most thoughtful of these studies have been handicapped by the absence of such key variables as creditworthiness (BCGH 1994b).

In summary, while comparing average default rates seems simple and has intuitive appeal, it cannot disprove the existence of discrimination unless the populations being considered have the same distributions of economic characteristics. Discrimination occurs when minority applicants are turned down more frequently than white applicants with the *same* characteristics and likelihood of default. Average default rates mix together many applicants with very different characteristics, and thus reveal very little about how individual minority and white applicants with the same characteristics are treated.

Omitted Variables

The possibility of omitted variables is a problem in almost all regression analyses, and the Boston Fed study is no exception. Closely related to this issue are questions of specification and goodness of fit, which are addressed in the following sections.

If an important explanatory variable is excluded from a regression, the researcher risks drawing a false conclusion because the influence of the omitted variable may be wrongly attributed to some other variable that was included in the regression. In studies of discrimination, the concern is that the omission of a variable that is correlated with race, for example, income, could lead to a finding that discrimination is taking place, when income is really responsible for the outcome. At the same time, because discrimination is such an important and controversial issue, a finding that race affects the outcome tends to set in motion a search for alternative explanations, or "missing variables"; and a danger exists that variables that reduce the influence of race will be treated as the "true" explanation, without sufficient regard to their theoretical justification.

In a sense, the Boston study was motivated by a search for omitted variables. As already noted, when the HMDA data showing applicants' race were first released, the racial disparities in denial rates were seen by many as evidence of discrimination, particularly as the one economic variable collected, income, did not account for these patterns. The lending industry's response was that the disparities could be explained by missing variables, most particularly loan-to-value and obligation ratios and applicants' credit histories.

Thus, the Boston Fed set out to determine whether these omitted variables really were the answer. Could including them explain the correlation between race and loan denial? At the outset, this was expected to be a fairly simple task, as the same few variables were always cited in media accounts of the industry's position. After talks with lending industry representatives, regulators, and academics, however, the list of variables was lengthened considerably. Every effort was made to collect everything of relevance, precisely to avoid charges of omitting key variables.

In a sense, the Boston study was motivated by a search for omitted variables, and it made every effort to collect everything of relevance.

Because the list was so comprehensive, critics have not come forth with many suggested additions. Assertions about omitted variables have been fairly numerous, but the same ideas tend to be repeated. In a number of cases, the so-called missing variables are not missing at all, but appear in the alternative regression specifications presented in the study's appendix; in other cases, they are well proxied by other explanatory variables that were included in the study. For example, the study has been faulted for not taking account of the presence of co-signers and local economic conditions (Zandi 1993). In fact, a co-signer variable appears in one of the equations in the appendix and economic conditions are represented by industry unemployment rates in Massachusetts, housing values (in the loan-to-value ratio), housing appreciation rates, neighborhood foreclosure rates, and various applicant financial characteristics.

Probably the two most frequently repeated charges of omitted variables involve funds available for closing and the dollar amount of gifts received, both of which were first mentioned by Horne (1994a). In fact, the study collected information on the applicants' liquid assets, the variable that lenders told us was most relevant as an indicator of funds available for closing, as well as information on total assets and liabilities. The survey also asked "Does a gift or grant account for any part of the down payment?"

Regressions including the answer to the gift or grant question, as well as the applicant's net wealth (assets minus liabilities) and liquid assets appear in the study. Liquid assets is not statistically significant; nor is net wealth. If liquid assets has no effect, it is hard to see why funds available for closing, a slightly different rendition of the same concept, would change the results materially. The presence of a gift or grant was found to reduce the likelihood of denial, but it was not significant at the 5 percent level. The influence of race is not affected by its inclusion.[7]

That liquid assets and net wealth did not have a significant effect on the probability of getting a mortgage has caused some to question the plausibility of the study's results. The Boston Fed researchers were also surprised; but as pointed out in the study, loan originators had already told us not to bother with asset information as they paid it little attention. The problem is verification. The value of many assets is difficult to determine. A clear case in point is the value of equity in the applicant's existing home, which has also been suggested as a potentially important omitted variable (Horne 1994a). Ideally, an estimate of home equity should be reflected in the answer to the question on total assets; but until the house is actually sold, the precise value of the owner's equity is unknown. The value of a self-employed applicant's business can also be difficult to pin down.

Two other variables warrant discussion—the presence of unverifiable information, and whether the applicant's credit history met the institution's loan policy guidelines for approval. Questions about both were asked as part of the Boston Fed study, but they do not appear in any of the Boston Fed's regressions. The information was made available to regulators and researchers, however; and some analysts have included the responses in their regressions (Day and Liebowitz 1993; Horne 1994b; and Schill and Wachter 1994). The answers to both questions are correlated with minority status. As can be seen in Table 3, taking account of the presence of unverifiable information does not have much effect on the race coefficient. Including the credit history/loan policy guidelines variable reduces the size of the coefficient on race by about one standard deviation, but it remains significant beyond the 1 percent level.

[7] As Engelhardt and Mayer (1994) showed in subsequent work with this data set, the probability of receiving a gift or a grant is not correlated with the applicant's race. It should also be noted that the primary effect of a gift is to enable the applicant to make a larger down payment; thus, the influence of gifts is captured in the loan-to-value and obligation ratios, which do appear in the Boston Fed regressions and are statistically significant. Apart from the effect on the down payment, the effect of a grant or gift is somewhat ambiguous, according to some of the loan originators consulted. They would prefer to see borrowers accumulate funds on their own.

Table 3
Original Regression[a] Results, and Results Adding Certain Additional Variables

Variable	Base Equation	Denial Regression with Unverifiable Information	Denial Equation with Credit History Guidelines	Dependent Variable: Credit History Guidelines
Constant	−7.70 (−15.25)	7.88 (−14.64)	−6.40 (−11.73)	−9.37 (−14.64)
Housing Expense/Income	.48 (3.12)	.49 (2.93)	.37 (2.06)	.64 (3.34)
Total Debt Payments/Income	.05 (6.62)	.05 (5.79)	.04 (5.33)	.02 (3.55)
Net Wealth	.0001 (1.28)	.0001 (1.58)	.00002 (.22)	.0001 (1.36)
Consumer Credit History	.32 (9.26)	.31 (8.45)	.008 (.16)	.69 (15.63)
Mortgage Credit History	.32 (2.69)	.28 (2.13)	.10 (.69)	.61 (4.02)
Public Record History	1.15 (6.43)	1.28 (6.77)	.23 (.98)	1.86 (9.36)
Industry Unemployment Rate	.08 (2.85)	.08 (2.67)	.07 (2.15)	.07 (1.81)
Self-Employed	.52 (2.74)	.50 (2.51)	.58 (2.71)	.24 (.93)
Loan/Value	2.01 (4.53)	2.06 (4.39)	1.63 (3.28)	2.23 (4.01)
Denied Private Mortgage Insurance	4.54 (9.17)	4.54 (9.00)	4.53 (9.00)	2.24 (6.65)
Rent/Value in Tract	.68 (3.51)	.62 (3.06)	.65 (3.12)	.54 (1.92)
Two- to Four-Family Home	.48 (2.89)	.51 (2.83)	.71 (3.74)	−.12 (−.54)
Unverifiable Information		3.03 (13.10)		
Credit History Guidelines			3.50 (15.55)	
Race	.62 (4.37)	.57 (3.76)	.47 (2.85)	.78 (4.51)
Log of Likelihood	−838.84	−746.19	−683.04	−537.02
Number of Observations	2925	2925	2922	2922

Figures in parentheses are t-statistics.
[a]The base equation in Table 3 is different from that found in the most recent version of the Boston Fed study, forthcoming in *The American Economic Review*. This specification is used here to ensure comparability with the critics' work.

These variables differ from all the other variables collected in that the answers are not based on objective criteria. They cannot be found in any of the boxes on the standard loan application form. They involve judgment by the individual completing the survey. This is particularly a problem with the question on loan policy guidelines, where the respondent is being asked to make an evaluation after the fact, knowing whether the application was approved or denied. Although the question refers specifically to credit history, respondents appear to have interpreted it to mean "did the sum total of applicant characteristics meet the institution's guidelines for approval?" Thus, unsuccessful applicants with virtually no credit problems, according to the objective measures of credit history, are recorded as failing to meet credit history guidelines.

That the answers to this question are, in large part, proxies for the lending decision is apparent from the fourth column of Table 3, which presents a regression in which the failure of credit history to meet loan policy guidelines replaces loan outcome as the dependent variable. As can be seen, the lender's evaluation of credit history is a function of variables having nothing to do with credit history—loan-to-value and obligation ratios—and is very strongly correlated with race.

Why were these questions asked, if the responses are so problematic? Both questions date from the early planning of the study and were suggested by people with an examination perspective. In particular, the phraseology of the question on loan policy guidelines was taken directly from examiners, who can look at loan files themselves to confirm that applicants' credit histories are indeed consistent with the institution's policy guidelines. Its implications for the current purpose, where the researchers would not have direct access to loan files, were not recognized.

A final possibility, which seems to underlie much of the skepticism about the Boston Fed's results and which some observers may think is supported by the work of Stengel and Glennon (1995), is that the race effect is attributable, not to one or two missing variables, but to many idiosyncratic or near-idiosyncratic factors. In other words, every borrower is unique; something always exists to distinguish one borrower from another. This view amounts to a rejection of statistical tests of discrimination. Although idiosyncratic factors may appear important to individual applications, quantifying their importance is impossible since, by definition, each factor affects so few applications. Thus, including idiosyncratic factors in any regression analysis would be equivalent to having dummy variables for individual observations. Moreover, even if idiosyncratic factors are important, no explanation has yet been offered for why the idiosyncratic features of minority applicants would be less favorable than those of white applicants *with similar economic characteristics, purchasing similar properties.*

Misspecification

Regressions are said to be misspecified not only when variables are missing, but also when variables are included in a manner that does not accurately reflect their influence on the dependent variable. In the case of the lending decision, for example, the obligation ratio may be influential only at values above a certain threshold. If the researcher forces the variable to have the same coefficient at all values, the role of the variable will be misstated.

Alternative specifications of the Boston Fed model almost invariably confirm that race affects the probability of mortgage denial.

Considerable effort has gone to trying to rework and recombine the Boston study variables, but a careful reading of the relevant studies indicates that these efforts to improve the specification do not change the results appreciably. As noted in the discussion of missing variables, in addition to the base equation presented in the main text, the study includes a lengthy appendix with a number of alternative specifications. These alternatives answer most of the specification questions that have been asked. For example, some researchers have critiqued the weighting given the credit variables in the base equation, arguing that such a weighting was subjective and restrictive (Horne 1994a). To simplify exposition, the authors did assign a ranking to the seriousness of credit transgressions and collapse the answers to nine questions pertaining to consumer credit and mortgage credit history into two variables. But an alternative treatment of credit history that imposes no such restrictions appears in the appendix. The weights estimated by this equation are consistent with the ranking in the base equation. And the coefficients for other variables, including race, are virtually unchanged.

In a similar vein, the Boston study has been faulted for including applicants who were denied private mortgage insurance in its base equation. The argument is that those denied private mortgage insurance were not rejected by the lenders, but by the mortgage insurers and, therefore, it is unfair to lenders to include these applicants in the analysis (Horne 1994a and 1994b). This issue receives extensive discussion in the Boston Fed's study. Indeed, it could be said that the Boston study highlighted the important role played by mortgage insurance and drew attention to the question of its treatment. As shown in the appendix to the study, however, excluding from the analysis those denied private mortgage insurance does not alter the effect of race on the probability of being turned down for a mortgage.

Endogeneity

A very different specification criticism has been made by Yezer, Phillips, and Trost (1994) (YPT). They argue that the loan application process is actually a negotiation, with some of the explanatory variables modified in response to the probability of denial. For example, upon learning that they were going to be rejected, applicants might respond by increasing their down payments, thereby lowering their loan-to-value ratios. In YPT's view, the mortgage decision and the setting of loan terms occur simultaneously, and failing to treat this process as a simultaneous system imparts a downward bias to the estimate of the loan-to-value coefficient and could bias upwards the race coefficient if the loan-to-value ratio depends upon minority status.

How serious an issue is this? YPT try to demonstrate that the potential bias could be quite large using a pseudo data set (their terminology) based on the Boston Fed's data. Despite the link to the Boston Fed's data, this is a theoretical exercise; the pseudo data are constructed using only a few of the many explanatory variables used in the Boston Fed's analysis.

YPT do not present any evidence on the prevalence or extent of negotiation and there is reason to doubt that it is widespread. Negotiation requires that borrowers have the flexibility to respond to the information that they face a high probability of denial. But many borrowers, white and minority, will be constrained from negotiating by their financial circumstances. In the case of the loan-to-value ratio, many prospective homebuyers will be unable to increase their down payments because they were already planning to buy as much house and make as large a down payment as their savings could support. Data from the Survey of Income and Program Participation (SIPP) indicate that home-owners who have recently purchased possess very little in the way of liquid assets and, therefore, presumably did not have an excess that could have been used to bolster their down payment. Specifically, nationwide, among home-owners who owned their homes for less than two years, the median liquid assets in 1990 were only \$1036.[8]

A more fundamental objection to the YPT argument is that negotiation need not imply simultaneity; indeed, true simultaneity that would bias the race coefficient is difficult to envision. No simultaneity problem exists if borrowers anticipate in advance that a weak credit history or other problem reduces their chance of getting a loan and compensate by increasing their down payment. Nor does it present a problem if the lender provides counseling to that effect, as long as the loan-to-value ratio is determined before the final decision is made. The lending process in such a case is sequential and the system of equations is considered block recursive. For any bias to result, the loan decision and the loan-to-value ratio must be determined at the same time—simultaneously. This means that the applicant must know the approval or denial outcome as the loan-to-value ratio is being determined; or in econometric terms, the error term in the mortgage denial equation must be realized and must affect the loan-to-value ratio.

The standard econometric approach to dealing with simultaneity issues is instrumental variables.[9] This technique involves replacing the variable believed to be simultaneously determined, here the loan-to-value ratio, with the estimated values from a regression on variables that are not simultaneously determined but are still correlated with loan-to-value. Possible candidates for instruments in this case include income and potential experience (age less years of schooling).[10] As can be seen from Table 4, replacing the loan-to-value ratio with the fitted values of an equation on these variables and all the variables in the base denial equation except loan-to-value leaves the statistical significance of race unaffected.

Table 5 presents another test of whether simultaneity between the denial decision and the loan-to-value ratio is responsible for the finding that race affects the mortgage outcome. If the race coefficient is actually picking up the effect of the loan-to-value ratio, increasing or decreasing the influence of the loan-to-value ratio should alter the race coefficient substantially. Table 5 tests the sensitivity of the race coefficient to bias in estimating the loan-to-value coefficient by constraining the loan-to-value coefficient to be approximately two standard deviations above and below the estimate in the base equation. As can be seen, these drastic changes have little effect on the size or significance of the coefficient on race, suggesting that any bias in the coefficient for the loan-to-value ratio has little effect on the race coefficient.

In sum, while one cannot dismiss the possibility of some feedback from lender to loan applicant, a truly simultaneous determination of loan terms and mortgage denial seems doubtful on both conceptual and econometric grounds. One irony is that while YPT believe that negotiation precludes a finding of racial discrimination, others have hypothesized that negotiation itself is an important source of discrimination, with lenders offering white applicants more opportunity and guidance to improve their applications. Such coaching is popularly referred to as the "thicker file" phenomenon, with coached white applicants having thicker files than their black and Hispanic counterparts because of explanatory letters and revised applications. Contrary to some impressions, the Boston Fed study shed no light on the existence of coaching.

[8] Calculated using the U.S. Bureau of the Census, *Survey of Income and Program Participation (SIPP) 1990 Panel, Wave 4 Core Microdata File* (1991).

[9] Rachlis and Yezer (1993) reject the instrumental variables approach to correcting for simultaneity on the grounds that not just loan-to-value but almost all the terms in the denial equation that are related to the mortgage contract are potentially endogenous.

[10] The choice of instruments is somewhat limited. The instruments should not be in the denial equation, but most of the variables in the Boston Fed's data set were collected because they were thought to affect denials. Potential experience, or age less years of schooling, was not mentioned by loan officers as a factor considered in approving mortgages, but might have some bearing on the loan-to-value ratio as more experienced applicants would have had more time to save up their down payments. Income was collected primarily as a fallback in case obligation ratios from the lenders' worksheets were not available. As noted in the section on omitted variables, the Boston Fed's study was undertaken largely because income did not explain the racial disparities in mortgage denials. Moreover, lenders said it should not—that their real concern was obligation ratios.

Table 4
Applying Instrumental Variables to the Loan/Value Ratio with Income and Potential Experience as Instruments

Variable	Base Equation	Instrumenting for Loan/Value
Constant	−7.70 (−15.25)	−5.47 (−4.38)
Housing Expense/Income	.48 (3.12)	.47 (3.02)
Total Debt Payments/Income	.05 (6.62)	.05 (6.71)
Net Wealth	.0001 (1.28)	.00005 (.63)
Consumer Credit History	.32 (9.26)	.33 (9.43)
Mortgage Credit History	.32 (2.69)	.44 (3.09)
Public Record History	1.15 (6.43)	1.22 (6.50)
Industry Unemployment Rate	.08 (2.85)	.08 (2.83)
Self-Employed	.52 (2.74)	.46 (2.41)
Loan/Value	2.01 (4.53)	−1.51 (−.77)
Denied Private Mortgage Insurance	4.54 (9.17)	4.88 (9.16)
Rent/Value in Tract	.68 (3.51)	.68 (3.47)
Two- to Four-Family Home	.48 (2.89)	.57 (3.09)
Race	.62 (4.37)	.82 (4.46)
Log of Likelihood	−838.84	−839.22
Number of Observations	2925	2893

Table 5
Testing for Simultaneity by Altering Loan/Value Coefficient

Variable	Low Loan-to-Value Coefficient	High Loan-to-Value Coefficient
Constant	−6.86 (−18.36)	−8.19 (−21.46)
Housing Expense/Income	.47 (3.06)	.48 (3.09)
Total Debt Payments/Income	.05 (6.67)	.05 (6.46)
Net Wealth	.00009 (1.46)	.0001 (1.68)
Consumer Credit History	.31 (9.21)	.31 (9.24)
Mortgage Credit History	.33 (2.83)	.29 (2.44)
Public Record History	1.17 (6.58)	1.14 (6.40)
Industry Unemployment Rate	.08 (2.88)	.08 (2.92)
Self-Employed	.47 (2.51)	.51 (2.68)
Loan/Value	1.00[a]	2.80[b]
Denied Private Mortgage Insurance	4.57 (9.32)	4.47 (9.00)
Two- to Four-Family Home	.54 (3.31)	.49 (2.93)
Race	.70 (5.06)	.62 (4.40)
Log of Likelihood	−846.48	−845.58
Number of Observations	2925	2925

[a]Constrained to be 1.00 (about two standard deviations below base estimate).
[b]Constrained to be 2.80 (about two standard deviations above base estimate).

Goodness of Fit

Some of those who assert that the Boston Fed study is misspecified have tried to support their claims by arguing that the model does not explain the data very well. In particular, Horne (1994a) has argued that simply assuming that every applicant is approved would result in a correct prediction for 85 percent of the outcomes, since 85 percent of applications were approved. Of course, the 15 percent of the applications that one is most interested in explaining—the denials—would be 100 percent wrong.

For equations that estimate probabilities, no simple goodness of fit measures exist that are comparable to the familiar R^2 associated with ordinary least squares regressions. The reason is that the actual outcome is discrete, in this case, whether the application was denied or not, while the estimated outcome is a probability of denial. In other words, the equation

Table 6
Goodness of Fit—Probability and Actual Frequencies of Denials

Probability Range	Predicted Denial Frequency	Actual Denial Frequency
0%–10%	4.5	4.4
10%–20%	13.8	12.2
20%–30%	24.4	24.5
30%–40%	34.5	37.6
40%–50%	45.2	56.1
50%–60%	55.2	62.2
60%–70%	63.9	64.7
70%–80%	73.6	77.8
80%–90%	85.5	80.0
90%–100%	96.8	92.8

does not predict whether a particular applicant will be approved or denied, but the fraction of applicants with those characteristics that will be denied. This is a critical but difficult distinction, as many people instinctively view an estimated probability of denial of less than 50 percent as a prediction that the application will be approved. Thus, Horne (1994a) faults the Boston Fed model "because two-thirds of the applications that were [actually] denied were predicted to be approved on the basis of a 50 percent probability threshold."

An estimated probability of denial of less than 50 percent is *not* a prediction of approval.[11] Rather, it is a prediction of an approval (or denial) *rate*. Thus, a probability of denial of, say, 20 percent is a prediction that four out of five applicants with certain characteristics will be approved—not that any individual application will be approved and certainly not that all will be approved.

Table 6 breaks the applications down according to the model's estimated probabilities of rejection. It then compares the actual incidence of rejection with that predicted according to the model, taking a 10 percent probability of denial to mean that one of 10 is denied and a 50 percent probability of denial to mean that one out of two is denied. As can be seen, the predicted denial rates fit the data well.

[11] It should be acknowledged that the logit regressions produce, as a routine matter, an estimate of "percent correct predictions," which is based on a 50 percent threshold, and the Boston study did present this measure in its regression tables. The Boston Fed did not refer to this statistic in evaluating the performance of the model, however. We pointed out the difficulty of assessing fit and compared actual with estimated denial rates according to obligation ratio.

Part of the same issue is the claim, again associated most closely with Horne, that the model gives insufficient weight to what he considers to be serious application weaknesses. The model cannot be very good, he argues, because liquid assets was not significant in the regressions and because credit history and obligation ratios, while important and significant, were not so important that individuals with very poor credit histories were automatically disqualified from getting mortgages.

There are two problems with this reasoning. One is the critics' insistence on characterizing estimated probabilities of denial of less than 50 percent as predictions of approval. In their minds, the model does not say an application is weak unless the probability of denial is greater than 50 percent. The second problem is that their claims are contradicted by the data. Bankruptcies and public records of credit problems are not automatic deal-breakers, nor are very high loan-to-value ratios. This is confirmed by examiners' reviews of some of the loan files used in the Boston Fed study. Even though these examiners sometimes felt that the model was not placing enough weight on certain variables, they were often able to find approved applications with what seemed like deal-breaker problems.[12] The regression estimates reflect the fact that exceptions are made.

Data Errors

The study has been criticized for "data errors", with the implication that these errors account for the influence of race. It is important to recognize that transcription or other random errors in the explanatory variables would not normally impart a bias to the race coefficient. The most likely consequence would be large standard errors and reduced statistical significance. A very thorough and objective review of the Boston Fed's data by Carr and Megbolugbe (1994) concluded that, despite some suspected errors, the Boston Fed's results held up.

Most of the charges of errors stem from Liebowitz (1993) and Horne (1994a and 1994b), both of whom misuse the term "data error." Both apply it to observations that clearly are not data errors. Horne uses the term data error to describe the action taken on rejected counteroffers. These are classified as denials according to regulation and appear as such in the lenders'

[12] Source: Private communication between representatives of one of the federal supervisory agencies and Boston Fed researchers.

HMDA submissions and the Boston study. According to Horne, however, such rejected counteroffers should really be viewed as approvals. (Horne (1994b) had not been published at the time of writing but has been widely circulated and quoted in the press.)

Horne also characterizes as data errors the action taken on applications that were turned down because the applicant was rejected for private mortgage insurance or did not meet the qualifications for a special program. The influence of private mortgage insurance and special programs should be taken into account, as indeed they were in the Boston Fed's study; but there is no getting around the fact that these applications were denied. The applicants did not get the loan. Yet Horne would have these loans treated as approvals. He characterizes the present treatment as a data error and even goes so far as to re-code these outcomes as approvals in some of his regressions (Horne 1994b).

Liebowitz labels any application with a loan-to-value ratio exceeding 80 percent and no application for mortgage insurance a data error. Because Fannie Mae generally requires mortgage insurance on high loan-to-value loans, the existence of such applications, he argues, is proof of error. But while the secondary market usually requires mortgage insurance on such loans, exceptions can be made. More importantly, many of these applications were denied and others were kept in the lenders' portfolio and, thus, not subject to secondary market guidelines.

Liebowitz also characterizes as data errors any observation that looks unusual. Thus, he cites as obvious examples of errors applicants who were approved for loans despite having negative net worth. This is an effective rhetorical technique since, at first glance it does seem odd that someone with negative net worth would be approved for a loan. On reflection, however, one can posit many reasons for why a negative net worth would not preclude receiving a loan, particularly as the net worth figures do not include the value of human capital.

Table 7
Effect of Removing Outliers

	Base Equation	Excluding Outliers[a]
Constant	−7.70	−8.39
	(−15.25)	−(14.53)
Housing Expense/Income	.48	.38
	(3.12)	(2.33)
Total Debt Payments/Income	.05	.06
	(6.62)	(6.94)
Net Wealth	.0001	.0001
	(1.28)	(1.22)
Consumer Credit History	.32	.32
	(9.26)	(8.97)
Mortgage Credit History	.32	.30
	(2.69)	(2.33)
Public Record History	1.15	1.27
	(6.43)	(6.88)
Industry Unemployment Rate	.08	.08
	(2.85)	(2.82)
Self-Employed	.52	.64
	(2.74)	(3.27)
Loan/Value	2.01	2.30
	(4.53)	(4.60)
Denied Private Mortgage Insurance	4.54	4.75
	(9.17)	(8.59)
Rent/Value in Tract	.68	.70
	(3.51)	(3.39)
Two- to Four-Family Home	.48	.47
	(2.89)	(2.64)
Race	.62	.62
	(4.37)	(4.14)
Log of Likelihood	−838.84	−766.27
Number of Observations	2925	2741

[a]Loan-to-value ratio is between 10% and 150%. Total obligation ratio is between 10% and 80%. Net wealth is positive.

Boston Fed researchers were concerned about extreme values distorting our results and ran many regressions with and without unusual observations. The findings hold up. Table 7 is an example of the type of test that was performed. Carr and Megbolugbe (1994) and Glennon and Stengel (1994) have essentially replicated the Boston Fed's results, making adjustments for what they considered to be extreme values.

The question remains: should extreme values be treated as errors? We think not. So while we tested the sensitivity of our results to these observations, we chose to leave them in the data set. As noted previously, lenders were called to verify unusual values. In addition, lenders had an incentive to be accurate, since they were informed that their supervisory agencies would have access to their responses.

Many of the extreme observations pertain to assets and liabilities, which appear as net worth in the data set made available to the research community. As noted in the Boston Fed's study and earlier in this article, the loan officers consulted prior to the study recommended against collecting these data. The value of assets, in particular, was said to be hard to verify

Table 8
Results of Horne's Alterations to the Data

	Full Sample		FDIC Sample	
	Boston Fed Data	Horne's Data	Boston Fed Data	Horne's Data
Constant	−7.50	−7.83	−7.76	−8.13
	(−15.25)	(−15.40)	(−10.64)	(−10.74)
Housing Expense/Income	.48	.43	.52	.39
	(3.12)	(2.77)	(2.21)	(1.61)
Total Debt Payments/Income	.05	.05	.04	.05
	(6.62)	(6.93)	(4.51)	(4.99)
Net Wealth	.0001	.00002	.0001	.00002
	(1.28)	(.47)	(1.47)	(.58)
Consumer Credit History	.32	.31	.32	.30
	(9.26)	(8.93)	(5.94)	(5.28)
Mortgage Credit History	.32	.32	.56	.57
	(2.69)	(2.60)	(3.15)	(3.04)
Public Record History	1.15	1.24	1.23	1.47
	(6.43)	(6.97)	(4.08)	(4.91)
Industry Unemployment Rate	.08	.08	.06	.05
	(2.85)	(2.80)	(1.25)	(1.16)
Self-Employed	.52	.55	.19	.24
	(2.74)	(2.89)	(.66)	(.80)
Loan/Value	2.01	2.07	1.77	1.98
	(4.53)	(4.59)	(2.62)	(2.80)
Denied Private Mortgage Insurance	4.54	4.59	4.57	4.67
	(9.17)	(9.29)	(6.97)	(7.17)
Rent/Value in Tract	.68	.72	−.13	.21
	(3.51)	(3.75)	(−.18)	(.37)
Two- to Four-Family Home	.48	.53	.59	.66
	(2.90)	(3.13)	(2.26)	(2.50)
Race	.62	.55	1.06	.91
	(4.37)	(3.78)	(4.96)	(4.08)
Log of Likelihood	−838.84	−817.51	−357.19	−336.77
Number of Observations	2925	2925	1379	1379

and, consequently, the loan officers did not pay attention to assets. One claimed never to look at the back page of the application. Their thinking was borne out by our regression analysis, which found that the wealth variables had no effect on the decision to deny an application.[13]

Some obligation ratios (housing expense to income) look suspiciously low; but these typically are associated with two- to four-unit properties and reflect the lenders' treatment of rental income. Statistical tests do not justify splitting the sample according to type of property purchased; but omitting two- to four-unit properties from the sample and, thus, eliminating these low obligation ratios does not affect the results, as will be shown in the next section.

The third concentration of extreme values occurs in the lenders' original HMDA submissions. This information was reproduced for the lenders and they were instructed to check it for accuracy. But the HMDA variables which provoked the most criticism, income and whether the mortgage was sold, were not used in the analysis of mortgage lending by either the Boston Fed or its critics. A few observations appear to have loan amounts that were too high and thus some loan-to-value ratios may be in error. As noted already, however, the sensitivity of our results to extreme values was tested extensively.[14]

Interestingly, most of Horne's claims of data errors have not involved the unusual observations noted by Liebowitz. Horne's criticisms have been given particular credence because he had access to the notes of FDIC examiners who followed up the Boston Fed's study by

[13] This suggests several explanations for the extreme values of some of the assets and liabilities figures. One possibility is that the numbers accurately represent both what is recorded on the application form and the applicant's circumstances, but that loan officers disregard this information—perhaps because they, like the researchers, distrust extreme values. A second possibility is that the extreme values accurately reflect the application form but not the applicant's true circumstances because the loan officer, who planned to ignore these data, did not bother to ensure accuracy. Even if one could determine the applicant's true circumstances from the loan files, however, second-guessing the lender by claiming that this was the information considered rather than the information on the final loan application seems presumptuous.

[14] Discrepancies between the applicants' incomes reported on the HMDA submissions and in the Boston Fed data set do not represent data errors, contrary to assertions by Liebowitz (1993). The wording of the HMDA question pertaining to income is somewhat ambiguous and could be interpreted in different ways. This ambiguity was one of the reasons why the Boston Fed research team decided to collect monthly income figures and, more generally, to use the standard loan application form as the primary template upon which to base its survey.

looking at some of the loan files.[15] According to Horne, these notes cited numerous instances of data errors. When we obtained his corrected data set, however, his corrections were generally small or pertained to variables that were not important to the Boston Fed's results. They involved differences of a few dollars in income or assets or errors in gender. Table 8 shows the effect of Horne's changes on the results for the entire sample and for the FDIC institutions-only sample used in Horne's own regressions (1994b). The race coefficient remains large and statistically significant.

In summary, critics have been liberal with the term data errors, using it to describe what are clearly not errors.

Sample Size

The Boston Fed study's findings are very robust. Not only do the results hold up for all the alternative specifications presented in the study itself, but these findings have also been verified by others. Indeed, a careful look at our critics' own analyses shows that reasonable specifications and samples confirm the finding that race affects the mortgage decision. (For critics, see Day and Liebowitz 1993 and Horne 1994b. Among others, see Carr and Megbolugbe 1994.) Only the inclusion of the credit history guidelines response, which amounts to putting the dependent variable on the right-hand side, or removing minority observations from the sample, seriously undermine this conclusion. If the sample contains very few minority applications, finding a racial effect will be difficult.

The removal of minority observations from the sample will not be obvious to the casual observer because the total number of observations in the sample is large. But out of a total sample of roughly 3,000, only 700 were black and Hispanic applications, of which 200 were rejections. Thus, sample sizes that are large enough to assess the effect of, say, loan-to-value ratios may not be large enough to detect the influence of race, particularly if the paring down of the sample has occurred in such a way as to remove minority applications disproportionately.

Two arguments have been advanced for reducing the sample. The first is that lenders might view different categories of applicants differently. Purchasers of single-family homes might be treated differently from those buying condominiums, in which case splitting the sample could be justified (Liebowitz 1993 and Day and Liebowitz 1993). It is also possible that black applicants are treated differently from Hispanic, in which case the two groups should not have been lumped together as "minorities" in the Boston Fed study.

Table 9
Distinguishing Between Black and Hispanic Applicants

Variable	Base Equation	Splitting Black and Hispanic Applicants
Constant	−7.70 (−15.25)	−7.70 (−15.23)
Housing Expense/Income	.48 (3.12)	.48 (3.13)
Total Debt Payments/Income	.05 (6.62)	.05 (6.60)
Net Wealth	.0001 (1.28)	.0001 (1.28)
Consumer Credit History	.32 (9.26)	.32 (9.21)
Mortgage Credit History	.32 (2.69)	.32 (2.69)
Public Record History	1.15 (6.43)	1.15 (6.43)
Industry Unemployment Rate	.08 (2.85)	.08 (2.86)
Self-Employed	.52 (2.74)	.51 (2.73)
Loan/Value	2.00 (4.53)	2.00 (4.52)
Denied Private Mortgage Insurance	4.54 (9.17)	4.54 (9.16)
Rent/Value in Tract	.68 (3.51)	.68 (3.51)
Two- to Four-Family Home	.48 (2.89)	.48 (2.90)
Race	.62 (4.37)	
Black		.63 (4.02)
Hispanic		.58 (2.71)
Log of Likelihood	−838.84	838.81
Number of Observations	2925	2925

[15] FDIC examiners pulled a sample of the files for applications to FDIC-regulated institutions that were denied but for which the Boston Fed's primary regression estimated a probability of denial of less than 50 percent. There were approximately 100 such applications.

Table 10
Distinguishing Among Types of Property

Variable	Base Regression	Single-Family	Two- to Four-Family and Condos	Single-Family and Condos	Two- to Four-Family	Condos
Constant	−7.70 (−15.25)	−7.57 (−11.15)	−7.82 (−10.30)	−7.31 (−13.43)	−10.45 (−6.63)	−7.04 (−7.14)
Housing Expense/Income	.48 (3.12)	.54 (2.60)	.35 (1.54)	.52 (3.11)	.24 (.60)	.47 (1.60)
Total Debt Payments/Income	.05 (6.62)	.04 (4.03)	.06 (5.53)	.04 (5.28)	.08 (4.13)	.06 (3.71)
Net Wealth	.0001 (1.28)	.0001 (0.91)	.0001 (1.07)	.0001 (0.81)	.0004 (2.01)	.00005 (.50)
Consumer Credit History	..32 (9.26)	.32 (6.78)	.31 (6.15)	.31 (8.13)	.35 (4.14)	.28 (4.43)
Mortgage Credit History	.32 (2.69)	.27 (1.75)	.37 (1.91)	.27 (2.02)	.55 (1.87)	.21 (.73)
Public Record History	1.15 (6.43)	1.10 (4.53)	1.22 (4.57)	1.11 (5.61)	1.54 (3.34)	1.19 (3.44)
Industry Unemployment Rate	.08 (2.85)	.09 (2.62)	.08 (1.79)	.10 (3.11)	.02 (.27)	.12 (1.68)
Self-Employed	.52 (2.74)	.57 (2.28)	.51 (1.77)	.50 (2.40)	.56 (1.12)	.44 (1.12)
Loan/Value	2.01 (4.53)	2.23 (3.55)	1.81 (2.92)	1.80 (3.80)	4.50 (3.08)	1.08 (1.45)
Denied Private Mortgage Insurance	4.54 (9.17)	4.71 (7.35)	4.36 (5.65)	4.65 (8.40)	4.18 (3.87)	4.49 (4.01)
Rent/Value in Tract	.68 (3.51)	1.19 (1.86)	.61 (2.93)	.57 (2.48)	1.34 (3.48)	.49 (1.93)
Two- to Four-Family Home	.48 (2.89)					
Race	.62 (4.37)	.86 (4.34)	.46 (2.40)	.75 (4.75)	.08 (.25)	.51 (2.00)
Log of Likelihood	−838.84	−457.49	−381.18	−689.27	−142.62	−228.90
Number of Observations	2925	1782	1143	2532	393	750

These are empirical questions that can be resolved only by looking at the data. In its preliminary work, the Boston Fed looked at these and other breakdowns and found no statistical basis for dividing up the sample. These results were not presented in the study but, as can be seen in Tables 9 and 10, no statistical basis exists for distinguishing between black and Hispanic applicants, or between those buying single-family and less traditional homes.[16]

The second argument for splitting the sample is to test for the influence of one or two prominent lenders. Are the results being skewed by one "bad apple?" This is a reasonable question, but the answer is no. The Boston Fed ran the regressions excluding different subsets of lenders. Two of these regressions are presented in the study, one for the six largest lenders to minorities collectively and one for the sample exclud-

[16] A chi-square test of the log likelihood does not reject (1) that lenders treat multifamily and condo applications the same as single-family applications and (2) that lenders treat multifamily applications the same as single-family and condo applications.

ing the six institutions. For both groups, race affects the probability of being denied a mortgage.[17] Thus, the conclusions about the influence of race are not sensitive to the activities of one or two organizations.

Press reports have made much of Horne's (1994b) claim that two institutions that actively serve minority applicants account for the finding that race affects the mortgage decision. (See, for example, Passell (1994) and Macey (1994).) The implication is that aggressive outreach to more marginal applicants explains the influence of race. This interpretation is wrong. The Boston Fed's study took into account applicants' economic circumstances; thus, if "community outreach" resulted in a weaker pool of applicants, this weakness would be accounted for by the applicants' economic variables. The two institutions are important to the results only because they represent a large fraction of all minority applicants in Horne's subsample.

Horne's study looks at FDIC-regulated institutions only. Restricting the analysis in this way cuts the number of observations in half. Moreover, within the FDIC sample, the two institutions in question account for roughly one-half of the minority applications. Thus, the seemingly innocuous removal from the FDIC subsample of two institutions that actively serve minorities reduces the number of minority observations in the analysis to one-quarter its original size. Even so, race remains economically and statistically significant, as can be seen in Table 11. The deletion of additional observations, the invalid inclusion of the credit history guidelines variable, and a series of other changes, only some of which we have been able to replicate, are necessary to make the significance of the race coefficient fall below 5 percent, even in this subsample.

Table 11
Sample Excluding Horne's Two Influential Lenders

	Full Sample		FDIC Sample	
	Boston Fed Data	Horne's Data	Boston Fed Data	Horne's Data
Constant	−8.14 (−15.10)	−8.19 (−15.20)	−8.80 (−10.15)	−8.89 (−10.27)
Housing Expense/Income	.48 (2.99)	.44 (2.68)	.47 (1.73)	.36 (1.28)
Total Debt Payments/Income	.05 (6.26)	.05 (6.63)	.04 (3.96)	.05 (4.56)
Net Wealth	.0001 (1.39)	.0001 (1.71)	.0001 (1.89)	.0001 (1.97)
Consumer Credit History	.32 (9.01)	.31 (8.74)	.32 (5.33)	.31 (4.98)
Mortgage Credit History	.29 (2.37)	.29 (2.34)	.51 (2.67)	.51 (2.64)
Public Record History	1.18 (6.40)	1.20 (6.47)	1.34 (4.11)	1.38 (4.17)
Industry Unemployment Rate	.08 (2.78)	.08 (2.75)	.05 (1.00)	.05 (.94)
Self-Employed	.59 (3.06)	.58 (2.96)	.42 (1.39)	.36 (1.15)
Loan/Value	2.58 (5.45)	2.52 (5.29)	3.29 (3.99)	3.22 (3.84)
Denied Private Mortgage Insurance	4.66 (8.39)	4.69 (8.47)	4.85 (6.00)	4.91 (6.06)
Rent/Value in Tract	.71 (3.65)	.72 (3.71)	−.66 (−.57)	−.55 (−.48)
Two- to Four-Family Home	.53 (2.99)	.53 (2.99)	.70 (2.30)	.73 (2.37)
Race	.44 (2.87)	.44 (2.86)	.70 (2.70)	.66 (2.49)
Log of Likelihood	−771.51	−761.81	−289.37	−281.71
Number of Observations	2799	2799	1253	1255

Another example of this whittling down of minority observations occurs in Liebowitz (1993). Liebowitz first splits the sample according to type of property being purchased, and then further splits the applications for single-family homes into those applying for private mortgage insurance and those who do not need it. He then focuses on the last group, which he characterizes as the "core" sample and for which the race coefficient remains highly significant, and asserts that removing "six extremely influential applications," all minority rejections, causes all evidence of discrimination to vanish. What he neglects to say is

[17] The appendix of the Boston Fed study also includes a regression in which each lender is represented by a dummy variable to allow for differences in lending standards. The race variable remains statistically significant.

that his "core" sample had already been reduced to 14 minority rejections and he had to remove almost half of them to achieve this outcome.[18]

Summary and Conclusion

The Boston Fed's 1992 study of mortgage denials has attracted a great deal of attention, some of which has been critical. This criticism has focused on racial differences in default rates, missing variables, misspecifications of the model, and data integrity. Of these, the default issue has perhaps received the greatest attention in the media, despite a number of articles pointing out that simple comparisons of average default rates cannot disprove the existence of discrimination in the mortgage approval process.

The specification issue may have received the most attention from professional economists. Various teams have tried alternative specifications of the Boston Fed model. Almost invariably these confirm that race affects the probability of denial. Only if the sample is split up and the number of minority observations sharply reduced does the statistical significance of race go away. The argument that the mortgage decision involves negotiation may have some validity, but a truly simultaneous determination seems unlikely.

The most frustrating criticism, from the Boston Fed's viewpoint, has been the charge that data errors undermine the study. Many of the alleged data errors are not errors at all. In other cases, outliers are called errors with no evidence. The Boston Fed made many calls back to lending institutions to confirm that suspicious values were, indeed, what was on the application form. We also tested the sensitivity of our findings to extreme values. Researchers who have worked with the publicly available data base can confirm that questionable observations do not affect the finding that race influenced the lending decision.

Some of the criticism of the Boston Fed study has been scholarly; but many of the critiques appear to be motivated or at least energized by opposition to policies the study may have inspired. The study advocated no policies, although the Boston Fed subsequently published a guide for lenders suggesting ways to ensure fair treatment of applicants of different races and cultures. This guide, entitled *Closing the Gap*, has been very well received by the lending industry, with more than 80,000 copies distributed as of mid-1995.

The study's primary contribution was to tell the lending industry that it had to face up to the task of ensuring fair lending and stop treating racial disparities in loan outcomes as simply reflecting secondary market guidelines. Although the study confirmed that the economic factors cited by lenders did indeed explain much of the racial disparity in mortgage denials, it also showed that discrimination could occur. It showed that the lending decision involves substantial discretion; that many applications are approved despite weaknesses; that examiners cannot readily detect discrimination because reasons frequently exist that could justify a denial; and that in one major market, with lenders not so very different from those elsewhere, minority applicants faced a higher probability of being turned down than their white counterparts after taking into account all the economic explanations that lenders had proposed.

---

[18] Researchers at Abt Associates have applied influence statistics to the Boston Fed data base to test whether this might be a useful technique for identifying minority applications that should be reviewed for evidence of possible discrimination (Rodda and Wallace 1995). This process involved ranking minority denials and white approvals according to their influence upon the coefficient on race. In so doing, they observe that the removal of about 20 observations, mainly minority denials, from the sample would cause the coefficient on race to become insignificant. While some might interpret this as meaning the Boston Fed's results are not robust, removing these observations from the sample is very misleading.

These observations are not outliers according to any criteria established in advance. Quite the contrary, they represent the most likely victims of discrimination: They are applications that appeared to have a low probability of denial according to their economic characteristics, but nevertheless were denied. Moreover, removing them severely distorts the sample, since "good" minority applications that were denied are dropped while any similarly "good" white applications that were denied remain in the sample. Thus, within this range of characteristics, white applicants appear to have been treated less favorably than minorities—since some whites were denied, while the comparable minority denials no longer appear.

Appendix
Summary of Primary Critiques and Point-by-Point Rebuttals

The following are summaries of the primary critiques of the Boston Fed's study and point-by-point rebuttals. They are listed in chronological order. Critical comments that simply cite others' criticisms are not included.

Brimelow and Spencer (January 1993)

Criticism: *The "default" data in the Boston Fed study show that lenders were not discriminating.*

Brimelow and Spencer point to information on foreclosure rates in Boston City neighborhoods, which do not show a racial pattern, and claim that discrimination would produce lower rates in minority neighborhoods. They also quote the study's lead author, Alicia Munnell, as saying "I do not have evidence....No one has evidence" in response to a question about the existence of discrimination.

Response: *Comparing average default rates for white and minority borrowers cannot disprove the existence of discrimination. In addition, neighborhood foreclosure data are a dubious proxy for minority default rates.*

The default argument confuses the experience of the individual with that of the group. As discussed at length in the text of this article and at greater length in Tootell (1993), discrimination occurs when applicants with the same expected default probabilities are treated differently. If the distributions of expected default probabilities are different for minority and white applicants, comparing the default experience of the two groups cannot disprove the presence of discrimination. In particular, if the proportion of minority applicants with very low default probabilities is smaller than the white proportion, the default rate for minority borrowers as a group may be the same as or higher than that for white borrowers, even though discrimination is occurring and individual minority applicants have been turned down more frequently than white applicants with the same default probabilities.

The neighborhood foreclosure data presented in the Boston Fed study are not a good indicator of the default experience of home-owners of different races, because they include non-owner-occupied properties and because minority home-buyers frequently purchase in non-minority areas.

With respect to Ms. Munnell's quote, which has sometimes been presented as an acknowledgment that the Boston Fed study was flawed, Ms. Munnell believed she was responding to a question about the availability of information on the default experience of minority and white borrowers.

Becker (April 1993)

This article makes the same point as Brimelow and Spencer above. Becker's discussion clarifies the reasoning behind the default argument, emphasizing that discrimination is assumed to take the form of requiring minority applicants to meet more stringent standards of creditworthiness than white applicants. It is further assumed that lenders are good predictors of defaults. If discrimination takes other forms or if lenders are not good predictors of default or care about other profit considerations, comparisons of default rates cannot reveal much about discrimination.

Zandi (August 1993)

Criticism: *The Boston Fed study omits important variables. There are errors in the data and the regression should have been run over a "matched" sample.*

The study does not take into account the "state of the economy and housing markets in Boston during 1990," particularly the fact that home prices fell much more for low-priced homes than for mid-range or high. It also does not include whether the applicant's credit history met the institution's standards, whether data could be verified, the presence of a co-signer, and loan amount.

Response: *Some of the variables mentioned were included in the study; others should not be included. Regression analysis obviates the need for a matched sample. Zandi's example of an error was not an error.*

The study included a variety of measures of the state of the economy and housing markets, including unemployment rates by industry, data on applicants' financial situation, and housing values. Contrary to Zandi's assertion about the decline in value of low-priced homes, prices fell less for low-priced homes than for high in 1990, although by the end of 1991 the decline was similar and, eventually, prices of high-priced homes turned up before prices of low-priced houses.

The Boston Fed did look at the influence of co-signer and loan amount; including these variables does not affect the results. The co-signer equation appears in the appendix. Including responses to the verification and credit history standards questions does indeed reduce the race coefficient, although as Zandi himself notes, the race effect remains large. As discussed in the text, however, the Boston Fed does not believe these variables should be in the analysis, as both involve an ex post judgment by the respondent. All the other variables are based on objective criteria. Moreover, the credit history question appears to be a proxy for denial, as it is a function of loan-to-value and other variables that have nothing to do with credit history.

The issue of data errors is addressed under Liebowitz, Day and Liebowitz, and Horne, below; however, the specific observation that Zandi used to support his claim of errors had been checked with the respondent institution, which had rejected the application precisely because the variable in question (loan-to-value) was extremely high.

Liebowitz (September 1993)

Criticism: *The data contain many errors. The sample mixes up different types of applications. The results are sensitive to a few applications.*

Liebowitz characterizes as errors applications with large negative net worth, applications with low or negative interest rates, applications with loan-to-value ratios in excess of 80 percent for which the applicant did not seek mortgage insurance, and applications where the yearly income does not match the monthly income.

Liebowitz claims that condominiums, two- to four-family properties, and single-family homes with down payments above and below 20 percent should be analyzed separately. He also asserts that the coefficient on race for applicants purchasing single-family homes with loan-to-value ratios below 80 percent is due to "six extremely influential loan applications."

Response: *Most of Liebowitz's "errors" are not errors.*

Liebowitz's statements about errors are based on intuition; he has no information about the contents of loan files. Most of his "errors" are not errors at all. Thus, he characterizes as errors all loans with high loan-to-value ratios and no mortgage insurance, on the grounds that these loans could not be sold on the secondary market. While insurance is generally required by the federally sponsored agencies, exceptions are made; and insurance is not required if the loan is kept in the lending institution's portfolio. Of these loans, 40 percent were denied, 40 percent were held in portfolio and 20 percent were sold to a combination of private and public entities. Far from being proof of errors, this is a reasonable outcome.

Another misstatement is categorizing as errors observations where yearly and monthly income figures do not agree. The yearly figures are from the lenders' original HMDA submissions. They were not part of the Boston Fed survey nor were they used by the Boston Fed, although they (along with other HMDA data) were made available to researchers as part of the public data set. The Boston Fed did not use the HMDA income figures and instead requested the monthly income figures from the loan application form because the latter were more precisely defined. Liebowitz had been informed of this prior to writing his article.

With respect to interest rates, the Boston Fed did not request information on interest rates. Liebowitz apparently estimated interest rates from the obligation ratios. Such a procedure is necessarily imprecise, especially with multi-unit properties; and, as is pointed out under the discussion of Day and Liebowitz (1993), most of the low interest rate estimates are associated with two- to four-family houses. Some of the net worth values do look peculiar. The Boston Fed researchers also had questions about these variables, and it was partly for this reason that we instituted procedures to inspect and verify suspect and missing values. Errors were, of course, corrected. Regressions were also run excluding observations with unusual values for net worth, as well other variables. The results are unaffected. Others have confirmed this (Carr and Megbolugbe 1993).

We chose not to exclude unusual observations from the data base because we had no standard, other than intuition, for what were reasonable values. This was an original data base and some of the unusual observations did not appear so unusual upon closer examination. For example, there were physicians with very large assets and even larger liabilities. Other researchers can choose to drop these observations (and will find it makes no difference to the results); but they are discarding data that were in the loan officer's information set. Whether the lenders used the information is another matter. The lenders consulted prior to undertaking the study said they did not pay much attention to net worth because assets were so hard to verify. This was confirmed by the regression analysis.

Response: *Splitting the sample is not justified. The results are not sensitive to unusual values or specific applications, as long as the sample has not been so reduced that most of the minority observations are eliminated.*

Splitting the sample is warranted only if the relationships are statistically different for the different groups. The Boston Fed performed chi-squared tests to see whether dividing the sample was appropriate. We considered whether different types of property should be treated separately and also whether black and Hispanic applicants should be split. These tests provided no basis for splitting the sample.

One consequence of splitting the sample is that the number of observations, particularly the number of minority observations, is reduced. Since the total sample contains only 700 black and Hispanic applicants and only about 200 rejected minority applicants, splitting the sample into smaller and smaller groups, with fewer and fewer minority rejections, automatically undermines the power of statistical tests to determine the effect of race.

Even so, when we run the study's primary regressions over Liebowitz's subsamples, race is statistically significant except for the smallest group—those applying to purchase two- to four-family properties. Liebowitz's statements to the contrary, race is significant for applicants buying condominiums and applicants buying single-family homes with private mortgage insurance. Moreover, among those purchasing single-family homes without mortgage insurance, whom Liebowitz characterizes as the "core" sample, race is highly significant despite the fact that only 14 minority rejections remain out of the original 200. Liebowitz then must remove almost half of these rejections before the effect of race disappears. Far from proving the sensitivity of the Boston Fed's results to key observations, this analysis demonstrates how robust they are.

Day and Liebowitz (December 1993)

Criticism: *Repeats criticism in Liebowitz (1993) above, specifically that the data contain errors, the sample should be split, and the results are sensitive to key observations. Day and Liebowitz also argue that important variables have been omitted and question the specification of the regression.*

The claims of data errors in this paper and in materials distributed at a session at the Federal Reserve Bank of Dallas in December 1993 focus on Day and Liebowitz's (henceforth D&L) estimates of interest rates. D&L argue that approximately 90 interest rates that they imputed from the housing expense variable are unreasonably high or low; obligation ratios appear wrong in 15 cases. These and other "errors" that they have identified (denied loans that were sold) cast doubt on the integrity of the entire data base.

Response: *Almost all of D&L's apparent inconsistencies in the data can be explained or involve variables that were not used in the Boston Fed regressions. Nor are the results sensitive to these observations.*

D&L estimate interest rates using the housing expense variable, loan term, and loan amount. This is a rough technique and will not work for multi-unit properties or properties for which the mortgage loan is small in relation to other elements of housing expenses. Almost all of the imputed rates that they conclude are too low involve two- to four-unit properties, for which the housing expense is reduced by rental income. Almost all of the imputed rates that they find are too high involve properties for which the loan-to-value ratios are very low (less than 35 percent). In a few cases, the term of loan may be incorrect, throwing off imputations of interest rates. Term appears in one of the regressions in the appendix to the Fed study.

The obligation ratios used in the Boston Fed's regressions were not based on the housing expense numbers upon which D&L made their interest rate imputations, but on the obligation ratios shown in the lenders' worksheets. The latter ratios were given the closest scrutiny and regressions were run excluding any unusual values, with no effect on the results.

D&L also note that some rejected mortgage applications were apparently sold. Both action taken and sale came from the original HMDA survey; the Boston Fed confirmed that the action taken (rejection) was probably correct from looking at other data elements, such as whether reasons for denial were included in the HMDA data for that application. We did not use data on loan sales and did not try to validate these figures.

Criticism: *D&L argue that the presence of unverified information and the applicant's failure to meet the institution's policy guidelines for credit history, as well as several other variables, should be included in the regression. They also argue for a slightly different specification and claim the predicted probabilities do not explain actual outcomes very well.*

D&L run regressions including variables for unverified information and failure to meet credit history guidelines. Unverified information does not change the results. The credit history variable reduces the magnitude and significance of the race variable, but race remains statistically significant. Some changes to the specification further reduce the coefficient for race, but it remains statistically significant.

D&L also claim that applicants' past customer relationships with lenders should be recognized, while industry unemployment rates should not since minorities seem to work in industries with relatively low unemployment rates. D&L further fault the study for failing to include information pertaining to "functional illiteracy, unemployment, drug use, criminal activity and a whole host of other social pathologies" (p. 16).

Response: *D&L's regressions for the most part confirm the robustness of the Boston Fed results. The credit history guidelines variable should not be included because it is a proxy for loan denial. Comments about missing variables confuse the individual applicant with group stereotypes.*

The race variable remains large and statistically significant when unverified information is included. Adding a variable for meeting credit history guidelines reduces the race coefficient but it remains statistically significant. The credit history guidelines variable should not be in the equations, however. In contrast to other variables, it is not objective. Regressions in which meeting credit history guidelines is used as the dependent variable show it to be a function of loan-to-value ratios, obligation ratios, and other explanatory factors that have nothing to do with credit history. This issue is discussed more fully in the text.

With respect to missing variables, the issue of past borrower relationships was broached with the lenders consulted prior to the study, who responded that it was a rare and idiosyncratic factor. With respect to the question of drug use and criminal activity, the people applying for mortgages appear to be solid economic citizens, as can be seen from the summary data presented in both the Boston Fed study and D&L. Minority applicants, on average, are not as financially strong as white applicants, but they do not look any more like people with literacy or drug problems than their white counterparts.

Finally, it should be noted that the estimation procedure used by D&L produces imprecise results, which happen to overstate the explanatory power of the economic variables in some of their regressions. A more precise regression package yields different results.

Criticism: *The sample should be split into different property types.*

Response: This issue is discussed above under Liebowitz (1993). *There is no statistical justification for splitting the sample.*

Moreover, D&L's regressions highlight one of the dangers of doing so; not only is the significance of race reduced in some of the samples, but so too is the significance of variables that almost everyone has cited as important, specifically, loan-to-value and obligation ratios.

Criticism: *Results are sensitive to influential observations.*

D&L try to find the observations that have the greatest influence on race and then they throw them out until race is insignificant. They claim that removing 15 observations will make the race coefficient go negative.

Response: *This is a questionable regression procedure in that observations are rejected only because of their influence, not for any objective criterion.*

Discarding information precisely because it is influential is a very different and much more questionable concept than removing observations because they appear to be outliers by some objective criterion. Unless there is clear evidence that the observations in question are erroneous, there is no justification for removing them. D&L repeatedly suggest that the regression is not working very well because some of their influential applications (with relatively low probabilities of denial) have blemishes. Even a cursory review of the data, however, shows that there are few "deal-breakers" in the mortgage business. The one exception is being denied private mortgage insurance. But people who have gone through bankruptcies get approved. People with very high loan-to-value and obligation ratios are approved. These flaws increase the likelihood of being turned down, but they do not guarantee that outcome.

D&L's regression also includes the credit history guideline variable. As discussed in the text, the credit history guideline variable is subjective and, to a large degree, a proxy for denial. It should not be included.

Horne (1994a)

At the request of the federal supervisory agencies, the Boston Fed identified those rejected loan applications for which the model estimated a probability of denial of less than 50 percent. Horne's criticisms are based on FDIC examiners' reviews of the loan files at their institutions. Because the FDIC examiners actually saw the loan files, Horne's comments, especially those about data errors, have been given considerable weight. The Boston Fed did not have access to the loan files, but did see a summary of the FDIC examiners' findings.

Criticism: *The data contain errors.*

Horne states "Overall, 57 percent of the applicant files contained serious data errors. . . ." These "errors" seem to fall into three categories: errors in responding to the Boston Fed survey, discrepancies between application forms and the underlying information in the files, and the classification of certain loan outcomes. Horne devotes most of his attention to the third of these.

Response: *Many of Horne's errors are not data errors at all. Most of the genuine data errors the examiners found seem to be minor.*

Horne himself acknowledges in the body of his article that he is using the term data error very loosely and that information included in his claim that "57 percent" of the applications have serious errors is actually correct. Horne cites only one or two examples of a true data error. One example is that a couple whom the Boston Fed data show having a two-year work history had a work history of two months. In a subsequent paper, discussed below, Horne makes various "corrections" to the Boston Fed data base, and these still do not alter the Boston Fed's findings.

Horne's "errors" that are not errors fall into two categories. Horne takes exception to the fact that applicants who reject a counteroffer or who subsequently reapplied and were accepted are considered to have been denied a loan according to Regulation B and HMDA reporting requirements. He also does not think that applicants who were rejected because they were overqualified for special programs or who were rejected for private mortgage insurance should be treated as rejected.

While Horne may have a point in suggesting that counteroffers, overqualified applicants, and private mortgage insurance call for some caution in interpreting the results, they in no way represent data errors. Indeed, the mortgage insurance issue was highlighted by the Boston Fed study and its interpretation is discussed at length.

Horne is also concerned about discrepancies between the data on the application form and information in the files and between information in the files and the true nature of the applicants' circumstances. This is an interesting issue, but to characterize the Boston Fed's use of application information as a data error is wrong. The relevant information for assessing the determinants of the loan denials is the information set available to the loan officer, and the best source of that information is the application form at the date of decision. The alternative that Horne appears to be suggesting—that examiners attempt to second guess the application information and even the underlying information in the files—boils down to making up data.

As for Horne's claim that rejected applications probably present too favorable a picture of the applicants' circumstances, others have claimed just the opposite. To the degree that accepted applicants have had the opportunity to straighten out errors in their credit history files or to pay down debts, their final applications will appear more favorable relative to the initial application than those that were rejected at the outset.

Criticism: *The Boston Fed study omitted some variables.*

Horne claims in his introduction that "a number of important factors influencing the ability to purchase a home were misspecified or insignificant." The discussion in the text is much less assertive; he does not offer alternative specifications and his candidates for omitted variables are limited to the value of home equity and the dollar amount of gifts. He also thinks it would be desirable to have only verified financial assets in the analysis, but acknowledges that this is not possible for applications that were rejected before the assets were verified. Horne doubts the Boston Fed finding that liquid assets do not affect the lending decision. More generally, Horne appears skeptical of the ability of statistical models to capture the underwriting process.

Response: *Horne's suggestions for omitted variables are already encompassed by included variables. He does not propose any alternative model specifications; stripped to its basics, his argument is that the model must be misspecified because he doubts the results.*

Horne would like to see the dollar amount of gifts and the value of home equity included as explanatory variables. Home equity, as he acknowledges, is included in net wealth. Horne accepts the argument that net wealth is not influential because assets are difficult to verify, but thinks home equity alone would be. Home equity is a good example, however, of why assets are difficult to verify, since its precise value will not be known until the existing home is already sold. The Boston Fed did recognize whether the applicant had previously been a home-owner by including as a dummy variable, in addition to the credit history variables, whether the applicant had a prior mortgage history; it had no effect on the results.

The Boston Fed included a dummy variable for whether gifts or grants accounted for part of the down payment. It appears in one of the appendix regressions and is of borderline statistical significance; it has no bearing on the race coefficient. Horne argues that the dollar amount of gifts would be preferable, but he does not really say why, other than to suggest that it might be an indicator of parental resources. He acknowledges that the presence of gifts is already reflected in the loan-to-value ratios.

Much of Horne's discussion of omitted variables is a discussion of the idiosyncratic features of the loan applications. He sees the presence of such idiosyncratic features as evidence that the Boston Fed model is misspecified, and perhaps as a general condemnation of all statistical models. He presents nothing to suggest that the idiosyncracies are systematically related to race, however. Moreover, the presence of idiosyncracies could be interpreted as an argument in favor of statistical models, since the alternative of file-by-file reviews will be constrained in the number of files examined and thus may fail to identify those determinants of denial that are common to many files.

Criticism: *The model is misspecified and does not fit the data very well.*

Horne argues the model is misspecified because certain variables he thinks are important (liquid assets) were found not to affect the mortgage decision and other variables, which are important, were not important enough. In particular, he thinks that poor credit history should guarantee or almost guarantee denial, whereas the model says it substantially increases the chances of denial, but in Horne's example still leaves the denial probability at roughly 30 percent. He also mentions other financial weaknesses that he thinks do not carry sufficient weight in the model.

Horne also claims the model does not fit very well, because "two-thirds of the applications that were [actually] denied were predicted to be approved on the basis of a 50 percent probability threshold." He further claims that a naive prediction that all the applications were approved would result in 85 percent of the outcomes being correctly predicted.

Response: *The model fits the data quite well and claims that the model is misspecified because certain variables do not work as Horne expects are disproved by the data and the FDIC examination reports.*

The Boston Fed data show that apart from being denied mortgage insurance, there are very few deal breakers. People with poor credit histories and even bankruptcies were approved. People with very high debt ratios were approved. This was confirmed by the summary of the FDIC examiners' findings; although the examiners might conclude that an application was appropriately rejected because of a poor credit history or large obligation ratio, they often could find approved applicants with an equally poor credit history or large obligation ratio. The approved applicant might have some compensating factors, but the credit history or obligation ratio was clearly not an absolute bar to approval. If people are getting approved despite serious weaknesses, the model will reflect this even if the conventional wisdom is that such individuals will always be denied.

As for fitting the data, using a 50 percent cutoff to predict approvals and denials, as Horne does, is not a good test. Horne characterizes all probability estimates below 50 percent as predictions of approval and concludes that the model is predicting too many approvals. But an estimated probability of denial of, say, 40 percent should not be interpreted as a prediction of approval, but instead a prediction that, among similar applications, the fraction denied will be roughly three times the fraction in the total sample. The appropriate test appears in Table 6 of the text of this article, which shows the fraction of applications actually denied according to the estimated probabilities of denial. As can be seen, the estimates match reality quite well. The naive prediction that all applicants were approved would not be very useful, as it would be 100 percent wrong for the applications of greatest concern.

Yezer, Phillips, and Trost (1994)

Criticism: *Single-equation models of rejection, like that used by the Boston Fed, produce biased estimates because borrowers can negotiate loan-to-value ratios and other loan terms to reduce the probability of rejection. Because minority applicants are "economically disadvantaged and are less able to increase down payments in order to avoid possible rejection," the minority coefficient will be biased upwards, leading to a false or at least overstated finding of racial discrimination.*

Yezer and his co-authors (YPT) argue that loan terms are the result of negotiations between lender and borrower. Thus, loan-to-value ratios, obligation ratios, presence of co-signer, and some other explanatory variables are endogenous and estimates of their coefficients will be biased. In addition, because minorities are economically disadvantaged and cannot as easily respond to lenders' requests for larger down payments or the presence of a co-signer, the coefficient for race estimated by the single-equation model will be biased upwards.

In support of this argument, YPT construct a three-equation model and a pseudo data set from the Boston Fed's data base. In generating their pseudo data set, the coefficient on race in the rejection equation is set equal to 0, while the coefficient on loan-to-value is large and positive. They then estimate a single-equation rejection regression using the pseudo data set and find that the coefficient on race is positive and significant and that on loan-to-value is negative.

The YPT article repeats criticisms also made in Rachlis and Yezer (1993). YPT also argue that single-equation default regressions produce biased estimates.

Response: *YPT present no evidence on the prevalence of negotiation in the lending decision and, thus, the potential for bias.*

Bias only arises if the borrower receives new information about the probability of rejection during the application process and responds by adjusting some of the loan terms, for example, by making a larger down payment and reducing the loan-to-value ratio. One cannot say this never happens, but YPT have not established that negotiation is widespread. For many borrowers, of all races, down payments will be dictated by their financial circumstances and their ability to negotiate will be limited.

No bias exists if the borrowers *anticipate* a high probability of rejection in advance of applying and try to offset it with a larger down payment or other enhancements to their finances. The system in this case is recursive rather than simultaneous.

Instrumental variables is the standard econometric technique for addressing problems of simultaneity. Finding appropriate instruments to apply to the present problem is difficult because almost all the variables available were collected as potential explanations for denial. In addition, YPT assert that all loan terms are endogenous; so even if appropriate instruments could be found for some of these terms, YPT would still question the results. That said, using plausible instruments for the loan-to-value ratio, the focus of most of YPT's discussion, leaves the effect of race on the probability of denial economically and statistically significant.

Response: *The argument that minorities cannot/will not negotiate as much as white applicants cannot be distinguished from the claim that lenders are less inclined to negotiate with or coach minority applicants.*

It takes two to negotiate. Given that the analysis has controlled for applicants' liquid assets, wealth, and other economic advantages, the YPT argument is stripped of its economic rationale and comes down to the assertion that minority applicants will not negotiate but white applicants will. Such an outcome could be explained, however, by lenders' greater willingness to initiate negotiations with white applicants. There is no way to distinguish between the two possibilities, but since lenders would have a lot more experience negotiating mortgages than borrowers, one could plausibly expect them to be the initiating party.

Horne (1994b)

This paper, which had not been published at the time of writing, has been quoted in a number of press accounts, with particular attention given to the claim that dropping from the sample two banks specializing in serving minority populations eliminates the effect of race on the mortgage decision (Passell 1994 and Macey 1994). Any published version is likely to be somewhat different; but since the unpublished version has been widely circulated, a response is necessary. Horne's analysis is based on FDIC institutions only; he also omits applications for which the loan-to-value ratios were below 30 percent. *The resulting sample is about 40 percent of the original sample and includes somewhat less than 40 percent of the minority observations.*

Criticism: *The data contain errors. These errors cause the race coefficient to be overstated.*

Horne claims to have corrected the Boston Fed data pertaining to FDIC institutions based on examiner file reviews. He then runs regressions using the "corrected" data. He finds that the race variable is smaller using the "corrected" data, although still about the same size as the Boston Fed's estimate for the overall sample and still statistically significant.

Horne then re-codes a number of denied applications as approvals and finds that the race effect becomes insignificant. Horne argues that denials based on mortgage insurance, problems with titles, and rejections of counteroffers are not the lenders' fault and thus should not be treated as denials.

Response: *Horne's data corrections are minor; they do not support his claims of serious data errors. Re-coding denied applications as approvals is highly misleading.*

Horne's data corrections do not affect the finding that race has a large and statistically significant effect on the mortgage lending decision. This can be seen in his own regression results. One reason for this outcome is that most of Horne's changes are minor. This is suggested by Horne's table comparing the Boston Fed and Horne data sets and is confirmed by a review of his data. The Boston Fed obtained Horne's data tape and many of the changes were either small, involving, for example, a couple of dollars in monthly income or $100 in assets, or pertained to variables that were not used by the Boston Fed's analysis. Some of the data changes are clearly wrong; for example, the gender variable was meaningless in the Horne data set. Of greater consequence to his results, an application that was rejected because of title problems was reclassified as an acceptance. Another rejected application was reclassified as an acceptance on the grounds that the applicant subsequently submitted another application with a larger down payment, which was accepted; thus, we have two accepted applications instead of a rejection and an acceptance.

After making these "corrections," Horne then re-codes as approvals applications that were denied private mortgage insurance and applications where the applicant rejected a counteroffer. While a case might be made for excluding such observations from the analysis, to re-code them as approvals is wrong. The applicant was turned down for a loan at the terms requested.

The private mortgage insurance issue and questions of interpretation are extensively discussed in the Boston Fed study. Regressions were run excluding those who were turned down for insurance. The overall results and the effect of race are unchanged.

Criticism: *Changes to the specification reduce the race coefficient.*

Horne makes several changes to the equation specification that have the effect of reducing the race coefficient, although it remains economically and statistically significant. Horne then adds variables for unverified information and meeting credit history standards and the effect of race becomes insignificant.

Response: *The variable on meeting credit history standards is a subjective variable that is correlated with race and should not be in the analysis.*

The appropriateness of the credit history standards variable is discussed at length under Day and Liebowitz (1993) above. Horne's other specification changes are not unreasonable. The Boston Fed also adopted a nonlinear approach to loan-to-value in a later version of its study, but in contrast to Horne, we found that the coefficient on race was increased.

Criticism: *Removing two institutions that actively serve minorities reduces the race coefficient to insignificance.*

Response: *This is the most widely cited and most misleading of Horne's criticisms. The number of minority observations in Horne's sample is now only one-quarter its original size.*

Horne starts out with a sample that is less than half the size of the Boston Fed sample and with fewer than 300 minority observations and 90 minority rejections. The removal of the two institutions in question cuts his minority sample by half again.

Horne acknowledges this in footnotes, but the press reports of his findings present the removal of these institutions as an innocuous change. Even so, dropping these two institutions from the sample still does not remove the influence of race, as is shown in Table 11 of this article. Other sample reductions and specifications are necessary to achieve that result.

References

Becker, Gary S. 1993. "The Evidence Against Banks Doesn't Prove Bias." *Business Week*, April 19, p. 18.

Berkovec, James A., Glenn B. Canner, Stuart A. Gabriel, and Timothy H. Hannan. 1994a. "Race, Redlining and Residential Mortgage Loan Performance." *Journal of Real Estate Finance and Economics*, vol. 9, no. 3, pp. 263–94.

______. 1994b. "Discrimination, Default, and Loss in FHA Mortgage Lending." Federal Reserve Board of Governors Working Paper. November 1994.

Brimelow, Peter and Leslie Spencer. 1993. "The Hidden Clue." *Forbes*, January 4, p. 48.

Carr, James H. and Isaac F. Megbolugbe. 1994. "A Research Note on the Federal Reserve Bank of Boston Study on Mortgage Lending." Fannie Mae Office of Housing Research, Final Draft. January 3.

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

Engelhardt, Gary V. and Christopher J. Mayer. 1994. "Gifts for Home Purchase and Housing Market Behavior." *New England Economic Review*, May/June, pp. 47–58.

Galster, George C. 1993. "The Facts of Lending Discrimination Cannot Be Argued Away by Examining Default Rates." *Housing Policy Debate*, vol. 4, issue 1, pp. 141–46.

Glennon, Dennis and Mitchell Stengel. 1994. "An Evaluation of the Federal Reserve Bank of Boston's Study of Racial Discrimination in Mortgage Lending." Comptroller of the Currency, Economic and Policy Analysis Working Paper no. 94–2.

Horne, David K. 1994a. "Evaluating the Role of Race in Mortgage Lending." *FDIC Banking Review*, Spring/Summer, pp. 1–15.

______. 1994b. "Mortgage Lending, Race, and Model Specification." Preliminary draft. Federal Deposit Insurance Corporation, Division of Research and Statistics.

Liebowitz, Stan J. 1993. "A Study That Deserves No Credit." *The Wall Street Journal*, September 1, p. A14.

Macey, Jonathan R. 1994. "Banking by Quota." *The Wall Street Journal*, September 7, p. A14.

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

Passell, Peter. 1994. "Redlining Under Attack." *The New York Times*, August 30, p. D1.

Rachlis, Mitchell B. and Anthony M.J. Yezer. 1993. "Serious Flaws in Statistical Tests for Discrimination in Mortgage Lending," *Journal of Housing Research*, vol. 4, pp. 315–36.

Roberts, Paul C. 1993. "Banks in the Line of Fire." *The Washington Times*, March 12, p. F1.

Rodda, David and James E. Wallace. 1995. "Fair Lending Management: Using Influence Statistics to Identify Critical Mortgage Loan Applications." Paper for presentation at the mid-year meeting of ARUEA. May 30. Photocopy, Abt Associates Inc.

Schill, Michael H. and Susan M. Wachter. 1994. "Borrower and Neighborhood Racial and Income Characteristics and Financial Institution Mortgage Application Screening." *Journal of Real Estate Finance and Economics*, vol. 9, no. 3, pp. 223–39.

Stengel, Mitchell and Dennis Glennon. 1995. "Evaluating Statistical Models of Mortgage Lending Discrimination: A Bank-Specific Analysis." Comptroller of the Currency, Economic and Policy Analysis Working Paper no. 95–3.

Syron, Richard F. 1993. Testimony before the Committee on Banking, Housing and Urban Affairs, United States Senate, February 24, p. 5.

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

Yezer, Anthony M.J., Robert F. Phillips, and Robert P. Trost. 1994. "Bias in Estimates of Discrimination and Default in Mortgage Lending: The Effects of Simultaneity and Self-Selection." *Journal of Real Estate Finance and Economics*, vol. 9, no. 3, pp. 197–215.

Zandi, Mark. 1993. "Boston Fed's Study Was Deeply Flawed." *American Banker*, August 19, p. 13.