THE JOURNAL OF REAL ESTATE RESEARCH

Application of Reverse Regression to Boston Federal Reserve Data Refutes Claims of Discrimination

*Michael LaCour-Little**

***Abstract*. The topic of mortgage discrimination has received renewed interest since publication of the Boston Federal Reserve Bank study based on 1990 Home Mortgage Disclosure Act data. That study used traditional direct logistic regression to assess the influence of race on the probability of mortgage loan denial and reported the parameter estimate of race to be positive and significantly different from zero across several model specifications, thereby supporting contentions of discriminatory behavior. This paper develops an alternate approach, reverse regression, a method often used in the measurement of gender discrimination in labor markets. After discussion of theoretical issues regarding model choice, results of a reverse regression on the Boston Federal Reverse Bank study dataset are reported. Contrary to results using direct methods, reverse regression does not support contentions of mortgage discrimination in the Boston mortgage market. Rather, the lower overall qualifications of minority applicants are likely to account for disparities in application outcomes.**

Introduction

Since the 1992 publication by Munnell, Browne, McEneaney, and Tootell (MBMT) of the Boston Federal Reserve study (“the Boston Fed study”), the issue of alleged mortgage discrimination by lending institutions has received widespread press, and even congressional, coverage. Regulatory and even Justice Department actions have been predicated on its methodology: single-equation estimation of the probability of loan denial as a function of borrower and property covariates, using logit or probit methods. Yet considerable controversy over the adequacy of these methods persists. This paper reviews these topics and develops an alternative methodology, reverse regression, which has been used in measurement of gender discrimination in labor markets. Application of reverse regression to the Boston Federal Reserve study dataset does not support contentions of discriminatory behavior by Boston area lenders; if anything, it suggests reverse discrimination.

Since passage of the Home Mortgage Disclosure Act of 1975 (HMDA), researchers have sought to use HMDA data to understand the geographic distribution of mortgage credit. After 1989, when HMDA was amended and disaggregated data became available, researchers have focused on the determinants of differential outcomes among ethnic groups.

*Department of Real Estate and Urban Land Economics, School of Business, University of Wisconsin-Madison, Madison, Wisconsin 53706.
Date Revised—May 1995; Accepted—July 1995.

The plan of the present paper is as follows. First, the Boston Federal Reserve study is described. Second, criticisms of it are presented and discussed. Third, the rationale for use of an alternative approach, reverse regression, is developed. Fourth, results of application of reverse regression to the Boston Fed data are presented. Finally, conclusions and suggestions for further research are offered.

Boston Federal Reserve Study

The Boston Federal Reserve Study was undertaken in 1992 based on disaggregated 1990 HMDA data, augmented by survey data collected from Boston area lenders who were required to submit HMDA data. Each lender was asked to provide additional information on every application for credit made by blacks and hispanics during 1990 and for a random sample of 3300 applications made by whites. According to the Boston Fed study, "Substantial lender cooperation resulted in a very good response rate and high-quality data."[1] In addition, denial rate disparities were representative of national patterns; black and hispanic applicants had a 28% denial rate as compared to slightly over 10% for whites. The aggregate denial rate was 14%.

The Boston Fed methodology applied direct regression methodology to the measure of discriminatory lender behavior based on the following model of the mortgage loan decision,

$$\text{Prob}\,(y=1 \mid X, z)=b'X+az+e, \tag{1}$$

where $y=1$, if the loan application is denied, X is a set of borrower, property and neighborhood covariates, conceptually including all factors used in loan underwriting, and z=an indicator variable for the presence of the attribute against which lenders may discriminate; here $z=1$ if applicant is black or hispanic, $z=0$ otherwise. The variable e represents an additional unobserved random error term. The parameter of interest, then, is "a," and a positive value significantly different from zero is taken as a measure of discriminatory lender behavior.

The covariates used in the Boston Fed study are defined in Appendix 1. In the Fed study sample, minority (black and hispanic) applicants differ from white applicants in a number of respects. They tend to purchase less expensive properties, though those properties were more likely to be multifamily dwelling units. In addition, they have higher loan-to-value ratios and smaller net worth, relative to white applicants. Finally, minority applicants have objectively worse credit histories, as measured by *MORTPAY, CONSPAY* and *PUBREC,* and higher housing expenses and total debt burdens. On the positive side, minority applicants are less likely to be self-employed than were white applicants. (Lenders generally view income from self-employment as less stable than wage income.)

It is also useful to compare rejected applicants to approved applicants. Rejected applicants were more likely to be purchasing a multifamily dwelling unit, more likely to be self-employed, had substantially worse credit histories, as measured by *CONSPAY, MORTPAY* and *PUBREC,* and, finally, had higher ratios: higher housing expense-to-income, total debt-to-income, and loan-to-value. Thus, we see an essential problem: minority applicants appear to be systematically less qualified than are white applicants. This difference in average applicant qualifications may be modeled by the reverse regression procedure.

The authors of the Boston Fed study conclude as follows:

> The results of this study indicate that race does play a role as lenders consider whether to deny or approve a mortgage loan application . . . the higher denial rate for minorities in Boston is accounted for, in large part, by their having higher loan-to-value ratios and weaker credit histories than whites. They are also more likely to be trying to purchase a two-to-four unit property rather than a single-family home. Nevertheless, after taking account of such factors, a substantial gap remains. A black or Hispanic applicant in the Boston area is roughly 60% more likely to be denied a mortgage loan than a similarly situated white applicant.[2]

These conclusions are based on the sign and statistical significance of the dummy variable for race included in the model. While the race variable is not the most important determinant—denial of private mortgage insurance and public record credit defects have much larger magnitudes, it is consistently positive and significant (*t*-ratio of approximately 5) across a number of model variations, according to MBMT.

The signs of parameter estimates are generally as expected. Higher housing expense (*HEXP*), debt ratios (*TOTDEBT*), or loan-to-value (*LTV*) ratios increase the probability of loan denial. Likewise, a poorer credit history (*CONSPAY, MORTPAY*), and particularly public record (*PUBREC*) defects in borrower credit history increase the likelihood of denial. Interestingly, net worth (*NETW*) seems to have little impact on loan application outcome. On the other hand, being self-employed (*SELF*) or purchasing a multifamily property (*MFDU*) does increase probability of denial. Finally, minority status (*RACE*) does indeed increase the probability of denial, as concluded by MBMT. Among all variables, denial of private mortgage insurance (*PMI*) has the greatest effect on probability of loan denial, virtually assuring it; at least for the subset of the sample requiring *PMI* (those with *LTV* greater than 80%).

Criticisms of Fed Study Methodology

The Boston Fed study has been criticized on a number of grounds. Zandi (1993) argued that the study suffered from omitted variable bias since MBMT failed to include certain binary variables such as "meets lender credit guidelines" in their final model specification. Megbolugbe and Carr (1993) responded that judgmental variables such as "meets credit guidelines" are themselves correlated with race and that, in fact, whites and minorities will meet, or fail to meet, credit guidelines based on race. Liebowitz and Day (1993) claim that (1) there are so many miscodings in the dataset as to make any conclusions drawn from it doubtful, and (2) that a small number of highly influential observations drive parameter estimate results on *RACE.* Megbolugbe and Carr (1993) concur that there appear to be coding errors in the Fed study, identifying some fifty-three likely errors, but after a series of data scrubbing procedures, conclude that these errors are not necessarily responsible for the positive coefficient estimate on *RACE.* In addition, Megbolugbe and Carr reestimated the Fed model after deleting the twenty-seven most influential observations as well as the fifty-three probable miscodes, yet found the coefficient on *RACE* still positive and highly significant. LaCour-Little (1994) re-estimated that Boston Fed model after deleting 252 suspicious observations. While the parameter estimate on *RACE* was reduced from .73 to .63, with an associated t-ratio decline from 5.28 to 4.03; by most standards, these are still very robust results. The difficulty with these sorts of discussion is, of course, that there really is no way to objectively determine which observations are, in fact, miscodings that should be deleted. Accordingly, in the empirical section of the paper that follows, I use the entire Boston Fed dataset.

On a more theoretical basis, Rachlis and Yezer (1993) argue that single-equation models of the mortgage lending process oversimplify complex phenomena and that unbiased tests for discrimination require multiple equation models. Yezer, Phillips and Trost (YPT) (1994) argue that single-equation models ignore several crucial issues. First, since rational borrowers choose lenders so as to maximize their chances of loan approval and lenders screen borrowers for the same reason, the "final sample of approved applicants is selected in a fashion that is certainly not random" (YPT, 1994, p. 197). Second, they argue, borrowers take into account their own assessment of the probability of default in selecting loan terms. This assessment is, of course, unobservable, yet theoretically should enter into equations describing borrowers' choice of loan terms and lenders' accept or reject decision. Thus, they argue, a sufficient model of the mortgage lending process must include at least three equations: borrowers' choice of loan terms, lender's accept or reject decision, and borrowers' default decision. Noting that "in most cases the system of equations . . . is too complex to permit unambiguous analytical results" (PTY, 1994, p. 205), they show by means of a Monte Carlo simulation the effects of simultaneity and selection biases using a three-equation model. Results indicate that the parameter estimate on minority status is biased upward, indicating discrimination when none exists.

The argument in this paper, while consistent with YPT's view that the single-equation direct regression approach cannot accurately measure discrimination, is somewhat different. I argue that even if the borrowers' choice of loan terms and lender were exogenous, under a reasonable proxy-variables model of the mortgage lending process, direct regression will produce a biased estimate of the discriminatory effect. Moreover, reverse regression will provide an unbiased estimate. By way of preview, when reverse regression is applied to the Boston Federal Reserve data, the sign of the coefficient measuring discrimination changes, evidence for, if anything, reverse discrimination.

My purpose here is not to debate these issues further, so let us assume, for argument's sake, that loan terms are indeed exogenous and that evidence from direct regression, with or without data deletions, suggests discriminatory behavior by mortgage lenders. What does this imply? If lenders discriminate against minority loan applicants, then it is reasonable to expect that they will tend to accept only the most qualified minority applicants; that is, they will tend to resolve borderline cases against minority applicants and the marginal minority loan applications will be better qualified than the marginal white applicant. This notion, generalized, is sometimes expressed as follows: because of discrimination, minorities must be better qualified in every way than whites, simply to achieve similar outcomes.

Reverse Regression

An excellent overview of the relationship between direct and reverse regression appears in Chapter 7 of Leamer (1978). Use of reverse regression as a method of measurement of discrimination developed in the labor economics literature. Just as in the mortgage discrimination case, gender discrimination in wages is often demonstrated by means of a direct regression of wages on a vector of job qualification variables plus a dummy variable for female, with a significant negative coefficient on female taken as evidence of wage discrimination against women. In contrast, Kamalich and Polacheck (1982) and Conway and Roberts (1983) propose use of reverse regression in the measurement of wage discrimination based on gender. For instance, Kamalich and Polachek argue that "If discrimination exists, one would expect to find blacks and women to have higher mean qualifications for any given wage level" and "a pattern of mixed positive and negative coefficients . . . is consistent with nondiscrimination."[3] In the present case, we shall see that the estimate of discriminatory effect produced by reverse regression is, indeed, of the opposite sign as the direct regression estimate. Leamer (1978) shows formally that in an errors-in-variables context, direct and reverse regression estimates bound the true parameter value and that in a proxy-variables case, direct regression produces an upwardly biased estimate whereas the reverse regression estimate is unbiased.

Reverse regression methods have reportedly become common in class action litigation over gender differentials in the past fifteen years, partially because they will generally produce different results than do the analogous direct regression, since the slope of $E(Y|X)$ is not generally equal to that of $E(X|Y)$. In particular, they may produce results that indicate much less discrimination, perhaps none, than was suggested by direct regression methods. In such cases, then, the measures of discrimination derived from the direct and reverse regression will have the same sign, but the reverse regression coefficient will typically have a much smaller magnitude. By way of a preview, in the case at hand, reverse regression coefficients of discrimination are not merely smaller than those obtained by direct regression, they are negative.

The labor market analogy to mortgage discrimination is fairly immediate. We observe women to have, on average, lower salary levels than do men in comparable jobs and speculate that discrimination may play a role. Upon further analysis we determine that women have, on average, lower qualifications (as measured by education, years of experience, etc.) than do men. What measurement technique is appropriate? Similarly in mortgage markets, we observe minorities to have a lower approval rate on mortgage loans than do white applicants but determine that they, too, seem to have, on average, lower qualification levels (worse credit, higher ratios) as well. How can these differences be explained?

In certain cases, we may view reverse regression as a solution to an errors-in-variables problem. The dependent variable is measured precisely (we know whether the application was approved or denied or what actual salaries are) but measured covariates are, at best, an incomplete set of proxies for qualifications. In general we wish to run the direct regression,

$$y = b'X + az + e\,, \tag{2}$$

where y=subject of interest (salary level or mortgage loan approval rates), X=a matrix of measured qualification covariates, and z=the attribute allegedly discriminated against (gender or race). The variable e is a random error term with expectation of zero. In this approach, "a" is used as the measure of discriminatory behavior. Since X is a series of multiple indicators of qualifications, actually of inverse qualifications (higher values of the index means the applicant is less qualified, at least in the mortgage discrimination case), and "y" can be measured precisely, we instead run the reverse regression of the composite inverse qualifications index, $q=b'X$ upon y and z,

$$q=cy+dz \tag{3}$$

and then use $-d/c=a^*$ as the measure of discriminatory effect.

Goldberger (1984) develops this approach in the wage discrimination case showing that where measured covariates are merely imperfect proxies for qualifications, the direct regression coefficient estimator will be biased upward while the reverse regression estimator will be unbiased. Goldberger further claimed that coefficients obtained from any individual reverse regression, that is, regression on a single qualifications proxy variable, all must be proportional to one another to justify use of reverse regression. Dempster (1988), while arguing that attempts to measure discrimination by either method required a non-statistical foundation, showed that proportionality of coefficients was not, in fact, a prerequisite to use of reverse regression. Appendix 2 contains a proof, following Goldberger, of the conditions under which the reverse regression estimator, "a," is unbiased.

It may be objected at this point that the problem of mortgage discrimination does not conform to salary discrimination in at least one very important regard: salaries are a continuous variable whereas mortgage application outcome is binary. Consider, however, the following model of the mortgage decision process. The lender implicitly assesses the probability that the mortgage loan applicant will default on a loan. This assessment, call it Y^*, is the sum of the true default probability, p, and any discriminatory effect, αz:

$$Y^*=p+\alpha z\ , \tag{4}$$

where z=an indicator of the applicant class allegedly discriminated against, i.e., $z=1$ if applicant is minority, $z=0$, otherwise. Now Y^* is a continuous variable on the unit interval. If Y^* is greater than some threshold, θ, then the lender rejects the loan and we observe $Y=1$:

$$Y=1,\text{ if } Y^*\geq\theta;\ Y=0,\text{ if } Y^*<\theta\ . \tag{5}$$

Moreover, the probability of borrower default may, for a variety of reasons, be related to z, but subject to a random error term:

$$P=\mu z+u\ . \tag{6}$$

Finally, we have an imperfect set of indicators, X, of p:

$$X=\gamma p+\epsilon\ \cdot \tag{7}$$

Estimation of a binary response model (logit or probit) will reveal $[Y^*|X]$. We can then form the inverse qualifications index $q=b'X$ and estimate the reverse regression given in (3).

As shown in Appendix 2, the resulting value of $a^* = -d/c$ will be an unbiased estimate of α, the discriminatory effect we seek to measure.

Results of Reverse Regression

I begin this section with a review of the mechanics of the procedure. A report of results using the Fed study dataset is then presented.

The first step in reverse regression is to estimate a logit function of *ACTION* on the set covariates used in the direct regression model (*CONSPAY, HEXP,* etc.) but without the *RACE* variable. The results of that regression, together with the complete model including *RACE* for reference purposes, are shown in Exhibit 1. Results from the Full

Exhibit 1
Direct Logistic Regression Excluding Race*

Variable	Parameter Est-*ACTION* (W/O *RACE*)	Parameter-Est (Full Model)
Intercept	−6.44	−6.50
	(−16.5)	(−16.6)
HEXP	.49	.46
	(3.25)	(3.0)
TOTDEBT	.05	.05
	(6.60)	(6.5)
NETW	−.00006	.00009
	(.96)	(1.4)
CONSPAY	.33	.31
	(9.96)	(9.2)
MORTPAY	.39	.35
	(3.35)	(3.0)
PUBREC	1.28	1.20
	(7.32)	(6.8)
URIA	.07	.08
	(2.56)	(3.0)
SELF	.39	.46
	(2.10)	(2.5)
LTV	.66	.61
	(3.14)	(3.2)
MFDU	.68	.51
	(4.31)	(3.1)
PMI	4.63	4.61
	(9.50)	(9.41)
RACE	NA	.73
	NA	(5.3)
-2 Log Likelihood:	1,745.7	1,718.7
Goodness of fit:	3,088.7	3,135.4
Model *Chi*-square	688.2	715.3
Degrees of freedom:	11	12
Overall percent correct:	88.81%	89.05%
Number of observations:	2,932	2,932

*Dependent variable is *ACTION* (*ACTION* =1, if loan denied).
T-ratios are in parentheses beneath parameter estimates.

Model show the parameter estimate of *RACE* to be .73 (t-ratio of 5.3).[4] Given the logit method, this means that minority status (*RACE*=1) significantly increases the odds of loan turndown. Exact probabilities of turndown depend on the values of other variables contained in the model. Given these, predicted probabilities of loan denial for each observation based on the race-free regression equation also estimated are saved as part of this first step. These predicted probabilities may be viewed as a race-free inverse qualifications index for each observation, hence are renamed *Q-INDEX,* an additional variable now available for each observation. Keep in mind that higher values of the inverse qualifications index mean the applicant is less qualified, i.e., the probability of turndown is greater. The second step in the reverse regression uses ordinary least squares to regress the inverse qualifications index, *Q-INDEX,* on the pair of binary covariates *RACE* and *ACTION.* Results of this second step, the reverse regression, are shown in Exhibit 2. The positive coefficient of .057 on *RACE* reflects the excess qualifications associated with minority status. Since, in this context, excess qualifications are negative (i.e., the applicant is *less* qualified), we may interpret this quantity as the excess probability of default required to turn down a minority loan applicant.

Exhibit 3 compares results of direct and reverse regressions. The key result is the difference in signs between "a" and "a^*". Under direct regression, minority status has a negative impact on loan application outcome, i.e., increases probability of denial ("a" is positive). Again, in the logit framework, this operates through the odds ratio, i.e., the actual increase in the probability of denial depends on levels of other variables, so the typical OLS interpretation of the coefficient is not appropriate. Under reverse regression, however, minority loan applicants are shown to be approved with average qualifications 19% lower than white applicants ("a^*" is negative). Hence we cannot say that minority loan applicants must meet a higher standard than white applicants; rather, the reverse appears to be the case. One might argue that this constitutes reverse discrimination, since some whites with objectively higher qualifications would have mortgage loan applications

Exhibit 2
Reverse Regression (OLS)*

Variable	Parameter Estimate
Intercept	.0892
	(25.5)
ACTION	.296
	(34.3)
RACE	.057
	(7.92)
F-value:	711.33
R-squared:	.33
Std Error:	.16
N	2,932
$a^* = -.057/.296 = -.196$	

*Dependent variable is *Q-INDEX.*
T-ratios are in parentheses beneath parameter estimates.

Exhibit 3
Comparison of Direct and Reverse Regression Coefficients

Variable	Direct	Reverse		
	a	*c*	*d*	*a**
z	.73	.296	.057	−.193

Exhibit 4
Reverse Regression (OLS) Hold-Out

Variable	Parameter Estimate
Intercept	.087
	(17.6)
ACTION	.286
	(22.9)
RACE	.064
	(6.2)
F-value:	339.8
Adj. *R*-squared:	.316
Std Error	.162
N	1,466
$a^* = -.064/.286 = -.224$	

T-ratios are in parentheses beneath parameter estimates.

denied while minorities with lower qualifications were approved. Why might such a result occur? If lenders, fearful of charges of discrimination, consciously or unconsciously resolve all borderline cases in favor of minority applicants, such a result is certainly possible.

To assess robustness of these findings, a hold-out procedure was employed. First, the Boston Fed data was split randomly in half, using the first half of the dataset to estimate the parameters of *Q-INDEX* and the second half to compute the reverse regression, using predicted *Q-INDEX* values. Results of the hold-out analysis are shown in Exhibit 4. Parameter estimates are almost identical to those derived on the entire dataset, although standard errors are slightly larger (and, accordingly, *t*-ratios smaller) as would be expected using a smaller number of observations. The resulting value of a^* is $-.224$, very similar to the $-.196$ value derived from estimation using the entire dataset.

In summary, reverse regression has been shown to provide an alternative method for assessing mortgage discrimination, just as it has in the labor market studies of gender-based discrimination. In contrast to direct methods, reverse regression addresses a distinct question: given outcomes, how do qualifications vary with race? The simple answer is minority qualifications are systematically lower than white qualifications.

Conclusions

Reverse regression does not reinforce the contention that minority applicants were discriminated against by Boston area lenders during 1990; it may, however, help explain why minority turndown rates were twice those of the entire group. Race-free indices of qualifications show minority loan applicants to have qualifications between 19% and 22% lower than white applicants. Looking at the results another way, the lenders' assessment of default probability must be about 5.7% higher for lenders to reject a minority loan applicant, as compared to a white applicant. In other words, lenders appear to apply less stringent underwriting standards to minority loan applications; nonetheless, since minority loan applicants are, on average, so much less qualified than white applicants, average turndown rates are higher for minority applicants. This finding implies that public policy might better focus, then, upon improving minority qualifications, rather than ferreting out apparent discrimination, which may result merely from differential qualifications.

The Boston Fed dataset is, of course, only a single sample from a particular geographic market at a particular point in time. Whether we accept results of direct regression or reverse regression, we should not infer that these results necessarily hold in all markets over all time periods. Further research, in different markets and over different timeframes, but also based on expanded HMDA data, will tend to confirm or refute the parameter estimates of the Fed study and variations on it.

Appendix 1
Information on Boston Fed Study Data

Definition of Variables	
Housing Expense to Income (*HEXP*)	A dummy variable indicating HEXP in excess of .30
Total Debt to Income (*TOTDEBT*)	The ratio of debt-to-income
Net Wealth (*NETW*)	Applicant net worth
Consumer Credit History (*CONSPAY*)	Number of consumer credit late payments
Mortgage Credit History (*MORTPAY*)	Number of mortgage credit late payments
Public Record History (*PUBREC*)	A dummy variable indicating presence of any public record defects in the applicant's credit history (e.g., a judgment or bankruptcy)
Probability of Unemployment (*URIA*)	The unemployment rate in the applicant's occupation
Self-Employment (*SELF*)	A dummy variable indicating self-employment
Loan-to-Appraised Value (*LTV*)	Loan-to-value ratio
Denied Mortgage Insurance (*PMI*)	A dummy variable indicating applicant was denied private mortgage insurance
Two-Four Family Property (*MFDU*)	A dummy variable indicating purchase of a multifamily property
Minority (Black or Hispanic) Status (*RACE*)	A dummy variable indicating whether the applicant is black or hispanic

Appendix 2
Proof of Bias of Direct Regression Estimator

The following shows the bias of the direct regression estimator and the lack of bias in the reverse regression estimator, assuming a multiple indicators model. Notation and calculations closely follow Goldberger. Assume that:

$$Y^* = p + \alpha z \tag{1}$$
$$x = \gamma p + \varepsilon \tag{2}$$
$$p = \mu z + u \tag{3}$$
$$E[u|z] = 0 \tag{4}$$
$$V[u|z] = \sigma_u^2 \tag{5}$$
$$E[\varepsilon|p,\ z] = 0 \tag{6}$$
$$V[\varepsilon|p,\ z] = \Omega \tag{7}$$
$$E[x|z] = \gamma\mu z \tag{8}$$
$$V[x|z] = \gamma\gamma'\ \sigma_u^2 + \Omega \tag{9}$$
$$E[Y^*|z] = (\mu + \alpha)z \tag{10}$$
$$V[Y^*|z] = \sigma_u^2 \tag{11}$$
$$c[x,\ Y^*|z] = \gamma\sigma_u^2 \tag{12}$$

Then for direct regression,

$$b = \pi^*(\gamma'\Omega^{-1}\gamma)^{-1}\Omega^{-1}\gamma\,, \tag{13}$$

where,

$$\pi^* = \sigma_u^2\gamma'\Omega^{-1}\gamma / \ (1 + \sigma_u^2\gamma'\Omega^{-1}\gamma)$$

and the discrimination coefficient, a, is

$$a = \alpha + (1 - \pi^*)\mu\ , \tag{14}$$

which is biased upward.

In contrast, for the reverse regression,

$$c = \sigma_u^2\gamma\sigma_u^2 = \gamma\,, \tag{15}$$

$$d = \gamma\mu - c(\mu + \alpha) = -\gamma\alpha\ . \tag{16}$$

Then,

$$c = \pi^* \tag{17}$$

$$d = -\pi^*\alpha \tag{18}$$

$$a^* = -d/c = \pi^*\alpha/\pi^* = \alpha\,, \tag{19}$$

and a^* is an unbiased estimator of α.

Notes

[1]Munnell, Browne, McEneaney, and Tootell (1992), p. 1. See also Browne's response to Liebowitz in September 21, 1993 *Wall Street Journal* letter to the editor.
[2]MBMT (1992), pp. 43–44.
[3]Kamalich and Polacheck (1982), pp. 450, 459.
[4]Perceptive readers will note that while the coefficient on *RACE* is .73, the coefficients on *PUBREC* (an indicator of public record credit defects) and *PMI* (denial of *PMI* insurance) are even larger. Plainly, lenders consider public record credit defects, such as judgments or bankruptcies, very serious. The *PMI* effect is more complex. *PMI* is generally required for loans with *LTV*s greater than 80%, so downpayment-constrained minority households are more likely to need *PMI.* The

Boston Fed study (1992, p. 32) specifically addresses this issue. Moreover, in the Appendix to the Boston Fed study, Table B5 reports a separate analysis of the *RACE* effect in *PMI* denial. While they find a smaller, but still positive, *RACE* effect here, too, sample size is small (only 2.6% of all loan applications were denied *PMI*). The role of *PMI* certainly deserves further study and offers an interesting control case, since it is unlikely that *PMI* underwriters would ever have any direct contact with the loan applicant; although they may be in possession of the loan application itself, which does indicate race, it is hard to understand how racial prejudice might enter into their decisions.

References

Conway, D. and H. V. Roberts, Reverse Regression, Fairness, and Employment Discrimination, *Journal of Business and Economic Statistics,* January 1983, 1, 75–85.

Dempster, A., Employment Discrimination and Statistical Science, *Statistical Science,* 1988, 3:2, 149–61.

Goldberger, A., Reverse Regression and Salary Discrimination, *Journal of Human Resources,* 1984, 19:3, 293–318.

Kamalich, R. and S. W. Polacheck, Discrimination: Fact or Fiction?, An Examination Using an Alternative Approach, *Southern Economic Journal,* October 1982, 49, 450–61.

LaCour-Little, M., Mortgage Discrimination: The Boston Federal Reserve Study and Its Detractors, unpublished working paper, University of Wisconsin-Madison, 1994.

Leamer, E., *Specification Searches,* New York: John Wiley, 1978, Ch. 7.

Liebowitz, S., A Study That Deserves No Credit, *Wall Street Journal,* September 1, 1993, A14.

—— and T. Day, Mortgages, Minorities, and Discrimination, Working Paper, University of Texas at Dallas, 1993.

Megbolugbe I. and J. Carr, Federal Reserve Bank of Boston Study on Mortgage Lending Revisited, *Journal of Housing Research,* 1993, 4:2, 277–314.

Munnell, A. H., L. E. Browne, J. McEneaney, and G. M. B. Tootell, Mortgage Lending in Boston: Interpreting HMDA Data, Federal Reserve Bank of Boston Working Paper, 92–7, Boston, Mass.: FRB, 1992.

Rachlis, M. B. and A. M. J. Yezer, Serious Flaws in Statistical Tests for Discrimination in Mortgage Markets, *Journal of Housing Research,* 1993, 4:2, 315–36.

Yezer, A. M. J., R. Phillips and R. Trost, Bias in Estimates of Discrimination and Default in Mortgage Lending: The Effects of Simultaneity and Self Selection, *Journal of Real Estate Finance and Economics,* 1994, 9, 197–215.

Zandi, M., Boston Fed's Bias Study Was Deeply Flawed, *American Banker,* August 19, 1993, 13.

The author gratefully acknowledges the comments of Charles F. Manski, James D. Shilling, and two anonymous referees who read earlier versions of this paper, while retaining responsibility for any and all remaining errors.