Estimation and Evaluation of Loan Discrimination: An Informational Approach

Marsha Courchane, Amos Golan, and David Nickerson*

Abstract

Many recent studies have analyzed whether lending discrimination exists. In all previous studies, the researcher faces constraints with the available data or modeling problems. In this article, we use a new informational-based approach for evaluating loan discrimination. Given limited and noisy data, we develop a framework for estimating and evaluating discrimination in mortgage lending. This new informational-based approach performs well even when the data are limited or ill conditioned, or when the covariates are highly correlated. Because most data sets collected by bank examiners or banks suffer from some or all of these data problems, the more traditional estimation methods may fail to provide stable and efficient estimates.

This new estimator can be viewed as a generalized maximum likelihood estimator. We provide inference and diagnostic properties of this estimator, presenting both sampling experiments and empirical analyses. For two of the three banks analyzed, we observe some evidence of potential racial discrimination.

Keywords: Discrimination; Maximum entropy; Maximum likelihood; Mortgage lending

Introduction

Background and Objectives

In the past five years, many researchers have focused on investigating whether discrimination exists in home mortgage lending. The major objective of those investigations is to determine, using different estimation techniques and data from individual loan files, if persons of a specific race or minority group face lower probabilities of receiving a loan, controlling for underwriting practices. During the past four years, the Office of the Comptroller of the Currency (OCC) has used statistical analysis in conjunction with examiner resources to analyze home mortgage lending to determine if there exists reason to believe that a violation of the Equal Credit Opportunity Act or the Fair Housing Act has occurred. With the current available resources and data, the OCC is usually constrained to make this determination using a relatively small number of files sampled from the total population of loan applications at a given bank. These bank-specific data sets, obtained by extracting information

* Marsha Courchane is Principal Economist at Freddie Mac, Amos Golan is Visiting Professor at American University, and David Nickerson is Associate Professor at Colorado State University and a consultant to Freddie Mac.
The authors thank Andy Kaplowitz for research assistance and Judy Giles, Dennis Glennon, David Malmquist, and Jeff Perloff for their comments. Amos Golan thanks the Office of the Comptroller of the Currency (OCC) for supporting this research and for use of the data sets. The empirical analyses presented here, particularly any statements regarding discrimination, reflect the views of the authors and not the OCC, the Department of the Treasury, or Freddie Mac. Marsha Courchane collected the data while Senior Financial Economist at the OCC.

from loan files, are typically small (as a proportion of the population of applications), imperfect, and involve a large number of covariates that in most cases might be highly correlated. One reason for this multicollinearity is the large number of discrete variables that may be highly related on the right-hand side of the regression. Thus, traditional maximum likelihood (ML) as well as the different Bayesian approaches often fail to yield stable estimates. Without stability (low variances) of the estimates, it is extremely hard to draw inferences from the data. Further, in most cases the underlying distribution that generated the data is unknown. When the underlying distribution is unknown (a realistic assumption for relatively small data sets), the maximum likelihood approaches may yield poor estimates.

The objective of this article is to introduce, evaluate, and apply an efficient and consistent informational-based estimator for evaluating discrimination in mortgage lending. This new method, called generalized maximum entropy (GME), has its roots in information theory and builds on the classical maximum entropy formalism (ME) that was developed specifically for evaluating and processing the information of underdetermined, ill-posed problems and imperfect data. This ME formalism was generalized by Golan, Judge, and Miller (1996) and Golan, Judge, and Perloff (1996). This generalization is the basis for the statistical formulation used in this article. For review of both the classical ME and the GME approaches, see Golan, Judge, and Miller (1996).

After discussing the loan discrimination problem and framework, we develop and discuss a basic model in the following section. Then, an informational-based method for estimating the discrimination model is formulated and discussed. This model is an application of the formulation proposed by Golan, Judge, and Perloff (1996). After formulating the estimator, the necessary diagnostics, inference, and information measures are developed in the subsequent section. We then present some sampling experiments performed on data collected by the OCC that contrast the traditional ML with the new estimation approach and provide evidence that, for these small and highly correlated data, the new approach outperforms the ML method. These experiments establish the foundation for the empirical analyses. We then analyze and discuss detailed empirical examples taken from three different banks throughout the United States. Concluding remarks and directions for future work complete the article.

Mortgage Lending and Race: A Brief Review

Academics, bankers, and government regulators all have devoted much time and many resources to the study of the impact of race, or minority status, on the home mortgage lending decision. The main focus of this work is to determine whether discrimination exists (see Horne [1997] or Ladd [1998] for a detailed overview of the issues). More specifically, the research focuses on whether the underwriting guidelines are applied consistently for all applicants regardless of race or other prohibited status. Statistical methods are used to test the hypothesis that, given the same credit profile and similar demographic characteristics, minority applicants face the same probability of denial as that of all applicants. If probabilities vary, disparate treatment may exist. See, for example, Calem and Stutzer (1995), Courchane and Cobas (1995), Courchane et al. (1995), Ladd (1998), Munnell et al. (1996), and Stengel and Glennon (1999).

There are several unresolved modeling issues that have been discussed in the literature (e.g., Heckman 1998; Ladd 1998; Longhofer and Peters 1998; and Yinger 1996). These issues arise from the choice of how best to represent the approval decision process at the bank level as well as which econometric or statistical modeling approach is best able to capture differences in treatment. These choices may vary from bank to bank and are closely related to the availability of data. In this article, a relatively simple and applicable model that accommodates some of the special problems in the banking data is developed and analyzed. A summary of the basics of this model follows.

The most widely employed statistical procedure used to model a bank's lending decision is an unordered discrete choice modeling approach. The decision to approve or deny a loan to an applicant is based primarily on the individual's credit profile but also includes demographic, economic, and property-specific attributes. It is generally argued in the literature that the decision model should reflect the probability that an applicant will default—a conceptual framework that underlies the design of most mortgage credit scoring models. However, in most cases, the approval process involves judgmental decisions made by underwriters using established policy guidelines that are qualitatively related but not quantitatively linked to the likelihood of default. For example, it is generally accepted that the higher the debt-to-income (DTI) ratio, the greater the likelihood of default (a qualitative relationship). However, few banks know what impact an increase in the total DTI ratio from 32 to 36 percent (or 48 percent) has on the likelihood of default (a quantitative relationship). Under this type of underwriting process, it is possible that the underwriting (judgmental) guidelines may introduce differences in treatment of applicants. This may lead to a violation of the fair lending laws. For that reason, the purpose of the statistical model is not to determine the optimal weights an underwriter should use to assess the creditworthiness of the applicant but rather to assess if the (predetermined) underwriting guidelines are fairly applied. That is, the models will be used to test the hypothesis that minority applicants with profiles (e.g., credit, employment, wealth) similar to nonminority applicants face the same likelihood of approval.

The statistical models in this case should be designed to determine the relative importance (beyond that associated with random chance) of any observed difference in the likelihood of approval for these different racial (minority) groups. Unfortunately, this assessment of the fair application of the underwriting guidelines is difficult to achieve using conventional discrete choice modeling methods. For all except the largest mortgage lenders, the number of events is small. Therefore, the conventional maximum likelihood estimators may become very sensitive and unstable. This is even more difficult if the difference in treatment is concentrated in the segment of marginally qualified borrowers.

To summarize, the common representation of the lending procedure is the familiar unordered, discrete choice (multinomial) model. In this view of reality, the bank makes its decision to deny or approve a loan based on the individual's characteristics. Given these characteristics, the bank calculates (based on historical data) the probability that the individual will pay back the loan. The presence of rational discrimination would mean that the individual's race (or gender) affects this probability. In this article, we use a relatively simple and applicable model (e.g., Golan, Judge, and Perloff 1996) that accommodates some of the special problems in the banking data. This method, which has its roots in information theory, has much better small sample properties than the traditional estimators used in this area. To make the presentation simpler, we start by briefly constructing the traditional multinomial logit model, which is just a special case of our model.

The Traditional Model

Within the traditional multinomial discrete choice problem, consider an experiment consisting of T trials. In each experiment a binary random variable $y_{1j}, \ldots, y_{TJ}$ is observed where y_{ij} (for $i = 1, 2, \ldots, T$) takes on one of the J unordered categories $j = 1, 2, \ldots, J$. That is, on each trial i, one of the J categories is observed in the form of a binary variable y_{ij} that equals unity if and only if alternative j is observed and zero otherwise. Let the probability of alternative (choice) j, on trial i, be $p_{ij} = Prob(y_{ij} = 1)$ and assume the p_{ij} are related to a set of explanatory variables (the individuals' characteristics), X, via the nonlinear model

$$p_{ij} \equiv Prob(y_{ij} = 1 \mid \boldsymbol{x}_i, \boldsymbol{\beta}_j) = \boldsymbol{F}(\boldsymbol{x}_i'\boldsymbol{\beta}_j) > 0 \quad \text{for all } i \text{ and } j \tag{1}$$

where $\boldsymbol{\beta}_j$ is a $(K \times 1)$ vector of unknowns, $\boldsymbol{x}_i'$ is a $(1 \times K)$ vector of covariates and F(.) is a function linking the probabilities p_{ij} with the (linear structure) covariates $\boldsymbol{x}_i'\boldsymbol{\beta}_j$ such that

$\sum_j \boldsymbol{F}(\boldsymbol{x}_i'\boldsymbol{\beta}_j) = 1$. Traditionally, the likelihood function's specification depends on the researcher's assumptions (beliefs) of the underlying data generation process.

Given the above Xs, the likelihood function can be expressed as

$$L = \prod_{j=1}^{J} p_{1j}^{y_{1j}} p_{2j}^{y_{2j}} \ldots p_{Tj}^{y_{Tj}} \tag{2}$$

and the log-likelihood function is

$$\ln(L) \equiv \ell = \sum_i \sum_j y_{ij} \ln p_{ij} \quad . \tag{3}$$

Next, we choose a reasonable distributional form for $\mathbf{p}$, or $\boldsymbol{F}(\boldsymbol{x}_i'\boldsymbol{\beta}_j)$ above. To maintain generality, for a large J (greater than 3), it is not practical to choose the normal distribution (probit) unless one wishes to use a more cumbersome sampling (computational) approach. A more practical distribution for large J is the logistic distribution. In this case, after normalizing $\boldsymbol{\beta}_{ik} = \boldsymbol{0}$ (e.g., Greene 1999), we have

$$p_{ij} = \frac{\exp\left(\sum_k \beta_{jk} x_{ik}\right)}{\sum_j \exp\left(\sum_k \beta_{jk} x_{ik}\right)} \equiv \frac{\exp\left(\sum_k \beta_{jk} x_{ik}\right)}{1 + \sum_{j=2}^{J} \exp\left(\sum_k \beta_{jk} x_{ik}\right)} \equiv \frac{\exp\left(\sum_k \beta_{jk} x_{ik}\right)}{\Omega_i(\boldsymbol{\beta})} \quad \text{for } j = 2, \ldots, J \tag{4}$$

and

$$p_{i1} = \frac{1}{1 + \sum_{j=2}^{J} \exp\left(\sum_k \beta_{jk} x_{ik}\right)} \equiv \frac{1}{\Omega_i(\beta)} \quad \text{for } j = 1 \tag{5}$$

Substituting (4) into (3) yields

$$\ell = \sum_{ij} y_{ij} \ln \left[\frac{\exp\left(\sum_k \beta_{jk} x_{ik} \right)}{1 + \sum_{j=2}^{J} \exp\left(\sum_k \beta_{jk} x_{ik} \right)} \right] \tag{6}$$

$$= \sum_i \sum_j \sum_k y_{ij} \beta_{jk} x_{ik} - \sum_i \sum_j y_{ij} \ln \Omega_i(\boldsymbol{\beta})$$
$$= \sum_i \sum_j \sum_k y_{ij} \beta_{jk} x_{ik} - \sum_i \ln \Omega_i(\boldsymbol{\beta})$$

which is the ML multinomial logit. For $J = 2$ and $\boldsymbol{\beta}_{ik} = \boldsymbol{0}$, equation 6 reduces to the conventional binomial logit model.

An Informational-Based Estimator

In this section, we formulate and apply the GME model that was developed in Golan, Judge, and Perloff (1996) for analyzing discrimination among loan applicants. Imposing a parametric structure through the choice of distributional form for $\boldsymbol{p}$ is, in many cases, a very strong assumption. In practice, the underlying distribution is rarely, if ever, known. For that reason, we use an estimator for which such a strong assumption is not required. Moreover, we search for an estimator that is based on a *minimum* set of assumptions. But before formulating the GME model, we briefly discuss the traditional ME approach.

The ME approach was originally developed to solve an underdetermined model. Given a small number of observed moments of some unknown, discrete (and proper) probability distribution, the objective is to get the most conservative estimates of this distribution. Because the number of unknowns is much greater than the number of observed moments (data points), this problem is underdetermined, meaning there are infinitely many proper probability distributions that satisfy these data. Therefore, the estimation problem reduces to the choice of a criterion function that will consistently choose one of these solutions. The entropy function is such a criterion. Using the entropy criterion ensures that the resulting estimates are the most conservative in the sense that the estimated distribution is the most uniform out of all distributions that satisfy the data. For a detailed discussion, formulation, and extensions see Golan, Judge, and Miller (1996).

Going back to our estimation problem, instead of considering a specific $F(.)$, or a likelihood function, we view the data as noisy. Under this view (representation), the constraints do not hold exactly and the model is reformulated as $y_{ij} = \boldsymbol{F}(\boldsymbol{x}_i'\boldsymbol{\beta}_j) + e_{ij} = p_{ij} + e_{ij}$, where the p_{ij} are the unknown multinomial probabilities and the e_{ij} are the *natural* noise components for each individual (observation) and are contained in [–1, 1] for each individual. Rewriting y_{ij} as $\boldsymbol{y}=\boldsymbol{p}+\boldsymbol{e}$, it is clear that the covariates (individuals' characteristics) are not included in this formulation unless one specifies a certain p or $\boldsymbol{F}(.)$. To incorporate them, we follow the instrumental variable approach and premultiply each vector by the set of covariates X to get

$$\sum_i y_{ij} x_{ik} = \sum_i x_{ik} p_{ij} + \sum_i x_{ik} e_{ij}, \tag{7}$$

where there are $(T \times (J - 1)$ unknown probabilities but only $(K \times J)$ data points (constraints). This implies that regardless of the number of observations there are infinitely many solutions to this underdetermined (ill-posed) problem. In other words, there are infinitely many sets P that satisfy the restrictions of equation 7: the available data. Thus, the estimation problem reduces to choosing a specific criterion that picks a specific solution out of the infinitely many possible solutions. One such criterion is the entropy measure $H(\boldsymbol{p}) = -\sum_{ij} p_{ij} \ln p_{ij}$ with $0 \ln 0 \equiv 0$. This measure is a continuous measure between 0 (perfect certainty) and $Kln\ (J)$, which reflects perfect uncertainty, or maximum ignorance, for each observation i.

The entropy measure $H(\boldsymbol{p})$ is defined only over the proper probabilities p_{ij} in equation 7. This entropy measure can be extended to cover the unknown disturbances e_{ij} in equation 7. This is done by representing the uncertainty of each e_{ij} as a finite and discrete random variable with $2 \leq D << \infty$ possible outcomes. That is, the unknown disturbances are represented by a D-dimensional support space $\boldsymbol{v}$ and a D-dimensional vector of weights, $\boldsymbol{w}$. Thus, each error component is redefined as $e_{ij} \equiv \sum_d v_d w_{ijd}$ with $\sum_d w_{ijd} = 1$, where the D-dimensional errors' support $\mathbf{v} = (-1/\sqrt{T},..,0,..1/\sqrt{T})'$ for each e_{ij}. Under this formulation, the D-dimensional vector of weights (proper probabilities), $\boldsymbol{w}$, converts the errors from the [–1, 1] interval into a set of TD proper probability distributions. The extended entropy measure, $H(\boldsymbol{p}, \boldsymbol{w})$, defined below is a simple extension of $H(\boldsymbol{p})$ to incorporate the unknown $\{\boldsymbol{w}\}$ that represent the disturbances $\{\boldsymbol{e}\}$. The full set of unknown $(\boldsymbol{p}, \boldsymbol{w})$ can be recovered by maximizing $H(\boldsymbol{p}, \boldsymbol{w})$ subject to the data and the requirement that both $\{\boldsymbol{p}\}$ and $\{\boldsymbol{w}\}$ are proper probabilities. Specifically,

$$\underset{\mathbf{p},\mathbf{w}}{Max} \quad \left\{ H(\mathbf{p},\mathbf{w}) = -\sum_{ij} p_{ij} \ln p_{ij} - \sum_{ijd} w_{ijd} \ln w_{ijd} \right\} \tag{8}$$

subject to the KJ data points

$$\sum_i y_{ij} x_{ik} = \sum_i x_{ik} p_{ij} + \sum_{i,d} x_{ik} v_d w_{ijd} \tag{9}$$

and the requirements for proper probabilities

$$\sum_i p_{ij} = 1 \text{ and } \sum_d w_{ijd} = 1 \tag{10}$$

The GME estimators derived from the above optimization are

$$\hat{p}_{ij} = \frac{\exp\left(-\sum_k \hat{\lambda}_{kj} x_{ik}\right)}{1 + \sum_{j=2}^{J} \exp\left(\sum_k \hat{\lambda}_{kj} x_{ik}\right)} \equiv \frac{\exp\left(-\sum_k \hat{\lambda}_{kj} x_{ik}\right)}{\Omega_i} \tag{11}$$

and

$$\hat{w}_{ijd} = \frac{\exp\left(-\sum_k x_{ik}\hat{\lambda}_{jk}v_d\right)}{\sum_d \exp\left(-\sum_k x_{ik}\hat{\lambda}_{jk}v_d\right)} \equiv \frac{\exp\left(-\sum_k x_{ik}\hat{\lambda}_{jk}v_d\right)}{\Psi_{ij}(\hat{\lambda})} \tag{12}$$

Under this general criterion function, the objective is to minimize the joint entropy distance between the data and the state of complete ignorance (the uniform distribution). It is a dual loss function that assigns equal weights to prediction and precision. Equivalently, it can be viewed as a shrinkage estimator that simultaneously shrinks the data to the priors (or uniform distributions in the GME case). If one is interested in assigning nonequal weights to the two parts of the objective function (prediction and precision), such weights can easily be incorporated.

Given the basic GME model in its primal form, it is possible to reformulate the model as a dual unconstrained problem. This reformulation has two advantages. First, it is computationally superior. Second, it allows us to compare the above entropy-based formulation (GME) with the ML-type estimators. With this dual formulation, the foundation for the statistical properties necessary for estimating and evaluating the discrimination in loans is developed.

Starting with the Lagrangean of the model represented by equations 8 through 10 and substituting the posterior, or post-data of equations 11 and 12 for the p_{ij}'s in $H(\boldsymbol{p,w})$ while ignoring the last term of the Lagrangean (since the posteriors already satisfy the normalizations, or the requirements of proper distributions, equation 10) yields the unconstrained (or generalized log-likelihood) model

$$\begin{aligned} L(\lambda) &= -\sum_{ijk} y_{ij}x_{ik}\lambda_{kj} + \sum_i \ln\left[\sum_j \exp\left(-\sum_k \lambda_{kj}x_{ik}\right)\right] \\ &+ \sum_{ij} \ln\left[\sum_d \exp\left(-v_d \sum_k \lambda_{kj}x_{ik}\right)\right] \\ &= -\sum_{ijk} y_{ij}x_{ik}\lambda_{kj} + \sum_i \ln\Omega_i + \sum_{ij}\ln\Psi_{ij} \end{aligned} \tag{13}$$

Minimizing the *dual unconstrained (concentrated)* model, $L(\boldsymbol{\lambda})$, with respect to λ and equating to zero yields $\hat{\lambda}$, which in turn yields the estimates $\hat{p}_{ij}$ via equation 11. It is important to note that the dual equation 13 is computationally much more (exponentially more) efficient. For example, a 10-covariate binomial problem with 100 observations may take about 5,000 iterations with the basic primal model, while the above dual takes, at most, 30 iterations. Because there are exponent functions in the dual approach, a normalization of the data may be required. A simple normalization entails dividing each element of $\boldsymbol{x}_k$ by $Max\{\boldsymbol{x}_k\}$.

The first two components of the dual unconstrained model represented by equation 13 with $\boldsymbol{\beta} = -\boldsymbol{\lambda}$ are the same as the conventional logit model in equation 6. Thus, the ML-logit is just a special case of the GME estimator discussed here. Specifically, given the errors' support $\boldsymbol{v}$, it is easy to see that as $T \to \infty$, $\boldsymbol{v} \to \boldsymbol{0}$ and the $GME \to ML$. The information matrix and covariance matrix for the GME is therefore similar in structure to the one for the ML-logit model. Keeping in mind that all available bank-specific data sets are finite and (relatively) small, these two estimators yield different estimates. Further, it is emphasized that the covariance elements, in general, and the variances, in specific, are always lower than those for the ML-logit model. This implies the greater stability and higher efficiency of the GME estimator.

While advances in automated underwriting have resulted in more bank data kept in electronic form, the ability to keep all of the underwriting variables in a single electronic data set has not been developed. Hence, given examiner resource limitations, the regulatory review process to date typically uses statistical sampling to obtain data sets of just a few hundred observations when examining for disparate treatment.

The above statements concerning greater stability and higher efficiency of the GME estimator will be shown empirically in the sampling experiments and empirical examples discussed below.

Inference, Diagnostics, and Information

Before proceeding with the analysis, a short discussion of the Lagrange multipliers is necessary. In investigating both the primal and the dual, it is clear that the basic unknowns of interest are the Lagrange multipliers, λ, which are just $-\boldsymbol{\beta}$ of equation 1. It is important to remember that the basic number of unknowns remains fixed, $(K-1)\times J$, regardless of the dimension of D. Using the common normalization, $\lambda_1 = \beta_1 = 0$, the objective is to recover these $(K-1)\times J$ unknowns, which in turn yield the $T \times J$ matrix P. However, because we seek to evaluate the impact of the $\boldsymbol{x}_k s$ on each p_{ij}, the more interesting parameters are the marginal effects. That is,

$$\frac{\partial \hat{p}_{ij}}{\partial x_{ik}} = \hat{p}_{ij}\left(\hat{\beta}_{jk} - \hat{S}_{ik}\right) \tag{14}$$

where $\hat{S}_{ik} \equiv \sum_j \hat{p}_{ij}\hat{\beta}_{jk}$. These marginal effects reflect the impact of each covariate (characteristic) on each coefficient p_{ij} and are the basic parameters of interest. These marginal effects are traditionally evaluated at their means. For discrete (dummy) variables such as race, one should evaluate these marginal effects at both values: zero and one. That is,

$$partial(x_k = 0) = \frac{1}{T_k}\sum_i \frac{\partial \hat{p}_{ij}}{\partial x_{ik}}\bigg|_{x_{ik} = 0} \tag{15}$$

$$partial(x_k = 1) = \frac{1}{T - T_k}\sum_i \frac{\partial \hat{p}_{ij}}{\partial x_{ik}}\bigg|_{x_{ik} = 1} \tag{16}$$

where T_k is the number of observations where $x_{ik} = 0$. Finally, using the estimated values of $\beta_{kj}s$, and following Crawford, Pollack, and Vella (1998), one can calculate the lower and upper bounds for these partials as follows.

$$\text{Lower Bound: } \frac{1}{4}\left[\hat{\beta}_{jk} - Max\left(\hat{\beta}_{jk}\right)\right] \text{ for each category } j=1,2,\ldots,J-1 \tag{17}$$

$$\text{Upper Bound: } \frac{1}{4}\left[\hat{\beta}_{jk} - Min\left(\hat{\beta}_{jk}\right)\right] \text{ for each category } j=1,2,\ldots,J-1 \tag{18}$$

This yields

$$\frac{1}{4}\left[\hat{\beta}_{jk} - Max\left(\hat{\beta}_{jk}\right)\right] < \frac{\partial \hat{p}_{ij}}{\partial x_{ik}} < \frac{1}{4}\left[\hat{\beta}_{jk} - Min\left(\hat{\beta}_{jk}\right)\right] \text{ for each } j. \tag{19}$$

Within the entropy approach used here, one can investigate the amount of information in the estimated coefficients. Following Golan, Judge, and Perloff (1996), the normalized entropy (information) measure for the whole system is

$$S(\hat{P}) \equiv \frac{-\sum_i \sum_j \hat{p}_{ij} \ln \hat{p}_{ij}}{T \ln(J)} \tag{20}$$

with $S(\hat{P})$ between 0 and 1—with 1 reflecting uniformity (complete ignorance, or similarly, $\boldsymbol{\lambda=0}$) of the estimates and 0 reflecting perfect knowledge. Similar measures, $I(\hat{P})$, were developed by Soofi (1992) where $I(\hat{P}) = 1 - S(\hat{P})$.

Next, following our derivation, it is possible to construct an Entropy-Ratio (ER) statistic (which is similar in essence to the Likelihood Ratio statistic). Ignoring the error term for simplicity of presentation, let ℓ_Ω be the unconstrained problem (primal or dual with the data), and ℓ_ω be the constrained one where, say $\boldsymbol{\beta} = \boldsymbol{\lambda} = \mathbf{0}$ (or no data are included). Then $ER = 2(\ell_\omega - \ell_\Omega)$. Note that ℓ_Ω is just $Max(H)$ while $\ell_\omega = TlnJ$. Thus, the ER statistic for the GME is

$$ER_{GME} = 2\ln \ell_\omega - 2\ln \ell_\Omega = 2T\ln(J)\left[1 - S(\hat{P})\right]. \tag{21}$$

Under the null hypothesis, ER_{GME} converges in distribution to $\chi^2_{(K-1)}$. Finally, a Pseudo-R^2 measure (e.g., McFadden 1974) can be constructed as

$$Pseudo\ R^2 \equiv 1 - \frac{\ell_\Omega}{\ell_\omega} = 1 - S(\hat{P}). \tag{22}$$

Thus, we established a connection between the traditional ML goodness-of-fit measure and the normalized entropy (or information measure) *S (P)* of the system. There are alternative goodness-of-fit measures that could be examined (for example, see Greene 1999).

Under the GME approach, as with any other estimation rule, the issue is to investigate how far the data pull the estimates away from a state of complete ignorance (uniform distribution).

Thus, a high value of χ^2 (or, equivalently, a low normalized entropy measure) implies that the data reveal information about the estimates or contain valuable information. It is important to note that the GME approach offers a more conservative estimation rule than the ML method. One could make the case that the policy maker responsible for enforcing the law would generally prefer conservative estimates compared with nonconservative and unstable estimates when using the following description of conservative. If discrimination truly exists, the GME approach will pick it up at a slower rate than will the ML estimator. That is, when the GME method picks up a discrimination signal, the probability that it is the true (correct) signal is always higher than that for any ML signal. On the other hand, if there is no discrimination, the GME method picks it up much faster than does the ML method. Again, the probability of making a mistake is lower than with any ML signal. This means that overall the probability of making a mistake or reaching the wrong conclusion in either direction is smaller with the GME estimator. In the next section, these statements are confirmed experimentally.

Sampling Experiments

Using a banking framework and bank data, a number of sampling experiments are presented to investigate and demonstrate the small sample behavior of the GME estimator and analyze the main issues of discrimination analysis. Given a complete sample of data from a bank (e.g., bank 1, provided by the OCC), the estimated coefficients and their marginal effects are analyzed (see the next section for the empirical analysis). The data consist of nine characteristics (variables) and an intercept and include two racial minority groups (Hispanics and blacks). Using the GME estimated values for these 10 parameters together with the *same* X matrix, a number of experiments were performed, each consisting of 100 replications. The outcome (y = 0 or 1: accept or reject) for each individual observation was generated with the logistic (logit) distribution. This ensures that the correct underlying data generation process is logit and thus the ML-logit is the *correct* model to use. To make it more realistic, the parameters' values in the first experiment, case 1, are those from the estimated values of bank 1. The results for the GME and ML estimators are reported in the second column of table 1. The statistics reported in table 1 include the mean values of the parameters for the Hispanic and black covariates, their variances, the total variances of all the estimated parameters, the total mean squared error (MSE) of all βs, the number of times (out of 100) each of the race variables (H and B) is significant at the 5 and 10 percent significance levels, the percentage of correctly (in sample) predicted, the number of observations for each experiment, and the mean value (per experiment) of the $j = 1$ category: denial.

Even though the ML-logit is the correct model, the GME model outperforms it under all criteria. In column 3, case 2, the data were generated through the same process with the exception that $\beta_H = 1.0$ and $\beta_B = 0.0$. Thus, we get a better understanding of the GME and ML performance for different levels of discrimination. It is clear that where no discrimination exists, the GME model identifies this result. However, when the value is 1 (e.g., the Hispanic variable), the estimator identifies correctly at a 5 percent significance level only 59 percent of the time but is better overall than its ML competitor. Also, the mean value of the estimated beta is perfect. In the fourth column, the data are again generated using the same process and variables except that both minority variables are set to 0. Again, the results show that, at a 5 percent significance level more than 95 percent of the time, those variables are identified correctly and indicate no discrimination. Again, under all criteria the GME estimator yields superior results. The fifth column presents a similar case but with values of β_H = 1.0 and β_B = 2.0. Again, the GME results reveal that for the black case the variable is always identified correctly, with evidence of potential discrimination. For the Hispanic case (correct value of β_H = 1.0), the discrimination is identified at the 5 percent significance level 50 percent of the time, while the value of the estimated beta is identified correctly. The sixth column presents the case of high and zero discrimination respectively. The last column presents the case of a high level of discrimination for both minority groups. Again, the GME estimator proves to have the better performance.

Table 1. **Summary of Sampling Experiments**

	Case 1 β_H=0.580 β_B=0.566 GME (ML)	Case 2 β_H=1.0 β_B=0.0 GME (ML)	Case 3 β_H=0.0 β_B=0.0 GME (ML)	Case 4 β_H=1.0 β_B=2.0 GME (ML)	Case 5 β_H=2.0 β_B=0.0 GME (ML)	Case 6 β_H=2.0 β_B=2.0 GME (ML)
$\bar{\hat{\beta}}_H$	0.556 (0.593)	0.911 (1.023)	–0.070 (–0.072)	0.899 (1.011)	1.909 (2.105)	1.892 (2.090)
$\bar{\hat{\beta}}_B$	0.532 (0.551)	–0.064 (–0.800)	–0.063 (–0.094)	1.939 (2.109)	–0.070 (–0.074)	1.928 (2.103)
Var($\hat{\beta}_H$)	0.205 (0.231)	0.175 (0.229)	0.215 (0.280)	0.171 (0.220)	0.188 (0.261)	0.179 (0.245)
Var($\hat{\beta}_B$)	0.201 (0.223)	0.201 (0.252)	0.202 (0.259)	0.203 (0.261)	0.190 (0.235)	0.199 (0.257)
Var($\hat{\beta}$)	11.706 (13.430)	10.602 (13.319)	11.162 (14.964)	11.649 (15.078)	9.872 (12.341)	10.642 (13.655)
MSE($\hat{\beta}$)	12.004 (13.721)	10.757 (13.551)	11.378 (15.250)	11.804 (15.239)	9.973 (12.510)	10.775 (13.792)
Significant at α = 5%: H, B	20, 20 (21, 22)	59, 2 (53, 2)	1, 2 (2, 4)	49, 99 (53, 99)	100, 2 (100, 2)	100, 100 (100, 100)
Significant at α = 10%: H, B	32, 28 (34, 29)	63, 5 (67, 7)	7, 6 (7, 8)	62, 100 (66, 100)	100, 5 (100, 7)	100, 100 (100, 100)
% correctly predicted	83.7 (83.7)	83.4 (83.5)	85.6 (85.6)	82.7 (82.8)	82.3 (82.2)	81.5 (81.5)
Number observed	216	216	216	216	216	216
% denied	26.00	25.17	21.67	33.31	29.71	37.85

Note: Based on 100 replications, 10 variables, and 216 observations with a total of 52 denials.

The next set of experiments (see table 2) is performed with only a subset of the above data. Of the total 216 individual observations, only whites and Hispanics are included, providing a sample of 163 observations and nine variables. The GME estimator is used while generating the samples in a way similar to the previous set of experiments. In this set of experiments, the ML estimates are not presented. Because of the complex structure of the reduced covariate matrix X (i.e., a very high condition number) and the resulting small number of remaining denial files in the subset, the ML does not converge if the more traditional likelihood approaches (such as the ML-logit or probit) are used. Therefore, there is no existing solution. Column 2 presents the results of the case where the data were generated based on the estimated GME values of all nine parameters. The rest of the columns, again, present different levels of theoretical discrimination. Despite the ill-conditioned nature of the data, the outcomes of these experiments are very similar to the outcomes presented in table 1, reflecting the relatively high stability of the GME estimation rule. To summarize table 2, it is clearly observed that in the case of zero discrimination ($\beta_H = 0.0$) the GME estimator identifies it at a 5 percent significance level more than 95 percent of the time. Similarly, for a high level of discrimination ($\beta_H = 2.0$), the GME rule identifies this discrimination at a 5 percent significance level 99 percent of the time. Further, in both tables it is easily observed that there is basically no difference between the total variance of all the parameters and the MSE, implying there is practically no bias.

Table 2. **Summary of Sampling Experiments Using the GME Estimator**

	Case 1 β_H=0.646	Case 2 β_H=0.0	Case 3 β_H=1.0	Case 4 β_H=2.0
$\bar{\hat{\beta}}_H$	0.585	-0.088	0.938	1.929
Var($\hat{\beta}_H$)	0.196	0.219	0.169	0.201
Var($\hat{\beta}$)	14.239	15.457	13.745	13.049
MSE($\hat{\beta}$)	14.630	15.592	14.125	13.172
Significant at α = 5%: H	19	4	55	99
Significant at α = 10%: H	36	5	62	100
% correctly predicted	85.8	87.6	84.94	82.84
Number observed	163	163	163	163
% denied	20.58	17.74	22.29	28.18

Note: Based on 100 replications, 9 variables, and 163 observations.

These two tables clearly reveal that, for analyzing bank data, the GME is superior to the ML estimator under a large number of criteria. Another major difference of the GME method can be observed throughout these experiments (tables 1 and 2). The GME method is more conservative than its ML counterpart. This means that if there is no discrimination, the GME method produces a finding of "no discrimination" more often than does the ML method. On the other hand, if there is a low level of discrimination, the GME will find it less often than will the ML estimator. For a high level of discrimination (e.g., $\beta_H = 2.0$) the GME estimator finds it as often as does the ML estimator. When analyzing bank attitudes toward minorities, this provides a very good estimation rule. In most cases, policy makers might prefer a conservative rule. Essentially, the preference would be toward extreme care in claiming discrimination exists, as the reputation effect for banks could be devastating in situations of false claims of discrimination. The GME rule used here provides protection from this type of error.

Finally, it is worth noting that researchers should not concentrate solely on the statistical significance of the estimated parameters for the race variables but should include an evaluation of the marginal impact of these variables on the probability of denial. That is, we should also focus on whether an individual belonging to a minority group (e.g., race, gender) has a positive effect on the marginal probability of denial. It is a matter of regulatory policy to determine the threshold level of marginal effect at which there is reason to believe that a pattern or practice of discrimination exists. However, these marginal effects together with their lower and upper bounds are easily analyzed. A brief summary of these effects is given in table 3. The partial effects for the discrete (dummy) variables are evaluated at both zero and one, while those for the continuous variables are evaluated at their means. Finally, in addition to the bounds shown in table 3, it is interesting to note the bounds for different levels of the marginal effect of the discrimination variable. For example, if the correct level of discrimination is zero (e.g., case 2 of table 2), both lower and upper bounds on the marginal effect of the discrimination variable are 0. For case 3, β_H = 1.0, the lower bound is 0 while the upper bound is 0.25. For case 4, β_H = 2.0, the lower bound is 0 while the upper bound is 0.5.

Table 3. **Analysis of the Partial Effects and Their Lower and Upper Bounds Based on the Sample's Estimates**
(Case 1, Table 2)

Variable Number	β	Partial j=1	Partial j=0	Lower Bound	Upper Bound
2	2.123	0.3865612	0.1966117	0.0	0.531
3	0.041	0.0045624	0.0045624	0.0	0.0103
4	–0.005	–0.0005430	–0.0005430	–0.0013	0.0
5	1.959	0.3631349	0.1864956	0.0	0.490
6	–0.00002	–0.0000021	–0.0000021	–0.000005	0.0
7	1.877	0.3336891	0.1879775	0.0	0.469
8	–1.281	–0.1741109	–0.1307281	–0.320	0.0
9 (H)	0.646	0.0842892	0.0640765	0.0	0.162

Empirical Analysis

In this section, the GME estimates of loan data from different banks are presented and discussed. The objective here is to present the methodology and show that it easily can be applied to actual bank data. The data are taken from three banks and include a large set of variables collected and examined by the OCC. Even though the income, or wealth, variable is included in the data set, we do not include it in our final analysis. The reasons are twofold. First, it is argued (for example, see OCC Bulletin 97-24) that this variable is naturally related to race variables and thus banks should avoid using it in isolation when underwriting. This would imply that researchers rarely would want to include it as an independent variable when searching for possible discrimination. Second, in all cases analyzed here, the income variable is statistically insignificant at the 10 percent level. These two arguments convinced us to not include income in our final analysis. (The estimates with this variable are available upon request from the authors.)

In the first part of the analysis, we follow the traditional philosophy and analyze complete data sets for each bank. In that case, the race effect is represented by a dummy variable. Then, a more comprehensive analysis is done. In that case, the race effect of all the variables is analyzed and evaluated. In this comprehensive analysis, we take advantage of the robust estimation approach used here and analyze each subgroup within each bank independently. Finally, as with the ML-logit, we need to adjust the estimated intercepts according to the sampling weights of each group (Maddala 1983). This is because the analyzed samples are not random. This adjustment may affect the mean probabilities but in most cases (when adjustments are relatively small) does not affect the marginal values, which are the coefficients of interest. Comparing the adjusted (weighted) results with the nonweighted results reveals that the differences are minor for all three banks. Thus, all the tables presented here are not adjusted.

Finally, we note that, as with any other empirical work, the results and implications of the analysis greatly depend on the apriori assumption that we use the correct underlying model specification. Because this basic assumption can never be fully validated, there is always a possibility that, even if we find discrimination, there might be none. For that reason, demonstrations of the sampling experiments follow.

The Traditional Analysis

Bank 1. The GME results are presented in table 4. The variables chosen are consistent with those analyzed earlier by the OCC. A detailed description of the different variables is given in the appendix. The Pseudo-$\bar{R}^2$ = 0.40; $S(\hat{P})$ = 0.60. The mean value of the estimated probabilities is $\bar{\hat{p}}(0)$ = 0.722 and $\bar{\hat{p}}(1)$ = 0.278 and the percent correctly predicted is 85 (88 for Hispanics and 74 for blacks). The ER statistic for the whole model is 71.098. Additional statistics of interest are the different partial effects when the other discrete variables (e.g., excess DTI or bad credit) are held constant at 0 or 1. For example, holding excess DTI at 0, the partial effect for Hispanics is 0.07 and 0.12 respectively for H = 0 and 1. Similarly, holding excess DTI at 1 (DTI ratio higher than 41) yields partials of 0.14 in both cases. This implies that there is a significant effect of the race variable Hispanic when the DTI ratio is low (a good credit). When the DTI is very high, however, the race variable does not have an additional impact. Other partial effects are calculated but not presented here.

The summary statistics in table 4 reveal that the race variable Hispanic is significant at the 5 percent level. The partial effects (column 5) show that if all other characteristics are the same, Hispanic applicants have a 4 percent higher probability of having a loan denied.

Next, the stability of the estimates was checked. We used a different analysis, composed only of the white and Hispanic subsets of the data, yielding qualitatively the same results, with the Hispanic variable positive and significant. Finally, we checked the robustness of our estimates, analyzing two randomly chosen subsamples drawn from the complete data set (216 observations), with each subsample having 108 observations and moments similar to the whole sample. The GME method yielded the same estimates (to an accuracy of three digits) and partial effects as those reported in table 4. We note that such a robustness test or analysis of the different subsamples cannot be done with the more traditional ML approaches because of the ill-conditioned nature of the data (the X matrix has a condition number higher than 200).

Exclusions for Bank 1. During the course of the examination at this bank, subsequent analysis of the loan files indicated a list of 22 individuals (observations) from this sample that needed to be excluded because of either coding errors or "uniqueness." Investigating the characteristics of these 22 observations shows that these are not random observations. For example, just analyzing the moments of the full sample (238 observations) and comparing with

Table 4. **Estimates of Bank 1**

Variable	Parameter Estimate	ER Statistic χ^2	Change in Information Value	Partial Effect	Lower Bound of Partial Effect	Upper Bound of Partial Effect
Intercept	0.339			0.045	0.0	0.085
Excess DTI	1.788	21.62	0.058	0: 0.216 1: 0.318	0.0	0.447
LTV	0.037	5.42	0.012	0.005	0.0	0.009
Credit score	–0.009	7.24	0.018	–0.001	–0.002	0.0
Bad credit	1.740	15.52	0.042	0: 0.203 1: 0.327	0.0	0.435
Gift funds	–0.00002	1.32	0.004	–0.000003	–5e–6	0.0
Public record	1.388	8.80	0.024	0: 0.172 1: 0.258	0.0	0.347
Explanation (for credit derogatories)	–0.945	4.64	0.012	0: –0.112 1: –0.158	–0.236	0.0
Hispanic	0.799	3.94	0.009	0: 0.085 1: 0.122	0.0	0.200
Black	0.475	1.16	0.003	0: 0.050 1: 0.076	0.0	0.119

Note: 238 observations, 63 denied cases, 65 Hispanics, and 62 Blacks.

the smaller sample (216 observations) reveal that the basic variables (those with the highest ER-statistic level) have significantly different moments. Estimation of this subset of the data (216 observations), using the same GME approach, yields significantly lower estimates for both Hispanic and Black variables (0.58 and 0.57 respectively, or partial effects of 0.06 and 0.08 for H = 0, 1, and 0.06 and 0.09 for B = 0, 1 respectively), and both are not significant at the 5 percent level. This highlights the importance of ensuring that neither lack of data integrity nor consistency of loan type or product produces an incorrect indication of discrimination when in fact none exists.

Bank 2. The main results of bank 2's analysis are presented in table 5. Following the bank's guidelines and an initial analysis done by the OCC examiners, only the relevant variables are presented in the table, where even those with a low χ^2 are essential in the lending evaluation process and therefore are included here.

The Pseudo-R^2 = 0.523; $S(\hat{P})$ = 0.477; the mean value of the estimated probabilities is $\bar{\hat{p}}(0)$ = 0.739 and $\bar{\hat{p}}(1)$ = 0.261; the percent correctly predicted is 87.9; and the percent correctly predicted for Hispanics is 88. The ER statistic for the whole model is 130.298. The main result here is that the race variable plays no significant role in the evaluation process.

Bank 3. The main results are summarized in tables 6 and 7. The first table summarizes the results based on the full sample (case 1) and a subsample (case 2) consisting of white and Hispanic applicants. In this case, all applicants that were denied for "other" reasons or "credit application incomplete" were deleted from the sample. In the second table, table 7,

Table 5. **Estimates of Bank 2**

Variable	Parameter Estimate	ER Statistic χ^2	Change in Information Value	Partial Effect	Lower Bound of Partial Effect	Upper Bound of Partial Effect
Intercept	7.237					
DTI	0.935	6.68		0: 0.069 1: 0.103	0.0	0.234
LTV	0.016	0.60		0.001	0.0	0.004
Credit score	–0.017	18.68		–0.001	–0.0043	0.0
Bad credit	0.749	2.22		0: 0.063 1: 0.079	0.0	0.187
Gift funds	0.000007	0.0		0.0	0.0	1.8e–6
Public record	0.656	2.06		0: 0.046 1: 0.080	0.0	0.164
Explanation (for credit derogatories)	–0.477	1.14		0: –0.031 1: –0.059	–0.119	0.0
Hispanic	0.364	0.90		0: 0.022 1: 0.035	0.0	0.091
Front ratio	0.0186	0.0		0: 0.001 1: 0.002	0.0	0.0047
Insufficient funds	1.926	28.30		0: 0.146 1: 0.237	0.0	0.4815
Private mortgage insurance denied	2.871	11.08		0: 0.253 1: 0.108	0.0	0.7178
Relationship	0.134	0.16		0.012	0.0	0.0335
Low LTV (LTV < .75)	0.866	1.24		0.075	0.0	0.2165

Note: Based on 346 observations, 88 denied cases, and 99 Hispanics.

the analysis was repeated for the complete sample (427 applicants) and two subsamples: whites and Hispanics as well as whites and blacks.

Some of the main statistics for both cases are presented in table 6, where the numbers in parentheses relate to case 2, the subsample of whites and Hispanics. The Pseudo-R^2 = 0.509 (0.459); $S(\hat{P})$ = 0.491 (0.541); the mean value of the estimated probabilities is $\bar{\hat{p}}(0)$ = 0.671 (0.708) and $\bar{\hat{p}}(1)$ = 0.329 (0.292); the percent correctly predicted is 87 (87.6); the percent correctly predicted for Hispanics is 83.3 (84.5); and the percent for blacks is 86 (–). The ER statistic for the whole model is 143.133 (103.444). Additional statistics of interest are the different partial effects when holding the other discrete variables constant at 0 or 1 (e.g., credit score, bad credit, or insufficient funds for closing costs). For example, holding score at its good level 1, the partial effect for Hispanics is 0.06 (0.07) and 0.115 (0.125) respectively for H = 0 and 1. Similarly, holding the bad credit variable at 0 (applicant with perfectly clean credit record) yields partials of 0.06 (0.117) and 0.12 (0.190) respectively for H = 0 and 1. Further, holding bad credit at 1 yields partials of 0.12 (0.19) in both cases. This implies that there is a significant effect of the race variable Hispanic when the credit history is good. When the

Table 6. **Estimates of Bank 3 (Full Sample and a Subsample)**

	Case 1				Case 2		
Variable	Parameter Estimate	ER Statistic χ^2	Change in Information Value	Partial Effect	Parameter Estimate	ER Statistic χ^2	Partial Effect
Intercept	–7.131			–0.755	–6.198		–0.723
Credit score	–0.871	6.2	0.010	0: –0.117 1: –0.077	–0.632	2.74	0: –0.095 1: –0.063
LTV	0.023	3.24	0.004	0.002	0.019	2.46	0.002
DTI	0.101	71.0	0.119	0.011	0.082	39.48	0.010
Bad credit	1.234	11.4	0.019	0: 0.116 1: 0.181	1.269	9.96	0: 0.135 1: 0.208
Insufficient funds	4.943	71.0	0.122	0: 0.589 1: 0.093	3.735	36.26	0: 0.473 1: 0.126
Hispanic	0.827	3.92	0.007	0: 0.073 1: 0.119	0.813	4.14	0: 0.082 1: 0.128
Black	0.387	0.86	0.002	0: 0.034 1: 0.043			

Note: Case 1 is based on the entire sample: 384 observations, 125 denied cases, 84 Hispanics, and 86 blacks. Case 2 is based only on whites and Hispanics: 298 observations, 84 denied cases, and 84 Hispanics.

Table 7. **Estimates of Bank 3 (Full Sample and Two Subsamples)**

	Case 1				Case 2		Case 3	
Variable	Parameter Estimate	ER Statistic χ^2	Change in Information Value	Partial Effect	Parameter Estimate	Partial Effect	Parameter Estimate	Partial Effect
Intercept	–7.171			–0.872	–7.865	–0.954	–7.576	–0.835
Credit score	–1.224	16.26	0.025	0: –0.198 1: –0.118	–0.953	0: –0.162 1: –0.092	–1.025	0: –0.154 1: –0.095
LTV	0.025	5.00	0.006	0.003	0.030	0.004	0.028	0.003
DTI	0.101	96.8	0.15	0.012	0.105	0.013	0.103	0.011
Bad credit	1.198	13.6	0.02	0: 0.135 1: 0.185	1.363	0: 0.152 1: 0.224	1.098	0: 0.107 1: 0.170
Insufficient funds	1.827	33.4	0.053	0: 0.213 1: 0.258	1.421	0: 0.162 1: 0.221	2.022	0: 0.208 1: 0.291
Hispanic	0.628	2.86	0.004	0: 0.062 1: 0.102	0.717	0: 0.074 1: 0.120		
Black	0.627	2.96	0.004	0: 0.062 1: 0.086			0.653	0: 0.064 1: 0.090

Note: Case 1 is based on the entire sample: 427 observations, 140 denied cases, 95 Hispanics, and 99 blacks. Case 2 is based only on whites and Hispanics: 328 observations. Case 3 is based on only whites and blacks: 332 observations.

credit history is bad (based on the bank's underwriting criteria), however, the race variable does not have an additional impact. Other partial effects are calculated but not presented here. These results as well as simply the values of the Hispanic dummy variable together with the corresponding χ^2 level in each case of table 6 suggest potential discrimination.

Following are some of the main statistics for the cases reported in table 7 (where the numbers in parentheses relate to cases 2 and 3 respectively). The Pseudo-R^2 = 0.446 (0.448, 0.489); $S(\hat{P})$ = 0.574 (0.552, 0.511); the mean value of the estimated probabilities is $\bar{\hat{p}}(0)$ = 0.668 (0.711, 708) and $\bar{\hat{p}}(1)$ = 0.332 (0.289, 292); the percent correctly predicted is 83.8 (85.1, 84.3); the percent correctly predicted for Hispanics is 77.9 (80.0, –); and the percent for blacks is 83.8 (–, 80.8). The ER statistic for the whole model is 138.296 (106.429, 117.609).

Finally, but most important, the χ^2 (or ER statistic) value for Hispanics in case 2 is 3.70, while for blacks in case 3 it is 3.14. This implies that the race variable Hispanic in case 2 is significant at the 5 percent level, while the race variable Black in case 3 is significant at the 10 percent level.

To summarize the results of bank 3, it appears that the data (full sample, excluding special categories "other" and "withdrawn" of the denial variable as well as the different subsamples) indicate evidence of potential discrimination.

A Comprehensive Analysis

Having analyzed and investigated the dummy, or intercept effect, we turn to a more complete analysis. Given that a main advantage of the estimator used here is its ability to analyze small data sets consistently and efficiently, we analyze each of the racial groups separately. This analysis enables us to capture the total effect of discrimination if it exists. We start by analyzing each race independently, using the estimated coefficients to compare the probability of denial for the different groups. Tables 8 through 10 present these results. In the first set of rows (3 or 2, depending on how many racial groups are in the bank's data set), the estimated probabilities for each race are reported. These averages are achieved by calculating the probability of each individual in the group and averaging over all individuals. The next set of rows represents the case where each set of coefficients was used to calculate the mean probability over the whole sample (all the subgroups simultaneously). Thus, we can compare the effect of each estimate on the whole sample. The next set of calculations (which varies from bank to bank) compares the different mean probabilities for each subgroup based on the group's estimates, where some of the characteristics are held fixed. This allows us to isolate the potentially "good" or "bad" individuals in each group and get a more objective comparison. For each bank we considered some of the major variables: those that are significant at least at the 1 percent level. To increase fairness of the analysis for the nonminority group, we adjusted the relative weights of denial (e.g., Maddala 1983) whenever necessary. These weights did not alter our basic results.

What can one learn from these more comprehensive analyses? Starting with bank 1, it is clear that even for the smaller subset of 216 observations, Hispanic customers face a lower probability of receiving a loan. This lower probability is not captured when one simply analyzes the direct effect of the dummy variable (compare rows 4 and 5 in table 8). Further, looking at the differences in the probabilities where some of the important variables are held constant reveals that the discrimination among applicants is magnified. Finally, even though this table shows that potential discrimination affects blacks as well as Hispanics, we emphasize discrimination of Hispanics to highlight the different conclusion reached under the more complete and comprehensive analysis of the data.

Table 8. **Comparison of Probabilities for Bank 1, Based on Separate Estimates for Three Racial Subgroups (Whites [W], Hispanics [H], and Blacks [B])**

	Probability = 0	Probability = 1
Subsample - W (106 obs)[a]	0.835	0.165
Subsample - H (57 obs)[a]	0.737	0.263
Subsample - B (53 obs)[a]	0.578	0.422
Whole sample - W[b]	0.805	0.195
Whole sample - H[b]	0.719	0.281
Whole sample - B[b]	0.757	0.243
Subsample - W (DTI=0, 1)[c]	0.891, 0.597	0.109, 0.403
Subsample - H (DTI=0, 1)[c]	0.842, 0.343	0.158, 0.657
Subsample - B (DTI=0, 1)[c]	0.618, 0.406	0.382, 0.594
Subsample - W (BAD=0, 1)[c]	0.884, 0.561	0.116, 0.439
Subsample - H (BAD=0, 1)[c]	0.783, 0.543	0.217, 0.457
Subsample - B (BAD=0, 1)[c]	0.779, 0.316	0.221, 0.684

[a] The estimated probabilities are reported. These averages are calculated by computing the probability of each individual in the group and averaging across the group.

[b] Each set of coefficients was used to calculate the mean probability over the whole sample (all the different subgroups simultaneously). Thus, we can compare the effect of each estimate on the whole sample.

[c] Representing DTI and bad credit, these calculations (which vary from bank to bank) compare different mean probabilities for each subgroup based on the group's estimates, where some of the characteristics are held fixed to good or bad levels.

Next, we look at the comprehensive analysis done for bank 2 (table 9). Comparing the third and fourth rows reveals that the differences among the groups is marginal, and the outcome of the basic results remains unchanged as achieved in the partial analysis discussed in the previous section. However, when one of the basic variables is held at a constant value (i.e., analyzing only the "good" or "bad" subgroups within the two races), it may appear that some level of discrimination is observed. As these subsets are relatively small, we cannot change our previous conclusion that race plays no significant role in the bank's evaluation process.

Finally, looking at bank 3, it is apparent that the different treatment received by the Hispanics, captured in the partial analysis, is shown again with even larger discrepancies in treatment. That is, being a Hispanic applicant means that the a priori probability of denial is 10 percent higher than that of a white customer (compare rows 4 and 5 in table 10). This difference in treatment is also apparent when observing the numbers in the rows of the "good" and "bad" subgroups for each race (bottom part of table 10).

Table 9. **Comparison of Probabilities for Bank 2, Based on Separate Estimates for Two Racial Subgroups (Whites [W] and Hispanics [H])**

	Probability = 0	Probability = 1
Subsample - W (247 obs)[a]	0.808	0.192
Subsample - H (99 obs)[a]	0.614	0.386
Whole sample - W[b]	0.753	0.247
Whole sample - H[b]	0.733	0.267
Subsample - W (DTI = 0, 1)[c]	0.886, 0.668	0.114, 0.332
Subsample - H (DTI = 0, 1)[c]	0.712, 0.352	0.288, 0.648
Subsample - W (INSUFND = 0, 1)[c]	0.868, 0.500	0.132, 0.500
Subsample - H (INSUFND = 0, 1)[c]	0.810, 0.208	0.190, 0.792
Subsample - W (PMIFAIL = 0, 1)[c]	0.824, 0.172	0.176, 0.828
Subsample - H (PMIFAIL = 0, 1)[c]	0.641, 0.029	0.359, 0.971
Subsample - W (DTI, PMI, INSUF = 0)[c]	0.931	0.069
Subsample - H (DTI, PMI, INSUF = 0)[c]	0.859	0.141

[a] The estimated probabilities are reported. These averages are calculated by computing the probability of each individual in the group and averaging across the group.

[b] Each set of coefficients was used to calculate the mean probability over the whole sample (the different subgroups simultaneously). Thus, we can compare the effect of each estimate on the whole sample.

[c] Representing DTI, insufficient funds to close, and required private mortgage insurance but it was denied, these calculations (which vary from bank to bank) compare the different mean probabilities for each subgroup based on the group's estimates and where some of the characteristics are held fixed to good or bad levels.

Concluding Remarks

In this article, a new informational-based method for estimating and evaluating discrimination in mortgage lending was presented, discussed, and evaluated. This method is found to be consistent and efficient, outperforming the ML method, under certain criteria, for practically all finite samples. In this approach, one maximizes the entropies of the systematic (signal) and error components subject to data and theoretical restrictions. This estimation procedure involves a smaller set of a priori assumptions than traditional methods, is nonparametric in the distribution, and performs well even when data sets are small or highly collinear.

A main critique of the ML-type methods for evaluating discrimination in mortgage lending is that they may yield systematic biases, may be inconsistent, and/or may be unstable because the data sets are small, imperfect, and may involve measurement errors. The method discussed here overcomes some of these problems. It is more efficient than the ML method and can always be used—even if the data set is small or ill conditioned. A further critique is that within the ML approach it is hard to incorporate additional (theoretical) economic/statistical knowledge. Under the approach taken here, this knowledge is easily incorporated as an additional set of constraints. Further, if for some reason the researcher has some prior knowledge or beliefs, those are easily incorporated into the generalized entropy objective used here. Finally, this different approach is more conservative than the common-

Table 10. **Comparison of Probabilities for Bank 3, Based on Separate Estimates for Three Racial Subgroups (Whites [W], Hispanics [H], and Blacks [B])**

	Probability = 0	Probability = 1
Subsample - W (233 obs)[a]	0.789	0.211
Subsample - H (95 obs)[a]	0.526	0.474
Subsample - B (99 obs)[a]	0.526	0.474
Whole sample - W[b]	0.724	0.275
Whole sample - H[b]	0.629	0.371
Whole sample - B[b]	0.631	0.369
Subsample - W (BAD=0, 1)[c]	0.827, 0.609	0.173, 0.391
Subsample - H (BAD=0, 1)[c]	0.589, 0.274	0.411, 0.726
Subsample - B (BAD=0, 1)[c]	0.617, 0.342	0.383, 0.657
Subsample - W (DTI<30, >50)[c]	0.964, 0.367	0.036, 0.633
Subsample - H (DTI<30, >50)[c]	0.896, 0.107	0.104, 0.893
Subsample - B (DTI<30, >50)[c]	0.911, 0.083	0.089, 0.917
Subsample - W (DTI<30, BAD=0)[c]	0.978	0.022
Subsample - H (DTI<30, BAD=0)[c]	0.906	0.094
Subsample - B (DTI<30, BAD=0)[c]	0.939	0.061

[a] The estimated probabilities are reported. These averages are calculated by computing the probability of each individual in the group and averaging across the group.

[b] Each set of coefficients was used to calculate the mean probability over the whole sample (all the different subgroups simultaneously). Thus, we can compare the effect of each estimate on the whole sample.

[c] Representing DTI and bad credit, these calculations (which vary from bank to bank) compare the different mean probabilities for each subgroup based on the group's estimates and where some of the characteristics are held fixed to good or bad levels.

ly used ML method. Therefore, the probability of incorrectly finding discrimination when it does not exist is smaller for this approach than for the ML method. This could be particularly advantageous for policy makers.

The main empirical findings indicate potential discrimination at two of three banks analyzed. In one case, the more traditional analysis of simply investigating the intercept (or race variable) reveals no discrimination, while the more complete analysis reveals potential discrimination. Subsequent analysis of individual loan files at these banks resulted in the OCC determining not to refer the banks for disparate treatment. We strongly emphasize that while the statistical modeling can provide evidence of disparate treatment, a comprehensive analysis, of which the statistical modeling is one part, must be conducted before one can assert that a bank is discriminating.

We believe that this estimator offers significant gains in the research efforts used to determine whether mortgage lending practices represent fair lending. The technique applies to most types of data sets used for fair lending analysis. It minimizes both types of errors relative to more traditional methods. Overall, the probability of making a mistake or reaching the wrong conclusion, in either direction, is smaller with the GME estimator.

Appendix

Variable Description

Excess DTI: A variable valued at one if DTI value exceeds 41 percent; otherwise, the value is zero.

Loan-to-Value: A continuous valued LTV variable.

Credit Score: A variable derived from bank 1's underwriting guidelines manual. Bank 1 collects up to three credit bureau scores and chooses the middle score of both the applicant and the coapplicant. If there are two credit bureaus reporting, the bank takes the lower of the scores. The bank compares the applicant and coapplicant scores and calculates a score variable that is the lower of the two.

Bad Credit: This variable is derived from information on clean credit record and is bank specific. For bank 1, if there are no public records of bankruptcy, foreclosure, unpaid judgments or collections, no late mortgage or rent payments, etc., for the past 24 months, this variable has a value of zero. If bad credit records are observed, this variable has a value of one.

Gift Funds: The sum of gifts and grants.

Public Record: A variable with public record information that does not overlap with the bad credit variable.

Explanation: A dummy variable equal to one if there was any type of derogatory explanation and equal to zero otherwise.

Hispanic: A dummy variable equal to one if the applicant is of Hispanic origin, and equal to zero otherwise.

Black: A dummy variable equal to one if the applicant is of black origin and to zero otherwise.

Front Ratio: A dummy variable equal to one if the housing DTI ratio exceeds the individual program guidelines that the application falls under.

Insufficient Funds: A dummy variable equal to one if there were not sufficient funds to close.

Fails PMI: A dummy variable equal to one if the applicant applied for private mortgage insurance and was denied by the insurance company.

Relationship: A dummy variable equal to one if applicant has any type of relationship with the bank (deposit or previous loan at the bank).

Low LTV: A dummy variable equal to one if the applicant has an LTV less than or equal to 75.

References

Calem, Paul, and Michael Stutzer. 1995. The Simple Analytics of Observed Discrimination. *Journal of Financial Intermediation* 4(3):189–212.

Courchane, Marsha, and MariaGloria Cobas. 1995. Using Statistical Techniques to Evaluate Fair Lending Practices at National Banks. *OCC Quarterly Journal*. 14(2):13–9.

Courchane, Marsha, William Lang, Dennis Glennon, and Mitch Stengel. 1995. Using Statistical Models to Identify Unfair Lending Behavior. In *Fair Lending Analysis: A Compendium of Essays on the Use of Statistics*, ed. A. Yezer. Washington, DC: American Bankers Association.

Crawford, David L., Robert A. Pollack, and Francis Vella.1998. Simple Inference in Multinomial and Ordered Logit. *Econometric Review* 17(3):289–99.

Golan, Amos, George Judge, and Douglas Miller. 1996. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. New York: Wiley.

Golan, Amos, George Judge, and Jeffrey Perloff. 1996. A Maximum Entropy Approach to Recovering Information from Multinomial Response Data. *Journal of the American Statistical Association* 91(434):841–53.

Greene, William H. 1999. *Econometric Analysis*. Upper Saddle River, NJ: Prentice-Hall.

Heckman, Jean J. 1998. Detecting Discrimination. *Journal of Economic Perspectives* 12(2):101–16.

Horne, David. 1997. Mortgage Lending, Race and Model Specification. *Journal of Financial Services Research* 11(1/2):43–68.

Ladd, Helen. 1998. Evidence on Discrimination in Mortgage Lending. *Journal of Economic Perspectives* 12(2):41–62.

Longhofer, Stanley D., and Stephen Peters. 1998. Self-Selection and Discrimination in Credit Markets. Working Paper No. 9809. Federal Reserve Bank of Cleveland.

Maddala, George S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.

McFadden, Daniel. 1974. The Measurement of Urban Travel Demand. *Journal of Public Economics* 3(4):303–28.

Munnell, Alicia, Geoffrey Tootell, Lynne Browne, and James McEneaney. 1996. Mortgage Lending in Boston: Interpreting HMDA Data. *American Economic Review* 86(1):25–53.

Soofi, Ehsan S. 1992. A Generalized Formulation of Conditional Logit with Diagnostics. *Journal of the American Statistical Association* 87(419):812–16.

Stengel, Mitch, and Dennis Glennon. 1999. Evaluating Statistical Models of Mortgage Lending Discrimination: A Bank-Specific Analysis. *Real Estate Economics* 27(2):299–334.

Yinger, John. 1996. Discrimination in Mortgage Lending: A Literature Review. In *Mortgage Lending, Racial Discrimination, and Federal Policy,* ed. John Goering and Ron Wienk, 29–74.Washington, DC: Urban Institute Press.