

Mortgage Lending Discrimination and Racial Differences in Loan Default

Stephen L. Ross*

Abstract

This article analyzes the use of racial differences in loan default to test for mortgage lending discrimination—an approach whose underlying theory has not been carefully explored. The loan approval process creates a selected sample of loans. If mortgage lenders hold minorities to a higher standard, the minority sample should have greater selection bias than the majority sample and so should exhibit lower default rates.

This article examines the default approach in this framework and observes that the approach suffers from heteroskedasticity bias caused by loan defaults as a discrete dependent variable. In addition, the correlations between race and the unobservable variables in the loan approval and default equations bias the default approach away from a finding of discrimination. The article concludes that tests based on the loan approval approach should be accepted as the principal technique for determining whether mortgage lending discrimination exists.

Keywords: mortgage; discrimination; default; selection

Introduction

Theoretical analyses of discrimination (Arrow 1973; Becker 1971; Peterson 1981) imply that if lenders hold minority loan candidates to a higher standard, discrimination against minorities in the loan approval process may lower default rates on loans to minorities or make such loans more profitable. But empirical evidence suggests that minorities experience higher loan default rates. As a result, many economists (Becker 1993; Brimelow 1993; Brimelow and Spencer 1993; Roberts 1993) reject studies that find evidence of mortgage lending discrimination based on such alternative methods as racial differences in loan approval (Munnell et al. 1992).

Berkovec et al. (1994) formalize the default approach to testing for discrimination by specifying a stochastic model of lender behavior and borrower default. In this manner, they avoid the problems inherent in examining average rates of default and prove that holding minority borrowers to a higher standard unambiguously lowers the propensity of minorities to default on loans. They estimate a default equation, find that minority borrowers have a higher propensity to default, and conclude that minority default rates are not consistent with the theoretical predictions that arise from models of mortgage lending discrimination.

The default approach faces substantial criticism from Yinger (1993) and Galster (1993), both of whom emphasize that many borrower characteristics are unobservable to the

* Stephen L. Ross is an Assistant Professor in the Department of Economics, University of Connecticut. The author would like to thank Chris Mayer, Steve Miller, and James Berkovec for their helpful suggestions.

lender. They argue that if the unobservable characteristics explain loan defaults and are correlated with race, the default approach cannot rule out the existence of discrimination. Even in a world with discrimination, the race coefficient in the default equation could be nonnegative. Berkovec et al. (1994) acknowledge this point by stating that the default approach can detect only discrimination that arises from prejudice. It cannot detect statistical discrimination, in which lenders use race as a signal for unobservable characteristics that explain default probability.

This article examines both the default approach and the key criticisms of it in the context of a stochastic model of the loan approval and loan default decisions.

Loan Approval and Default Analyses

Loan default ($D = 1$) or continued payment ($D = 0$) by a borrower is represented by

$$D = \begin{cases} 1 & \text{for } D^* > 0, \\ 0 & \text{for } D^* < 0, \end{cases} \quad (1)$$

where $D^* = b_d \mathbf{X} + e_l + e_u$; \mathbf{X} is a vector of exogenous loan, borrower, and housing unit characteristics that explain loan default and profitability;¹ b_d is a vector of behavioral parameters; e_l is an error term composed of unmeasured characteristics observed by the lender at the time of loan approval; e_u is an error term composed of unobservable characteristics; and the variance of $e_l + e_u$ is initialized to 1 for the sample of loan applicants because default is a discrete choice specification.

Loan approval ($A = 1$) or denial ($A = 0$) by a lender is represented by

$$A = \begin{cases} 1 & \text{for } A^* > 0, \\ 0 & \text{for } A^* < 0, \end{cases} \quad (2)$$

where $A^* = \beta_a \mathbf{X} - c_r R + \gamma e_l + \epsilon_o$; β_a is a vector of behavioral parameters; R is a dummy variable that is 1 if the applicant is a minority and 0 otherwise; c_r is a profit premium that must be met before approval of a loan for minorities; γ is a constant that explains the role of default risk in loan profitability; ϵ_o is an error term composed of unmeasured characteristics that influence lender decisions but not loan default;² and the variance of $\gamma e_l + \epsilon_o$ is initialized to 1.³

¹ \mathbf{X} is used in both the default and loan specifications. This does not, however, imply that the approval and default equations contain the same variables. Any differences between equations can be modeled by setting coefficients equal to zero.

² These characteristics must be observed by the lender. Otherwise, they could not influence loan approval.

³ The linear specification of loan approval actually encompasses a more general specification of loan approval. The net profitability of a loan is defined as the expected profitability of the loan minus the product of the probability of default and the expected loss from loan default. If the expected profitability and the probability of default are linear functions of \mathbf{X} , and the expected loss from loan default is a constant (γ), the specification in equation (2) holds. Note that this specification implies that any variable in the default equation also belongs in the approval equation. In addition, with minor modifications, this specification allows expected loss from loan default to be a function of \mathbf{X} . The vector \mathbf{X} must be expanded to include the square and interaction terms for all variables that influence loan default, and the coefficients of these terms are set to zero in the default equation. In addition, γ is now a function of \mathbf{X} and varies by individual. All results hold with these modifications.

The lender loan approval decision takes place before the borrower loan default decision, and only approved loans enter the pool of possible defaults, thus creating a sample selection bias in default estimates (Rachlis and Yezer 1993). A loan that appears to the researcher to be a poor risk might have been approved only because of favorable borrower characteristics that remain unmeasured but were observable to the lender. This bias provides the basis for the default approach. If minorities are held to a higher standard for loan approval, the sample selection bias is greater for minority loans such that the minority propensity to default should fall further.

Specifically, the distribution of $e_l + e_u$ is incidentally truncated by the loan approval process. It is assumed that e_l , e_u , and e_o are uncorrelated with race, \mathbf{X} , and each other and that $(e_l + e_u, \sigma_l e_l + e_o)$ is distributed as a standard bivariate normal with a correlation of $\gamma\sigma_l^2$, where σ_l^2 is the variance of e_l .⁴

The resulting distribution of $e_l + e_u$ for the sample of approved loans is normal, with mean and variance of

$$\tilde{\mu}(R, \mathbf{X}) \equiv E[\epsilon_l + \epsilon_u \mid A = 1, R, \mathbf{X}] = \gamma\sigma_l^2\lambda(-\beta_a\mathbf{X} + c_rR), \quad (3)$$

and

$$\tilde{\sigma}(R, \mathbf{X}) \equiv \text{var}[\epsilon_l + \epsilon_u \mid A = 1, R, \mathbf{X}] = 1 - (\gamma\sigma_l^2)^2\delta(-\beta_a\mathbf{X} + c_rR), \quad (4)$$

where $\delta(\alpha) = \lambda(\alpha)(\lambda(\alpha) - \alpha)$ and λ is the inverse Mills ratio (Greene 1990).⁵

If minorities are held to a higher approval standard ($c_r > 0$), they should exhibit, on average, lower propensities to default in the selected sample, as evident from equation (3). The inverse Mills ratio is a positive, monotonically increasing function. In addition, σ is assumed to be negative because loan defaults decrease profitability and, other things being equal, a high probability of default should decrease the probability of loan approval. This finding is consistent with the intuition of Becker (1993) and the theoretical derivation of Berkovec et al. (1994) and might seem to imply that the coefficient on R in a default estimation should be negative. However, no analysis of the default approach has taken into account that the sample selection process also affects the variance of the error distribution. As demonstrated by equation (4), the loan approval process decreases the variance of the unobservable in the default equation. The sample of approved loans is heteroskedastic because the decrease in variance depends on borrower and loan characteristics. In addition, if minorities are held to a higher standard in approval, minority variances will fall further and the variance will be correlated with race.

Yatchew and Griliches (1985) examine the problem of heteroskedasticity in probit models. They find that traditional estimators are inconsistent if the variance of the unobservable variable is correlated with the explanatory variables, as it must be in this case because the heteroskedasticity is created by a selection process for borrower and loan characteristics. Applying Yatchew and Griliches's findings to the default model

⁴ These assumptions are consistent with the assumptions used by Berkovec et al. (1994) in deriving the predictions underlying the default approach.

⁵ This derivation ignores possible selection bias due to the household choice to apply for a loan as well as the household choice to reapply when a loan application is denied.

specified above shows that the first-order approximation of the maximum likelihood estimates $(\hat{\beta}_d + \hat{\beta}_r)$ is

$$\text{plim} \begin{bmatrix} \hat{\beta}_d \\ \hat{\beta}_r \end{bmatrix} \cong (\chi' \Omega^{-1} \chi)^{-1} \chi' \Omega^{-1} \chi * \begin{bmatrix} \beta_d + \tilde{\mu}(0, \mathbf{X}) \tilde{\mathbf{1}} \\ \tilde{\mu}(1, \mathbf{X}) - \tilde{\mu}(0, \mathbf{X}) \end{bmatrix}, \tag{5}$$

where χ and χ^* are created by stacking the row vectors $(1/\sigma) [\mathbf{X}'_i R_i]$ and $(1/\tilde{\sigma}(R_i, \mathbf{X}_i)) [\mathbf{X}'_i R_i]$, respectively, for all approved loans i ; $\tilde{\mathbf{1}}$ is a vector with the first element set to 1 for the intercept variable and all other elements set to zero; Ω^{-1} is a diagonal matrix with entries $f^{(*)}_2/[F^{(*)}(1 - F^{(*)})]$ with normal density (f) and cumulative distribution function (F) arguments of $(1/\sigma) (\beta_d \mathbf{X}_i + \beta_r R_i)$; and σ^2 is the constant variance that has been assumed incorrectly. Note that the second set of brackets in equation (5) contains the true values for the coefficients conditional on loan approval.

The resulting bias in the race coefficient depends not only on the values of c_r , \mathbf{X} , and R but also on the true values of all other parameters. A closed-form solution for $\hat{\beta}_r$ does not exist, and in general the sign of the race coefficient resulting from holding minorities to a higher standard cannot be determined. Under the assumption that \mathbf{X} is constant at $\bar{\mathbf{X}}$ over the entire sample, only the coefficient on R and the intercept are identified and b_d is a scalar. It can be shown that the coefficient vector converges in probability to

$$\text{plim} \begin{bmatrix} \hat{\beta}_d \\ \hat{\beta}_r \end{bmatrix} = \begin{bmatrix} \frac{\sigma}{\tilde{\sigma}(0, \mathbf{X})} (\beta_d + \tilde{\mu}(0, \mathbf{X})) \\ \left(\frac{\sigma}{\tilde{\sigma}(1, \mathbf{X})} - \frac{\sigma}{\tilde{\sigma}(0, \mathbf{X})} \right) (\beta_d + \tilde{\mu}(0, \mathbf{X})) + \frac{\sigma}{\tilde{\sigma}(1, \mathbf{X})} (\tilde{\mu}(1, \mathbf{X}) - \tilde{\mu}(0, \mathbf{X})) \end{bmatrix}. \tag{6}$$

With the intercept uncorrelated with the variance, the coefficient on the intercept is multiplied by a positive scalar only. The coefficient of R has two terms. The first term is the true intercept times a positive number because the variance of the minority sample is smaller if minorities are held to a higher standard. The second term is the true sample-selected race coefficient multiplied by a positive scalar and is negative. Since the second term is negative, the estimated race coefficient is unambiguously negative if the true intercept is negative. Therefore, the prediction of the default approach holds as long as the probability of default for approved majority loans is less than 50 percent, which is a very safe assumption.⁶

The finding that the default approach prediction holds for constant \mathbf{X} restores some confidence in the prediction for a general problem. For caution's sake, however, default analyses should be supported by simulations that investigate the bias in the race coefficient due to heteroskedasticity. In addition, because the relationship between the selection process and the estimated race coefficient is complex, simulation results from one sample and model may not hold for other samples and models. Therefore, the simulations should be conducted by using an actual sample of loan applications and reasonable estimates for both the loan approval model and a default model corrected for sample selection bias.

⁶ For constant \mathbf{X} , the same conditions can be derived by examining the first-order conditions for a probit estimation, as first suggested to me by James Berkovec.

Common Biases in the Default Approach

The assumption that e_l , e_u , and e_o are uncorrelated with race is currently a much debated topic. As stated earlier, Galster (1993) and Yinger (1993) both reject the default approach because they believe that unobserved characteristics affecting loan default (e_u) are correlated with race. In addition, Leibowitz (1993) and Zandi (1993)⁷ argue that the study by Munnell et al. (1992) using approval data is flawed because it omits, or includes only poor proxies for, key variables that might be correlated with race.⁸

The following subsections use the specification introduced in equations (1) and (2) to evaluate the default approach when unobservable characteristics are correlated with race. This section investigates only the selection effect on the mean of the error distribution or the true value of the race coefficient conditional on loan approval. The reason is that if heteroskedasticity is a major source of bias, the basic prediction of the default approach is no longer valid. Note that Ross (1995) derives results similar to those below for a continuous index of loan profitability.

Allowing Only e_l to Be Correlated with Race

The situation is most favorable for the default approach if only e_l is allowed to be correlated with race. If variables exist that are observable to the lender but unmeasured and correlated with race, the estimate of c_r in the loan approval model in equation (2) is biased. In this situation, it is reasonable to ask whether examining default might yield an unbiased result while analyses based on acceptance data are not robust.

The intuition behind such an idea is that nondiscriminating lenders should appear to hold minorities to a higher standard when controlling for \mathbf{X} because they are taking unmeasured variables into account. If a race dummy variable (R) is included in the default model (equation [1]), the higher standard should bias the resulting coefficient upward, canceling out the negative bias in equation (1) arising from the fact that e_l is negatively correlated with R .

However, this intuition is not correct. Let

$$c_e = \gamma \left(E[\epsilon_l | R = 0] - E[\epsilon_l | R = 1] \right) > 0, \quad (7)$$

which means that minority loan applications represent less profitable loans on average and implies that approval-based tests are biased toward finding discrimination as proposed by Leibowitz (1993) and Zandi (1993). Therefore, equation (2) can be rewritten as

⁷ Carr and Megbolugbe (1993) assert that Leibowitz's and Zandi's concerns are unfounded.

⁸ Endogeneity critiques, such as that by Rachlis and Yezer (1993), can also be reduced to an omitted-variable problem. For example, if the loan-to-value ratio is endogenous, it can simply be replaced with all variables that explain loan-to-value ratio in the approval specification. The resulting race coefficient represents the entire effect of race on approval both directly and through loan-to-value ratio. To the extent that lender behavior creates racial differences in the loan-to-value ratio, this specification provides a reasonable test for mortgage lending discrimination. However, this test suffers from omitted-variable bias if unobservable borrower characteristics that influence approval through the loan-to-value ratio are also correlated with race.

$$A = \begin{cases} 1 & \text{for } A^* > 0, \\ 0 & \text{for } A^* < 0, \end{cases} \quad (8)$$

where $A^* = \beta_a \mathbf{X} - (c_r + c_e)R + \gamma \tilde{\epsilon}_l + \epsilon_o$ and $\tilde{\epsilon}_l$ is an adjusted version of e_l with a mean independent of race. Then equation (1) can be rewritten as

$$D = \begin{cases} 1 & \text{for } D^* > 0, \\ 0 & \text{for } D^* < 0, \end{cases} \quad (9)$$

where $D^* = \beta_d \mathbf{X} + \tilde{\beta}_r R + \tilde{\epsilon}_l + \epsilon_u$ and $\tilde{\beta}_r$ is shifted upward by c_e/γ from the correlation between ϵ_l and R and shifted downward by the effects of c_e and c_r on the composition of the approved sample.

In fact, the net shift in the race coefficient varies with \mathbf{X} . Therefore, the true value of $\tilde{\beta}_r$ for an approved loan is a function of \mathbf{X} . That is,

$$\begin{aligned} \tilde{\beta}_r &= E[\epsilon_l + \epsilon_u \mid A = 1, R = 1, \mathbf{X}] - E[\epsilon_l + \epsilon_u \mid A = 1, R = 0, \mathbf{X}] \\ &= \left(-\frac{1}{\gamma} c_e + E[\tilde{\epsilon}_l + \epsilon_u \mid A = 1, R = 1, \mathbf{X}] \right) - E[\tilde{\epsilon}_l + \epsilon_u \mid A = 1, R = 0, \mathbf{X}] \\ &= -\frac{1}{\gamma} c_e + \gamma \sigma_l^2 \left(\lambda(-\beta_a \mathbf{X} + c_r + c_e) - \lambda(-\beta_a \mathbf{X}) \right). \end{aligned} \quad (10)$$

Under the assumption of no discrimination ($c_r = 0$),

$$\tilde{\beta}_r = \frac{1}{\gamma} \left(-c_e + \gamma^2 \sigma_l^2 \left(\lambda(-\beta_a \mathbf{X} + c_e) - \lambda(-\beta_a \mathbf{X}) \right) \right) > 0. \quad (11)$$

The default approach, which assumes that $\tilde{\beta}_r$ is zero if no discrimination exists, is biased away from finding discrimination. The slope of the inverse Mills ratio is always less than 1, so the difference between the inverse Mills ratios is less than c_e . The term $\gamma^2 \sigma_l^2$ is a portion of the approval equation variance and therefore is also bounded by 1. The omitted-variable bias term dominates, and $\tilde{\beta}_r$ is positive under the assumption of no discrimination.

Allowing Only e_u to Be Correlated with Race

If only e_u is allowed to be correlated with race, approval-based tests for discrimination are unbiased. The correlation of e_u with race may create an incentive for a rational lender to discriminate, and approval-based analyses cannot distinguish between discrimination based on this incentive and discrimination due to prejudice. Regardless of the reasons for it, discrimination is illegal.

Moreover, the performance-based tests are biased. To demonstrate, let

$$\bar{\beta}_r = E[\epsilon_u \mid R = 1] - E[\epsilon_u \mid R = 0] > 0, \quad (12)$$

which means that minority loan applications represent, on average, loans with a higher propensity to default and, as proposed by Galster (1993) and Yinger (1993), implies that lenders have an economic incentive to discriminate against minorities. Then equation (1) can be rewritten as

$$D = \begin{cases} 1 & \text{for } D^* > 0, \\ 0 & \text{for } D^* < 0, \end{cases} \tag{13}$$

where $D^* = \beta_d \mathbf{X} + \tilde{\beta}_r R + \epsilon_l + \tilde{\epsilon}_u$ and $\tilde{\epsilon}_u$ is an adjusted version of ϵ_u with a mean independent of race. Therefore, the coefficient on a race dummy variable for an approved loan is

$$\begin{aligned} \tilde{\beta}_r &= E[\epsilon_l + \epsilon_u \mid A = 1, R = 1, \mathbf{X}] - E[\epsilon_l + \epsilon_u \mid A = 1, R = 0, \mathbf{X}] \\ &= (\bar{\beta}_r + E[\epsilon_l + \tilde{\epsilon}_u \mid A = 1, R = 1, \mathbf{X}]) - E[\epsilon_l + \tilde{\epsilon}_u \mid A = 1, R = 0, \mathbf{X}] \\ &= \bar{\beta}_r + \gamma\sigma_l^2 \left(\lambda(-\beta_a \mathbf{X} + c_r R) - \lambda(-\beta_a \mathbf{X}) \right). \end{aligned} \tag{14}$$

However, under the assumption of no discrimination ($c_r = 0$), $\tilde{\beta}_r$ equals $\bar{\beta}_r$, which is assumed to be positive, and the default approach is biased toward accepting the null hypothesis of no discrimination. Therefore, failure to find a lower propensity to default for minority loans relative to majority loans cannot be used to refute findings of mortgage lending discrimination by other methods such as examination of the loan approval decision.

Berkovec et al. (1994) attempt to defend the loan performance approach against this criticism by arguing that the default approach is useful for detecting discrimination arising from prejudice rather than statistical discrimination. If the unobservable loan characteristics are correlated with race and negatively correlated with loan profitability, profit-maximizing lenders will hold minorities to a higher standard, and the resulting sample selection bias may cancel out the effect of the correlation on the race coefficient. Any race effect that remains must represent discrimination due to prejudice.⁹

As in equation (11), however, the competing effects of sample selection and the correlation between ϵ_u and R may not cancel out. Statistical discrimination occurs when lenders recognize that ϵ_u is correlated with race and then use race as a signal of the propensity to default. The traditional model of loan approval assumes that nondiscriminating lenders hold all borrowers to the same standard on an index of variables that are observed by the lender even though some of these variables may be unobserved by the researcher. A reasonable extension that allows for statistical discrimination is the assumption that nonprejudiced lenders hold applicants to the same standard after incorporating into their index the effect of race on the propensity to default.

A^* in equation (2) must then be rewritten to include ϵ_u because ϵ_u contains information on loan profitability. In other words,

⁹If only some lenders engage in statistical discrimination, discrimination due to prejudice might be hidden by the portion of lenders who do not statistically discriminate. Therefore, the default approach might be interpreted as a test of whether the total amount of discrimination creates sufficient sample selection bias to cancel existing racial differences in default.

$$\begin{aligned}
A^* &= \boldsymbol{\beta}_a \mathbf{X} - c_r R + \gamma(\epsilon_l + \epsilon_u) + \epsilon_o \\
&= \boldsymbol{\beta}_a \mathbf{X} - (c_r - \gamma\bar{\beta}_r)R + \gamma(\epsilon_l + \tilde{\epsilon}_u) + \epsilon_o,
\end{aligned} \tag{15}$$

where c_r is now a profit premium to compensate lenders for their prejudice and $\tilde{\epsilon}_u$ can be dropped from the second line because the mean-adjusted unobservable error term contains no information on profitability.

From equation (14),

$$\tilde{\beta}_r = \bar{\beta}_r + \gamma\sigma_l^2 \left(\lambda(-\boldsymbol{\beta}_a \mathbf{X} + (c_r - \gamma\bar{\beta}_r)) - \lambda(-\boldsymbol{\beta}_a \mathbf{X}) \right). \tag{16}$$

Under the assumption of no discrimination ($c_r = 0$),

$$\tilde{\beta}_r = \frac{1}{\gamma} \left[\gamma\bar{\beta}_r + \gamma^2\sigma_l^2 \left(\lambda(-\boldsymbol{\beta}_a \mathbf{X} - \gamma\bar{\beta}_r) - \lambda(-\boldsymbol{\beta}_a \mathbf{X}) \right) \right] > 0. \tag{17}$$

This coefficient is positive because the first term arising from the correlation between ϵ_u and R outweighs the selection effect. Therefore, the loan performance approach is biased away from finding discrimination arising from prejudice even under the assumption that all lenders engage in statistical discrimination.

Default Approach Using Sample Selection Corrections

The problems with the default approach may be circumvented if a sample contains information on both loan approvals and loan profitability for approved loans. For example, an unbiased estimate of $\bar{\beta}_r$ may be obtained by using a specification for default that controls for sample selection,¹⁰ and the default approach may be followed by comparing the estimate of $\tilde{\beta}_r$ from the specification without a selection correction in equation (13) with the unbiased estimate of $\bar{\beta}_r$.

However, the default approach with sample selection correction produces the same answer as the loan approval approach. Using $\bar{\beta}_r$ from equation (14),

$$\begin{aligned}
\tilde{\beta}_r - \bar{\beta}_r &= \left[\bar{\beta}_r + \gamma\sigma_l^2 \left(\lambda(-\boldsymbol{\beta}_a \mathbf{X} + c_r R) - \lambda(-\boldsymbol{\beta}_a \mathbf{X}) \right) \right] - \bar{\beta}_r \\
&= \gamma\sigma_l^2 \left(\lambda(-\boldsymbol{\beta}_a \mathbf{X} + c_r R) - \lambda(-\boldsymbol{\beta}_a \mathbf{X}) \right).
\end{aligned} \tag{18}$$

The difference is positive if and only if c_r is positive. Since the selection correction is generated from the approval equation, the sign of the difference is equivalent to the sign of c_r in the approval specification. In fact, if the approval specification suffers from omitted-variable bias, a performance test that uses a selection correction will suffer from the same omitted-variable bias. It is important to note that, as in the previous section, these derivations ignore heteroskedasticity bias.

¹⁰Brueckner (1994) suggests estimating a default equation that corrects for selection bias but does not mention the difficulties created by the fact that loan default is a discrete dependent variable.

Conclusions

The estimation of single-equation default specifications cannot provide reliable insights into the existence of mortgage lending discrimination. For this reason, researchers with an interest in mortgage lending discrimination should focus on obtaining unbiased, precise estimates of racial differences in loan approval. Therefore, while the findings of current studies that use approval data may be the topic of considerable debate, tests based on the loan approval process should be accepted as the principal technique for determining whether mortgage lending discrimination exists.

Specifically, theory does not predict the effect of discrimination on the estimates of a race coefficient in default equations. Loan default is a discrete dependent variable, and the variance of the unobservable is unidentified. Discrimination in the loan approval process affects both the mean and the variance of the unobservable error term, and the net effect on the race coefficient cannot be determined. A default approach prediction might be obtained empirically by using a simulation, but the nonlinearity of the bias would raise doubts about the generality of simulation results unless they were based on a representative sample of loan applications and reasonable loan approval and default model estimates.

Even if the heteroskedasticity bias is small, the default approach is still a less robust test for mortgage lending discrimination than approval-based tests. The performance-based approach is biased away from finding discrimination if either unmeasured loan characteristics observed by the lender or unobserved characteristics affecting default correlate with race. Approval-based tests are unbiased only when unobserved characteristics correlate with race, because unobserved characteristics can influence lender choices only if the lender uses race as a signal for those characteristics, which would be a discriminatory act.¹¹

The most troubling aspect of performance-based approaches is the effect of unobserved characteristics. The bias due to unmeasured loan characteristics in both approval- and performance-based tests might eventually be eliminated by improved data collection and improved specification of the loan approval process. It is unlikely, however, that researchers will be able to gather and measure characteristics that are unobserved by the mortgage lender, since it is in the lender's interest to gather all information that affects profitability.

In addition, the default approach is not helpful in testing for discrimination due to prejudice alone. Even if all lenders statistically discriminate, the default approach is biased away from finding discrimination due to prejudice. In considering the possibility of using the default approach with a sample selection correction, it would appear that such a test yields the same results as approval-based tests. Finally, all the above conclusions except for heteroskedasticity bias apply to any performance-based test for mortgage lending discrimination.

¹¹ The one exception to this statement is when unobserved borrower characteristics influence a second endogenous variable, such as loan-to-value ratio, which in turn influences expected propensity to default and the probability of loan approval. This problem in approval-based tests might be overcome, however, by carefully investigating the process that determines loan-to-value ratio and then including the appropriate variables in the approval equation.

References

- Arrow, Kenneth. 1973. The Theory of Discrimination. In *Discrimination in Labor Markets*, ed. Orley Ashenfelter and Albert Rees, 3–42. Princeton, NJ: Princeton University Press.
- Becker, Gary S. 1971. *The Economics of Discrimination*. 2nd ed. Chicago: University of Chicago Press.
- Becker, Gary S. 1993. The Evidence against Banks Doesn't Prove Bias. *Business Week*, April 19.
- Berkovec, James A., Glenn B. Canner, Stuart A. Gabriel, and Timothy H. Hannan. 1994. Race, Redlining, and Residential Mortgage Loan Performance. *Journal of Real Estate Finance and Economics* 9(3):263–94.
- Brimelow, Peter. 1993. Racism at Work? *National Review*, April 12.
- Brimelow, Peter, and Leslie Spencer. 1993. The Hidden Clue. *Forbes*, January 4.
- Brueckner, Jan K. 1994. Unobservable Default Propensities, Optimal Leverage, and Empirical Default Models: Comments on “Bias in Estimates of Discrimination and Default in Mortgage Lending: The Effects of Simultaneity and Self-Selection.” *Journal of Real Estate Finance and Economics* 9(3):217–22.
- Carr, James H., and Isaac F. Megbolugbe. 1993. The Federal Reserve Bank of Boston Study on Mortgage Lending Revisited. *Journal of Housing Research* 4(2):277–313.
- Galster, George C. 1993. The Facts of Lending Discrimination Cannot Be Argued Away by Examining Default Rates. *Housing Policy Debate* 4(1):141–46.
- Greene, William H. 1990. *Econometric Analysis*. New York: Macmillan.
- Leibowitz, Stan. 1993. A Study That Deserves No Credit. *Wall Street Journal*, September 1, p. A14.
- Munnell, Alicia H., Lynn E. Browne, James McEneaney, and Geoffrey M. B. Tootell. 1992. Mortgage Lending in Boston: Interpreting HMDA Data. Working Paper 92-7. Federal Reserve Bank of Boston.
- Peterson, Richard L. 1981. An Investigation of Sex Discrimination in Commercial Banks' Direct Consumer Lending. *Bell Journal of Economics* 12:547–61.
- Rachlis, Mitchell B., and Anthony M. J. Yezer. 1993. Serious Flaws in Statistical Tests for Discrimination in Mortgage Markets. *Journal of Housing Research* 4(2):315–36.
- Roberts, Paul Craig. 1993. Banks in the Line of Fire. *Washington Times*, March 12, p. F1.
- Ross, Stephen L. 1995. Mortgage Lending Discrimination and Racial Differences in Loan Performance. Working paper. University of Connecticut.
- Yatchew, Adonis, and Zvi Griliches. 1985. Specification Error in Probit Models. *Review of Economics and Statistics* 67:134–39.
- Yinger, John. 1993. Discrimination in Mortgage Lending: A Literature Review and Recommendations for Future Research. Paper read at the Discrimination and Mortgage Lending: Research and Enforcement Conference, U.S. Department of Housing and Urban Development, May 18–19, Washington, DC.
- Zandi, Mark. 1993. The Boston Fed's Bias Study Was Deeply Flawed. *American Banker*, August 19.