University of Victoria
Department of Economics

Econometrics Working Paper EWP0604

ISSN 1485-6441

ON THE ROBUSTNESS OF RACIAL DISCRIMINATION FINDINGS IN MORTGAGE LENDING STUDIES

Judith A. Clarke
Nilanjana Roy

Department of Economics, University of Victoria
Victoria, B.C., Canada V8W 2Y2

&
Marsha J. Courchane

ERSGroup
Washington, D.C., U.S.A. 20037

Abstract

That mortgage lenders have complex underwriting standards, often differing legitimately from one lender to another, implies that any statistical model estimated to approximate these standards, for use in fair lending determinations, must be misspecified. Exploration of the sensitivity of disparate treatment findings from such statistical models is, thus, imperative. We contribute to this goal. This paper examines whether conclusions from several bank-specific studies, undertaken by the Office of the Comptroller of the Currency, are robust to changes in the link function adopted to model the probability of loan approval and to the approach used to approximate the finite sample null distribution for the disparate treatment hypothesis test. We find that discrimination findings are reasonably robust to the range of examined link functions, which supports the current use of the logit link. Based on several features of our results, we advocate regular use of a resampling method to determine p-values.

Keywords: Logit; Mortgage lending discrimination; Fair lending; Stratified sampling; Binary response; Semiparametric maximum likelihood; Pseudo log-likelihood; Profile log-likelihood; Bootstrapping.

Author Contact:

Judith A. Clarke, Tel.: +1 250 721 8542; fax.: +1 250 721 6214

Email address: jaclarke@uvic.ca

1. Introduction

An issue of continuing interest among regulators, economists, consumers and policy makers concerned with the U.S. housing market, is the feasibility of Congress's goal "that every American family be able to afford a decent home in a suitable environment"[1]. One potential obstacle is disparate treatment in the mortgage lending market against minorities. Discrimination can take many forms, including turning down a loan application, based on certain personal characteristics of the applicant such as race, age, and gender[2]. Such action is prohibited under U.S. laws.

Data collected by the Federal Financial Institutions Examination Council (FFIEC) under the Home Mortgage Disclosure Act (HMDA), enacted by the Congress in 1975, are designed to help regulators enforce fair lending laws. Results indicate that loan approval rates for minority applicants have been and continue to be lower than those of white applicants, but this evidence alone need not infer that lending discrimination exists, as we must account for differences in variables representing creditworthiness.

Statistical models provide one way to control for such variables. Indeed, several regulatory agencies (e.g., the Office of the Comptroller of the Currency) estimate bank-specific logit models that aim to approximate underwriting procedures. In each case, the outcome variable is the probability that a home mortgage loan is approved. Although the regulators do not base a finding of discrimination solely on statistical models, it is, nevertheless, vital to appreciate the extent that specification issues

[1] National Housing Act of 1934.

[2] Discrimination in mortgage lending can take other forms, e.g., prescreening, unfavorable terms for an approved loan and redlining. Our concern is with discrimination in the loan approval process.

associated with the regressions affect discrimination findings[3]. We contribute towards this understanding.

In this paper, we re-examine the statistical findings from five OCC bank-specific regulatory examinations of home-purchase mortgage lending; see Courchane *et al.* (2000b) for a detailed description of OCC review practice. We ask the question, "To what extent are the discrimination findings from the statistical models sensitive to the distribution adopted to model the probability function? " We also examine whether the test findings based on asymptotic approximations, used by the regulators to determine evidence of discrimination, differ when we adopt bootstrapping tools to approximate unknown finite sample null distributions. In essence, our study satisfies Commandment Ten from Kennedy's (2002) Ten Commandments of Applied Econometrics: **Thou shalt confess in the presence of sensitivity.**

Our examination is not the first to consider robustness of the discrimination outcomes from the OCC bank-specific models: Clarke and Courchane (2005) and Dietrich (2005a) consider the form of stratified sampling adopted; Blackburn and Vermilyea (2004) illustrate that combining results across several models leads to evidence of discrimination against blacks in mortgage lending, despite the lack of such evidence from the individual models; and Dietrich (2005b) considers how omitted variables affect underwriting models. However, to the best of our knowledge, no information exists on the sensitivity of the discrimination results in the two ways we explore: the assumed probability distribution and the approximation used to determine statistical significance. Like those before us, our findings assist regulators,

[3] Calem and Longhofer's (2002) finding that statistical analysis and the more-traditional comparative file reviews complement each other by balancing off some of the issues associated with each method further supports the importance of undertaking sensitivity studies.

bank officials and those bodies to which cases are referred (the Department of Justice and the Department of Housing and Urban Development) on the directions that may cause issue with statistical underwriting models.

Our approach is to compare the discrimination outcomes from logistic regressions, using five OCC bank-specific studies undertaken in the late 1990s, with those arising from three alternative link functions: probit, gompit and complementary log log; the latter two being examples of asymmetric links. Our move away from logit complicates estimation, as the OCC models are estimated with samples stratified on the basis of race and outcome, easily handled with a logistic regression but not so with the other links. We consider two consistent estimators: one estimator is user-friendly but, depending on the link choice, may be asymptotically inefficient, while the other estimator, the maximum likelihood estimator, has computational disadvantages. By adopting two estimation principles, we are able to ascertain for practitioners whether the computationally simpler estimator results in substantively the same discrimination finding as the maximum likelihood estimator.

This paper is organized into the following sections. Section 2 presents our model setup, including a discussion of the link functions; section 3 considers estimation methods and hypothesis testing procedures when the data are stratified both, endogenously, by the dependent variable and, exogenously, by our categorical race covariate; section 4 details our data, including particulars on covariates; section 5 provides the empirical results and section 6 concludes.

2. Binary response model, cdfs and link functions

Our adopted statistical models arise from bank-specific examinations that aim to model underwriting practices. A regression models whether a loan is approved or denied as a function of covariates such as race, loan-to-value ratio (LTV) etc[4]. More generally, for each bank, we assume a binary outcome dependent variable, y_j, which takes values $y_j = 0$, when a mortgage loan application is denied, and $y_j = 1$, when it is approved; j=1,...,N, the number of applicants whose loan applications have been denied or approved. There are K race categories (e.g., White, African American, Hispanic American) with a vector x_j, of dimension K, which contains categorical dummy variables that describe the race of an applicant: $x_{jk}=1$ if the j'th applicant belongs to racial group k (k=1,…,K), 0 otherwise; then, $x_j = [x_{j1}, x_{j2}, ..., x_{jK}]'$. There is an additional q-dimensional vector z_j containing other discrete and continuous variables describing characteristics of the loan applicant. Our aim is to estimate a binary response model of the form:

$$h(P_1(w_j;\beta)) = w_j'\beta \quad , \qquad j=1,2, ..., N \qquad (2.1)$$

where, for i=0,1, $P_i(w_j;\beta) = pr(y_j = i \mid w_j;\beta)$, $w_j' = [x_j' \; z_j']$, h(.) is the link function and β is a p-dimensional coefficient vector (p=K+q); $\beta=[\beta_1, \beta_2, \ldots, \beta_K, \beta_{K+1}, \ldots, \beta_p]'$. Having appropriately estimated (2.1), the regulator ascertains discrimination by testing whether the impacts of the racial categorical variables are equal; i.e., we test the K!/(2((K-2)!)) distinct null hypotheses, $H_0^m : \beta_m - \beta_k = 0$, m≠k, m, k=1,…,K; against, usually, a one-sided alternative hypothesis (e.g., that discriminatory treatment is against African Americans).

[4] Covariates are provided in Table 3.

We can equivalently write (2.1) as: $P_1(w_j;\beta) = h^{-1}(w_j'\beta) = F(w_j'\beta)$ where F(.) denotes a cumulative distribution function (cdf). Statistical analyses undertaken by fair lending regulators have, to our knowledge, exclusively considered a logistic cdf, which corresponds to the logit link function: $P_1(w_j;\beta) = \exp(w_j'\beta)/(1+\exp(w_j'\beta))$.

Another commonly used link function is the normit, which results in the probit regression: $P_1(w_j;\beta) = \Phi(w_j'\beta)$, where $\Phi(.)$ is the cumulative distribution function of a standard normal variate. The logistic cdf has fatter tails than the probit cdf, appoaching zero and one more slowly. The choice of a logit or a normit link can lead to different conclusions when (a) there are large numbers of observations or (b) many of the predicted probabilities are close to zero or one. The bank data sets we examine range from 145 to 420 observations, not particularly large compared to the thousands of observations often used in the estimation of binary response models, which may lead to little difference between logit and probit models. However, the percentage distribution of the predicted probabilities from logistic regressions for our five banks, denoted as Bank 1 to Bank 5 for confidentiality reasons, indicates that a significant percentage of the predictions are close to one for Banks 2, 3 and 4, supporting our exploration of probit; see Table 1.

One concern with using logit or probit models is that the probability $P_1(w_j;\beta)$ approaches zero and one at the same rate, as their links are symmetric. This may be a questionable assumption for the sub-populations of bank applications, which feature few denials compared to approvals. Incorrectly assuming a symmetric link might lead to substantial bias in our coefficient estimates and detrimentally affect our disparate treatment test. We consider two common asymmetric links: gompit and cloglog. The gompit model is $P_1(w_j;\beta) = \exp(-\exp(-w_j'\beta))$, with $P_1(w_j;\beta)$ approaching zero faster than one. The cloglog, or complementary log-log, model is $P_1(w_j;\beta) = 1-\exp(-\exp(w_j'\beta))$, with $P_1(w_j;\beta)$ approaching one faster than zero.

3. Estimation issues

In order to estimate (2.1), we need information on the y_j and w_j variables for the N applicants. For cost and efficiency reasons, the OCC draws a stratified choice based sample (SCBS) of size n from the N available. This enables the sample to contain information on a sufficient number of minority denied loans. Let $N_{i,k}$ be the number of applicants in racial class k with y_j=i, i=0,1, k=1,2,...,K; $\sum_{i=0}^{1}\sum_{k=1}^{K} N_{i,k} = N$. Under SCBS, $n_{i,k}$ applicants are drawn from the $N_{i,k}$ available, i=0,1, k=1,2,…,K; $\sum_{i=0}^{1}\sum_{k=1}^{K} n_{i,k} = n$.

Specifically, from each of the S=2K strata or classes, we sample $n_{i,k}$ units with y_j=i and x_j such that the case belongs to race k, which we denote by $x_j \in k$. The associated w_{ijk} values are subsequently recorded; the k subscript noting that the case belongs to the k'th race class, k=1,2,...,K, i=0,1, j=1,2,...,$n_{i,k}$. The likelihood function is:

$$L^{SCBS} = \prod_{i=0}^{1}\prod_{k=1}^{K}\prod_{j=1}^{n_{i,k}} pr(w_{ijk} \mid y_j = i, x_j \in k)$$

$$= \prod_{i=0}^{1} \prod_{k=1}^{K} \prod_{j=1}^{n_{i,k}} pr(y_j = i \mid w_{ijk}, x_j \in k) g(w_{ijk} \mid x_j \in k) / pr(y_j = i \mid x_j \in k)$$

$$= \prod_{i=0}^{1} \prod_{k=1}^{K} \prod_{j=1}^{n_{i,k}} P_i(w_{ijk}; \beta \mid x_j \in k) g(w_{ijk} \mid x_j \in k) / pr(y_j = i \mid x_j \in k) \qquad (3.1)$$

using Bayes' Rule and the same notations as in (2.1)[5]. As $pr(y_j=i)=\int P_i(w_{ij};\beta) dG(w_{ij})$, where G(.) denotes the appropriate marginal distribution function, we cannot separate out $g(w_{ij})$ when estimating β.

Estimation of the log-likelihood function from (3.1)

$$\ell = \sum_{i=0}^{1} \sum_{k=1}^{K} \sum_{j=1}^{n_{i,k}} \log P_i(w_{ijk}; \beta \mid x_j \in k) + \sum_{i=0}^{1} \sum_{k=1}^{K} \sum_{j=1}^{n_{i,k}} \log g(w_{ijk} \mid x_j \in k)$$

$$- \sum_{i=0}^{1} \sum_{k=1}^{K} \sum_{j=1}^{n_{i,k}} \log pr(y_j = i \mid x_j \in k) \qquad (3.2)$$

requires that we specify $P_i(.)$, in addition to modeling g(.). We use semiparametric maximum likelihood estimation, where the term "semiparametric" is taken to mean that we parametrically model $P_i(w_{ijk};\beta \mid x_j \in k)$ (for example, using one of the links provided in the previous section) and we nonparametrically model $g(w_{ijk} \mid x_j \in k)$; e.g., Scott and Wild (2001). The literature proposes two routes for solving for estimates for β using this semiparametric approach: maximizing either a profile log-likelihood or a pseudo log-likelihood. The former, considered in the next sub-section, leads to maximum likelihood estimates irrespective of the form of the link function, but is less user-friendly in the sense of not being straightforward to code in standard packages. The alternative path of maximizing a pseudo log-likelihood is

[5] We use the notation pr(.) to denote the probability function for our discrete outcome variable and the notation g(.) for the joint data density function associated with the regressors.

uncomplicated to code, but, for many common link functions, has severe computational issues. Accordingly, we consider a computationally simpler estimator, which is consistent, but not usually asymptotically efficient, that is available via the pseudo log-likelihood route.

3.1 A profile log-likelihood route

Without proof (see Scott and Wild, 2001), the profile log-likelihood for β $\ell_P(\beta) = \ell(\beta, \hat{g}(\beta))$, after nonparametrically modeling the density of w by replacing its (unknown) cumulative probability distribution with its empirical distribution[6], is:

$$\ell_P(\beta) = \ell^*(\beta, \rho(\beta)) = \sum_{j=1}^{n} \Big\{ (1-y_j)\log(1-P_1(w_j;\beta)) + y_j \log P_1(w_j;\beta) -$$

$$\sum_{k=1}^{K} \Big(S_{jk} \log\big[\mu_{0k} P_0(w_j;\beta) + \mu_{1k} P_1(w_j;\beta)\big] - \big(m_{1k}\rho_{1,k} -$$

$$(m_{0k} + m_{1k}) \log(1+\exp(\rho_{1,k})) \big) / n \Big) \Big\} \quad (3.3)$$

where: $m_{ik} = (N_{i,k}-n_{i,k})$; $\mu_{ik} = N_{+,k} - \frac{m_{ik}(1+\exp(\rho_{1,k}))}{(\exp(\rho_{1,k}))^i}$; S_{jk}=1 if the j'th applicant belongs to stratum k, 0 otherwise; i=0,1, k=1, …, K, j=1,2, …, n; $N_{+,k}=N_{0,k}+N_{1,k}$ and $\sum_{k=1}^{K} N_{+,k} = N$. Excluding variance-covariance matrix parameters, this objective function has (p+K) unknown parameters, p from β and K from $\rho_{1,1} \ldots \rho_{1,K}$, which arise from the nonparametric modeling of the density of w; these additional parameters relate to unconditional probabilities. Specifically, let $Q_{i,k}$ be the

6 The empirical distribution is the maximum likelihood estimate of an unknown distribution function; e.g., Kiefer and Wolfowitz (1956).

unconditional probability that y=i in stratum k with $\sum_{i=0}^{1} Q_{i,k} = 1$, then

$\rho_{i,k}=\log(Q_{i,k}/Q_{0,k})$.

The criterion (3.3) is highly non-linear in β and ρ (=[$\rho_{1,1} \ldots \rho_{1,K}$]), although, for fixed β, the ρ parameters are orthogonal, as each involves only observations from the relevant stratum. We apply the iterative routine suggested by Scott and Wild (2001, p.18) to solve for the maximum likelihood solutions, say $\hat{\beta}_{PR}$ and $\hat{\rho}_{PR}$; throughout this paper, a subscript "PR" will refer to a statistic or a p-value obtained by means of the profile log-likelihood. Specifically, the additional sub-population information on $N_{i,k}$ provides initial, consistent, estimates of $\rho_{1,1} \ldots \rho_{1,K}$, say $\overline{\rho}_{1,1} \ldots \overline{\rho}_{1,K}$, which are then used to maximize (3.3) for estimates of β, say β^*. With β fixed at β^*, we again maximize (3.3) to obtain new ρ estimates and so on until we converge to $\hat{\beta}_{PR}$ and $\hat{\rho}_{PR}$. When solving for β, our algorithm used the score vector and information matrix provided by Scott and Wild (2001, p.18). Convergence usually resulted in fewer than five such major loops, with ten major loops being the highest number required for our data sets.

3.2 <u>A pseudo log-likelihood route</u>

Without proof (e.g., Scott and Wild, 2001), when we model g(.) nonparametrically, maximizing ℓ is equivalent to maximizing the pseudo log-likelihood function:

$$\ell^* = \sum_{i=0}^{1} \sum_{k=1}^{K} \sum_{j=1}^{n_{i,k}} \log P_i^*(w_{ijk};\beta,\kappa_k) \tag{3.4}$$

with $\text{logit}\, P_i^*(w_{ijk};\beta,\kappa_k) = \text{logit}\, P_i(w_{ijk};\beta \mid x_j \in k) + \log \kappa_k$ defining $P_i^*(w_{ijk};\beta,\kappa_k)$.

The parameter κ_k is the ratio of the sampling rates for race class k:

$$\kappa_k = \left(\frac{n_{1,k}}{\text{pr}(y_j = 1 \mid x_j \in k)}\right) \Big/ \left(\frac{n_{0,k}}{\text{pr}(y_j = 0 \mid x_j \in k)}\right) \tag{3.5}$$

and

$$\text{logit}\; P_1(w_{1jk};\beta \mid x_j \in k) = \log\left(\frac{P_1(w_{1jk};\beta \mid x_j \in k)}{P_0(w_{0jk};\beta \mid x_j \in k)}\right). \tag{3.6}$$

The objective function (3.4) is termed a "pseudo log-likelihood" because in general it is not equal to the log-likelihood ℓ; they are equal at their maximums.

The parameters $\kappa_1,\ldots,\kappa_K$ are non-identifiable in a multiplicative intercept model, such as logit but are identifiable in a non-multiplicative intercept model, such as probit, gompit and cloglog, although there may be some multicollinarity issues that might cause convergence concerns. To further complicate computational matters, the stationary point of (3.4) occurs at a saddlepoint in the combined parameter space; Scott and Wild (2001).

This may suggest that it is preferable to avoid working with the pseudo log-likelihood but the supplementary information available on sub-population stratum totals enables us to consistently estimate κ_k; specifically:

$$\hat{\kappa}_k = \left(\frac{n_{1,k}}{N_{1,k}}\right) \Big/ \left(\frac{n_{0,k}}{N_{0,k}}\right) \tag{3.7}$$

is a consistent estimator of κ_k. Use of this rule for the logit link leads to the estimator of β examined by Clarke and Courchane (2005) in their fair lending study; this estimator is known to be in fact the maximum likelihood estimator of β[7]. That is, for the logit link, maximum likelihood estimates of all the parameters, except stratum constants, are obtained by estimating the model as if it were from a simple random sample; a minor adjustment provides the maximum likelihood estimates of stratum constants.

With non-multiplicative links, use of $\hat{\kappa}_k$ will lead to a consistent, but not necessarily asymptotically efficient, estimator of β - we denote this estimator as $\hat{\beta}_{PS}$ [8] - a pseudo log-likelihood one-step estimator; hereafter, a subscript "PS" will refer to a statistic or p-value obtained via the pseudo log-likelihood. Obtaining the maximum likelihood estimator requires iteration, taking account that we are locating a saddlepoint, which can be computationally difficult, compared to obtaining $\hat{\beta}_{PS}$. Comparing outcomes for our disparate treatment tests, using the (consistent but asymptotically inefficient) one-step pseudo log-likelihood estimator, $\hat{\beta}_{PS}$, and the maximum likelihood estimator obtained by iteration via the profile log-likelihood, $\hat{\beta}_{PR}$, is instructive, as the former is easier to code. It may be that the gains in efficiency do not lead to practical changes in test outcomes.

3.3 Variance-covariance matrix

Testing the null hypotheses of interest also requires variance-covariance matrices for our estimators obtained from the profile and pseudo log-likelihood routes. When using the pseudo log-likelihood procedure for either the logit link or another

[7] Indeed, this holds for multiplicative intercept models with a complete set of stratum constants.
[8] It is, in fact, one form of the Manski-McFadden (1981) estimator.

multiplicative intercept model, a consistent estimator of var($\hat{\beta}_{PS}$), say var$_{est}$($\hat{\beta}_{PS}$), is given by (e.g., Scott and Wild, 1986): var$_{est}$($\hat{\beta}_{PS}$) = var*($\hat{\beta}_{PS}$) - $\begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}$

where var*($\hat{\beta}_{PS}$) is the inverse of the pseudo-information matrix for $\hat{\beta}_{PS}$, assuming simple random sampling, and A is a (K×K) diagonal matrix with elements:

$$\left\{ a_k = \left[\left(\frac{1}{n_{0,k}} + \frac{1}{n_{1,k}} \right) - \left(\frac{1}{N_{0,k}} + \frac{1}{N_{1,k}} \right) \right] \right\}; \text{ k=1,2, …, K.} \qquad (3.8)$$

The first term is the reduction in variance from stratifying, while the second term is the increase in variance arising from using $\hat{\kappa}_k$ to estimate κ_k.

With a non-multiplicative intercept model, such as probit, gompit and cloglog, the one-step estimator $\hat{\beta}_{PS}$ is obtained by maximizing the pseudo log-likelihood (3.4) with $\hat{\kappa}_k$ as the estimator of κ_k. A consistent estimator of var($\hat{\beta}_{PS}$) is given by var*($\hat{\beta}_{PS}$), the inverse of the pseudo-information matrix; see, e.g., Scott and Wild (2001). The disparate treatment null hypotheses – $H_0^m : \beta_m - \beta_k = 0$, m≠k, m, k=1,…,K, are tested using $t_{PS}^m = (\hat{\beta}_{PS,m} - \hat{\beta}_{PS,k}) / \text{s.e.}(\hat{\beta}_{PS,m} - \hat{\beta}_{PS,k})$, where $\hat{\beta}_{PS} = [\hat{\beta}_{PS,1}, \hat{\beta}_{PS,2}, \ldots, \hat{\beta}_{PS,K}, \ldots, \hat{\beta}_{PS,p}]'$ and $\text{s.e.}(\hat{\beta}_{PS,m} - \hat{\beta}_{PS,k}) = \sqrt{\text{var*}(\hat{\beta}_{PS,m}) + \text{var*}(\hat{\beta}_{PS,k}) - 2\,\text{cov}(\hat{\beta}_{PS,m}, \hat{\beta}_{PS,k})}$. It follows (e.g., Scott and Wild, 2001), that the limiting null distribution for t_{PS}^m is standard normal (SN).

As we use the analytic score vector and Hessian matrix to solve for the maximum likelihood estimator, $\hat{\beta}$, by way of the profile log-likelihood, we estimate this estimator's asymptotic covariance matrix as the inverse of the information matrix, evaluated at the maximum likelihood estimates; see. e.g., Scott and Wild (2001, pp. 14-15).

3.4 Bootstrapped p-values

An alternative route to using an asymptotic S.N. distribution to approximate the null distribution is to use bootstrapping. We now describe that methodology. To allow for the finite sub-population of N applicants presenting at a bank and the use of SCBS to form the sample of n applicants, when forming our bootstrapped p-values we take the following steps, primarily suggested by Booth et al. (1994).

Step 1: The first step is to create an empirical subpopulation for a bank. Let $f_{i,k}=n_{i,k}/N_{i,k}$ so that $N_{i,k}=g_{i,k}n_{i,k}+s_{i,k}$, $0\leq s_{i,k}\leq n_{i,k}$, $g_{i,k}=\text{int}(1/f_{i,k})$, i=0,1, k=1,2,…,K. If $g_{i,k}$ is an integer for all i,k then we can create a unique empirical subpopulation by combining $g_{i,k}$ copies of the k^{th} stratum's sample; e.g., Gross (1980). More often than not, this is not possible, as, typically, one or more $g_{i,k}$ are not integers. Then, we create an empirical subpopulation by combining $g_{i,k}$ copies of the appropriate stratum's sample with a without replacement sample of size $s_{i,k}$ from the original sample.

Step 2: We draw B without replacement resamples of size n, stratified as per the original sample, from the empirical subpopulation; i.e., each resample has stratum denial ratios that match the original sample. For a particular link choice, we estimate the regression models for each resample, forming B values of the K!/(2((K-2)!)) test statistics to examine $H_0^d : \beta_m - \beta_k = 0$, m≠k, m,k=1,…,K, d=1,…, K!/(2((K-2)!)); denote the bootstrapped statistics as $t_1^d, \ldots t_B^d$. As our data may not have been drawn from a subpopulation that satisfies H_0^d, we follow the advice of Hall and Wilson (1991) by centering when forming these bootstrapped statistics, which has the effect of increasing power. That is, we form $t_{A,i}^d = \dfrac{(b_i^m - b_i^k) - (\hat{\beta}_{A,m} - \hat{\beta}_{A,k})}{se(b_i^m - b_i^k)}$ (i=1,…,B; A = PS or PR), where b_i^m is the estimate of β_m from the i^{th} bootstrap resample and so on[9].

Step 3: Let $t_{A,samp}^d$ be the statistic value from the original sample for testing H_0^d. The bootstrapped p-value is then the simulated number of rejections obtained by comparing $t_{A,1}^d \ldots t_{A,B}^d$ with $t_{A,samp}^d$; e.g., the bootstrapped p-value is

$p^d = (1/B)\sum_{i=1}^{B} I(t_{A,i}^d < t_{A,samp}^d)$ when the alternative hypothesis is $H_a^d : \beta_m - \beta_k < 0$.

Step 4: Repeat *Steps 2* and *3* for each bank using the other links.

We follow the pretesting method advocated by Davidson and MacKinnon (2000) to choose B, the number of bootstraps; typically, this led to B=99 for our chosen 5% nominal level of significance.

[9] As we are sampling from a finite subpopulation, we resample without replacement, rather than with replacement, as the latter would not be consistent with our original data collection.

4. Data

The data used in this research, collected by the OCC in the course of several fair lending examinations in the late 1990s, come from five separate national banks geographically distributed from the East to the West and the Midwest. Each statistical model, structured to best reflect banks' underwriting procedures in the approval of a mortgage application, uses a combination of explicit elements collected from bank loan files and variables created from the primary data to measure credit worthiness as independent variables. A list of regressors included in the model specifications of our five banks is given in Table 2 while Table 3 provides their brief broad meanings. The specific definition of each variable depends on bank-specific factors; e.g., DTI is a one/zero binary regressor with a threshold DTI ratio determining the switch for one bank, while it is the actual DTI ratio for another bank.

The use of samples stratified by race and loan outcome results in sample racial stratum denial rates that differ from those for the subpopulation of N applicants. We provide denial rates in Figure One. Racial groups are denoted as follows: Whites - k=1; African Americans - k=2; Hispanic Americans - k=3. We see that for Banks 1,4 and 5 there are three racial strata (K=3), while for Banks 2 and 3 there are only two (K=2). The subpopulation measures are denoted by "N", the sample measures by "n", and denial of a loan application by "0"; e.g., "N01" is the number of denied whites loans, "n2" is the number of African Americans in the sample, and so on. We observe denial rates for African Americans that always exceed those for Whites and, when present, the denial rates for Hispanic Americans fall between those for African Americans and Whites.

5. Results

We estimated the five bank-specific models, with the covariates summarized in Table 2, using the estimators $\hat{\beta}_{PS}$ and $\hat{\beta}_{PR}$ for the four links detailed in section 2; recall that these two estimators are equivalent for the logit link but not for the other three studied links. We used Gauss, with the MAXLIK sub-routine, to obtain the maximum likelihood estimates from the profile log-likelihood, and EViews, Stata and Gauss – to satisfy ourselves that results were similar across standard packages – to obtain the one-step pseudo log-likelihood estimates.

Prior to comparing p-values, we detail two measures of fit that may provide guidance on link preference. One way is via the value of the average log-likelihood function; Table 4 provides this information, with the measures given relative to the average log-likelihood value for the logit link; e.g., a number less than one indicates that the logit link has a smaller average log-likelihood value. The results suggest this measure is quite similar across the link functions, with the average log-likelihood values being different by at most 6%. This small difference could be arising due to finite sample bias.

As the logit link's average profile log-likelihood and pseudo log-likelihood values are identical, the numbers in Table 4 also provide one measure of loss, for the non-multiplicative links, in using the one-step pseudo log-likelihood approach over the profile log-likelihood method. For the banks we examine, the loss in average log-likelihood value is at most 5.2% with the average loss being 1.6%; this suggests that it may be practically reasonable to work with the computationally easier pseudo log-likelihood.

Another commonly reported measure of model performance is the percentage correctly predicted, obtained by comparing the predicted and observed outcomes of the binary response. Classification of the predicted probabilities into 0/1 outcomes is achieved by relating them to a chosen cutoff value and counting the matches of observed and predicted outcomes; a classification is "correct" when the model predicts the applicant's loan disposition. We provide this information in Tables 5a and 5b, using three cutoff values – the standard value of "0.5", a reasonable choice in samples with a balance of 1/0 outcomes, "sf", which is the frequency of y=1 observations in the sample, and "spf", which is the frequency of y=1 observations in the subpopulation; Table 5a presents the outcomes from the pseudo log-likelihood approach, while those from the profile log-likelihood route are given in Table 5b. As our subpopulations are unbalanced, as are also the samples despite the OCC's oversampling of denials, the "spf" and "sf" cutoffs are likely more realistic and sensible; e.g., Cramer (1999).

We observe only minor differences between the profile and pseudo log-likelihood percentages. For the few cases when there are practical differences, it is often less than two percentage points, although significant variations arise with the cloglog link. The influence of the cutoff value is evident; when it is "0.5" or "sf", the models do better at predicting approvals than denials, while their performance is more equitable with "spf". Then, the models do better at predicting denials than approvals. Overall, the models correctly classify, approximately, 65% to 90% of outcomes, irrespective of cutoff value.

We observe little difference with prediction abilities across links. Given its asymmetry, the gompit link predicts loan approvals better than the other links, with an associated minor loss (usually) in predicting denials. The logit link often correctly predicts more denied loans than the other links, although there is little difference between this link's ability and that of the cloglog link with the profile estimator.

In summary, using the two measures of fit, we find there is little practical gain in choosing one link over another for the banks under study. When comparing overall classification ability, irrespective of loan disposition, the computationally easier logit link is likely as good a choice as any of the other links examined here.

We now focus on the hypothesis tests for racial discrimination. Given our notation that β_k is the coefficient belonging to the k'th racial dummy variable with k=1 for Whites, k=2 for African Americans and k=3 for Hispanic Americans, the relevant null hypotheses tested by the OCC are: $H_0^1 : \beta_1 - \beta_2 = 0$, $H_0^2 : \beta_1 - \beta_3 = 0$ and $H_0^3 : \beta_2 - \beta_3 = 0$. Our alternative hypotheses corresponding to H_0^1 and H_0^2 are $H_a^1 : \beta_1 - \beta_2 > 0$ and $H_a^2 : \beta_1 - \beta_3 > 0$ to reflect our prior belief that disparate treatment, should it exist, is expected to favour White applicants; an exception is for Bank 3 for which we consider $H_a^2 : \beta_1 - \beta_3 < 0$ due to particular features for this bank. As we have no prior beliefs regarding discrimination between African Americans and Hispanic Americans, we examined a two-sided alternative with H_0^3, $H_a^3 : \beta_2 - \beta_3 \neq 0$.

In Table 6, we report p-values for t-ratios for testing the nulls using the standard normal (SN) distribution, the limiting null distribution, for both the profile and the one-step pseudo log-likelihood approaches. We also present p-values based on the bootstrap procedure, outlined in sub-section 3.4, for the computationally simpler one-step pseudo log-likelihood method. The legal standard for a statistically significant race effect is two or three standard deviations, which suggests a nominal 5% or 1% significance level[10]. Such a choice effectively gives the benefit of doubt to the bank, as it implies a belief in nondiscrimination unless the sample evidence is extreme in suggesting otherwise. We adopt a 5% level. A bold font highlights rejections at this level.

Examination of the SN p-values reveals that general similarities exist in the pattern of outcomes. In particular, out of the eleven cases (H_0^1 for Banks 1, 2, 4 and 5, H_0^2 for Banks 1, 3, 4 and 5, and H_0^3 for Banks 1, 4 and 5), only three cases (H_0^1 for Banks 1 and 4, and H_0^2 for Bank 3) have inconsistent results across the two estimation methods. Comparing the results across links, we find inconsistent findings again for only those three cases. In other words, irrespective of whether we use the profile or the one-step pseudo estimators and regardless of the link choice, the SN p-values suggest: Bank 1 favors Whites over Hispanic Americans; Bank 2 favors Whites over African Americans; Bank 4 does not discriminate between Whites and Hispanic American or between African Americans and Hispanic Americans; and Bank 5 does not discriminate.

Thus, our results show that there is usually no qualitative difference in test outcomes between the SN p-values, for the probit, gompit and cloglog models[11], from

[10] See, e.g., Kaye and Aicken (1986). LaCour-Little (1999) provides a useful commentary on this matter.
[11] Recall that there should not be any difference in the test outcome from the profile and one-step pseudo log-likelihood methods for the logit link.

the pseudo and profile routes[12]. This is a useful result for the practitioner, as obtaining estimates via the pseudo log-likelihood is substantially easier than from the profile log-likelihood.

In addition to the SN p-values, we provide bootstrapped p-values to test the null hypotheses of nondiscrimination, since tests based on bootstrapped p-values are generally believed to perform better than do those based on approximate asymptotic distributions. Considering the bootstrapped p-values, we find more consistency in test outcomes across links compared to those from examining the SN p-values. In particular, only one case (H_0^2 for Banks 3) exhibits inconsistency in test outcome across links. This result strengthens our observation that there is little practical difference in test outcomes across the various links considered for the five banks under study.

Moreover, we observe that the bootstrapped and SN p-values are quite similar and give consistent results for seven out of eleven cases. However, two of the cases lead to markedly different discrimination findings (H_0^2 for Bank 5 and H_0^3 for Bank 4); the bootstrapped p-values are usually much smaller than the SN p-values, which suggests a finite-sample null distribution for the t-ratio that is thinner tailed than the standard normal. Such a feature leads us to support the nondiscrimination null when using the SN p-values, for a given nominal level of significance, but to reject it (i.e., support discrimination) when using the bootstrapped p-values. This is evident even when using the logit link, as has been standard in the fair lending empirical literature.

[12] When comparing the SN p-values via these two methods, we do not automatically expect the profile SN p-values to be smaller than those from the one-step pseudo route, because, although the profile estimator has higher precision than the pseudo estimator, at least asymptotically, coefficient estimates also change, which may result in a smaller (in magnitude) test statistic.

Given that the aim is to determine whether banks are discriminating, our view is that it is better to err on the side of finding statistical support for discrimination at a given level of significance. Then regulators can look more closely at the cases where the statistical analysis suggests discrimination by using, for instance, the more-traditional comparative file reviews. We thus advocate the adoption of bootstrapping to generate p-values in statistical analysis for racial discrimination.

6. Summary and concluding remarks

Concerns regarding racial disparate treatment in mortgage lending have not abated over the years, despite legislation and efforts by regulators. Our contribution is to continue the examination of the statistical models adopted by regulators to answer the question "Is race a significant determinant of the likelihood of approval, after controlling for lender underwriting criteria?" Although statistical models do not form the sole tool used by regulators to ascertain bank specific discrimination, given the social, economic, political and legal ramifications of disparate treatment, it is important to understand any shortcomings of, and lack of robustness of outcomes from, the statistical models. The issue of link function has received little, if any, attention. Our study begins the exploration of this question by comparing the logit disparate treatment test outcomes with those from probit, gompit and cloglog links.

Our empirical evidence indicates that discrimination findings are practically robust to this choice. Assuming our use of representative banks, our results suggest that regulators, bank officials and others interested in testing for racial disparate treatment can be reasonably comfortable in estimating statistical models with logit links.

In addition, we observe qualitative disparate treatment test outcomes that are quite robust to use of the one-step pseudo log-likelihood estimator, a consistent, but asymptotically inefficient, coefficient estimator, or the profile log-likelihood estimator, which is maximum likelihood. This distinction is not relevant when using the logit link as then the two estimators are equivalent. However, for non-multiplicative links (e.g., probit) the two estimators vary, so that our finding has computational advantages for practitioners given that the one-step pseudo estimator is straightforward to code.

Although the discrimination test outcomes did not usually vary with whether we used standard normal or bootstrapped p-values, we still advocate that practitioners adopt resampling tools to form these p-values. This recommendation is based on our finding that sometimes the bootstrapped p-values can suggest evidence of discrimination when it is not detected via the standard normal p-values. Such a feature has important policy implications. As resampling p-values are generally more accurate than standard normal p-values, regulators, bank officials, consumers and court officials need to be aware that the latter may be significantly overstated.

Despite our use of consistent estimators of the parameter vector, finite-sample bias, known to be present, likely differs across the links and between the profile and pseudo methods. This might also possibly be contributing to some of our observed test outcomes. Benefits of adopting bias-reduction techniques, such as bootstrapping and jackknifing, would be worth exploring in future research. In addition, it would be of interest to undertake simulation experiments to ascertain the impact of link choice misspecification on the statistical properties of the discrimination hypothesis test and the pseudo and profile estimators.

Table 1: Distribution of Illustrative Predicted Probabilites of Loan Approval

Range for Predicted Probability									
0-<0.10	0.10-<0.20	0.20-<0.30	0.30-<0.40	0.40-<0.50	0.50-<0.60	0.60-<0.70	0.70-<0.80	0.80-<0.90	0.90-<1
Bank 1: N=7013, n=332									
5.1%	7.5%	6.3%	7.2%	11.7%	7.5%	8.1%	12.7%	19.9%	13.9%
Bank 2: N=2959, n=245									
5.7%	3.7%	2.4%	3.3%	2.0%	7.8%	6.1%	6.9%	12.2%	49.8%
Bank 3: N=939, n=340									
8.2%	4.1%	3.5%	2.1%	2.1%	3.2%	5.9%	5.0%	14.1%	51.8%
Bank 4: N=3550, n=420									
10.7%	3.8%	4.8%	4.0%	4.0%	5.5%	6.0%	6.7%	15.0%	39.8%
Bank 5: N=1976, n=228									
1.8%	1.8%	2.6%	2.2%	2.2%	11.0%	16.2%	29.8%	26.3%	3.1%

Notes: The probabilities are calculated from the logistic specifications adopted by the OCC for each bank.

Table 2: Explanatory Variables

	Bank				
Variable	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5
Credit score	×	×	×	×	×
LTV	×	×	×	×	×
Public record	×		×		×
Insufficient funds			×	×	
DTI	×	×	×	×	×
HDTI			×		
PMI			×		
Bad credit	×		×	×	×
Gifts/grants	×		×		
Relationship			×		
Income/savings				×	
Explanation	×		×		
Gender		×			
White	×	×	×	×	×
African American	×	×		×	×
Hispanic American	×		×	×	×

Table 3: Broad Variable Definitions

Variable	*Definition*
Credit Score	Derived from the bank's underwriting guidelines manual. Typically, a specified procedure is used to calculate a score variable, combining information across obtained credit bureau scores and the applicant and any co-applicant.
LTV	Loan-to-value ratio. May also be a dummy variable equal to 1 if the loan-to-value ratio exceeds specific guidelines; otherwise 0.
Public record	Public record information, created to be approximately uncorrelated with the bad credit variable.
Insufficient funds	Dummy variable equal to 1 if there were not sufficient funds to close.
DTI	Debt-to-income (gross) ratio. May also be a dummy variable equal to 1 if DTI value exceeds bank guidelines; otherwise 0
HDTI	House payment-to-income (gross) ratio
PMI	Dummy variable equal to 1 if the applicant applied for private mortgage insurance and was denied
Bad credit	Derived from bank specific information on credit records. Equal to 1 if a bad credit element is observed, or this variable may be number of derogatories or delinquencies depending upon the underwriting standards of the bank.
Gifts/grants	Sum of gifts and grants, which may provide down payment information.
Relationship	Dummy variable equal to 1 if the applicant has any type of relationship with the bank, such as deposits or previous loan at the bank.
Income/savings	Income and savings information
Explanation	Various dummy variables equal 1 if the bank asked for, received, or accepted explanations for credit bureau or other underwriting elements; 0 otherwise
Gender	Dummy variable equal to 1 if the applicant is Female; 0 otherwise
White	Dummy variable equal to 1 if the applicant is White; 0 otherwise
African American	Dummy variable equal to 1 if the applicant is African American; 0 otherwise
Hispanic American	Dummy variable equal to 1 if the applicant is Hispanic American; 0 otherwise

Table 4: Relative average log-likelihood values

Bank/ Method	Regression Model			
	logit	probit	gompit	cloglog
Bank 1				
profile	1	1.000	0.989	1.002
pseudo	1	1.003	0.999	1.014
Bank 2				
profile	1	0.999	1.000	0.999
pseudo	1	0.999	1.003	1.020
Bank 3				
profile	1	0.983	0.983	0.983
pseudo	1	1.008	1.007	1.025
Bank 4				
profile	1	1.002	0.989	1.004
pseudo	1	1.026	0.999	1.056
Bank 5				
profile	1	1.000	1.001	1.001
pseudo	1	1.005	1.004	1.007

Notes: The numbers provide average log-likelihood values relative to that for the logit link.

Table 5a: Percentage correctly predicted from pseudo log-likelihood route

Bank/ Cutoff Value	Loan Outcome								Overall			
	Denied (y=0)				*Approved (y=1)*							
	logit	**probit**	**gompit**	**cloglog**	**logit**	**probit**	**gompit**	**cloglog**	**logit**	**probit**	**gompit**	**cloglog**
Bank 1												
0.5	45.9%	45.9%	45.1%	42.1%	94.5%	95.0%	95.5%	95.5%	75.0%	75.3%	75.3%	74.1%
sf	57.9%	57.9%	57.1%	56.4%	89.9%	89.9%	90.5%	89.9%	77.1%	77.1%	77.1%	76.5%
spf	78.2%	78.9%	78.2%	80.5%	65.3%	64.3%	66.8%	61.8%	70.5%	70.2%	71.4%	69.3%
Bank 2												
0.5	41.7%	40.0%	40.0%	28.3%	96.2%	96.8%	96.8%	93.5%	82.9%	82.9%	82.9%	77.6%
sf	73.3%	73.3%	66.7%	61.7%	90.3%	87.6%	91.4%	81.1%	86.1%	84.1%	85.3%	76.3%
spf	86.7%	88.3%	86.7%	76.7%	77.8%	76.8%	78.4%	70.3%	80.0%	79.6%	80.4%	71.8%
Bank 3												
0.5	60.5%	58.1%	55.8%	58.1%	97.2%	97.2%	97.6%	97.2%	87.9%	87.4%	87.1%	87.4%
sf	76.7%	76.7%	72.1%	77.9%	90.6%	89.8%	92.5%	87.4%	87.1%	86.5%	87.4%	85.0%
spf	83.7%	83.7%	82.6%	83.7%	83.9%	82.3%	85.8%	79.1%	83.8%	82.6%	85.0%	80.3%
Bank 4												
0.5	42.1%	37.6%	44.4%	27.1%	100%	100%	100%	100%	81.7%	80.2%	82.4%	76.9%
sf	56.4%	54.1%	56.4%	45.9%	98.6%	99.3%	98.6%	100%	85.2%	85.0%	85.2%	82.9%
spf	82.0%	84.2%	80.5%	85.0%	80.1%	78.0%	80.1%	75.6%	80.7%	80.0%	80.2%	78.6%
Bank 5												
0.5	15.3%	9.7%	13.9%	6.9%	99.4%	99.4%	99.4%	99.4%	72.8%	71.1%	72.4%	70.2%
sf	29.2%	25.0%	23.6%	22.2%	96.8%	96.8%	96.8%	96.8%	75.4%	74.1%	73.2%	73.2%
spf	61.1%	61.1%	61.1%	65.3%	68.6%	67.9%	69.2%	64.1%	66.2%	65.8%	66.7%	64.5%

Table 5b: Percentage correctly predicted from profile log-likelihood route

Bank/ Cutoff Value	Loan Outcome								Overall			
	Denied (y=0)				*Approved (y=1)*							
	logit	**probit**	**gompit**	**cloglog**	**logit**	**probit**	**gompit**	**cloglog**	**logit**	**probit**	**gompit**	**cloglog**
Bank 1												
0.5	45.9%	46.6%	45.1%	41.4%	94.5%	95.0%	96.0%	95.5%	75.0%	75.6%	75.6%	73.8%
sf	57.9%	59.4%	57.1%	50.4%	89.9%	89.4%	91.5%	90.5%	77.1%	77.4%	77.7%	76.8%
spf	78.2%	78.9%	76.7%	80.5%	65.3%	63.8%	66.3%	60.8%	70.5%	69.9%	70.5%	68.7%
Bank 2												
0.5	41.7%	41.7%	30.0%	26.7%	96.2%	96.2%	92.4%	94.6%	82.9%	82.9%	77.1%	78.0%
sf	73.3%	73.3%	51.7%	66.7%	90.3%	89.2%	85.4%	82.7%	86.1%	85.3%	77.1%	78.8%
spf	86.7%	90.0%	78.3%	76.7%	77.8%	75.7%	70.8%	69.7%	80.0%	79.2%	72.7%	71.4%
Bank 3												
0.5	60.5%	58.1%	55.8%	58.1%	97.2%	97.2%	97.6%	97.2%	87.9%	87.4%	87.1%	87.4%
sf	76.7%	76.7%	72.1%	77.9%	90.6%	89.8%	92.5%	87.4%	87.1%	86.5%	87.4%	85.0%
spf	83.7%	83.7%	81.4%	83.7%	83.9%	82.7%	85.8%	92.5%	83.8%	82.9%	84.7%	90.3%
Bank 4												
0.5	42.1%	37.6%	37.6%	56.4%	100%	99.7%	99.7%	100%	81.7%	80.0%	80.0%	86.2%
sf	56.4%	52.6%	56.4%	46.6%	98.6%	98.3%	98.6%	99.7%	85.2%	83.8%	85.2%	82.9%
spf	82.0%	82.0%	79.7%	86.5%	80.1%	76.7%	80.5%	75.3%	80.7%	78.3%	80.2%	78.8%
Bank 5												
0.5	15.3%	9.7%	13.9%	6.9%	99.4%	99.4%	99.4%	99.4%	72.8%	71.1%	72.4%	70.2%
sf	29.2%	26.4%	23.6%	25.0%	96.8%	96.8%	97.4%	97.4%	75.4%	74.6%	74.1%	74.6%
spf	61.1%	62.5%	61.1%	63.9%	68.6%	67.3%	68.6%	63.5%	66.2%	65.8%	66.2%	63.6%

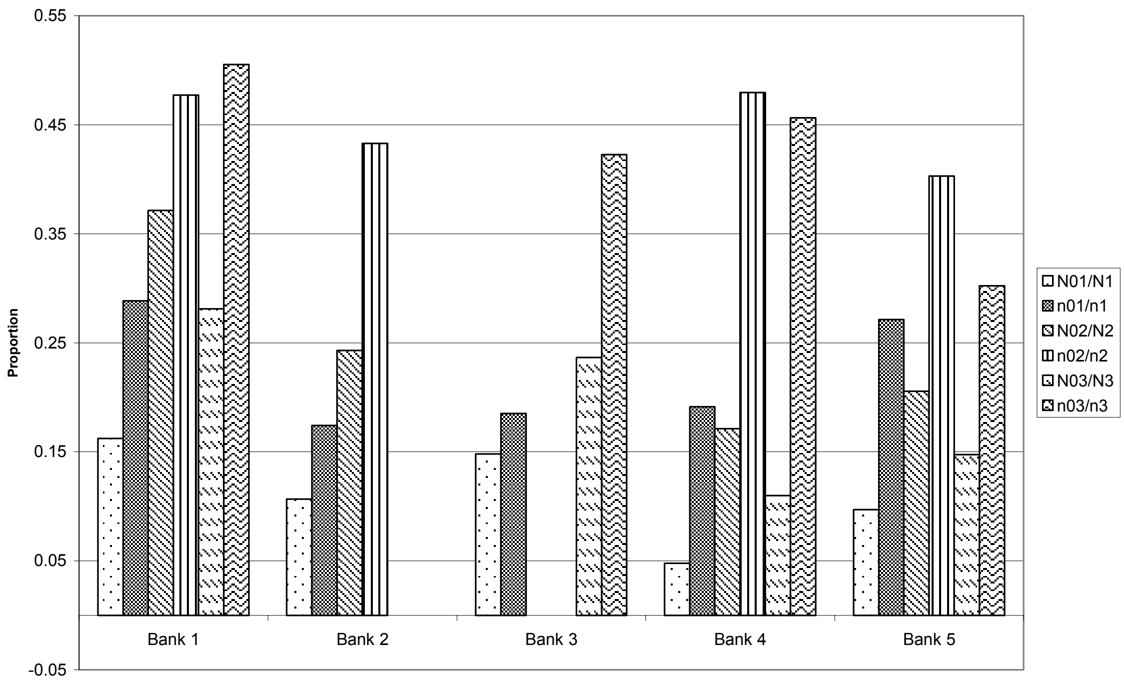
Table 6: P-values for testing for racial disparate treatment

Bank: p-value	**Regression Model** logit	probit	gompit	cloglog
	$H_0^1: \beta_1 - \beta_2 = 0$ vs. $H_a^1: \beta_1 - \beta_2 > 0$			
Bank 1: PR SN p-value	**0.000**	**0.000**	**0.008**	**0.000**
Bank 1: PS SN p-value	**0.000**	**0.040**	0.101	**0.032**
Bank 1: PS boot p-value	**0.000**	**0.000**	**0.000**	**0.000**
Bank 2: PR SN p-value	**0.000**	**0.000**	**0.000**	**0.000**
Bank 2: PS SN p-value	**0.000**	**0.004**	**0.001**	**0.022**
Bank 2: PS boot p-value	**0.000**	**0.000**	**0.000**	**0.000**
Bank 4: PR SN p-value	**0.036**	**0.031**	0.052	**0.006**
Bank 4: PS SN p-value	**0.036**	0.085	0.079	0.111
Bank 4: PS boot p-value	**0.010**	**0.010**	**0.000**	**0.000**
Bank 5: PR SN p-value	0.591	0.529	0.726	0.492
Bank 5: PS SN p-value	0.591	0.622	0.716	0.565
Bank 5: PS boot p-value	0.505	0.535	0.798	0.509
	$H_0^2: \beta_1 - \beta_3 = 0$ vs. $H_a^2: \beta_1 - \beta_3 > 0$ *			
Bank 1: PR SN p-value	**0.000**	**0.000**	**0.000**	**0.000**
Bank 1: PS SN p-value	**0.000**	**0.012**	**0.012**	**0.017**
Bank 1: PS boot p-value	**0.000**	**0.000**	**0.000**	**0.000**
Bank 3: PR SN p-value	**0.050**	**0.044**	0.136	0.689
Bank 3: PS SN p-value	**0.050**	0.248	0.156	0.287
Bank 3: PS boot p-value	**0.000**	0.122	**0.010**	0.145
Bank 4: PR SN p-value	0.411	0.349	0.455	0.223
Bank 4: PS SN p-value	0.411	0.397	0.424	0.361
Bank 4: PS boot p-value	0.616	0.283	0.419	0.343
Bank 5: PR SN p-value	0.285	0.238	0.285	0.246
Bank 5: PS SN p-value	0.285	0.214	0.229	0.364
Bank 5: PS boot p-value	**0.000**	**0.000**	**0.030**	**0.010**
	$H_0^3: \beta_2 - \beta_3 = 0$ vs. $H_a^3: \beta_2 - \beta_3 \neq 0$			
Bank 1: PR SN p-value	0.054	0.888	0.069	0.958
Bank 1: PS SN p-value	0.054	0.754	0.353	0.972
Bank 1: PS boot p-value	0.495	0.687	0.121	0.691
Bank 4: PR SN p-value	0.149	0.182	0.057	0.145
Bank 4: PS SN p-value	0.149	0.265	0.219	0.366
Bank 4: PS boot p-value	**0.000**	**0.020**	**0.000**	**0.030**
Bank 5: PR SN p-value	0.569	0.590	0.360	0.735
Bank 5: PS SN p-value	0.569	0.445	0.282	0.492
Bank 5: PS boot p-value	0.394	0.414	0.283	0.485

Notes: PS = pseudo log-likelihood; PR = profile log-likelihood; SN = standard normal; Boot = bootstrap

* The alternative hypothesis for Bank 3 is $H_a^2: \beta_1 - \beta_3 < 0$

Figure 1. Bank Denial Ratios



Notes: The subpopulation measures are denoted by "N", the sample measures by "n", approval (denial) of a loan application by "1" ("0"); e.g., "N01" is the number of denied whites loans, "n2" is the number of African Americans in the sample, and so on.

References

Blackburn, M.L. and T. Vermilyea (2004) Racial disparities in bank-specific mortgage lending models, *Economics Letters*, **85**, 379-83.

Booth, J.G., Butler, R.W. and P. Hall (1994) Bootstrap methods for finite populations, *Journal of the American Statistical Association*, **89**, 1282-89.

Calem, P.S. and S.D. Longhofer (2002) Anatomy of a fair lending exam: The uses and limitations of statistics, *Journal of Real Estate Finance and Economics*, **24**, 207-37.

Clarke, J.A. and M.J. Courchane (2005) Implications of stratified sampling for fair lending binary logit models, *Journal of Real Estate Finance and Economics*, **30**, 5-31.

Courchane, M., Nehbut, D. and D. Nickerson (2000) Lessons learned: Statistical techniques and fair lending, *Journal of Housing Research*, **11**, 277-95.

Cramer, J.S. (1999) Predictive performance of the binary logit model in unbalanced samples, *Journal of the Royal Statistical Society*, **C48**, 85-94.

Davidson, R. and J.G. MacKinnon (2000) Bootstrap tests: How many bootstraps?, *Econometric Reviews*, **19**, 55-68.

Dietrich, J. (2005a) The effects of sampling strategies on the small-sample properties of the logit estimator, *Journal of Applied Statistics*, **32**, 543-54.

Dietrich, J. (2005b) Under-specified models and detection of discrimination: A case study of mortgage lending, *Journal of Real Estate Finance and Economics*, **31**, 83-105.

Gross, S.T. (1980) Median estimation in sample surveys, *American Statistical Association Proceedings of the Survey Research Methods Section*, 181-84.

Hall, P. and S.R. Wilson (1991) Two guidelines for bootstrap hypothesis testing, *Biometrics*, **47**, 757-62.

Kaye, D.H. and M. Aicken (eds.) (1986) *Statistical Methods in Discrimination Litigation*, New York: Marcel Dekker.

Kennedy, P.E. (2002) Sinning in the basement: What are the rules? The ten commandments of applied econometrics, *Journal of Economic Surveys*, **16**, 569-89.

Kiefer, J. and J. Wolfowitz (1956) Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters, *Annals of Mathematical Statistics*, **27**, 887-906.

LaCour-Little, M. (1999) Discrimination in mortgage lending: A critical review of the literature, *Journal of Real Estate Literature*, **7**, 15-49.

Manski, C.F. and D. McFadden (1981) Alternative estimators and sample designs for discrete choice analysis, in *Structural Analysis of Discrete Data with Econometric Applications* (Eds.) C.F. Manski and D. McFadden, Cambridge, Mass.: MIT Press, pp. 2-50.

Scott, A.J. and C.J. Wild (1986) Fitting logistic models under case-control or choice based sampling, *Journal of Royal Statistical Society,* **B48**, 170-182.

Scott, A.J. and C.J. Wild (2001) Maximum likelihood for generalised case-control studies, *Journal of Statistical Planning and Inference*, **96**, 3-27.