Under-specified Models and Detection of Discrimination in Mortgage Lending

Jason Dietrich

Economic and Policy Analysis Working Paper

March 2003

Under-specified Models and Detection of Discrimination in Mortgage Lending

Jason Dietrich

Office of the Comptroller of the Currency
Economic and Policy Analysis Working Paper

March 2003

Abstract: Most empirical studies of discrimination in mortgage lending can be criticized for omitted variable bias. With access to data and policy guidelines typically unavailable to researchers, the OCC is in a unique position to assess the importance of omitted variables on fair lending models. This study examines how variables available to the OCC, but often unavailable to researchers, affect estimates from statistical models and identification of outliers for manual review.

The results show that omitted variables have an important impact on both the estimate of the effect of race and on the identification of outliers for review. Further, there appears to be no consistent patterns to the direction of these impacts. This suggests that it is inappropriate to make generalizations about the potential direction of bias based on assumptions about the correlations between omitted variables and race.

The views expressed in this paper are those of the author alone, and do not necessarily reflect those of the Office of the Comptroller of the Currency or the Department of the Treasury. The author would like to thank Amber Jessup, Hannes Johannsson, Shanzi Ke, Melissa McShea, Erik Larson, Dennis Glennon, Irene Fang, and Amy Millen for their insightful comments and editorial assistance.

Please address correspondence to Jason Dietrich, Economist, Risk Analysis Division, Office of the Comptroller of the Currency, 250 E Street, S.W., Washington, DC 20219 (phone: 202-874-5119; e-mail: jason.dietrich@occ.treas.gov)

I. Introduction

Researchers use statistical modeling tools during fair lending analyses to isolate, and estimate, the role race plays in banks' underwriting decisions. They achieve this by first controlling for the effects of all relevant economic factors, and then considering prohibited factors, such as race, to explain any remaining systematic patterns in the data. Unfortunately, data and information limitations make it difficult to control for the effects of all relevant economic factors. As a result, critics argue that statistically significant racial estimates do not indicate discrimination, but the effects of omitted economic factors that are correlated with race. Implicit in this criticism is the notion that the racial estimates would decline if these omitted factors were included in the model. If this notion is correct, then conclusions of no discrimination can be drawn reliably when the racial estimates in an under-specified, yet reasonable, model are not significant. Omitted factors therefore affect the reliability of statistical models by introducing bias into the racial estimates, and, depending on the theorized direction of this bias, affect the amount of resources expended gathering data during the analysis. Unfortunately, it is difficult to accurately assess the magnitude and direction of this bias empirically, precisely because limited data and information are the cause of the bias.

Banking regulators are in a unique position to evaluate the effects of omitted factors on fair lending models, because of access to data and information typically unavailable to academic researchers. During fair lending exams, regulators have access to bank policies, underwriters, and loan applications. This allows them to identify all of the economic factors a bank relies on during the underwriting decision-making process, as well as to determine how each factor is used. The opportunity is available, therefore, for regulators to diminish greatly problems of omitted variable bias. The tradeoff is the potential for large resource expenditures to gather, clean, and test the necessary data. This is especially true if data need to be entered manually. The Office of the Comptroller of the Currency (OCC) has chosen a strategy of replicating the bank's underwriting decision-making process to maximize the reliability of the models' racial estimates. In addition to providing reliable estimates for exam purposes, this strategy provides the data and information necessary to assess the effects of omitted variables on fair lending models.

This study uses data from 18 fair lending exams recently conducted by the OCC to examine the effects that omitted variables would have had on statistical analyses of banks' underwriting decisions.[1] Two specific questions are addressed. First, how would omitted variables have affected the estimated racial effects and their corresponding t-statistics? These results directly assess the magnitude and direction of omitted variable bias in fair lending models and address criticisms that omitted variable bias influenced previous studies showing evidence of discrimination. Specific attention is paid to differences in the effects of omitted variables between linear and non-linear estimators, because the type of estimator affects hypotheses about the direction of omitted variable bias. It is well known that omitting relevant variables introduces bias into linear estimators that is a function of the effect of the omitted variable on the dependent variable and the correlation of the omitted variable and each non-omitted variable.

[1] The data used in this study were collected as part of official OCC fair lending exams. They are therefore strictly confidential and cannot be released to the public or shared with other researchers.

Relatively little is known, however, about these effects on the non-linear logit estimator, which researchers typically use to estimate models of the underwriting decision. The second question this study examines is how do omitted variables affect the identification of outliers for manual review? Reviewing a model's outliers is an important part of any modeling exercise, because it identifies information potentially missed by the model and provides a second source of supporting evidence for the statistical analysis. Using an under-specified model to identify outliers to review can result in not reviewing outliers that would have been identified using a fully specified model and unnecessarily reviewing outliers that would not have been identified using a fully specified model. This study examines the extent of each of these potential problems using data from recent statistically modeled fair lending exams.

Section II begins with background information and a summary of the OCC's approach to statistically modeled fair lending exams. Section III examines the omitted variable bias problem in the logit estimator, summarizing the theoretical work on this topic presented in Lee (1980, 1982). Section IV examines the impact omitted variables would have had on estimates of the effect of race, and the corresponding t-statistic from statistical models used during past exams. Section V focuses on the impact omitted variables would have had on the predicted probability of denial and the identification of outliers for manual review for these same exams. Section VI concludes the discussion.

II. Background

The Financial Institutions Reform, Recovery and Enforcement Act (FIRREA) passed by Congress in 1989 required banks to report data on the underwriting decision on applications for home mortgage products, as well as on the race of the applicants. These data allowed the public one of the first opportunities to calculate denial disparities by race at the individual level. Denial disparities, which almost always indicate that minorities are denied at a higher rate than whites, have been, and continue to be used as indicators of lending discrimination. Although these data are useful for developing an initial characterization of mortgage lending, using them as evidence of discrimination is dangerous as it does not consider creditworthiness factors.

Munnell *et al.* (1992) is commonly cited as the first study to control thoroughly for applicants' creditworthiness in an individual-level analysis of discrimination in the underwriting process.[2],[3] Incorporating information on 38 factors that mortgage underwriters and lenders in Boston indicated were important in their decision-making process, the authors found that Black and Hispanic applicants were still approximately 60 percent more likely to be denied credit than similarly situated white applicants.

Munnell *et al* was a clear improvement over previous studies of discrimination in mortgage lending. However, even with the extensive list of variables the authors used to capture applicants' creditworthiness, one set of responses still focused on omitted variable bias. Critics identified three general sources of omitted information. The first source consists of specific factors in the Munnell *et al* dataset that were relevant to the underwriting decision and correlated with race (Liebowitz (1993), Zandi (1993), Harrison (1998), and Day and Liebowitz (1998)). The three mostly commonly cited omitted variables are net worth, the banks' assessment of whether the applicant met its

[2] The working paper was eventually published as Munnell *et al* (1996).
[3] Additional studies that examined individual-level rejection probabilities, but that did not control for as many factors as Munnell *et al* (1992) include Black, Schweitzer, and Mandell (1978); King (1980); Schafer and Ladd (1981); and Warner and Ingram (1987). LaCour-Little (1999) provides a comprehensive review of the fair lending literature.

credit standards, and whether the bank was able to verify information provided by the applicant. The second source consists of idiosyncratic factors that determine the underwriting decision, regardless of the applicant's creditworthiness (Horne (1994, 1997)). One common idiosyncratic factor is that the collateral did not meet policy guidelines. These factors can have a large impact on the model estimates, and are typically identified only by a thorough review of the application. The third source consists of bank specific policy information (Stengel and Glennon (1999)). Stengel and Glennon provide evidence suggesting it may be inappropriate to aggregate across banks when analyzing underwriting decisions, because banks consider unique sets of factors, and apply unique weights to these factors when making their underwriting decisions. For all three sources of omitted information suggested by critics of Munnell *et al*, when the additional information was incorporated, the effect of race was mitigated or eliminated.[4]

The work by Munnell *et al* and subsequent criticisms about omitted variables has shaped the OCC's current policies for using statistical models during fair lending examinations. The OCC employs a bank-specific approach that attempts to replicate the underwriting decision. The intuition behind this approach is to control for all economic factors the bank uses in its underwriting decision-making process, and then attribute any remaining systematic denial disparity to prohibited factor reasons. The first, and most important step in this process is to meet with the bank's underwriters and review the bank's underwriting policies. The goal is to identify all factors the bank considers in its decision-making process, and understand how these factors are used. The next step is data gathering and cleaning. Clearly, the model will only be as good as the data, so

[4] Browne and Tootell (1995) responds to many of the criticisms of Munnell *et al* (1992). Carr and Megbolugbe (1993) provide support for the robustness of the results in Munnell *et al* (1992) as well.

examiners conduct extensive data cleaning. The third step is to estimate an initial model and identify outliers for review.[5] The manual review of outliers is used to gather information potentially missed by the model, items such as fatal characteristics in the application, incorrect values in the data, and additional variables that should have been included in the original model. In addition, the review provides supporting evidence for the statistical model and identifies tangible examples when discrimination is found. The amount of information gathered during the manual review of outliers, and how that information is used, is related to the method for gathering the initial data and the extent of data cleaning. If data are gathered manually from the loan applications, less information is typically gained from the outlier review, because the examiners identify much of this information during data entry. In this case, information from the outlier review typically adds little to the model, and the outlier review is used more for supporting evidence than to improve the model. Alternatively, if a bank has extensive electronic data, more information is typically gained from the outlier review, because these electronic datasets rarely contain information on idiosyncratic factors. In this case, information from the outlier review is incorporated back into the data to improve the preliminary model. As the number of banks maintaining electronic data has increased, information gathered during the outlier review has become increasingly important to the reliability of the models. The final step in the process is to estimate a final model specification.[6] If no additional information is obtained during the outlier review, the preliminary and final

[5] The manual review of outliers is typically the first instance during an exam in which prohibited factors are considered as a possible explanation for an underwriting decision. Economists occasionally include prohibited factors in the preliminary model to help identify applications that have a large impact on the model estimates, but this is the exception rather than the rule.

[6] During some fair lending exams, information from an outlier review are incorporated back into the model, and additional outliers are identified and reviewed. This iterative process can continue until the model captures all additional information gained from reviewing outliers.

model specification will be identical. Conclusions about the role of race during the underwriting decision-making process are based on both the final statistical model results and the manual review of outliers.

III. Omitted Variable Bias: Theory

Using statistical models for fair lending exams provides an objective estimate of the pattern and practice of discrimination, and allows regulators to quantify the probability that the true racial effects are different from zero. Omitted variables diminish these benefits, because the racial estimate will capture a portion of the effects of these omitted variables. For the OLS estimator, it is well known that omitted variable bias is the product of the effect of the omitted variable on the outcome, and the correlation of the omitted variable and the included variable. Although we typically cannot explicitly calculate the bias, because we rarely have data for the omitted variable, we can generally determine the direction of bias based on theorized signs of the two component effects. For fair lending analyses, the omitted variables are typically measures of creditworthiness. With an assumption that minorities are generally less creditworthy, one could argue that the omitted variable bias in the racial estimate will typically be positive.[7] This is an important assumption, because if we cannot reject the null hypothesis of no racial effect using an under-specified model, then one could argue that we would not be able to reject the null hypothesis in a fully specified model either, because the racial

[7] Munnell *et al* (1996) presents evidence suggesting minorities are less creditworthy than whites. The authors summarize their findings on page 31 by stating, "As reported in other surveys, black and Hispanic applicants have considerably less net wealth, liquid assets, and income than whites and they have weaker credit histories."

estimate will only get smaller.[8] This is one of the main arguments for estimating an initial under-specified model for fair lending exams using only data available in electronic form. A second argument is a potential cost savings since additional data do not need to be gathered or cleaned. This savings could be particularly large if data need to be entered manually.

On theoretical grounds, the argument for using under-specified models is flawed, because it is based on results from the OLS estimator, and not the non-linear estimators that are typically used for fair lending analyses. Relative to the standard omitted variable bias result for the OLS estimator, little work has examined how omitted variables affect non-linear estimators. Lee (1980, 1982) are two studies that have looked specifically at these issues. Lee shows that for a multinomial logistic probability model where the omitted variable (*z*) is dichotomous and the included variable (*r*) is discrete, the coefficient on *r* in the under-specified model will be biased upward, if the effect of *z* on the dependent variable (*y*), and the correlation of *r* and *z* **conditional on** *y* have the same sign; biased downward, if the effect of *z* on *y* and the correlation of *r* and *z* **conditional on** *y* have opposite signs; and unbiased, if *z* has no effect on *y*, or *r* and *z* are independent **conditional on** *y*.

Table 1 compares Lee's findings of omitted variable bias for the non-linear logit estimator to those for the OLS estimator. To simplify the notation, both *y* and *r* in Lee's model are restricted to be dichotomous. For the OLS estimator, the effect of *z* on *y* (α_2), and the correlation of *r* and *z* **unconditional of** *y* are the two components that determine the direction of omitted variable bias. These two components, and the values they can take on, make up the vertical and horizontal dimensions of the table. The results in the

[8] This argument ignores the effects of omitted variable bias on the standard error estimate.

"bias in OLS estimator" columns convey the standard omitted variable bias findings for a linear estimator. The results in the "bias in logit estimator" columns convey the possible directions of bias in the non-linear logit estimator if the two components that determine the direction of omitted variable bias for the OLS estimator are used. These results show that estimates of the effect of *r* in an under-specified model may be biased downward,

Table 1: Direction of Omitted Variable Bias for OLS and Logit Estimators Using Correlations Unconditional of y.

Under-specified Model: $\ln \frac{P(y=1 \mid r,z)}{P(y=0 \mid r,z)} = \alpha_0 + (\alpha_1 * r)$

where, y = dependent variable (0/1)
r = included variable (0/1)
z = omitted variable (0/1)
α_2 = population parameter for z

	Corr (r,z) = 0		Corr (r,z) < 0		Corr (r,z) > 0	
	Bias in OLS Estimator	Bias in Logit Estimator	Bias in OLS Estimator	Bias in Logit Estimator	Bias in OLS Estimator	Bias in Logit Estimator
$\alpha_2 = 0$	None	None	None	None	None	None
$\alpha_2 < 0$	None	+ if $\alpha_1 < 0$ - if $\alpha_1 > 0$	+	- or +	-	- or +
$\alpha_2 > 0$	None	+ if $\alpha_1 < 0$ - if $\alpha_1 > 0$	-	- or +	+	- or +

even if the effect of *z* on *y* and the correlation between *r* and *z* **unconditional of** *y* are of the same sign; biased upward, even if the effect of *z* on *y* and the correlation between race and *z* **unconditional of** *y* are of opposite signs; and biased either upward or downward, if *r* and *z* are independent **unconditional of** *y*. For the latter case, the bias will be upward if the effect of *r* on *y* is negative and downward if the effect of *r* on *y* is positive. These findings suggest it is inappropriate to apply omitted variable results for linear estimators when theorizing the direction of omitted variable bias for the logit estimator.

The preceding analytic results are specific to a model with two independent variables, a discrete variable that is included and a dichotomous variable that is omitted. Fair lending models typically have many discrete, dichotomous and continuous independent variables, so application of these results is somewhat limited. Lee does provide two extensions to his results that may be useful for the typical fair lending models. First, the previous results are still valid if additional independent variables are included in the model, as long as the omitted variable z and these additional variables are independent conditional on r and y. Second, if the omitted variable z is continuous instead of dichotomous, the effect on the previous results depend on the distribution of z conditional on r and y. If z, conditional on y and r, is normally distributed, then omitting z from the multinomial logistic probability model will yield the same results as above.[9]

IV. Omitted Variable Bias

It is difficult to apply the analytic results from Lee (1980, 1982) to the fair lending models the OCC estimates, because the number of independent variables in the model and the number of omitted variables is considerably larger than the models Lee analyzed. Further, it is also impossible to make reliable generalizations about the correlations between omitted variables and race needed to determine the direction of omitted variable bias. As an alternative, this study uses results from 18 statistically modeled fair lending exams the OCC recently conducted to demonstrate how omitted variables would have

[9] All of these findings for the logit estimator do not necessarily hold for the probit estimator. See Yatchew and Griliches (1984) for omitted variable bias results specific to the probit estimator.

affected the results for those exams. Assuming the final model specification used for an exam reflects the true data generating process and is thus the fully specified model, this study estimates various under-specified models and compares the coefficient estimates and t-statistics for each racial variable in these models to their corresponding values from the fully specified models.[10] Loan-to-value ratio (LTV), debt-to-income ratio (DTI), and credit score variables, as well as all HMDA data, are assumed to be available during every fair-lending exam, so these variables were never omitted as long as the bank considered them in its decision-making process. For this study, these variables will be called the core variables, and the remaining variables in the fully specified models will be called the omittable variables. Using the omittable variables, under-specified models are created in two ways. First, only one variable is omitted at a time from the fully specified model. Second, the entire group of omittable variables is omitted at the same time.[11],[12] Information on idiosyncratic characteristics identified during a manual review of outliers is another potential source of omitted information that will affect the statistical model. These effects are discussed separately in Appendix A.

Four different sets of results measuring the effects of omitted variables are presented: 1) effects on model fit using Likelihood Ratio tests (LR); 2) effects on the magnitude of coefficient estimates; 3) effects on the magnitude of t-statistics; and, 4) tests

[10] Implicit in this assumption is that the model and error distribution specifications used during the fair lending exams are correct as well.

[11] Regressors with limited variation are another source of omitted variable bias that is a problem even though data are available for all relevant regressors. Typically, the small number of observations in which the decision is affected by such a factor are identified during the file review and eliminated from the sample.

[12] Occasionally, categorical variables were used in place of continuous measures during a fair lending exam. For example, instead of using a continuous measure or reserves, the economist constructing the model may have created four indicator variables measuring quartiles of reserves to capture potential non-linearities in the reserves effect. For this study, the reserves categorical variables would be treated as one variable, “Reserves,” when omitting variables to create the under-specified models.

of the null hypothesis that the omitted variable bias equals zero. All of these results are based on the final dataset and model specification used for the particular exam.

Model Fit

The first item of interest is to determine the statistical importance of the omittable variables for recent fair lending exams. Specifically, do these variables matter? To answer this question, this study conducts LR tests to examine how omitting variables affects models' goodness of fit. Table 2 presents the results for 18 statistically modeled fair lending exams the OCC recently conducted. Each test uses a 95 percent confidence level and tests the null hypothesis that the model fit does not deteriorate when omitting variables.[13] The first column of the table lists the exam number. The second column indicates the number of omittable variables used in the final model specification for the particular exam. Column 3 shows the number of times the null hypothesis of no deterioration of goodness of fit can be rejected when excluding the omittable variables separately with replacement. For example, for exam 12 there were eight omittable variables, and therefore eight under-specified models. For seven of these under-specified models, the LR test rejected the null hypothesis, suggesting that the model's fit worsened when each of these seven variables were omitted. Column 4 shows the LR test statistic when all omittable variables are excluded from the fully specified model simultaneously. Because only one under-specified model was estimated for each exam for column 4, the actual LR test statistic values are presented. Finally, the last column shows the LR test

[13] Although model fit is used to identify the optimal model specification, it is not the primary reason for including or excluding a particular variable. Because the OCC's modeling strategy is to replicate the underwriting decision-making process, the signs, and occasionally the magnitudes, of the coefficient estimates are just as, if not more, important. Therefore, the likelihood function value will not necessarily increase when omitting variables merely because of the modeling process used.

Table 2: Likelihood Ratio Tests (LR) of Under-Specified Models for Past Fair Lending Exams (95 percent confidence level)

Exam	# of omittable variables	LR test results from excluding one omittable variable at a time (# of rejections)	LR statistics from excluding all omittable variables at once	LR statistics from excluding core variables
1	4	3 of 4	116.484*	22.002*
2	4	3 of 4	88.857*	84.972*
3	2	1 of 2	7.860*	103.969*
4	8	7 of 8	339.107*	45.597*
5	4	3 of 4	25.673*	81.105*
6	7	2 of 7	59.298*	127.666*
7	2	2 of 2	57.214*	184.318*
8	3	3 of 3	43.856*	132.037*
9	8	5 of 8	41.509*	330.636*
10	9	4 of 9	178.075*	63.871*
11	8	4 of 8	154.107*	179.645*
12	8	7 of 8	137.746*	28.759*
13	4	3 of 4	44.418*	129.593*
14	2	2 of 2	26.770*	51.323*
15	7	4 of 7	78.393*	887.888*
16	14	7 of 14	100.621*	152.103*
17	10	3 of 10	40.717*	96.989*
18	8	5 of 8	393.326*	678.385*

* Indicates rejection of the null hypothesis that the under-specified model is not a worse fit than the fully specified model. The LR statistic is distributed as $\chi 2$ with degrees of freedom equal to the number of restrictions applied to the model.

statistic comparing a model with only the core variables and one with only a constant. This column thus indicates the statistical importance of the core variables, for comparison purposes. On the whole, excluding both the group of omittable variables and the group of core variables cause the model fit to deteriorate, suggesting these variables are statistically important. The null hypothesis of no deterioration can be rejected at the 95 percent confidence level for every exam, as shown in the table by the asterisks. Although the core variables appear to have the greatest impact on model fit, the omittable variables have a larger impact for five of the 18 exams. This is somewhat surprising, because the core variables are often believed to be the main drivers of the underwriting decision. Also, the results for the omittable variables are conditional on the core variables already being included in the model. The LR test results from excluding one omittable variable at a time are not as strong as those from dropping the omittable variables as a group or dropping the core variables. Of the 112 under-specified models created by excluding one omittable variable, the null hypothesis could be rejected in only 68 (60.7 percent) instances. This suggests that the addition of particular variables did not always improve the model's fit. This is not overly surprising, because the OCC places more emphasis on the bank's policies than on statistical importance when deciding whether to include variables in the model. Overall, the evidence from table 1 suggests that the variables included in the final model specifications for these 18 fair lending exams improved the fit of the model, and were therefore statistically important as well as theoretically important.

Coefficient Estimates

Table 3 shows how omitted variables would have affected the racial estimates for 18 statistically modeled fair lending exams the OCC recently conducted.[14] Using exam 12 as an example, the final model specification for that exam included three racial variables and eight variables this study defined as potentials for omission. Dropping each of these variables separately from the fully specified model yields eight under-specified models and a total of 24 racial estimates. Columns 4 and 5 show that 11 of these racial estimates were higher and 13 were lower than the corresponding estimates from the fully specified model. Omitting all eight variables at the same time yields a 9th under-specified model and three more racial estimates for a total of 27. Columns 6 and 7 show that two of these racial estimates were higher and one was lower than the corresponding estimates from the fully specified model.

The main result in table 3 is that the racial estimates are more likely to be higher for the under-specified models. Looking at the results in which only one variable is omitted at a time, 55.6 percent of the racial estimates are larger for the under-specified model than for the fully specified model. Only four exams — 5, 10, 11, and 12 — had fewer racial estimates that were larger for the under-specified model. The results from excluding all omittable variables at once are even stronger. The racial estimates from these under-specified models are larger than the corresponding racial estimates from the fully specified model almost 75 percent of the time. These findings provide some support for the argument that excluding such measures tend to impose an upward bias on the racial estimate. However, the percentages of times the racial estimates decreased in

[14] Non-proportional choice-based sampling was used for 16 of the 18 exams used in this study. All of the coefficient and t-statistic results presented are appropriately adjusted to account for bias this sampling approach introduces into the racial estimates (see Dietrich (2001)).

Table 3: Effects of Omitted Variables on Racial Coefficient Estimates from Past Fair Lending Exams

			Only one omittable variable excluded at a time		All omittable variables excluded at the same time	
Exam	# of races*	# of omittable variables	**Under-specified model estimate > fully specified model estimate**	**Under-specified model estimate < fully specified model estimate**	**Under-specified model estimate > fully specified model estimate**	**Under-specified model estimate < fully specified model estimate**
1	2	4	5	3	2	0
2	2	4	5	3	2	0
3	1	2	1	1	1	0
4	1	8	6	2	1	0
5	2	4	3	5	0	2
6	1	7	5	2	1	0
7	2	2	2	2	1	1
8	2	3	3	3	0	2
9	3	8	15	9	3	0
10	1	9	3	6	1	0
11	2	8	7	9	1	1
12	3	8	11	13	2	1
13	2	4	4	4	1	1
14	2	2	2	2	2	0
15	1	7	5	2	1	0
16	2	14	17	11	2	0
17	1	10	7	3	1	0
18	1	8	4	4	1	0
Total	31	112	105 (55.6%)	84 (44.4%)	23 (74.2%)	8 (25.8%)

* Column does not include the excluded race variable in the count.

response to omitting variables (44.4 percent and 25.8 percent) are too high to make this generalization during future fair lending exams.

T-statistic Estimates

In addition to affecting coefficient estimates, omitted variables also affect standard error estimates. As a result, it is not enough to look only at changes in coefficient estimates to determine how omitted variables would affect conclusions drawn for an exam; the effects on t-statistics need to be examined as well. Using data from the 18 statistically modeled exams, table 4 presents the effects omitted variables would have had on the t-statistics used to test the null hypothesis that the population racial parameter equals 0. The format is similar to table 3. Using exam 12 again as an example, the final model specification for that exam included three racial coefficients and eight variables this study defined as potentials for omission. Dropping each of these variables separately from the fully specified model yields eight under-specified models and a total 24 racial estimates. Columns 4 and 5 show that 15 of these racial estimates were higher and nine lower than the corresponding estimates from the fully specified model. Omitting all eight variables at the same time yields a 9th under-specified model and three more racial estimates for a total of 27. Columns 6 and 7 show that all three of these racial estimates were higher than the corresponding estimates from the fully specified model. As this example illustrates, the results for the coefficient estimates and t-statistics, while generally similar, do differ on occasion.

Similar to the coefficient results from table 3, the t-statistics are more likely to be higher for the under-specified models. Looking at the results in which only one variable

Table 4: Effects of Omitted Variables on t-statistic Estimates from Past Fair Lending Exams						
			Only one omittable variable excluded at a time		All omittable variables excluded at the same time	
Exam	# of races*	# of omittable variables	**Under-specified model estimate > fully specified model estimate**	**Under-specified model estimate < fully specified model estimate**	**Under-specified model estimate > fully specified model estimate**	**Under-specified model estimate < fully specified model estimate**
1	2	4	7	1	2	0
2	2	4	5	3	2	0
3	1	2	1	1	1	0
4	1	8	7	1	1	0
5	2	4	3	5	1	1
6	1	7	2	5	1	0
7	2	2	3	1	1	1
8	2	3	4	2	1	1
9	3	8	9	15	2	1
10	1	9	5	4	1	0
11	2	8	10	6	1	1
12	3	8	15	9	3	0
13	2	4	4	4	1	1
14	2	2	2	2	1	1
15	1	7	2	5	1	0
16	2	14	12	16	2	0
17	1	10	9	1	1	0
18	1	8	4	4	1	0
Total	31	112	104 (55.0%)	85 (45.0%)	24 (77.4%)	7 (22.6%)

* Column does not include the excluded race variable in the count.

is omitted at a time, 55.0 percent of the t-statistics are larger for the under-specified model than for the fully specified model. Only five exams — 5, 6, 9, 15, and 16 — had fewer t-statistics that were larger for the under-specified model. The results from omitting all omittable variables at once are even stronger. The t-statistics from these under-specified models are larger than the corresponding racial estimate from the fully specified model 77.4 percent of the time. Overall, the results again provide some support for the argument that a statistically insignificant race estimate in an under-specified model will be statistically insignificant in a fully specified model. However, the percentages of times the t-statistic estimates decreased in response to omitting variables (45.0 percent and 22.6 percent) are too high to make this generalization for fair lending examination purposes.

Tests of Hypothesis that Omitted Variable Bias = 0

All of the coefficient estimates and t-statistics discussed in tables 3 and 4 are sample estimates of the true population parameters. As such, each estimate is merely one data point somewhere within the sampling distribution of the particular estimator. When comparing magnitudes of sample estimates, as was done in tables 3 and 4, the difference between the under-specified and fully specified estimates can be either positive or negative regardless of whether the true omitted variable bias is positive or negative. The direction of the difference depends on the shape and centrality of the sampling distributions, as well as where the two sample estimates fall within the sampling distributions. For example, suppose the true omitted variable bias is positive, so the sampling distribution of the estimator for the under-specified model is shifted right of the sampling distribution of the estimator for the fully specified model. If the estimate from the under-specified model is in the left tail of the sampling distribution and the estimate from the fully specified model is in the right tail, the under-specified model estimate may be lower even though the bias is positive. Therefore, a more appropriate test of the effects of omitted variables than comparing the magnitudes of sample estimates is to test for differences in the means of the sampling distributions of the estimators for the under- and fully specified models. The mean of the sampling distribution for the under-specified model estimator is merely the mean of the sampling distribution for the fully specified model estimator plus an omitted variable bias term. Therefore, this is merely a test of whether the omitted variable bias equals zero.

This study uses a t-test to test the hypothesis that omitted variable bias equals zero against the alternative hypotheses that the bias is positive and negative. One shortcoming of this approach is that it assumes the sampling distributions of the two estimators are independent, which is clearly not the case. However, it is reasonable to assume that, if the racial estimate from the fully specified model is large, the racial estimate from the under-specified model will be large as well. In other words, the two estimators are positively correlated. If this assumption is correct, the standard deviation of the difference between the two coefficient estimates will be over-stated if this correlation is not accounted for. Therefore, using the t-test, the percentage of instances when the null hypothesis of no omitted variable bias is rejected in favor of the alternative hypotheses of either positive or negative bias will provide only a lower bound of the true rejection percentages.

Table 5 presents the results from these t-tests, using a 95 percent confidence level, for the 18 fair lending exams. Using exam 12 as an example once again, columns 1-3 are identical to those in tables 3 and 4. Columns 4-6 present the t-test results using under-specified models created by dropping one omittable variable at a time. The null hypothesis was rejected in favor of the alternative hypothesis of positive bias for 10 of 24 estimates. Similarly, the null hypothesis was rejected in favor of the alternative hypothesis of negative bias in 10 of 24 estimates as well. For the remaining four estimates, the null hypothesis could not be rejected. Columns 7-9 present the t-test results using under-specified models created by omitting all omittable variables at once. The null hypothesis was rejected in favor of the alternative hypothesis of positive bias for two of three estimates and in favor of the alternative hypothesis of negative bias for the remaining estimate.

Similar to the results for coefficients and t-statistics previously presented, the results in table 5 suggest that the omitted variable bias is non-negative. The null hypothesis of no bias was rejected at the 95 percent confidence level in favor of the alternative hypothesis of negative bias in only 24.3 percent of the under-specified models created by dropping single omittable variables and 22.6 percent of the under-specified models created by dropping all omittable variables. Rejecting the null hypothesis in favor of the alternative hypothesis of negative bias was the most prevalent outcome for only two exams (5 and 8). For more than half of the tests using under-specified models created by dropping one omittable variable, the null hypothesis of no bias could not be rejected at the 95 percent confidence level. This compares to only 6.5 percent for the tests using under-specified models by dropping all omittable variables. As previously

Table 5: T-tests of the Null Hypothesis that Omitted Variable Bias Equals Zero Using a 95 Percent Confidence Level								
			Only one omittable variable excluded at a time			All omittable variables excluded at the same time		
Exam	# of races*	# of omittable variables	H_a: bias > 0	H_a: bias < 0	H_0 could not be rejected	H_a: bias > 0	H_a: bias < 0	H_0 could not be rejected
1	2	4	4	1	3	2	0	0
2	2	4	4	1	3	2	0	0
3	1	2	1	0	1	1	0	0
4	1	8	5	1	2	1	0	0
5	2	4	1	5	2	0	2	0
6	1	7	2	0	5	1	0	0
7	2	2	2	1	1	1	1	0
8	2	3	2	3	1	0	1	1
9	3	8	8	5	11	2	0	1
10	1	9	3	3	3	1	0	0
11	2	8	4	6	6	1	1	0
12	3	8	10	10	4	2	1	0
13	2	4	3	3	2	1	1	0
14	2	2	2	2	0	2	0	0
15	1	7	4	1	2	1	0	0
16	2	14	10	1	17	2	0	0
17	1	10	4	1	5	1	0	0
18	1	8	3	2	3	1	0	0
Total	31	112	72 (38.1%)	46 (24.3%)	71 (51.9%)	22 (71.0%)	7 (22.6%)	2 (6.5%)
* Column does not include the excluded race variable in the count.								

noted, however, both of these percentages are likely inflated given that the test used does not account for correlation between the estimators. Although these results, again, generally support the argument that omitted variable bias will be positive, the results are not consistent enough to apply during fair lending exams.

The effects of each potential omitted variable were examined across exams to determine if generalizations could be made about the direction of bias in the race estimate if a particular variable was omitted. For example, number of mortgage delinquencies was used in six of the 18 models examined in this study. Among the under-specified models created by dropping this variable, 53 percent of the race estimates and 46 percent of the t-statistics were higher than the corresponding estimates from the fully specified models. Clearly, if regulators knew that a mortgage delinquency variable was omitted from the model, they could not determine the expected direction of bias in the race estimates with much certainty. Overall, of the variables that were used during at least two exams, only one showed the same effect on the race estimates for all under-specified models — a positive bias for compensating factors —, and only one showed the same effect on the t-statistic estimate for all under-specified models – a positive bias for employment verification. These results suggest that generalizations about the direction of omitted variable bias cannot be made with much certainty about specific variables.

V. Outlier Analysis

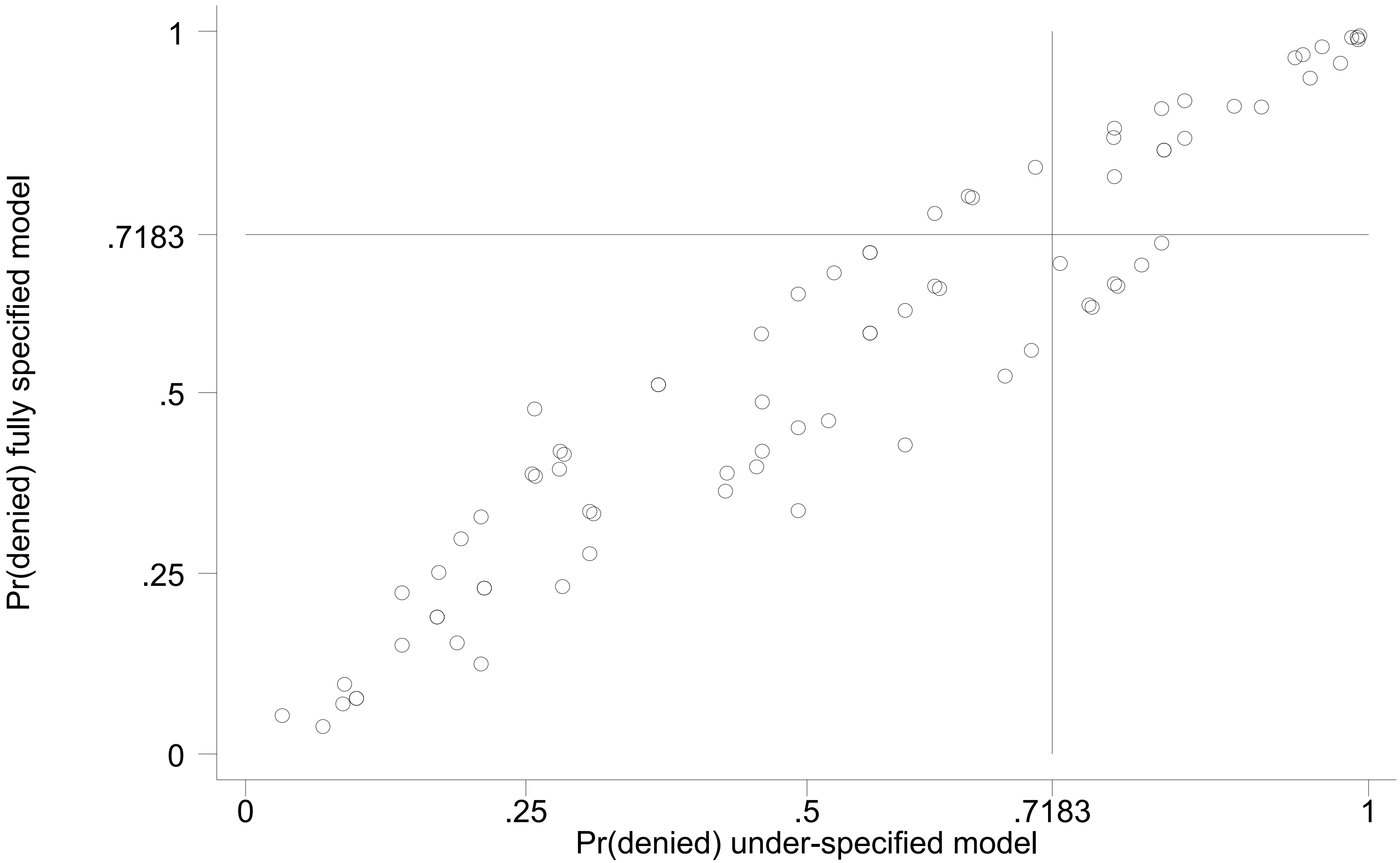
In addition to providing an estimate of pattern and practice of disparate treatment, the models the OCC estimates also identify outliers for manual review. Similar to a credit score, statistical models provide a predicted probability of denial for each applicant, which rank orders the applicants' risk of default based on the factors included in the model. Applicants with a high predicted probability of denial that were approved, and applicants with a low predicted probability of denial that were denied are questionable applications that need to be reviewed manually.[15] To identify outliers, some cutoff probability first needs to be specified. A common approach is to use the actual approval rate for the population, because from the bank's perspective, the applicants near this cutoff represent the marginal applicants. During fair lending exams, denied applications are typically over-sampled, so using the population approval rate as the cutoff value will overstate the number of denied outliers and understate the number of approved outliers. As a result, this study uses the sample approval rate as the cutoff for each exam.[16]

A fully specified model will identify accurately all of the questionable applications that require an in-depth manual review. The question examined here is whether an under-specified model will lead to the same rank ordering of risk and identify the same set of outliers. There are two potential errors that can occur when using an under-specified model to identify outliers, inappropriate outliers can be identified, and appropriate outliers can be missed. The first error leads to inefficient resource expenditures as applications are reviewed that do not need to be reviewed. This error will not affect the fair lending conclusions. The second error will affect the fair lending

[15] In addition, standardized differences in regression coefficients due to dropping individual observations can also be used to identify those outliers where race appears to have had a large impact in the decision.
[16] A cutoff of .5 is commonly used for fair lending exams as well.

conclusions, however, because questionable applications that should be reviewed manually are not reviewed. Graph 1 depicts these two errors for denied outliers using data from exam 4. The graph presents a scatter plot of the predicted denial probabilities for denied applicants from the fully specified model versus the predicted denial probabilities for denied applicants from the under-specified model created by omitting all omittable variables. The fully specified results are plotted on the vertical axis. The sample approval rate for this exam, 71.83 percent, is used as the cutoff to identify outliers. The denied outliers for the fully specified model will be those applications with a low predicted denial probability, i.e., those below the 71.83 percent line. Looking now at the predicted denial probabilities for the under-specified model, the denied outliers will be those to the left of the 71.83 percent line. Therefore, applications in the upper left quadrant would have been reviewed manually using the under-specified model even though they were not questionable. This measures the degree of inefficiency from using an under-specified model. Applications in the lower right quadrant would not have been manually reviewed using the under-specified model even though they were questionable. This measures the degree of error in the fair lending conclusion from using an under-specified model. An analogous graph could also be created for approved outliers. With such a graph, applications with predicted denial probabilities above and to the right of the cutoff lines for the fully- and under-specified models, respectively, would constitute outliers. The upper left quadrant would now contain the questionable applications that would have been missed had an under-specified model been used and the lower right

Graph1: Denied Outliers From Fully and Under-specified Models



quadrant would contain the applications that would have been reviewed manually even though they were not questionable.

Table 6 presents these results for all 18 exams analyzed in this study. Similar to the omitted variable bias results stated previously, all of these results are based on the final dataset and model specification used for the particular exam. Race is not included in any of the estimations, and sample approval rates for each exam are used as cutoffs in determining the outliers. The first column of the table lists the exam number. Columns 2 and 3 show the number of outliers identified using the fully specified model and the under-specified model created by dropping all omittable variables, respectively. Column 4 contains the number of outliers in the fully specified model that were not outliers in the under-specified models. The results are also shown as a percentage of the outliers from the fully specified model to convey more effectively the magnitude of questionable applications that would have been missed had the under-specified model been used. Column 5 contains the number of outliers in the under-specified model that were not outliers in the fully specified model. The results are also shown as a percentage of the applications that were not outliers from the fully specified model to convey more effectively the magnitude of applications that would have been reviewed unnecessarily had the under-specified model been used. Finally, columns 6 and 7 show the number of applications that were outliers in both models and the number that were not outliers in either model, respectively.

Table 6: Outlier Comparison Between Fully and Under-specified Models (Sample Approval Rate Is Used as the Cutoff)						
Exam	Total outliers for fully specified model	Total outliers for under-specified model	Applications that were outliers in the fully specified model but not under-specified model[1]	Applications that were outliers in the under-specified model but not the fully specified model[2]	Applications that were outliers in both models	Applications that were not outliers in either model
1	49	82	20 (40.8%)	53 (25.2%)	29	157
2	69	100	21 (30.4%)	52 (17.7%)	48	241
3	62	54	12 (19.4%)	4 (1.8%)	50	218
4	91	144	5 (5.5%)	58 (9.3%)	86	567
5	41	51	4 (9.8%)	14 (7.3%)	37	177
6	43	66	3 (7.0%)	26 (8.6%)	40	276
7	64	81	11 (17.2%)	28 (7.8%)	53	333
8	57	76	8 (14.0%)	27 (10.0%)	49	242
9	59	68	10 (16.9%)	19 (4.3%)	49	418
10	61	125	10 (16.4%)	74 (22.4%)	51	257
11	60	112	21 (35.0%)	73 (18.3%)	39	325
12	103	159	25 (24.3%)	81 (22.6%)	78	277
13	57	59	13 (22.8%)	15 (6.0%)	44	235
14	55	63	5 (9.1%)	13 (7.3%)	50	165
15	201	214	22 (10.9%)	35 (2.5%)	179	1377
16	55	87	11 (20.0%)	43 (13.4%)	44	278
17	26	42	5 (19.2%)	21 (12.4%)	21	149
18	128	217	48 (37.5%)	137 (12.6%)	80	947
Total	1281	1800	254(19.8%)	773 (10.4%)	1027	6639

1 The denominator for the percentages is the number of outliers in the fully specified model.
2 The denominator for the percentages is the number of applications that were not outliers in the fully specified model.

The number of outliers is larger for the under-specified model for every exam, except exam 3. This is expected because fewer relevant factors are included in the model to explain variation in the dependent variable. The percentage of outliers from the fully specified model that would not have been reviewed manually had the under-specified model been used ranges from 5.5 for exam 4 to 40.8 for exam 1. The average across all exams is 19.8 percent. The number of outliers that would have been manually reviewed had the under-specified model been used, but not reviewed, if the fully specified model had been used ranges from 1.8 percent for exam 3 to 25.2 percent for exam 18. Across all exams, 773 applications (10.4 percent) would have been unnecessarily reviewed, an average of nearly 43 outliers per exam. In summary, using an under-specified model instead of a fully specified model would lead to missing an average of 19 percent of the questionable applications that should be reviewed and reviewing an average of 43 applications per exam that were not questionable and therefore that should not have been reviewed. The uncertainty caused by the first and the resource expenditures caused by the second provide further support for the use of fully specified models during fair lending exams.

VI. Conclusion

A long-standing criticism of studies testing for discrimination in mortgage lending is that omitted variables introduce an upward bias into estimates of the effect of race on the underwriting decision, thereby creating the illusion of statistical evidence of discrimination. With access to bank policies, underwriters, and all data that underwriters report that they use in their decision-making process, regulators are in a unique position to diminish greatly problems of omitted variable bias. Access to this information also allows regulators to assess the impact of omitted variables on fair lending models.

Using data from 18 statistically modeled fair lending exams the OCC recently conducted, this study examined the effects of omitting variables available to regulators, but often unavailable to researchers. There were two main findings. First, omitted variables introduced bias into the racial estimates, but the direction of bias was not consistently positive as commonly thought. The null hypothesis of no bias was rejected at the 95 percent confidence level in favor of the alternative hypothesis of negative bias in 24.3 percent of the under-specified models created by dropping single omittable variables and 22.6 percent of the under-specified models created by dropping all omittable variables. Although omitted variables were considerably more likely to introduce no bias or positive bias, the results are not strong enough to be able to make these generalizations with a high degree of certainty. Second, omitted variables have an important effect on the identification of outliers to review manually. Estimating an under-specified model instead of a fully specified model resulted in missing an average of 19 percent of the applications that should have been reviewed, and reviewing an average of 43 applications per exam that should not have been reviewed. These results provide some indication of the inefficient use of resources and the uncertainty introduced into conclusions from estimating an under-specified model.

Overall, the findings in this study suggest that variables often omitted by researchers do have an important impact on both the estimate of the effect of race and on the identification of outliers to review. They also show that it is dangerous to make generalizations about the potential direction of bias based on assumptions about the correlations between omitted variables and race. From a regulatory perspective, although estimating fully specified models may increase resource costs during the exam, especially if data for numerous factors relevant to the underwriting decision must be entered manually, estimating fully specified models is the best strategy to take to ensure the most reliable conclusions.

VII. References

Black, Harold A., Robert Schweitzer, and Lewis Mandell. 1978. “Discrimination in Mortgage Lending,” *American Economic Review*. 68(2):186-91.

Browne, Lynn E., and Geoffrey Tootell. 1995. “Mortgage Lending in Boston: A Response to the Critics,” *New England Economic Review*, Sept./Oct.:53-78.

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.

Day, Theodore E., and S.J. Liebowitz. 1998. “Mortgage Lending to Minorities: Where’s the Bias?,” *Economic Inquiry*. 36(1): 3-28.

Dietrich, Jason. 2001. “The Effects of Choice-based Sampling and Small-sample Bias on Past Fair Lending Exams,” Economic and Policy Analysis Working Paper 2001-2, Office of the Comptroller of the Currency.

Harrison, Glenn W. 1998. “Mortgage Lending in Boston: A Reconsideration of the Evidence,” *Economic Inquiry* 36(1):29-38.

Horne, David K. 1994. “Evaluating the Role of Race in Mortgage Lending,” *FDIC Banking Review*. 7(1):1-15.

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

King, Michael A. 1980. “An Econometric Model of Tenure Choice and Demand for Housing as a Joint Decision,” *Journal of Public Economics*. 14:137-59.

LaCour-Little, Michael. 1999. “Discrimination in Mortgage Lending: A Critical Review of the Literature,” *Journal of Real Estate Literature*. 7:15-49.

Leibowitz, S. 1993. “A Study That Deserves No Credit,” *Wall Street Journal*. September 1, A1.

Lee, Lung-Fei. 1980. "Specification Error in Multinomial Logit Models: Analysis of the Omitted Variable Bias," Discussion Paper No. 80-131, Center for Economic Research, Department of Economics, University of Minnesota.

Lee, Lung-Fei. 1982. "Specification Error in Multinomial Logit Models: Analysis of the Omitted Variable Bias,” *Journal of Econometrics*. 20:197-209.

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.

Munnell, Alicia H., Lynn E. Browne, James McEneaney, and Geoffrey M.B. Tootell. 1996. "Mortgage Lending in Boston: Interpreting HMDA Data," *American Economic Review*. 86(1):25-53.

Schafer, Robert, and Helen F. Ladd. 1981. Discrimination in Mortgage Lending. Cambridge, MA: MIT Press.

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

Warner, Alfred, and James Ingram. 1987. "A Test of Discrimination in a Mortgage Market," *Journal of Bank Research*. 13:116-24.

Yatchew, Adonis, and Zvi Griliches. 1984. "Specification Error in Probit Models," *The Review of Economics and Statistics*. 66:134-39.

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

Appendix A: Effects of Information Obtained During Manual Review of Outliers.

All of the results presented in tables 1-6 showed the impact of omitting variables from the final model specification used for the 18 fair lending exams. Data gathered during the outlier review constitutes a second source of information that may affect the statistical model, but that is often omitted from fair lending analyses. As noted earlier, for some fair lending exams, information gathered during the outlier review was incorporated back into the dataset prior to estimating the final model specification. For these exams — 3, 6, 11, 13, 16, 17, and 18 — the necessary data is therefore available to examine the effects of this second source of omitted information. This analysis addresses the criticism of Munnell *et al* (1992) that idiosyncratic factors unaccounted for in the model affected their results. These factors, which can have a large impact on the model estimates, are typically found only during a manual review of applications, and Munnell *et al* did not have direct access to the applications during their analysis.

Table A1 shows the effects information obtained during the manual review of outliers had on the statistical analysis for seven of the 18 exams used in this study. Columns 1 and 2 show the exam number and number of race variables included in the model for each exam, respectively. Column 3 shows the LR test statistic comparing the fully specified model incorporating information from the outliers, and the fully specified model not incorporating information from the outliers. Columns 4-6 are similar to those in table 6, conveying t-test results of testing the null hypothesis of no omitted information bias against alternative hypotheses of positive or negative bias. Other than exam 11,

Table A1: Effects of Omitted Information on Racial Estimates from Past Fair Lending Exams					
			Omitted Variable Bias due to Exclusion of Information Gained During Outlier Review		
Exam	# of races	LR statistics from excluding information from outlier review	H_a: bias > 0	H_a: bias < 0	H_0 could not be rejected
3	1	36.511*	1	0	0
6	1	96.928*	1	0	0
11	2	-45.532*	1	1	0
13	2	66.385*	2	0	0
16	2	78.650*	1	1	0
17	1	46.186*	1	0	0
18	1	689.697*	1	0	0
Total	10		8 (80%)	2 (20%)	0 (0%)

* Indicates rejection of the null hypothesis that the under-specified model is not a worse fit than the fully specified model. The LR statistic is distributed as $\chi 2$ with degrees of freedom equal to the number of restrictions applied to the model.

incorporating information gathered during the outlier review greatly improved the fit of the model. The null hypothesis of no deterioration can be rejected at the 95 percent confidence level for every exam, as shown in the table by the asterisks. Exam 18 showed the largest effect, an increase of nearly 700 points in the LR test statistic when information gained during the outlier review was excluded. These large effects are not surprising, because the primary modifications resulting from the outlier review are elimination of applications with idiosyncratic characteristics. The result for exam 11 is different than the other six exams primarily because a data entry problem was identified during the review of outliers that increased the number of useable observations by 39. This is not typical of the type of adjustments made using information from the outlier review. The hypothesis test results also suggest that information gathered during the outlier review is important with the null hypothesis of no bias being rejected in each of the 10 tests. In 8 of the 10 cases, the bias is positive, so the estimated racial effects decline when this additional information is incorporated. This is considerably higher than the corresponding percentages in table 6, highlighting the importance of information gained during the outlier review.