Forthcoming *Applied Economics*

On the Robustness of Racial Discrimination Findings in Mortgage Lending Studies

Judith A Clarke*
Department of Economics, University of Victoria, PO Box 1700, Victoria, BC, CANADA V8X 5A3. E-mail: jaclarke@uvic.ca

Nilanjana Roy
Department of Economics, University of Victoria, PO Box 1700, Victoria, BC, CANADA V8X 5A3. E-mail: nroy@uvic.ca

Marsha J Courchane
ERSGroup, 2100 M Street NW, Suite 810, Washington, DC, U.S.A., 20037. E-mail: mcourchane@ersgroup.com

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 the 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. Our evidence, of discrimination findings that are reasonably robust to the range of examined link functions, suggests that regulators and researchers can be reasonably comfortable with their 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.

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

* Corresponding author.

I. 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, assist 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 give 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 aiming to approximate underwriting criteria, with the outcome variable being the probability of approval of a home mortgage loan. Although regulators do not rely solely on such models, it is important to appreciate how specification issues with the regressions affect discrimination findings[3].

---

[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.
[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.

We contribute towards this understanding by examining the sensitivity of the conclusions from five bank specific regulatory examinations of home-purchase mortgage lending; see Courchane *et al.* (2000) 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 test results 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) of applied Econometrics: **Thou shalt confess in the presence of sensitivity.**

Several methods have been adopted by researchers to examine for discrimination in mortgage lending. Ross and Yinger (2002) is an excellent source that presents an in-depth discussion of the mortgage lending discrimination literature and reanalysis of existing data to take into account the changes that are occurring in mortgage markets. A critical review is provided by LaCour-Little (1999).

As LaCour-Little (1999) points out, the empirical studies can be divided along various dimensions: focusing on neighborhood rather than borrower characteristics; the type of data used; and techniques adopted to test for discrimination. While single-equation response probability models are most prevalent, other approaches examine default rates, matched-pair audits, mortgage flows[4] and mortgage choice.

[4] This body of research focuses on neighborhood, rather than borrower, characteristics and, hence, will not be discussed herein. See, e.g., LaCour-Little (1999).

The premise behind studies using default rates is that lower loan default rates should result for minorities in the presence of racial discrimination[5]. As noted by Becker (1993, p. 389), the argument is: "discriminating banks would be willing to accept marginally profitable white applicants who would be turned down if they were black". In other words, discriminating banks use higher standards for minority loan applicants resulting in lower default rates. Empirical works include Van Order and Zorn (2001), Berkovec *et al.* (1996, 1994) and Ferguson and Peters (1995, 2000).

Matched-pair audits provide another way to examine for racial discrimination in housing markets; e.g., Yinger (1994, 1986) and Fix and Struyk (1993). The approach is to compare mortgage loan outcomes between minority and nonminority applicants with similar characteristics. Courchane and Nickerson (1997), for instance, compare the overages charged by a bank's loan officer to her borrowers from different races (black or white) but with otherwise similar characteristics using matched-pairs.

Mortgage choice studies are based on the idea that government-insured loans (e.g., from the US Federal Housing Administration) have nondiscriminatory underwriting criteria. This suggests that a sub-group facing discriminatory underwriting criteria for conventional loans should choose government-insured loans more often than expected with nondiscrimination. Empirical examples are Gabriel and Rosenthal (1991) and Shear and Yezer (1983, 1985).

The most frequent method of ascertaining discrimination in mortgage lending involves the use of single-equation response probability model, where the likelihood of approval (or denial) for a loan is allowed to potentially vary with racial class,

[5] This research should be contrasted with that which aims to determine the causes behind residential mortgage defaults; e.g., Feinberg and Nickerson (2002).

controlling for other characteristics of the borrower and the underwriting criteria. Should this probability vary (at least on average) with racial group, then the possibility of disparate treatment is deemed to exist. Examples include Clarke and Courchane (2005), Courchane *et al.* (2000), Stengel and Glennon (1999), Calem and Stutzer (1995) and Munnell *et al.* (1992, 1996). The Munnell *et al.* "Boston Fed" study, which provided support for discrimination against minorities, led to a number of follow-up studies, including Ross and Yinger (2002), Harrison (1998), Horne (1997), and Liebowitz and Day (1992, 1998). We also use such models.

An unordered binary logistic regression is most common. Some econometric issues associated with these models have been explored, including the implications of omitted variable bias (e.g., Dietrich, 2005b), modelling with multiple equations rather than a single equation (e.g., LaCour-Little, 2001; Maddala and Trost, 1982), the benefits of combining information from bank-specific regulatory models (e.g., Blackburn and Vermilyea, 2004) and the choice of sampling strategy to obtain sample data (e.g., Clarke and Courchane, 2005; Dietrich, 2005a). However, to the best of our knowledge, no information exists on the sensitivity of discrimination findings in the two ways we explore: the link choice[6] and the approximation used to determine statistical significance. That is, we take the OCC's covariates as given and compare discrimination outcomes from changing the choice of link. Like those before us, our study assists regulators, 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.

[6] Other researchers in different contexts have examined the issue of link choice; e.g., Jin *et al.* (2005) study the question of logit versus probit when modelling crop insurance fraud.

We examine three alternative links: 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 use samples stratified by 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 link choice, may be asymptotically inefficient, while the other estimator, the maximum likelihood estimator, has computational disadvantages. By adopting two estimation principles, we can ascertain whether the computationally simpler estimator results in substantively the same findings as the maximum likelihood estimator.

This paper is organized into the following sections. Section II presents our model setup, including a discussion of the link functions; Section III 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 IV details our data, including particulars on covariates; Section V provides the empirical results and Section VI concludes.

II. 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[7]. More generally, for each bank, we assume a binary outcome dependent variable, y_j, which takes values $y_j = 0$, when an application is denied, and $y_j = 1$, when it is approved; $j=1,\ldots,N$, the number of applicants. There are K race categories (e.g., White, African

[7] Covariates are provided in Table 3.

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:

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

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); $\beta=[\beta_1, \beta_2, \dots, \beta_K, \beta_{K+1}, \dots, \beta_p]'$. The regulator ascertains discrimination by testing whether the impacts of the race 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).

Equation (1) can be equivalently written 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 has the logit link function: $P_1(w_j;\beta) = \exp(w_j'\beta)/(1+\exp(w_j'\beta))$.

Another common link is the normit, leading to a 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 binary response models, likely to lead to little difference between logit and probit models. However, the percentage distribution of the predicted probabilities from logistic regressions for our banks, denoted as Bank 1 to Bank 5 for confidentiality reasons, shows a significant percentage of predictions 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 coefficient estimates and detrimentally affect the 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.

III. Estimation issues

To estimate expression (1) we need information on y_j and w_j 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, to ensure 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 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:

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

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

$$= \prod_{i=0}^{1}\prod_{k=1}^{K}\prod_{j=1}^{n_{i,k}} P_i(w_{ijk}; \beta \mid x_j \in k) g(w_{ijk} \mid x_j \in k) / pr(y_j = i \mid x_j \in k) \qquad (2)$$

using Bayes' Rule and the notation from (1)[8]. As $pr(y_j{=}i) = \int P_i(w_{ij};\beta) dG(w_{ij})$, where G(.) is the marginal distribution function, we cannot separate out $g(w_{ij})$ when estimating β.

Estimation of the log-likelihood function from (2)

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

$$- \sum_{i=0}^{1}\sum_{k=1}^{K}\sum_{j=1}^{n_{i,k}} \log pr(y_j = i \mid x_j \in k) \qquad (3)$$

[8] 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.

requires 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 $P_i(w_{ijk};\beta|\ x_j \in k)$ (e.g., using one of the links provided in the previous section) and we nonparametrically model $g(w_{ijk}|\, x_j \in k)$; 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 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.

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[9], is:

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

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

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

$$(m_{0k} + m_{1k})\log(1+\exp(\rho_{1,k})))/n)\}\qquad (4)$$

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, we have (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 and relate to unconditional probabilities. Specifically, let $Q_{i,k}$ be the 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 (4) 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 used to maximize (4) for estimates of β, say β^*. With β fixed at β^*, we again maximize (4) to obtain new ρ estimates and so on until we converge to $\hat{\beta}_{PR}$ and $\hat{\rho}_{PR}$. 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.

A pseudo log-likelihood route

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

$$\ell^* = \sum_{i=0}^{1} \sum_{k=1}^{K} \sum_{j=1}^{n_{i,k}} \log P_i^*(w_{ijk};\beta,\kappa_k) \tag{5}$$

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

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

$$\kappa_k = \left(\frac{n_{1,k}}{\operatorname{pr}(y_j = 1 \mid x_j \in k)}\right) \Big/ \left(\frac{n_{0,k}}{\operatorname{pr}(y_j = 0 \mid x_j \in k)}\right) \tag{6}$$

and

$$\operatorname{logit}\ P_1(w_{1jk};\beta \mid x_j \in k) = \log\left(\frac{P_1(w_{1jk};\beta \mid x_j \in k)}{P_0(w_{0jk};\beta \mid x_j \in k)}\right). \tag{7}$$

The objective function (5) 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 multicollinarity issues that cause convergence concerns. In addition, the stationary point of (5) 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:

$$\hat{\kappa}_k = \left(\frac{n_{1,k}}{N_{1,k}}\right) / \left(\frac{n_{0,k}}{N_{0,k}}\right) \tag{8}$$

is a consistent estimator of κ_k. Use of this rule for the logit link leads to the estimator of β used by Clarke and Courchane (2005) in their fair lending study; this estimator is known to be in fact the maximum likelihood estimator of β[10]. 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 as $\hat{\beta}_{PS}$[11] - 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.

---

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

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 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:

$$\left\{ a_k = \left[\left(\frac{1}{n_{0,k}} + \frac{1}{n_{1,k}} \right) - \left(\frac{1}{N_{0,k}} + \frac{1}{N_{1,k}} \right) \right] \right\}; \text{ k=1,2, …, K.} \qquad (9)$$

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 (5) with $\hat{\kappa}_k$ as the estimator of κ_k. A consistent estimator of var($\hat{\beta}_{PS}$) is 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^{m}_{PS} is standard normal (SN).

As we use the analytic score vector and Hessian matrix to obtain the maximum likelihood estimator, $\hat{\beta}$, via 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).

Bootstrapped p-values

Bootstrapping provides an alternative route to using an asymptotic S.N. distribution to approximate the null distribution. 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 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 Hall and Wilson's (1991) advice by centering when forming these bootstrapped statistics, which has the effect of increasing power; i.e., 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[12].

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 Davidson and MacKinnon's (2000) pretesting method to choose B, the number of bootstraps; typically, this gave B=99 for our 5% significance level.

[12] 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.

IV. Data

Our data, collected by the OCC in the course of several fair lending examinations in the late 1990s, comes from five separate national banks geographically distributed from the East to the West and the Midwest. Each statistical model, structured to 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 is given in Table 2 while Table 3 provides 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.

To provide an indication of the characteristics of covariates, Table 4 presents sample means (adjusted for the stratification) of the regressors for Banks 1 and 3 separately by race; details for the other banks are available on request from the first author. To ensure confidentiality, each sample statistic is presented as an index, with the base of 100 being the sample mean for the full sample of applicants for that bank; e.g., the average LTV ratio for Bank 3's Whites is 1.7% lower than for the full sample, whereas that for Hispanic Americans is 6.1% higher. The following points are evident. Whites have a higher credit score, on average, than either minority race for both banks, with Hispanic Americans out scoring African Americans for Bank 1. Similarly, White applications have cleaner credit, on average, than either African or Hispanic Americans, and Hispanics have better credit, on average, than African Americans. A higher average LTV ratio results for Whites from Bank 1 than for either minority group, whereas the average LTV ratio for Whites from Bank 3 is lower than for Hispanic Americans from Bank 3 with the latter group of Whites having a substantially higher proportion of loans with a LTV ratio less than 75%.

Using samples stratified by race and loan outcome leads to 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: Whites - k=1; African Americans - k=2; Hispanic Americans - k=3. There are three racial strata (K=3) for Banks 1,4 and 5, 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.

V. Results

We estimated the five bank-specific models 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 only for the logit link. To obtain maximum likelihood estimates from the profile log-likelihood we used Gauss, with the MAXLIK sub-routine, whereas EViews, Stata and Gauss – to satisfy ourselves that results were similar across standard packages – were used find one-step pseudo log-likelihood estimates.

Given our objective of examining the sensitivity of discrimination outcomes to link choice and adopted approximation for determining statistical significance, our discussion here focuses on testing the hypothesis of discrimination (or nondiscrimination). However, for readers interested in estimation details and marginal effects, an appendix (Appendix A) provides some illustrative results.

Another appendix (Appendix B) addresses a specification concern that might be present with our use of a stratified sample: strata heteroskedasticity. Signs of misspecification are detected for the OCC's model for Bank 3. We do not pursue any adjustment for this bank for two reasons. First, hypothesis tests for strata heteroskedasticity cannot distinguish between this concern, coefficients that vary with strata, heterogeneity arising from unobserved variables that change with strata or some combination of one or more of these factors. An exploration of the cause is beyond our scope but the results do suggest that regulators routinely check for such specification issues. Second, our goal is to understand the sensitivity of the OCC's statistical findings, for the models they specify, to their assumed logit link; we would be unable to achieve this objective if we changed the model for Bank 3 to accommodate the signs of strata heterogeneity. Accordingly, we proceed as is, but note that care is needed with interpreting further test outcomes for Bank 3.

Prior to comparing outcomes from the disparate treatment hypothesis tests, we detail two measures of fit to provide some guidance on link preference. One gauge of fit is the value of the average log-likelihood function, reported in Table 5, with quantities given relative to the logit's average log-likelihood value; e.g., a number less than one indicates that the logit link has a smaller average log-likelihood value. We observe similar fit across links, with average log-likelihood values being different by at most 6%. This small difference could be arising from finite sample bias.

As the logit link's average profile log-likelihood and pseudo log-likelihood values are identical, the numbers in Table 5 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 our banks, 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 performance measure 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 6a and 6b, 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 6a presents the outcomes from the pseudo log-likelihood approach, while those from the profile log-likelihood route are given in Table 6b. 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).

The profile and pseudo log-likelihood percentages are similar. 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, there is little practical gain in choosing one link over another. 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 7, we report p-values for t-ratios for the nulls using the standard normal (SN) distribution for both the profile and the one-step pseudo log-likelihood approaches. We also present bootstrap p-values 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[13]. Such a choice effectively gives the benefit of doubt to the bank, as we support nondiscrimination unless the 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 broad similarities 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 (H_0^1 for Banks 1 and 4, and H_0^2 for Bank 3) give rise to inconsistent results across estimators. Comparing across links, dissimilar findings again only arise for these 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 similar test outcomes with the SN p-values, for the probit, gompit and cloglog models[14], from the pseudo and profile routes[15]. This is a useful

[13] See, e.g., Kaye and Aicken (1986). LaCour-Little (1999) provides a useful commentary.
[14] 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.
[15] 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

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 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. Specifically, only one out of eleven cases (H_0^2 for Bank 3) results in a discrimination finding that varies with link choice. If we ignore Bank 3, our bootstrapped examples suggest that the choice of link function does not matter for the 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 suggesting 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 with the logit link, the standard choice in fair lending work. Given a goal of ascertaining discriminating banks, we view it preferable to err on the side of finding statistical support for discrimination at a given level of significance. Regulators can then look more closely at cases where the statistical analysis suggests discrimination using, for

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.

instance, more-traditional comparative file reviews. We thus advocate the adoption of bootstrapping to generate p-values in statistical analysis for racial discrimination.

VI. Summary and concluding remarks

Concerns regarding racial disparate treatment in mortgage lending have not abated, 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 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 using two consistent estimators.

We observe qualitative disparate treatment test results 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 with logit as the two estimators are equivalent for this choice. However, for non-multiplicative links (e.g., probit) the two estimators vary, so 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.

Our empirical evidence indicates that discrimination findings are robust to the choice of link function for the majority of cases, irrespective of the approximation used to determine statistical significance. Specifically, the bootstrapped p-values lead to only one inconsistent discrimination conclusion out of the eleven cases considered, with this exception arising with the OCC's model for Bank 3, which we believe is likely misspecified. In other words, except for this one case, the disparate treatment findings, from the OCC's statistical models, are not sensitive to which link is used in estimation. Given that researchers and regulators testing for discrimination in mortgage lending have almost exclusively used logistic specifications, perhaps because of its computational simplicity even under complex sampling designs such as stratification or ease of accounting for individual lender effects (Chamberlain, 1980), we do not have any evidence to suggest that they move away from this practice.

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. 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%

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

Table 2. Explanatory variables

Variable	Bank				
	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. Sample means* for Bank 1 and Bank 3 regressors

Variable	Bank 1 Whites	Bank 1 Af. Amer.	Bank 1 Hisp. Amer.	Bank 3 Whites	Bank 3 Hisp. Amer.
Credit score	100.5	93.4	98.4	100.9	96.8
LTV	100.5	97.8	95.9	98.3	106.1
LTV dummy**				158.5	76.7
Public record	100.6	101.9	89.0	91.3	130.4
Insufficient funds				86.4	149.4
DTI	97.8	105.9	119.3	101.2	95.6
HDTI				96.9	110.8
PMI				84.0	160.0
Bad credit	92.8	226.1	106.1	93.5	123.4
Gifts/grants	90.8	118.8	190.9	116.8	39.0
Relationship				105.2	81.4
Explanation	92.8	164.0	137.7	93.5	124.3
Whites***	0.024			0.330	
Af. Amer.***		0.251			
Hisp. Amer.***			0.173		0.478
# sample obs.	149	88	95	243	97
# population obs.	6115	350	548	736	203

* The means, adjusted for the stratified sampling, are reported as indices relative to the full sample means.

** Equals 1 if the applicant has a LTV ratio ≤ 75%, 0 otherwise.

*** Sampling ratio

Table 5. 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

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

Table 6a. 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 6b. 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%

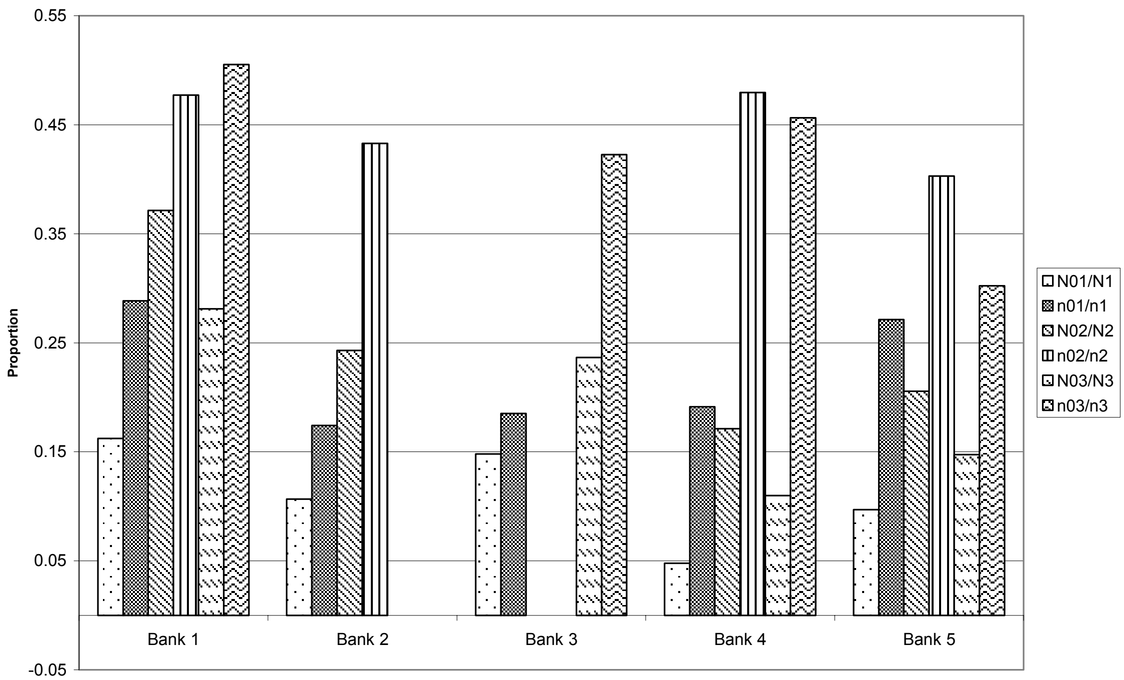
Table 7. P-values for testing for racial disparate treatment

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

Note: 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



Note: 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.

Acknowledgments

We gratefully acknowledge the OCC for permitting use of the data, collected while the third author was with the OCC. We also thank the referee and editor for comments and suggestions on an earlier version of this paper.

References

Allison, P.D. (1999) Comparing logit and probit coefficients across groups, *Sociological Methods & Research,* **28**, 186-208.

Amemiya, T. (1985) *Advanced Econometrics*, Harvard University Press, Cambridge, MA.

Becker, G.S. (1993) Noble Lecture, *Journal of Political Economy*, **101**, 386-89.

Berkovec, J.A., Canner, G.B., Gabriel, S.A. and Hannan, T.H. (1994), Race, redlining, and residential mortgage loan performance, *Journal of Real Estate Finance and Economics*, **9**, 263-94.

Berkovec, J.A., Canner, G.B., Gabriel, S.A. and Hannan, T.H. (1996) Response to critiques of "Mortgage discrimination and FHA loan performance", *Cityscape: A Journal of Policy Research*, **21**, 49-54.

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

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

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

Calem, P.S. and Stutzer, M. (1995) The simple analytics of observed discrimination in credit markets, *Journal of Financial Intermediation*, **4**, 189-212.

Chamberlain, G. (1980) Analysis of covariance with qualitative data, *Review of Economic Studies*, **47**, 225-38.

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

Courchane, M., and Nickerson, D. (1997) Discrimination resulting from overage practices, *Journal of Financial Services Research*, **11**, 133-52.

Courchane, M., Nehbut, D. and Nickerson, D. (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 MacKinnon, J.G. (1984) Convenient specification tests for logit and probit models, *Journal of Econometrics,* **25**, 241-262.

Davidson, R. and MacKinnon, J.G. (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.

Ferguson, M.F. and Peters, S.R. (1995) What constitutes evidence of discrimination in lending? *Journal of Finance*, **50**, 739-48.

Ferguson, M.F. and Peters, S.R. (2000) Is lending discrimination always costly? *Journal of Real Estate Finance and Economics*, **45**, 474-89.

Feinberg, R.M. and Nickerson, D. (2002) Crime and residential mortgage default: An empirical analysis, *Applied Economics Letters*, **9**, 217-20.

Fix, M. and Struyk, R. (1993) *Clear and Convincing Evidence*, Urban Institute, Washington, DC.

Gabriel, S.A. and Rosenthal, S. (1991) Credit rationing, race, and the mortgage market, *Journal of Urban Economics*, **29**, 371-79.

Greene, W.H. (2003) *Econometric Analysis,* 5th edn., Prentice-Hall, NJ.

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

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

Harrison, G.W. (1998) Mortgage lending in Boston: A reconsideration of the evidence, *Economic Inquiry*, **36**, 29-38.

Horne, D.K. (1997) Mortgage lending, race, and model specification, *Journal of Financial Services Research*, **11**, 43-68.

Jin, Y., Rejesus, R.M. and Little, B.B. (2005) Binary choice models for rare events data: A crop insurance fraud application, *Applied Economics*, **37,** 841-48.

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

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 Wolfowitz, J. (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.

LaCour-Little, M. (2001) A note on identification of discrimination in mortgage lending, *Real Estate Economics*, **29**, 329-35.

Liebowitz, S. and Day, T. (1992) Mortgages, minorities, and discrimination, Working Paper, University of Texas, Dallas.

Liebowitz, S. and Day, T. (1998) Mortgage lending to minorities: Where's the bias? *Economic Inquiry*, **36**, 3-28.

Maddala, G.S. and Trost, R. P. (1982) On measuring discrimination in loan markets, *Housing Finance Review*, **1**, 245-68.

Manski, C.F. and McFadden, D. (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, MIT Press, Cambridge, MA., pp. 2-50.

Munnell, A.H., Browne, L.E., McEneaney, J. and Tootell, G.M.B. (1992) Mortgage lending in Boston: Interpreting HMDA data, Federal Reserve Bank of Boston, Working Paper, 92-7.

Munnell, A.H., Tootell, G.M.B., Browne, L.E., and McEneaney J. (1996) Mortgage lending in Boston: Interpreting HMDA data, *American Economic Review*, **86**, 25-53.

Ross, S. and Yinger, J. (2002), *The Color of Credit*, MIT Press, Cambridge, MA.

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

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

Shear, W.B. and Yezer, A. (1983) An indirect test of discrimination in housing finance, *Journal of the American Real Estate and Urban Economics Association*, **10**, 405-20.

Shear, W.B. and Yezer, A. (1985) Discrimination in urban housing finance: An empirical study across cities, *Land Economics*, **61**, 293-302.

Stengel, M. and Glennon, D. (1999) Evaluating statistical models of mortgage lending discrimination: A bank-specific analysis, *Real Estate Economics*, **27**, 299-334.

Van Order, R. and Zorn, P. (2001) Performance of low-income and minority mortgages: A tale of two options, Unpublished manuscript, Freddie Mac, Washington, DC.

Verlinda, J.A. (2006) A comparison of two common approaches for estimating marginal effects in binary choice models, *Applied Economics Letters,* **13**, 77-80.

Wooldridge, J.M. (2002) *Econometric Analysis of Cross Section and Panel Data,* MIT Press, Cambridge, MA.

Yinger, J. (1986) Measuring racial discrimination with fair housing audits: Caught in the act, *American Economic Review*, **76**, 881-93.

Yinger, J. (1994) *Ethnicity: Source of Conflict? Source of Strength?*, State University of New York Press, Albany.

Appendix A: Illustrative Estimation Results

Although it is not feasible to report our thirty-five regressions[16], in this appendix we detail illustrative estimation results. Table A1 provides coefficient estimates and (asymptotic) standard errors for two representative banks and links: Banks 1 and 3 with the logit and gompit models estimated via the one-step pseudo log-likelihood approach.

For both banks and both links, the dummy variables representing race are statistically significant at the nominal 5% level of significance or better. Note that we cannot compare coefficient estimates across links (e.g., Greene, 2003, p. 675). All coefficient estimates, except for the variable “LTV dummy”, are statistically significant under both link choices for Bank 1. In contrast, a number of explanatory variables are not statistically significant with the OCC specification for Bank 3. Given our goal of ascertaining sensitivity of the discrimination test outcomes to link choices, given the specification adopted by the OCC, we work with their models irrespective of statistical significance of individual explanatory variables.

Estimated marginal effects and associated asymptotic standard errors, for these representative banks, are given in Tables A2 and A3. Since the marginal effects depend on the values of w_j, which vary among the individuals, we fix the explanatory variables at their stratified sample means for each racial group[17], which provides an indication of the range of marginal effects. We calculate marginal effects for continuous variables using derivatives and for dummy variables as discrete changes in the estimated probabilities. The reported asymptotic standard errors for these marginal effects are calculated using the linear approximation method; e.g., Greene (2003, pp. 674-675). One may reasonably

[16] All regression results are available on request from the first author.
[17] As opposed to taking a sample average of the marginal effects calculated for each loan case; see, e.g., Verlinda (2006) for a thoughtful discussion on these two approaches of reporting marginal effects.

argue that testing for discrimination by comparing the difference between the marginal effects of race dummies is preferable to testing for equality of race coefficients. While supportive of this view, as our goal is to replicate the OCC, apart from link choice, we follow them by testing for discrimination using the coefficients.

Table A1. Logit and gompit equations for Banks 1 and 3 using one-step pseudo log-likelihood.

	Bank 1				Bank 3			
Variable	logit		gompit		logit		gompit	
	coeff.	asy. se	coeff.	asy. se	coeff.	asy. se	coeff.	asy. se
Credit score	0.0080	0.003*	0.0080	0.002*	0.020	0.005*	0.017	0.004*
LTV	-0.0020	0.009	-0.0005	0.008	-0.024	0.022	-0.024	0.019
LTV dummy+					-1.042	0.851	-0.787	0.695
Public record	-1.3336	0.414*	-1.0172	0.316*	-0.410	0.595	-0.233	0.471
Insufficient funds					-2.143	0.443*	-1.790	0.353*
DTI	-1.5074	0.402*	-1.2716	0.327*	-0.248	0.427	-0.252	0.363
HDTI					-0.308	0.430	-0.234	0.369
PMI					-3.752	1.051*	-1.963	0.563*
Bad credit	-1.5949	0.365*	-1.3099	0.286*	-0.834	0.577**	-0.445	0.428
Gifts/grants++	0.0482	0.032***	3.8930	2.779***	0.005	0.022	0.007	0.018
Relationship					-0.202	0.395	-0.158	0.331
Explanation	0.7696	0.363**	0.5933	0.296**	0.434	0.625	0.266	0.488
White	-3.7777	2.199**	-3.1651	1.648**	-9.087	4.022**	-6.808	3.422**
African American	-4.4639	2.168**	-3.6275	1.619**				
Hispanic American	-4.5936	2.205**	-3.8536	1.654*	-8.756	4.000**	-6.460	3.364**

Note: +Equals 1 if the application has a LTV ratio ≤ 75%, 0 otherwise; ++$'000;
*Significant at the nominal 1% level; **Significant at the nominal 10% level;
***Significant at the nominal 5% level.

Appendix B: A Specification Issue

Aside from exploring the robustness of fair lending determinations to the choice of link, we work with the OCC's specifications of the response probability models adopted in their individual bank studies. Although it is beyond the scope of this paper to undertake a detailed study of relaxing other dimensions of their models, an obvious question arising from their use of stratified sampling when we write the response probability model in its common latent variable formulation is: Is there homoskedasticity of the error terms across the strata?[18] As noted by Wooldridge (2002, p.479), heteroskedasticity in such a framework is equivalent to altering the probability model's functional form, which in the case of strata or group heteroskedasticity leads to separate stratum response probability models.

Specifically, following Davidson and MacKinnon (1984), Allison (1999) and Hole (2006), amongst others, we assume an underlying latent variable y_j^* that gives rise to the observed binary variable y_j, with y_j^* generated by the p-regressor linear model $y_j^* = w_j'\alpha + \sigma\varepsilon_j$, where the random disturbance ε_j is assumed to be independent of the w variables and has mean zero and constant variance dependent on assumptions (e.g., $\pi^2/3$ with the logit link, one with the probit model). The parameter σ is a convenient way to allow for adjustments to the variance. Then, e.g., Amemiya (1985, p.269), the binary response model $h(P_1(w_j;\beta)) = w_j'\beta$ with $\beta_i = \alpha_i/\sigma$ (i=1,…,p) arises. This formulation highlights that we are unable to separately identify σ and the elements in α, a feature that complicates an analysis for strata heteroskedasticity.

[18] We thank the referee for raising this matter.

Given this framework, the heteroskedastic model is $y_j^* = w_j'\alpha + \sigma_j\varepsilon_j$ with

$\sigma_j = 1/\sum_{k=1}^{K}\delta_k x_{jk}$, x_{jk} has a value of 1 if individual j is in strata k and 0 otherwise. The corresponding response probability model is $h(P_1(w_j;\beta_j)) = w_j'\beta_j$ with $\beta_{ij} = \alpha_i/\sigma_j$ (i=1,…,p) or, equivalently,

$$h(P_1(w_j;\beta_j)) = \sum_{k=1}^{K}\delta_k x_{jk}(\alpha_k + \alpha_{K+1}z_{j1} + \ldots + \alpha_p z_{jq})$$

$$= \sum_{k=1}^{K}(\beta_k x_{jk} + \beta_{k,K+1}x_{jk}z_{j1} + \ldots + \beta_{k,p}x_{jk}z_{jq}) \tag{B1}$$

where $\beta_k = \delta_k\alpha_k$ and $\beta_{k,i} = \delta_k\alpha_i$, i=K+1,…,q. Thus, this heteroskedastic response probability model corresponds to a pooled model with a dummy variable for strata membership and strata membership interaction terms for the other variables proposed to explain loan determination. A likelihood ratio test that compares this representation with the model that constrains all coefficients (except for intercepts) to be common across strata, provides a straightforward way to examine for strata heteroskedasticity.

A major complication, as pointed out by Wooldridge (2002, p479) and Allison (1999) among others, is that the response probability model (b1) also arises when the underlying α coefficients differ across strata with σ constant or when both σ and the α's differ across strata. Accordingly, the hypothesis test cannot distinguish between differences in strata variances and differences in strata coefficients; rejection of the null hypothesis must be regarded as a sign of one or more possible specification errors, rather than as a direct indication of strata heteroskedasticity.

We provide the likelihood ratio statistic sample values for testing for strata homoskedasticity, along with the corresponding chi-square p-values, in Table B1 when the response probability models are estimated using one-step pseudo maximum likelihood. Results show that test outcomes are robust across the links with no evidence of the explored strata effects, except for Bank 3 where the calculated chi-squared statistics are significant at the nominal 5% significance level. Rejection could be arising from strata heteroskedasticity, coefficients that vary across strata, heterogeneity arising from unobserved omitted factors or some combination of all of these effects. As this study is not about such issues, despite their importance, but instead the goal is to explore the sensitivity of the OCC discrimination findings to the assumed link function, we do not pursue this matter further. We proceed by noting that care is needed when interpreting conclusions from the given OCC model specification for this bank.

Table B1. Likelihood ratio tests for strata homoskedasticity

Case	*Bank 1* *m=14*	*Bank 2* *m=4*	*Bank 3* *m=12*	*Bank 4* *m=12*	Bank 5 m=8
logit – LR	3.613	2.400	22.546	16.568	2.163
(χ^2(m) p-value)	(0.997)	(0.663)	(0.032)	(0.167)	(0.976)
probit – LR	2.835	2.574	25.344	16.562	2.354
(χ^2(m) p-value)	(0.999)	(0.631)	(0.013)	(0.167)	(0.968)
gompit – LR	4.990	2.066	23.584	19.579	3.428
(χ^2(m) p-value)	(0.986)	(0.724)	(0.023)	(0.075)	(0.905)
cloglog – LR	3.343	4.457	24.514	16.498	3.380
(χ^2(m) p-value)	(0.998)	(0.348)	(0.017)	(0.169)	(0.908)

Note: The degrees of freedom of the LR test is denoted by m.