

Detecting Discrimination by the Numbers

Remarks by

Lawrence B. Lindsey

to

The Boston Bar Association

Boston, MA

June 7, 1994

## Detecting Discrimination by the Numbers

Thank you. I am deeply honored to be here today to accept the Boston Bar Association's award for distinguished public service. I consider myself quite fortunate to be involved with so many of the challenging issues that confront our country. I am also fortunate to have been exposed during my education to some of the best minds in this country -- most notably here in the Boston area -- and so be able to apply their teachings to the issues with which I deal.

My responsibilities currently lead me into the midst of a policy dilemma embroiling our nation for which, it seems, no amount of formal education or policy experience can provide a clear or easy answer. At its root, this dilemma is one of conflicting values -- including some of the most fundamental principles on which our republic is based. At issue is the proper role of statistical analysis in detecting illegal discrimination in our society. Perhaps no other issue in bank regulation is as contentious or has wider ramifications for how we view ourselves and our society today.

Let me be quite clear at the outset about my views on the subject of discrimination. Neither I, nor the institution I represent, the Federal Reserve, will stand second to anyone in opposing unfair practices in the provision of financial services. Economic decisions which are based on irrelevant criteria such as race, gender, religion or other protected characteristics have no place in any part of our society. The use of these irrelevant

criteria not only offend our sense of propriety and our democratic values, they also undermine the efficient functioning of markets. And as a believer in democratic capitalism and a society based on economic and political liberty, I find the use of irrelevant personal characteristics abhorrent.

Let me also say that my general attitude toward statistics is easy -- the more the better. I love statistics. I make my living with statistics. But statistics can be, and often are misleading. Indeed, some skeptics even claim that my profession makes its living arguing both sides of any number. But anyone who deals with numbers as much as I do should keep in mind the limitations of the data with which one is working, as well as its uses. Statistics can be used to confuse as well as to enlighten, and so in the field of public policy we need to treat them with particular care.

Those who have followed the regulatory process of fair lending enforcement should be quite familiar with the inherent limitations of statistics. The regulatory agencies have been repeatedly criticized by community organizations and some members of Congress for not finding more hard evidence of illegal discrimination at financial institutions. On the other hand, some in the financial services industry and elsewhere have been highly critical of some of our recent findings of discrimination. But a careful consideration of the issues involved suggests to me that this controversy really centers on the role of statistics in the discrimination area. To illustrate, let me give a bit of

history as to how fair lending enforcement has been performed over the years.

Most of the history of fair lending enforcement has been directed at detecting what is best termed as "unfairness". Specifically, examiners were trained to look at the loan files of rejected applicants and seek out those who were rejected for no apparent reason other than their race, or gender, or some other prohibited basis. All of us would agree that, if detected, such practices are unacceptable. They are simply wrong. The individual involved should be compensated for the damages suffered. The institution should be compelled to take remedial actions regarding its lending policies and, in some cases, pay punitive damages as well.

But while such cases of discrimination are clearly wrong, our experience has taught us that they are also extremely rare, or at least very difficult to detect. With rare exceptions, rejected loan applicants had some characteristic for which they could be legitimately rejected. It might have been poor credit history or lack of income or inadequate job tenure -- but there was almost always a legitimate reason on which to pin the rejection. This long history of evidence, these "statistics" if you will, led some to conclude that discrimination in the financial services industry was extremely rare. This finding meshed well with the notion that no banker would turn down a loan on which he or she expected to make money.

Yet this interpretation of "the facts" turned out to be

incorrect. What we came to realize over time, in addition to our earlier findings, was that many applicants with sound reasons to be rejected were actually accepted. For example, although the bank might have a policy that said that persons who had been delinquent on credit card payments in the last two years were not to be granted mortgages, sometimes bankers found it profitable to lend to such people. Why? Often, applicants had a good reason for credit history flaws -- a temporary financial setback, a medical emergency, etc. In short, the lending officer involved used his or her judgment to determine that these individuals would be good risks, in spite of the flaws in their credit histories that would be captured in most statistical computations.

Evidence began to accumulate, furthermore, that the use of such judgment was sometimes correlated with the racial or ethnic background of the applicant. Our procedures were thus challenged with the problem of detecting evidence that even though qualified applicants were not rejected because of race, marginally qualified applicants who happened to be white were often approved while marginally qualified applicants who weren't white were rejected. This is not an easy task.

One early technique developed to detect this potential discrimination was the "matched pair" approach. The case was made that if a rejected applicant from a minority background was compared with an accepted white applicant who had similar economic characteristics, then illegal discrimination probably

occurred.

On its face, matched pair analysis would seem to be quite straightforward. But, again, statistical comparisons are never simple. One problem has to do with omitted applicants. Matched pair analysis usually involves two similar applicants, one a minority applicant who was rejected, and the other a white applicant who was accepted. But what if there are four similar applicants: one minority who was accepted, one minority who was rejected, one accepted white and one rejected white. In that case, a finding of discrimination based on race is not so clear by looking at all, instead of some, of the statistics.

Thus, the analytic process involved in selecting matched pairs is crucial. The process of selecting rejected minority applicants and seeking a match with an accepted white applicant is not sufficient by itself to indicate discrimination. In a sense, the statistics you see cannot be viewed in isolation -- their interpretation depends on the statistics you don't see. To succeed, the matched pair examination procedure requires a much more sophisticated examination of loan applications, with examination of both acceptances and rejections from all racial and ethnic classifications.

Further complicating this process is the fact that perfectly matched pairs are very hard, perhaps impossible, to come by. What may seem at first blush to be analytically a very straightforward procedure in fact requires a good deal of examiner judgment. In essence, the examiner, after the fact,

matches his or her judgment with that of the bank's loan officers.

These problems have led us to an even more statistically sophisticated approach and the use of regression analysis. I'm sure that most of you are familiar with the study of home lending practices done here at the Boston Federal Reserve. In many ways that study provides the model for this latest development in the statistical analysis of loan data. The problems I associated with the matched pair approach are eliminated because the entire set of information regarding all applicants, or at least a large sample of them, is used. Rather than matching individuals, a statistical model of applicants is created, and individuals are compared with that broad based model.

This statistical system, in essence, takes the bank's decision making system and reduces it to a single equation, based on how the bank treats white applicants with a wide variety of characteristics. From that equation, minority applicants are assigned a probability of approval based on their creditworthiness. Some regulatory agencies have used this analytic approach as the basis for their statistical examination for discrimination. There are those who conclude that if a minority applicant receives a probability of approval greater than 50 percent, but is rejected, then that applicant may be considered a victim of discrimination. In at least one actual case, applicants with approval probabilities greater than 50 percent have received monetary damages as compensation for having

been rejected.

What of this approach to statistical examination of loan applications? While I think it is an important analytic tool, I must admit to being somewhat troubled by the amount of faith the enforcement process is placing in statistics, and interpretations of model results. For example, consider the use of the 50 percent probability cutoff for determining who is the victim of discrimination. What that cutoff means is that, regardless of race, someone with those economic characteristics was literally a toss of the coin regarding loan approval. More specifically, half of white applicants with the same characteristics of a supposed victim of discrimination were also rejected for a loan. Statistically speaking, a probability based model cannot be used to say anything conclusive about a single individual. Surely this approach cannot form the basis of what we mean by illegal discrimination.

Similarly, consider the implications for those applicants who got scores of less than 50 percent, say 40 percent. The statistics mean that the bank approved 40 percent of the applicants with those loan characteristics. But, the implications of standardizing the use of the 50 percent threshold is that those people at the 40 percent threshold really should not have gotten their loans. Is this really the signal we as regulators want to be sending?

And, of course, all of this assumes that we got the model "right", whatever that means, in the first place. For example, I

have enormous respect for the Boston Fed study I mentioned earlier. It was an important, seminal work in this very complicated area. But, surely it should not be considered flawless. I have never run across a serious bit of academic statistical research that could make that claim. Indeed, the key role of academics is to constantly criticize and refine analysis. No work is ever considered flawless, and the Boston study has received its fair share of criticism. The research process, after all, always pursues truth -- it never finds it in any definitive sense of the word. As regulators, we should not confuse an analytic process which pursues the truth with truth itself.

These limitations led us at the Federal Reserve to develop a statistical approach using regression analysis that will improve our examiner's ability to detect potential lending discrimination. We do not expect this regression model to identify definitively who might be a victim, but it will improve significantly our ability to detect potential discrimination. Our regression analysis indicates whether race is statistically significant in a model of a lender's decisions on mortgage loan applications. To support a finding of credit discrimination, however, statistical evidence of apparent discrimination discovered through the program would have to be supplemented by analyzing the lender's treatment of individual loan applicants. In this regard, the program also identifies matched pairs of rejected and accepted loan applicants that examiners use for a

loan by loan review. Thus, a finding of credit discrimination would include both statistical evidence as well as evidence obtained from actual loan files. At this point, this approach would seem to strike a proper balance between the power of statistics and the flexibility inherent in human judgment.

So what we did at the Federal Reserve is bring to bear the full power of statistical analysis on the issue of lending discrimination and found that statistics alone could not solve this problem. But of course, we did not view this as a signal to give up. The issue we were dealing with is much too serious. Instead, we viewed this as a signal to stop and reflect on the approach we were taking.

Let me take this opportunity to express some of my concerns regarding the increasing use of statistics in determining that discrimination has occurred. Let me also explain why I am so troubled about some of the difficulties that may result from the ever increasing reliance on statistics.

First, I am troubled at the prospect of how the use of statistics will fare in a judicial setting. Do we really want to have the nuances of regression procedure examined carefully by a jury? Do we really expect our judges, learned though they may be, to be deciding statistical points that are normally debated only in the most esoteric of academic journals? To date, cases have been settled out of court. But what will the judicial process do when confronted with batteries of opposing statisticians and economists? However remunerative this approach

may be for members of my profession to play their roles as expert witnesses, doesn't this prospect really trivialize the very important issue of illegal discrimination? Couldn't we be approaching a point where the statistics will soon be obscuring the facts?

Second, I think we should consider very carefully what the logical extension of the legal and regulatory use of statistics really amounts to. Statistical second guessing of loan decisions, with punitive consequences, may soon mean that loan decisions themselves will be statistically based. For example, if I were a banker who thought that I would owe damages to rejected loan applicants who received scores of 50 percent or higher from a statistical model, I would soon make sure that I got a copy of that model and approved everyone with a score of 50 or higher.

Indeed, we would be naive to think individual bankers would behave differently. In fact, such a statistics-based appraisal already exists. It is called credit scoring. It will continue to gain broader use as regulatory forces, in pursuit of a laudable objective, seek ever more sophisticated statistical means to detect discrimination.

Is that such a bad result? Maybe not. One could imagine a world in which the judgment of loan officers and boards of directors is eliminated and credit scoring models make all the loan decisions. There is no doubt that such a result would be conceptually non-discriminatory. The software could easily be

brought into compliance with the very latest statistical advances made by the regulatory agencies. With certitude, no one would be accepted or rejected based on their race, gender, or other characteristic, because it would not be a factor in the computer's scorecard.

What would we lose by such a development? Only the human judgment that loan officers now bring to the process. And it is unclear what, if anything, that is worth. Under current practice, the loan officer makes the loan to some of the individuals with 40 percent scores and rejects some of the applicants with 60 percent scores. There must be reasons for such non-standard judgments.

Those reasons might, in some cases, be both inappropriate and illegal. They might, for example, be based on the race or gender of the applicant. If this is the case, our public policy decision is easy: the loan officer should be fired and replaced with a computer. Getting rid of human judgment would be appropriate.

The reasons might also be tied to some factor not picked up by a computer's statistical model -- let's call it a "hunch" -- one not based on an illegal or inappropriate factor such as race or gender. Then the question is whether the outcome of hunches is correlated with actual loan performance. If the loan officer's hunches turn out to have no bearing on loan performance, or even worse, are negatively correlated with loan performance, then we again have an easy call. The loan officer

should be fired and replaced with a computer.

If the loan officer's hunches are, however, positively related to loan performance, then replacing the loan officer would entail a cost. Replacing this loan officer with a computer would mean more delinquent loans and more missed opportunities, greater losses for the bank, and a less efficient allocation of resources from society's point of view.

However, this becomes a tougher public policy call if the loan officer's hunches turn out to be pretty good regarding loan performance, but occasionally marred by the individual's innermost prejudices. This most difficult public policy case, what I call the expert but flawed human, is probably the most accurate description of our current loan officer situation. Credit scoring, therefore, is really an alternative to this expert but flawed human being. The computer will make fair loans -- both in performance and in being devoid of discrimination. The human will have better performance, but may, occasionally discriminate in socially unacceptable ways.

Before making our public policy decision between the fair computer and the expert but flawed human, let us also consider another aspect of the human's hunches. Some of the judgment calls the human makes are out of sympathy. For example, the couple who skipped credit card payments in order to buy medicine for an ailing child, would be viewed sympathetically by the loan officer, but considered a deadbeat by the computer. Or consider the young person who may have grown up on the proverbial wrong

side of the tracks and has little capital or credit history, but whose teachers said, "this kid's a go-getter". That person might get a small business loan from our expert but flawed human, but would surely receive zip from the computer.

The concept that I am trying to engender is that we humans also have a sense of justice which transcends, and is distinct from, the statistical sense of fair treatment which the computer provides. I believe therefore, that in our policy choice between the expert but flawed human and the scrupulously fair computer, we are also making a choice between what our laws are decreeing as fair and what we humans know to be right and just.

That is why I am so troubled by the policy dilemma our country now faces in the area of lending discrimination. We clearly have a responsibility to maximize the positive aspects of the human loan officer and to minimize his or her flaws. But let us not exaggerate this option. Though vast, we humans do have a limit to our capacity for self-improvement. And the policy process is demanding answers to the challenge of discrimination far sooner than any amount of education and increased self-awareness could provide in the time available.

As a result, I see the current orientation of policy making as driving us rapidly and inexorably toward the computer based approach. Under current policy conditions, I would expect credit-scoring type procedures to be overwhelmingly dominant by the end of the decade. We will obtain the fairness of the machine, but lose the judgment, talents, and sense of justice

that only humans can bring to decision making.

Ultimately the pendulum will swing back. As a society we will become dissatisfied with the real world of statistical fairness. New institutions will develop to circumvent the restrictions on lending to those who are not up to the computer's statistical standards. We will also realize, though too late, that statistically based procedures may actually work against those who need opportunity the most and who have the fewest credentials to offer at present. The intended beneficiaries of our drive for fairness, may in fact, be those who suffer the most. And, as is so often the case, it may be that cleaning up the unintended consequences of well-intentioned policy actions taken today, that will be the biggest challenge for tomorrow's policy makers.