NEW EVIDENCE ON THE RELATIONSHIP BETWEEN RACE AND MORTGAGE DEFAULT: THE IMPORTANCE OF CREDIT HISTORY DATA

by

Robert F. Cotterman
Unicon Research Corporation
1640 Fifth Street, Suite 100
Santa Monica, CA 90401

Prepared for
Office of Policy Development and Research
U.S. Department of Housing and Urban Development
451 Seventh Street, S.W.
Washington, D.C. 20410-3000

May 23, 2002

This work was funded under by the Department of Housing and Urban Development under Contract DU100C000018484. The author is grateful to the HUD Policy Development &Research staff for help in obtaining and understanding the data, as well as for offering comments and suggestions. Thanks are also due a reviewer of an earlier draft.

SECTION 1

INTRODUCTION

The traditional method for testing for lender discrimination has involved the estimation of a mortgage rejection equation to determine whether there is an independent race effect, after controlling for credit risk factors that lenders typically consider when underwriting a loan. In the early 1990s, the now widely cited Boston Federal Reserve study (described below) used this methodology to provide what many consider to be convincing evidence of discrimination against black and Hispanic mortgage applicants in the Boston metropolitan area. In the mid-1990s, a series of papers by Berkovec, Canner, Gabriel, and Hannan (BCGH) offered an alternative model to test for lender discrimination based on the performance of the mortgage following origination. In their model, lenders discriminate by holding minorities to higher credit standards, a practice that suggests minorities would have lower default probabilities than non-minorities for given values of other default-related factors. BCGH argue that their empirical results, which show higher Federal Housing Administration (FHA) default probabilities for blacks, can be interpreted as evidence against mortgage discrimination – a conclusion inconsistent with that of the Boston Fed study. As explained in Section 2 below, BCGH's empirical findings as well as their default-based methodology for testing discrimination have been hotly debated among economists and others in the fair lending field.

One problem with the BCGH work was that their estimated default equation did not control for the credit history of the FHA borrower. Recent studies have found credit history (*e.g.*, past record of making monthly payments on time) to be an important determinant of mortgage default that is also highly correlated with race, with minorities, on average, exhibiting below-average past credit records. Thus, BCGH's finding that black borrowers had higher FHA default probabilities could be traceable to their not controlling for the effects of credit history.[1]

The limited twofold purpose of this study is to report the findings obtained by including a measure of borrower credit history in a model of FHA defaults that is similar to the BCGH default model and to demonstrate the bias attributable to omitting such data. As shown in Section 4, controlling for credit history substantially reduces the estimated coefficient for black borrowers, often rendering it insignificant (and in some cases even turns the effect negative). Thus, the findings from this study contrast sharply with those of the BCGH study, which consistently found a large, positive, and statistically significant estimate of the black/white differential in (the log odds of) default. Even for those who believe that studies like this can reveal lending discrimination, the results here do not consistently favor a position either supporting or rejecting discrimination in mortgage lending.[2]

Those who were convinced by the arguments presented by BCGH should find the current results of special interest. In addition, those who believed that the only important defect in the

[1] As explained in Section 2, BCGH acknowledged this limitation and in a second study (BCGH, 1994) attempted to assess the overstatement of the black coefficient resulting from this omitted variable bias.

[2] As noted in Section 2, the omitted variable (credit history) problem that is addressed in this paper is only one of the criticisms that have been made of the BCGH model. Other observers have argued that there are fundamental flaws in using mortgage performance to test for lender bias, and indeed some have argued that there are serious problems with virtually all existing empirical studies of mortgage default. Within the context of this controversy, this paper has a more limited purpose – simply to demonstrate that empirical findings here, and by extension, presumably those reported by BCGH, are heavily influenced by the inclusion or exclusion of credit history controls.

BCGH analysis was the omission of credit history controls may also find the current analysis interesting, though there will inevitably be questions of whether the characterization of credit history adopted here is adequate. We also introduce additional controls for post-origination changes in housing markets, but, as explained below, these controls are not ideal. Given the limited purpose of this study, we make no attempt to modify the fundamental econometric framework used by BCGH, nor do we rehash the arguments for or against that framework. Those who have questioned the appropriateness of the BCGH analytic procedure will find similar defects here.

The outline of the remainder of this paper is as follows. Section 2 describes the rationale for the default model as a test of lender bias and summarizes BCGH's major empirical findings. The issue of omitted variables is also briefly discussed. Section 3 discusses the data used as the basis for the empirical findings. The estimation samples consist of FHA-insured loans from three application years (1992, 1994, and 1996), together with an assortment of explanatory variables derived from FHA and other sources. Section 4 presents empirical findings, emphasizing the contrast between models that control for credit history and those that do not. Finally, Section 5 offers a few conclusions.

SECTION 2

MOTIVATION AND BACKGROUND: THE BCGH DEFAULT MODEL

Several studies of mortgage lending activity have documented large and persistent racial disparities, particularly in the approval and rejection of mortgage applications.[3] The most comprehensive of these studies was conducted by four economists at the Boston Federal Reserve Bank (Munnell, Browne, McEneaney, and Tootell, 1992) who analyzed mortgage rejection data for a sample of lenders in the Boston area. The authors found that minority applicants were 60 percent more likely to be denied a loan, after controlling for the credit characteristics of the applicant. What distinguished this so-called "Boston Fed study" from other efforts to test for discrimination in the mortgage market was that it controlled for a wide range of underwriting variables that lenders say they consider when underwriting a loan, including, for example, the loan-to-value ratio, product characteristics (term, fixed rate), individual characteristics (age, gender, marital status), years of job experience and tenure on current job, the monthly payment-to-income ratio, and the past credit history of the applicant. Because this study finds significant race effects despite the inclusion of a wide variety of underwriting factors, the Boston Fed study has been widely cited in the economics literature and popular press as providing evidence of lender bias in the mortgage market.[4]

In a series of articles appearing after the Boston Fed study, Berkovec, Canner, Gabriel, and Hannan (BCGH, 1994, 1996, 1998) proposed using a mortgage default estimation scheme to test for lender bias as an alternative to the traditional mortgage rejection test for discrimination. The concept behind the BCGH default model is simple and appealing. BCGH assume that discriminating lenders first rank applicants by their creditworthiness and then apply a higher standard of creditworthiness to minority borrowers. The result is that the marginal (lowest ranking) minority borrower who is approved would be more qualified than the marginal (lowest ranking) white borrower who is approved. One would thus expect to observe lower levels of default[5] for marginally qualified minority borrowers than for marginally qualified white borrowers. Empirical data on mortgage defaults could be used to test for lender bias --- a significant, negative minority race effect in an estimated mortgage default equation that controlled for underwriting factors (and other factors related to mortgage default) would be consistent with lender bias while a significant, positive minority race effect would be inconsistent with lender bias. That is, holding constant other observable factors, an estimated negative minority race effect on mortgage defaults suggests that minorities on the margin are

[3] For examples of mortgage rejection models, see in particular Munnell *et al.* (1992, 1996) and Hunter and Walker (1996). A different technique, paired testing at the pre-application stage, has also recently been used to test for lender bias; see The Urban Institute (1998) for an analysis of paired testing.

[4] Not all are convinced by the Boston Fed study. Numerous studies have been written reacting to the Boston Fed's methodology and data analysis, and the authors of the Boston Fed study have responded to the various critiques. For discussion of the various issues, see Browne and Tootell (1995), Carr and Megbollugbe (1993), LaCour-Little (1998), Rachlis and Yezer (1993), Ross and Yinger (1998), Tootell (1993), Van Order and Zorn (1995), Yezer and Trost (1994), and Yinger (1996, 1998).

[5] Although the model can be phrased in terms of equating (on the margin) actual white default probabilities with actual black default probabilities less a discrimination premium (*i.e*, blacks must offer lower default probabilities to compensate discriminatory lenders), equilibration could occur in some other metric, such as expected losses (also examined in BCGH, 1994).

more qualified, which is consistent with lender bias, while an estimated positive minority race effect suggests that minority borrowers on the margin are less qualified, which is inconsistent with lender bias.[6]

Naturally, these conclusions hinge on variety of other conditions being satisfied. Among these is the assumption that there exists at least one default-related factor observed by underwriters but not by analysts. Discriminatory underwriters demand that black applicants with identical values of observable factors have (on net) superior values of these unobserved factors to compensate for their race.[7] As a consequence, black borrowers will be found to default less frequently than observationally equivalent (as viewed by analysts) white borrowers. Absent such unobservables, however, noneconomic discrimination would not show up as a race differential in estimated default models, but it would in principle show up as, say, racial differences in the upper limit of predicted default probabilities. That is, the least qualified (but still acceptable to underwriters) black borrower would be a better risk than the least qualified (but still acceptable to underwriters) white borrower.

BCGH estimate their model using Federal Housing Administration (FHA) data for single-family mortgages that were originated during 1987-89. FHA also provided BCGH with performance information (*i.e.*, whether or not the loan had resulted in a foreclosure) and with numerous characteristics of the loan. Section 3 below provides a list of specific variables BCGH included in their default model. As indicated there, BCGH included in their default equation the types of variables that had appeared in previous studies of mortgage default, including characteristics of the property (*e.g.*, house value), loan (*e.g.*, loan-to-value ratio, front and back-end ratios), and borrower (*e.g.*, age, first-time borrower status), as well as characteristics of the neighborhood in which the property was located (*e.g.*, median income of families in the census tract). To test for discrimination, BCGH also included as separate explanatory variables the race and ethnicity of the borrower (Black, Hispanic, Asian American, American Indian).[8]

The main result of the BCGH model concerns the coefficient for black borrowers, which BCGH found to be positive and highly significant for each book of FHA business that they analyzed (1987, 1998, and 1989). BCGH characterized their results as follows:

> Results of the analysis fail to find evidence of better performance on loans granted to minority borrowers. Indeed, black borrowers are found, all else being equal, to exhibit a higher likelihood of mortgage default than other borrowers. These findings argue against allegations of substantial levels of bias in mortgage lending. (BCGH, 1996, page 9)

---

[6] Blacks may be able to escape the effects of discrimination if there are enough nondiscriminatory lenders. Empirical analysis of the kind described above will show, at best, only the bias among lenders making loans to blacks.

[7] Actions of discriminatory lenders will ensure that the unobservable factor(s) is correlated with race among marginal borrowers, but there may not be any correlation between the unobservable factor(s) and race in the population at large.

[8] BCGH emphasize that inferences about discrimination must distinguish between *average* and *marginal* behavior. BCGH (as well as others) note that average default rates for minorities will be higher than for non-minorities even in the presence of discrimination because many default-related factors are distributed less favorably for minority borrowers than for non-minority borrowers. The default theory assumes that discrimination against minorities should be apparent at the margin, affecting those who are near the borderline for creditworthiness.

BCGH considered their results quite robust, as they found a positive and significant black/white default differential in a number of subsamples of the FHA data.

BCGH's results generated much controversy and criticism among economists and other researchers in the field of mortgage lending. Critiques centered on both specific shortcomings of BCGH's estimation of their default model as well as on general methodological flaws of using a model of mortgage performance to test for discrimination. With respect to former, the default model estimated by BCGH was plagued by "omitted variables" problems.

One of the most serious omissions was information on borrower credit history, which was not available for BCGH to include as an explanatory variable in their empirical models. Credit history data could include indicators of an applicant's overall creditworthiness -- such as an applicant's financial experience (*e.g.*, how long since an applicant opened his or her first credit account), an applicant's past record in paying revolving credit accounts (*e.g.*, the number of accounts that have been delinquent for more than 60 or 90 days), and an applicant's current credit balance relative to his or her approved credit limits. Recent studies have shown that credit history is one of the main determinants of mortgage default, and it is perhaps the most important variable in the new automated mortgage scorecard systems that have recently spread throughout the mortgage market (Bunce, *et al.*, 1999). Individuals with a poor credit history are much more likely to default on their mortgages. It is also well documented that minority borrowers tend to have a poorer credit history than non-minority borrowers; black borrowers, in particular, and Hispanic borrowers, to a lesser extent, tend to score lower on measures of creditworthiness than do other borrowers.[9] Thus, the BCGH analysis excluded an important determinant of mortgage default that is distributed differently across racial and ethnic groups. This omission in itself tends to cause the estimated black/white differential to be overstated because the black borrower variable is picking up the effects of poorer credit history, which in turn biases findings in favor of no discrimination.

BCGH recognized this omission in a second version of their model (BCGH, 1994) where they attempted (quite creatively) to use data from the Boston Fed study to adjust for the effects of omitting credit history on their coefficient estimate for black borrowers. According to BCGH, while their adjustment reduces the estimated coefficient of black borrowers by about 40 percent, the downward adjustment is not enough to change the positive direction of the coefficient or influence its statistical significance. BCGH conclude that the significant and positive black default differential remains even after adjusting for credit history.[10]

As explained in Sections 3 and 4, this paper estimates a default model that includes a summary measure of credit history for FHA borrowers who applied in 1992, 1994, and 1996. Our analysis finds that the sign and significance of estimated black default differentials varies widely from specification to specification, which suggests that BCGH’s conclusions may well have been shaped by their omission of credit characteristics.

---

[9] As shown below in Table 2 (Panel B), the average FICO score (a summary indicator of credit history for which higher scores indicate a better credit record) by race and ethnicity for FHA loans in 1996 were as follows: Black (627), Hispanic (655), White (668), and Asian American (672).

[10] There appears to be an error in the BCGH bias calculation. Their derivation of bias implicitly assumes that the variance of the random component is the same for the model that correctly includes credit history as for the model that incorrectly excludes credit history. This assumption will not be valid unless the omitted credit history variable(s) can be expressed as an exact linear combination of the remaining explanatory variables.

We note in passing that earlier critiques of the BCGH model also touched on what several economists consider to be fundamental methodological flaws of using a traditional model of mortgage performance to test for discrimination. These observers conclude that a fully specified mortgage rejection model, such as the Boston Fed study, provides the best methodology for testing lender bias. The details of this debate are included in a 1996 issue of *Cityscape*, a publication of HUD's Office of Policy Development and Research, and an analysis by Stephan Ross and John Yinger, "The Default Approach to Studying Mortgage Discrimination: A Rebuttal," in *Mortgage Lending Discrimination: A Review of Existing Evidence*, a report by The Urban Institute (1998) for HUD. Readers are referred to these publications for a discussion of the complex methodological issues and for an interesting exchange between BCGH and others engaged in this debate.[11]

[11] Although there have been numerous criticisms of their approach, BCGH in the 1996 *Cityscape* issue continued to stand by it. They stated: "In our opinion, these concerns do not invalidate the use of loan performance data to investigate discrimination. Loan performance studies may not provide the complete and final answers to all research issues in lending discrimination, but they should be an integral part of the overall research program designed to understand discrimination and its effects." (BCGH, 1996, page 49) In the same issue, others disagreed strongly with BCGH's characterization of the usefulness of the default model to investigate discrimination (see, Ross, Galster, and Yinger, 1996).

SECTION 3

DATA SOURCES AND VARIABLE CONSTRUCTION

3.1. Estimation Samples

As indicated above, the estimates reported below are based on FHA-insured loans drawn from borrowers who applied in the years 1992, 1994, and 1996. Because important pieces of information (*e.g.,* loan-to-value ratios) are unavailable for those who applied for funding through the FHA streamline refinance program, streamline refinances are excluded from the analysis. Because the streamline refinance program was a major source of applications identified as refinances, this restriction serves to reduce substantially the representation of refinances.

Sampling from the universe of loans helped to reduce the cost of obtaining credit history information from commercial vendors (see below) and to reduce the estimation burden. Because default is a relatively rare event, defaults were oversampled relative to nondefaults.[12] Given the unequal sampling probabilities, the statistical procedures employ weighting according to the sample stratum (application year and default status) from which the loan was drawn. Approximately 60,000 to 75,000 loans are contained in each yearly sample. Additional information on the characteristics of these samples is provided below.

3.2. Analogs to the BCGH Variables

Table 1 lists the explanatory variables used in the major portion of this study. With the exception of the variables that are added in our attempt to overcome some of the data-related deficiencies of the BCGH analysis, the data sources and variables used here are generally chosen to match those in the BCGH analysis.[13] Variables that attempt to duplicate the BCGH constructs generally adopt the names used by BCGH and are listed first in Table 1. The underlying data are drawn from two principal sources: FHA data on FHA-insured loans and Census Bureau data from the 1990 Decennial Census. Both kinds of data have been supplied by HUD. The FHA loan files contain a variety of loan, borrower, and property characteristics measured at loan origination or application, as well as information on the claim status of the loan at the time that the files were constructed (July 2000) and the dates of critical events in the life of the loan (*e.g.,* the date of loan origination and the date of default, if any). The Census Bureau data characterize the census tract in which the property is located.

As may be seen in Table 1, BCGH included in their default equation the types of variables that typically appear in studies of mortgage default, including characteristics of the property (*e.g*., house value), the loan (*e.g.*, loan-to-value ratio, front- and back-end ratios,

---

[12] More precisely, loans from application years 1992 and 1994 were stratified according to claim status as of May 1997, when the FHA data files used in this analysis were originally generated. Although claim status and related default activity have been updated, the weights are calculated according to the original stratification scheme. The loans from application year 1996 were not selected according to claim status.

[13] There is not a single BCGH analysis, but instead a family of three analyses with nearly identical variable lists. We choose to use the essential form of the *Cityscape* version as our base model. This choice is essentially arbitrary, but one advantage of this choice over the *Review of Economics and Statistics* version is that the latter contains an interaction between the Herfindahl and the black indicator, which in turn complicates the calculation of racial impacts. In addition, much of the commentary centered on the *Cityscape* version.

mortgage term), and the borrower (*e.g.*, age, first-time homebuyer status), as well as characteristics of the neighborhood in which the property is located (*e.g.*, median income of families in the census tract). Race and ethnicity of the borrower (Black, Hispanic, Asian American, American Indian) are included to support tests of discrimination. The omitted group is largely white, though it contains a small fraction composed of races other than those explicitly listed.

A few of the variables used by BCGH were unavailable for this study, and others may be defined somewhat differently. Variables used by BCGH but not available here (*e.g.*, PCBINC --- the percentage of household income earned by the coborrower) are indicated in Table 1 by an "(NA)" following the variable name. Although we cannot be certain, we doubt that these omissions materially affect the results presented here. Several other variables, while available in principle for the statistical analysis, were omitted from some analyses because they were collinear with other included variables or, in the case of some indicator variables, one value of the variable was associated with a single outcome (default or nondefault) and thus perfectly "explains" the outcome.

Among the variables defined through the FHA data is the default indicator. For purposes of this analysis, defaults were defined to include only those loans that (a) defaulted on or before April 30, 2000, and (b) for which a claim[14] was recorded by the time the data were extracted in July 2000. In particular, loans that entered default status but subsequently cured are not included under this definition of default. Although this particular definition of default seems fairly consistent with the spirit of the BCGH default definition, it has the potential defect of treating defaults inconsistently. More specifically, defaults leading to claim are not captured in the FHA data files until the claim has been recorded. For this reason, defaults that will eventually lead to a claim, but which occur close to the time that the data are extracted, will be missed if the claim is not processed prior to data extraction. One problem is that the latter event is more likely to occur in states featuring slower foreclosure processing, leading to a relative understatement of claim rates in such states; the inclusion of state indicators, however, is expected to provide an adequate remedy to the problem of uneven coverage of judicial foreclosure. Yet, as explained below, there remain other more serious and less easily resolved problems associated with the failure to identify defaults that have in fact occurred. We can avoid the latter difficulties, as well as provide an additional check on the robustness of our findings, by using a different default definition. For these reasons, we later consider an alternative default definition that will serve to check on the findings derived with the primary definition.

3.3. Additional Variables

Variables used by BCGH are augmented with four different kinds of variables: a simple indicator for ARMs, variables measuring the relative price of the home at origination, variables measuring changes in house prices after mortgage origination, and variables measuring past credit performance. The ARM indicator is introduced to recognize the popularity and possible differential default performance of ARMs in the 1990s.

A set of four variables is used to measure relative house prices. The variable HPrelPW is the ratio of the sales price of the home relative to the reference home price in that area, as given by the PricewaterhouseCoopers median home price series for the application year.[15] When the

---

[14] Loans entering the assignment program are treated as defaults even though such loans technically remain active.

[15] The PricewaterhouseCoopers median home price series for metropolitan areas is briefly described in the *MMI*

PricewaterhouseCoopers median price series is unavailable, we set HPrelPW to zero and instead measure relative house prices with the variable HPrelLL.[16] The latter construction follows the HPrelPW calculation except that the area reference house price is defined as the area FHA loan limit for the year divided by 0.95. Because FHA loan limits are intended to be 95 percent of the area median house price, HPrelLL is effectively sales price divided by the area median house price. For those relatively rare loans in areas for which (a) the PricewaterhouseCoopers series is unavailable, and (b) the FHA loan limit is at the legislative maximum or minimum, and is thus constrained so that it may no longer accurately measure median area house prices, we set both HPrelPW and HPrelLL to zero; in such a case, we also set an indicator (LLmax or LLmin) to unity.

The basis for including the set of relative house price measures is to control for the size of the relevant housing market, which could affect the ease of sale and thus default. FHA is concentrated in the lower-priced portion of the aggregate housing market. Homes that are relatively low priced within these samples of FHA-insured loans are likely to be nearer the low end of the full local house price distribution where the market is thin. In contrast, relatively high priced homes within these samples are likely to be closer to the heart of the full local house price distribution. The consequence is expected to be a negative relationship between relative house price and likelihood of default.

A set of three variables is used to measure changes in house prices after mortgage origination. The variable HPcMSA00 is calculated as the proportional change in the Freddie Mac MSA-level house price index from the quarter of mortgage origination to the second quarter of 2000. When the latter Freddie Mac data are unavailable, we set the variable HPcMSA00 to zero and instead use the variable HPcST00, the post-origination proportional change in the Freddie Mac house price index at the state level. For such cases we also set an indicator variable STdata00 to unity.[17]

The latter variables are introduced to recognize post-origination changes in housing markets that might otherwise contaminate the estimated race effect. If, for example, blacks tend to live in areas that suffer lower house price growth, and if we fail to control for post-origination house price changes, we might find black default probabilities exceed those of whites only because of this spurious correlation.[18] That is, even though the purpose of the analysis is to estimate the role of default-related factors that are observable at the time of underwriting, the outcome being analyzed (default) is a product of not only these factors, but also of default-related events that occur after origination. If the occurrence of these events is correlated with race, estimates of racial impacts will be affected by failing to control for these post-origination

---

Fund Analysis FY 1998, an actuarial review by PricewaterhouseCoopers LLP.

[16] More precisely, HPrelPW is set to zero when HPrelLL is used to measure relative price, and HPrelLL is set to zero when HPrelPW is used to measure relative house price.

[17] When the Freddie Mac MSA-level house price indices are available, the variables HPcST00 and STdata00 are set to zero.

[18] Depending on the precise areas affected, this kind of bias might show up in estimated impacts of tract racial composition, rather than individual race. Note also that not all agree that post-origination changes in housing markets should be recognized in estimating a default model designed to uncover possible discrimination in mortgage lending. On this point, see BCGH (1994), page 265, which seems to defend the exclusion of post-origination events, and Yezer (1996), page 72.

events.[19]

Although the house price change measures used here may proxy the desired effects, they are very likely to be defective for at least two reasons. First, these variables are measured at the MSA or state level, ignoring the realistic possibility that house price growth varies across much narrower geographic areas.[20] Second, these measures calculate house price growth from origination to the second quarter of 2000. Stopping the calculation at the second quarter of 2000 is consistent with the choice of default interval but is otherwise arbitrary and could be misleading. In particular, house price growth from mortgage origination to other intermediate dates may be strongly negative and may contribute to default; yet house price growth as measured here may be positive. Part of the difficulty is that the logit specification is a clumsy tool for estimating effects of time-varying covariates like house price changes, for it forces us to parameterize house price changes with only a few variables.[21]

The most crucial change to the BCGH specification is the inclusion of credit history data in the form of the FICO score. As noted, higher FICO scores are intended to reflect better credit histories. Although the FICO score is a commonly employed summary of an individual's credit history, FHA underwriting guidelines did not specify the use of the FICO score or any other summary score as a measure of past credit performance during the time interval spanned by the loan applications that are the subject of this paper. The focus instead was on a set of specific credit characteristics, such as the nature and recency of any bankruptcies, the size of payments on credit cards, outstanding judgments, collections, delinquencies, and the recency of any foreclosures. The FICO score recognizes many of these same factors, though not necessarily in the same form, but it may include numerous other credit history characteristics that Fair, Isaac and Company --- the score originators --- have found to be predictive of future credit performance. Because the precise formulation of the FICO score is proprietary, its nature is not completely known to outsiders; yet it seems reasonable to believe that the FICO score contains many, if not all, of the aspects of credit history of interest to FHA underwriters, but it reflects additional features of credit history as well. As such, it is perhaps best viewed as a proxy for a host of credit history characteristics used by FHA underwriters.[22]

FICO scores used in this project were obtained retrospectively long after the loans were made, but the scores were drawn from credit data archives dated at the approximate time of loan application. To obtain FICO scores for borrowers and coborrowers, we sent identifying

---

[19] One advantage of the application rejection approach to the identification of lender discrimination is that in that approach there is no need to account for post-origination events that might affect default activity. The application rejection model does, however, place strong demands on information in other dimensions. In particular, the application rejection model demands not only all variables that are correlated with race and that influence underwriting decisions, but also knowledge of the way in which these variables are used. It may not, for example, suffice to know that LTV enters underwriters' evaluations; it also matters how LTV is used (*e.g.,* it may enter nonlinearly in a wide variety of ways).

[20] For some evidence of variation in house price growth across tracts in the Chicago MSA, see Cotterman (2000).

[21] A hazard specification would permit us to analyze house price growth on a more time-disaggregated basis. See, for example, Cotterman (2000).

[22] In this sense, introducing the FICO score replaces a missing variable problem with a proxy variable problem. In contrast to the typical "errors-in-the-variables" scenario in which data errors are assumed to be uninformative, in this case the proxy may contain some elements that are predictive of default but which were unknown to FHA underwriters. Hence, the proxy may control for more than what was recognized in FHA underwriting.

information (name, Social Security number, address) for samples of the 1992, 1994, and 1996 applications to Trans Union and Equifax; the score data were subsequently merged with FHA loan files that had been stripped of all identifying information that could be used to link a loan to a specific individual.

Although we attempted to obtain FICO scores for all individuals from both repositories, scores were sometimes unavailable from one or both sources.[23] When an individual (borrower or coborrower) has more than two FICO score readings, we arrive at a single "operational" score for that person by taking the minimum of the two FICO scores. We use these operational FICO scores to construct three FICO variables that correspond to alternative borrower/coborrower configurations and patterns of missing FICO scores. For loans with a lone borrower (no coborrower), we use the variable FICOsb, which is the operational FICO for that single borrower. For borrower/coborrower pairs in which only one party has an operational FICO score, we use the variable FICO2b1, the operational score for the one borrower of the pair who has a FICO score reading. When both the borrower and coborrrower have one or more FICO score readings, we use the variable FICO2b, defined as the average of the operational scores for borrower and coborrower.

This formulation, which resulted from some experimentation, permits, but does not force, different FICO effects for alternative configurations.[24] That is, one might expect the link between an individual's past credit performance, as measured by the FICO, and future default propensities to be different for a single decision maker (*i.e.*, a lone borrower) than for a joint decision making unit composed of a borrower/coborrower pair. Moreover, one might arguably anticipate a different link for borrower/coborrower pairs in which we can observe FICOs for both parties than for borrower /coborrower pairs in which only one party has a FICO reading.

Table 2, Panel A, provides (weighted) sample means and standard deviations by application year for the variables listed in Table 1. For variables for which zero values are inserted to replace missing values (*e.g.*, the inapplicable two of the three FICO variables on each loan), the reported sample means are calculated after removing these zero values. Panel B of Table 2 supplements these figures with average default rates and average FICO scores by racial/ethnic group.[25] Note that the large default rate for Hispanics is probably heavily influenced by the downturn in the California housing market in the 1990s. The intergroup differences in average FICO scores are also noteworthy. In each yearly sample average FICO scores are highest for Asians, second highest for whites, and lowest for blacks; the average FICO scores for the remaining three groups tend to be close together, and the ordering changes from sample to sample.

---

[23] Samples used here are restricted to loans for which at least one FICO score is available on the borrower or coborrower.

[24] Limited experimentation seemed to indicate that estimated racial impacts would be largely unaffected by alternative specifications of the FICO-related variables.

[25] Calculations of average FICO scores use whichever FICO variable is appropriate for a particular loan.

SECTION 4

FINDINGS

4.1. Basic Findings

Table 3 reports logit coefficient estimates and asymptotic normal (z) statistics for four different specifications run separately on the three samples from the years 1992, 1994, and 1996. Each single-page panel contains results from one of the yearly samples. The four specifications vary according to whether the FICO variables are included and whether the relative house price and house price change variables are included. Variables are ordered within the table so that the original BCGH variables follow the non-BCGH variables.[26] Although we devote some attention to estimated effects of other variables, the discussion below centers on the estimated impact of the variable BLACK.

The first pair of columns in each panel is labeled "BCGH Model" and reports findings from our approximation to the original BCGH specification, differing from the latter because of the inclusion of an ARM indicator and because of the omissions noted above. Note that the coefficient on the black indicator measures the estimated difference in the log odds of default for blacks relative to that of the omitted group (which is essentially whites), holding constant the remaining controls. In all three samples, differentials are positive and statistically significant at conventional levels. Note also that the estimated BLACK coefficients in the 1992 and 1994 samples closely match the corresponding coefficients reported for the 1987 and 1988 loans in the BCGH *Cityscape* paper, though the coefficient estimate for 1996 is smaller. Most of the remaining estimated effects for this first specification appear to be reasonable with respect to sign.

The specification reported in the second pair of columns, headed "BCGH Enhanced," adds the variables reflecting relative house prices and post-origination house price growth. The coefficient pattern on these additional variables generally tends to show that more expensive homes (relative to the area reference price) and higher post-origination house price growth are associated with lower default probabilities. Including this array of variables, however, seems to have only a trivial impact on the estimated BLACK coefficients, perhaps in part a reflection of the above-mentioned defects in the house price growth variables in particular.[27]

The specification in the third pair of columns, labeled "BCGH with Credit History," adds the FICO variables to the BCGH specification. The coefficient estimates on the FICO variables are always statistically significant by any reasonable standard. The estimated credit history impact appears to be somewhat larger for borrower/coborrower pairs in which both parties have scores (FICO2b), but no tests for differences were performed. Notice that including the credit history measure has an important impact in reducing the size of the estimated BLACK coefficient. In the 1992 sample, the introduction of the FICO variables leads to an estimated BLACK effect that is less than one-fourth of its original size and renders the estimate insignificant by conventional standards. For the 1994 sample, the estimated BLACK coefficient is less than one-third of its original value and is now of marginal statistical significance. In the

[26] All specifications include state indicators (as in BCGH), but the corresponding coefficient estimates are suppressed.

[27] In the 1992 and 1994 samples, the estimated impact of tract racial composition (CTBLACK) seems more strongly affected by the introduction of these controls, as might perhaps be expected. This pattern is reversed in the 1996 sample, however.

1996 sample, the estimated BLACK coefficient changes to negative and meets some conventional standards for statistical significance.

The fourth pair of columns, headed "BCGH Enhanced with Credit History," adds the variables measuring relative house price and post-origination house price growth, as well as the credit history measures, to the basic BCGH specification. As in the comparison between the first and second pairs of columns, a comparison of the third and fourth pairs of columns shows that adding the relative house price and house price growth variables has little impact on the estimated BLACK coefficient.

These findings demonstrate the importance of including credit history controls when assessing the differential default experience of black borrowers. Before offering additional discussion of this point, however, it is worth exploring this phenomenon using a different definition of default.

4.2. Additional Explorations with an Alternative Default Definition

As discussed earlier, the method of identifying defaults used above will likely lead to the omission of defaults occurring near the end of the observation window. To avoid this difficulty and to provide a useful check on the results above, we introduce a second default definition and rerun selected logit specifications on the 1992 and 1994 samples. Under this alternative definition, defaults include only those loans that (a) defaulted on or before April 30, 1998, and (b) within 48 months of amortization start, and (c) for which a claim[28] was recorded by the time the data were extracted in July 2000. This alternative closes the default window more than two years prior to when the data extraction occurs so that sufficient time remains for all those who have defaulted to be observed in claim status by the date of data extraction. The latter feature is expected to result in a more consistent treatment of defaults. The four-year maximum on time-to-default restricts defaults to those of the short-to-medium term variety, and the choice of a fixed four-year default horizon may facilitate cross-year comparisons.[29]

Before proceeding with additional analyses, it is worth providing a bit more discussion to motivate this additional work. The focus on four-year defaults will serve mainly as a check on the Table 3 results. The Table 3 logits utilize samples for which the default horizon is determined solely by the length of the observation window, and as a consequence each set of findings pertains to a different application year and a different length of default horizon. There is no way to assess the separate influence of differing application years and different default horizons, nor is there any assurance that our findings are not the product of the particular combinations of application years and default horizons used in the Table 3 samples. By holding fixed the default horizon at four years in a new analysis, we will be able not only to remove one potential source of differences in findings across years, but more importantly we will also provide an additional check on the robustness of the Table 3 findings. Although we do not

---

[28] As before, loans entering the assignment program are treated as defaults even though such loans technically remain active.

[29] Notice, however, that there is not complete comparability across even the 1992 and 1994 samples because some of the 1994 loans do not have a full four years of potential exposure to default. Indicator variables for different amounts of exposure in the 1994 sample yielded coefficient estimates that were both individually and jointly statistically insignificant by conventional standards; these indicator variables were not retained in the models presented here. Loans in the 1996 sample have only about two years of potential exposure under this alternative default definition; the 1996 sample is thus excluded from the analysis.

expect this new analysis to reveal that the changes in racial differences with the introduction of credit controls in Table 3 are traceable to the particular combinations of default horizons and years of application used there, it seems preferable to confirm that the findings continue to hold for alternative default horizons.

Next consider the omission of defaults occurring near the end of the observation window. As noted earlier, defaults are recorded in these data only after the claim is processed, and thus lags between default and the payment of the claim prevent us from observing defaults that occur near the time when the observation window closes. One might imagine the impact of omitting these "late" defaults to fall with equal proportion on all race groups. We find, however, that the time interval between the occurrence of default and the processing of the claim tends to be longer on average for blacks than for whites, and thus omission of defaults at the end of the observation window is likely to result in an understatement of the actual black default rate relative to the actual white default rate. (The basis for these arguments is discussed in some detail in the appendix to this paper.[30]) This racial differential in understating default rates may in turn affect our findings with regard to black/white default differentials both before and after the introduction of credit history controls. Closing the default window more than two years prior to the close of the observation window is an effective and simple way to avoid this potential problem.

We now proceed with the analysis. Because the emphasis is now on four-year default rates, we redefine a few variables for the purpose of this analysis. In particular, the variables HPcMSA00 and HPcST00 now measure house price growth from the quarter of origination to 16 quarters later, rather than through the second quarter of 2000.

Table 4 presents logit estimates for the "BCGH Enhanced" specifications, both with and without credit history controls. The first four columns use the 1992 sample; the last four columns use the 1994 sample. Our focus again is primarily on the estimated BLACK coefficient, the estimated differential between blacks and whites[31] in the log odds of default, other things the same. Comparing the first to the second pair of columns, as well as the third to the fourth pair of columns, we see that the addition of credit history controls again induces substantial declines in the estimated BLACK coefficients. Now, however, the estimated differential in the 1994 sample remains positive and statistically significant by most conventional measures.

The evidence from Tables 3 and 4 demonstrates a consistent pattern in that the introduction of credit history controls dramatically reduces the estimated black differential in log odds ratios of default, but there is substantial variation in the resulting estimates. We see cases in which the final estimates (which include credit history controls) cannot be statistically distinguished from zero, as well as estimates that are positive or negative and statistically significant at conventional levels. For those who believe that this kind of analysis can reveal discrimination in lending, the negative and significant estimates support the notion of racial discrimination in lending. Insignificant estimates or positive and significant estimates fail to support that notion. Hence, we are left with ambiguity.

To give a more complete picture, and possibly to help resolve this ambiguity, we consider one more set of estimates.

---

30 The appendix also shows the BLACK coefficient estimates that result if the Table 3 logits are rerun using only defaults that occur by April 30, 1998.

31 Again, the omitted (comparison) group is technically composed of whites and others, but is dominated by whites.

4.3. Estimation Results by Risk Class

As a further check on results and to gain additional insight, we follow the BCGH (1994) suggestion of reestimating default models within risk classes.[32] The idea is that if loan qualification demands better risk characteristics for blacks in order to compensate discriminatory lenders, then the superior risk characteristics for observationally equivalent (to us) blacks might be expected to show up more clearly among the riskiest borrowers. Indeed, borrowers in the best risk classes may be so far from the margin of loan rejection that all such borrowers are accepted regardless of race; for these borrowers there may effectively be no racial discrimination in lending, and none would be revealed in estimated default differentials. We might then expect to find that black/white default differentials tend to favor blacks in the worst risk classes (*i.e.*, the black default probability is less than that of observationally equivalent whites in the worst risk classes). Depending on the distribution of unobservable (to us) factors, differentials may gradually erode as one moves up to better risk classes, or they may go to zero and remain there.[33]

Risk classes are defined here in two different ways. One alternative relies on the apparently strong relationship between the FICO and the occurrence of default. Under this method the FICO variable for each loan is used to rank loans in order and categorize each loan into one of the following four risk classes

1. FICOs up to 595 (about 15 percent of the loans in the aggregate of the three samples)
2. FICOs between 595 and 625 (about 15 percent of the loans)
3. FICOs between 625 and 660 (about 20 percent of the loans)
4. FICOs over 660 (about 50 percent of the loans).

This method of categorizing risk was, of course, unavailable to BCGH.

The second risk ranking uses the enhanced BCGH model with credit history controls to predict the log odds of default for each loan. Loans are then ranked from highest (riskiest) to lowest predicted log odds of default.[34] The risk classes are defined as the worst 15 percent of the risks, followed by classes of 15 percent, 20 percent, and 50 percent. BCGH use a similar methodology relying on quartiles of the estimated risk distribution.

The advantage of the latter method of assigning risk classes is that the predicted log odds of default is a more comprehensive measure of default risk than is the FICO score alone. That is, the predicted log odds of default is a linear combination of the explanatory variables in which numerous factors are weighted according to their ability to jointly predict the occurrence of default; the FICO score is simply one component of this linear combination. Assigning full weight to a single component, such as the FICO score, is likely to result in a poorer assessment of risk --- and, more importantly, presumably a poorer representation of risk classes as viewed by FHA underwriters. Notice, however, that neither of these methods of assigning risk classes is likely to match perfectly a risk classification that would be induced by applying an exact

[32] Estimation in this section reverts to the default measure used in Table 3.

[33] The assumption is that the fraction of black borrowers who are near the margin might be expected to decline as one moves to better and better risk classes. With sufficient ingenuity, however, one could presumably find distributional assumptions that would make a variety of patterns feasible.

[34] We also tried removing the estimated BLACK effect before calculating predicted log odds of default, but the results were essentially unchanged.

replication of FHA underwriting criteria. Perfection --- with regard to the formulation of the model or the number of risk classes --- is surely not required for current purposes, however. As long as the proportion of marginal borrowers (as viewed by FHA underwriters) is larger in the higher risk categories, we might expect to find differences in the BLACK coefficient across these categories. For this purpose, a perfect sorting by risk is not required, though of course a more accurate sorting would be expected to sharpen the differences across classes.

We next rerun the enhanced BCGH model with credit history controls within each risk class and examine the estimated black/white differentials.[35] The appropriate basis for these comparisons may not be completely clear. The use of the logit as a basis for estimating racial differentials suggests that we compare the estimated black coefficients across risk classes, for each such coefficient is a consistent estimate of the difference in black and white log odds ratios of default, other things the same (provided the model is correct). BCGH seem to suggest, however, that ratios of, or differences in, black and white default probabilities are a more appropriate basis for comparisons across risk classes.[36] Presumably, the correct comparisons would follow from the specification of a complete model of discrimination. Here we present a full complement of comparisons.

Table 5 presents the actual default rate and the estimated BLACK coefficient and asymptotic normal statistic for each risk class, as well as the ratio of black to white default probabilities and the difference between black and white default probabilities within each risk class.[37] Looking first at the size and pattern of estimated BLACK coefficients, we see that in three of the six rankings, the riskiest class contains the lowest estimated BLACK effect, and in two cases (the two rankings for 1996) the estimates in the riskiest class are negative and statistically significant by at least some commonly accepted standards, as might be predicted from the lending discrimination argument above. In the remaining three rankings, the lowest estimated effect occurs in one of the middle two risk classes, and in only one of the six cases is the ordering of estimated effects monotonic.[38] Indeed, only a few of the estimated effects within a risk class would meet common standards of statistical significance, and one could legitimately argue that almost all of these estimated effects cannot be statistically distinguished from zero, thereby failing to offer evidence in favor of or against lending discrimination. Thus, while the pattern of estimated black/white differentials in log odds ratios seems in some ways consistent

---

[35] Notice that we run separate logits within each class in anticipation of possible differences in coefficients across classes. Such differences are suggested by the underlying theoretical arguments stating that conditional distributions of unobservables (given loan approval) would be expected to vary across risk classes, implying differences in coefficients on observables as well. Using a pooled logit and testing for differences in coefficient estimates across risk classes showed that numerous coefficients would in fact differ across classes (at conventional significance levels).

[36] BCGH (1994) argue that the logit coefficients are not comparable across risk classes because logit estimates are subject to a variance normalization. Although we accept the fact that logit estimates are subject to a variance normalization that makes it impossible to extract true structural parameters, logit coefficient estimates are still consistent estimates of partial effects on the log odds of default (assuming the model is correct).

[37] Here we follow the BCGH (1994) procedure of calculating all default probabilities using characteristics of whites. That is, we calculate average predicted default probabilities for whites; we then calculate average predicted default probabilities for these same loans under the assumption that the borrower is black instead. Repeating these calculations using characteristics of black borrowers rather than white borrowers yields very similar results.

[38] Note that the differences in estimated effects across classes have not been subjected to statistical testing.

with the pattern that might be expected to result from lending discrimination, and two of the estimated BLACK effects in the riskiest classes are negative and statistically significant, the evidence seems far too weak to be compelling evidence either in favor of or against the existence of discrimination in lending.[39] In contrast, BCGH found that, with one exception, estimated BLACK effects were positive and significant by conventional standards in all risk classes and cohorts, a finding that again suggested the absence of lending discrimination.

The patterns of black/white ratios of estimated default probabilities and black/white differences in estimated default probabilities again offer no convincing support either for or against lending discrimination. There does appear to be some tendency for differentials (ratios or differences) to favor blacks in the riskiest classes, but there are exceptions, and the orderings are rarely monotonic across risk classes.[40] BCGH (1994) similarly find little consistent pattern in default differentials across risk classes.

[39] We make no claim that the breakdown into risk classes is in any way optimal, nor have we experimented with alternative categorizations. In particular, we cannot be certain that we have defined “marginal” borrowers narrowly enough even in our most risky classes.

[40] We have made no attempt to calculate the statistical significance of any of these ratios or differences within risk categories, nor have we conducted any tests of differences across categories.

SECTION 5

CONCLUSIONS

The evidence in this paper supports at least one important conclusion: for the data used here, the introduction of credit history into a logit model of mortgage default has an important impact on the estimated black/white differential in the log odds of default. The result is a substantially different set of estimated race effects than those presented in the BCGH papers. In all samples and default definitions examined here (as well as in other estimations that have not been presented), the introduction of a measure of past credit performance substantially reduces the estimated log odds differential favoring whites. In contrast, introducing measures of post-origination changes in house prices at the state or MSA level has no important effect on estimated black/white differentials in the data examined here, perhaps because of defects in these price growth measures.

More substantive conclusions, particularly with regard to implications for the existence of racial discrimination in mortgage lending, are hazardous. Those who believe that default models can reveal nothing of substance regarding lending discrimination will presumably be unswayed by the evidence offered here in any case. There continue to be important unresolved issues in econometric methodology, and there are surely important default-related factors that have been omitted from the analysis or measured incorrectly. Even those who found the earlier BCGH framework convincing, and those who believe that the only important defect in that work was the lack of credit history controls, will find it difficult to draw clear inferences from these new results regarding discrimination in lending. What is clear is that the consistent results reported by BCGH are lost. Logit estimates above sometimes show that blacks have statistically significantly lower log odds of default than do whites, thereby offering empirical support to the hypothesis that there is racial discrimination in lending. Differentials are sometimes statistically significant in the opposite direction, however, and sometimes the differentials cannot be statistically distinguished from zero. The latter two kinds of findings fail to support the hypothesis that there is discrimination in mortgage lending. Variation in default differentials across samples and default definitions appears to be an important phenomenon, and this variation points to possible model deficiencies, including defects in the econometric framework or in the list of explanatory variables. There is some evidence of differentials in log odds of default favoring blacks in the riskiest loan categories in the 1996 sample, again supporting the hypothesis of discrimination in lending, but there are no corresponding statistically significant differentials in the riskiest loan categories for the 1992 and 1994 samples. Similarly, the evidence from patterns of black/white differentials in default probabilities across risk categories seems less than compelling.

It may be worth offering additional commentary on the nature of this evidence. The basis for the series of BCGH studies was to use information on racial differentials in mortgage default to draw inferences regarding possible discrimination in mortgage lending. The evidence presented above is constructed to parallel the BCGH studies, thus (a) demonstrating the importance of controlling for credit history, and (b) supplementing those studies with additional information that believers in this methodology might use to help determine whether or not there is discrimination in mortgage lending. The use of credit history controls, like the FICO score, seems to be widely accepted in the context of loan qualification, and it may be perfectly reasonable for lenders to adopt such measures in assessing the likelihood of mortgage default. One should be clear, however, about the interpretation of evidence that relies on such controls to assess differential default performance. In particular, suppose that past credit performance tends to be poorer on average for blacks than for whites[41] and one finds that there is no significant difference in racial default differentials once one controls for past credit performance. One interpretation of this evidence is that, on average, blacks perform no worse than whites in mortgage default once one controls for their poorer average credit performance in the past. It is in this limited way that past credit performance "explains" future default behavior and eliminates "unexplained" racial differentials in default activity.

[41] This statement assumes that the FICO measures past credit performance alone or that interracial differences in FICO scores are traceable to such differences in past credit performance.

REFERENCES

Becker, Gary S. 1993. "The Evidence Against Banks Doesn't Prove Bias." *Business Week*, April 19, p. 18.

_____. 1971. *The Economics of Discrimination,* 2nd Edition. Chicago: University of Chicago Press.

Berkovec, James A., Glenn B. Canner, Stuart A. Gabriel, and Timothy H. Hannan. 1998. "Discrimination, Competition, and Loan Performance in FHA Mortgage Lending." *Review of Economics and Statistics 80*, 241-250.

_____.1996. "Response to Critiques of 'Mortgage Discrimination and FHA Loan Performance'." *Cityscape: A Journal of Policy Development and Research 2 (1):* 49-54.

_____.1996a. "Mortgage Discrimination and FHA Loan Performance." *Cityscape: A Journal of Policy Development and Research 2 (1):* 9-24.

_____.1996b. "Race and Residential Mortgage Defaults: Evidence from the FHA-insured Single-Family Loan Program." Conference on Discrimination and Mortgage Lending. Washington, D.C.

_____.1994. "Race, Redlining, and Residential Mortgage Loan Performance." *Journal of Real Estate Finance and Economics 9* (3): 263-294.

Browne, Lynn E. and Gregory M. Tootell. 1995. "Mortgage Lending in Boston: A Response to the Critics." *New England Economic Review* (Sept./Oct.): 53-78.

Bunce, Harold L., William J. Reeder, and Randall Scheessele. 1999. "Understanding Consumer Credit and Mortgage Scoring: A Work in Progress at HUD." U.S. Department of Housing and Urban Development, Unpublished Paper, (June).

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): 277-314.

Cotterman, Robert F. 2000. *Neighborhood Effects in Mortgage Default Risk.* Report prepared for the U.S. Department of Housing and Urban Development, Office of Policy Development and Research, (September 30, 2000).

Galster, George C. 1996. "Comparing Loan Performance between Races as a Test for Discrimination." *Cityscape: A Journal of Policy Development and Research 2* (1): 33-40.

_____. 1993. "The Facts of Lending Discrimination Cannot Be Argued Away by Examining Default Rates." *Housing Policy Debate 4 (1):* 141-46.

_____. 1996. Galster, George C., John Yinger, and Stephen Ross. "Rejoinder to Berkovec, Canner, Gabriel, and Hannan." *Cityscape: A Journal of Policy Development and Research 2 (1): 55-58*

Goering, John M., Editor. 1996. "Race and Default in Credit Markets: A Colloquy." *Cityscape: A Journal of Policy Development and Research 2 (1).*

Hunter, William C., and Mary Beth Walker. 1996. "The Cultural Affinity Hypothesis and Mortgage

Lending Decisions." *Journal of Real Estate Finance and Economics 13* (July): 57-70.

LaCour-Little, Michael. 1998. "Application of Reverse Regression to Boston Federal Reserve Data Refutes Claims of Discrimination." *Journal of Real Estate Research 11: 1-12.*

Ladd, Helen F. 1998. "Evidence on Discrimination in Credit Markets." *Journal of Economic Perspectives 12* (Spring): 41-62.

Munnell, Alicia H., Lynn E. Browne, James McEneaney, and Geoffrey M.B. Tootell. October 1992. “Mortgage Lending in Boston: Interpreting HMDA Data.” Working Paper No. 92-7, Federal Reserve Bank of Boston.

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

Phillips, and Anthony M. Yezer. 1995. "Self Selection and the Measurement of Bias and Risk in Mortgage Lending: Can You Price the Mortgage If You Don't Know the Process." Paper presented at the American Real Estate and Urban Economics Association Meetings, Washington, D.C., January.

Quercia, Roberto.G. and Michael A. Stegman. 1992. "Residential Mortgage Default: A Review of the Literature." *Journal of Housing Research 3:* 341-380.

Rachlis, Mitchell.B. and Anthony M. Yezer. 1993. "Serious Flaws in Statistical Tests for Discrimination in Mortgage Markets." *Journal of Housing Research 4* (2): 315-336.

Ross, Stephen L. 1998. "Mortgage Lending, Sample Selection, and Default." Working Paper, University of Connecticut.

_____. 1997. "Mortgage Lending Discrimination and Racial Differences in Loan Default: A Simulation Approach." *Journal of Housing Research 8:*

_____. 1996a. "Mortgage Lending Discrimination and Racial Differences in Loan Default." *Journal of Housing Research 7 (1):* 117-126.

_____. 1996b. "Flaws in the Use of Loan Defaults to Test for Mortgage Lending Discrimination." *Cityscape: A Journal of Policy Development and Research. 2(1): 41-48.*

Ross Stephen L., and John Yinger. 1998. “Does Discrimination Exist?: The Boston FED Study and its Critics.”
In The Urban Institute, *Mortgage Lending Discrimination*: *A Review of Existing Evidence.* Report prepared for the U.S. Department of Housing and Urban Development, Office of Policy Development and Research, (February 1998).

_____. 1998. “The Default Approach to Studying Mortgage Discrimination: A Rebutal.” In The Urban Institute, *Mortgage Lending Discrimination*: *A Review of Existing Evidence*. Report prepared for the U.S. Department of Housing and Urban Development, Office of Policy Development and Research, (February 1998).

The Urban Institute. 1998. *Mortgage Lending Discrimination*: *A Review of Existing Evidence*. Report prepared for the U.S. Department of Housing and Urban Development, Office of Policy Development and Research, (February).

Tootell, Geoffrey M. B. 1993. "Defaults, Denials, and Discrimination in Mortgage Lending,*"* *New England Economic Review*. (September/October).

_____. 1996. "Turning a Critical Eye on the Critics." In *Mortgage Lending, Racial Discrimination, and Federal Policy,* edited by John Goering and Ron Wienk (Washington D.C.: Urban Institute Press), pp. 143-182.

Van Order, Robert and Peter Zorn. 1995. "Testing for Discrimination: Combining Default and Rejection Data." In *Fair Lending Analysis: A Compendium of Essays on the Use of Statistics,* edited by Anthony M. Yezer (Washington D.C.: American Bankers Association), pp. 105-112.

Yezer, Anthony M. 1995. "Editor's Notes." In Fair *Lending Analysis: A Compendium of Essays on the Use of Statistics,* edited by Anthony M. Yezer (Washington D.C.: American Bankers Association), pp. iii-vi.

Yezer, Anthony M., R. Phillips, and R. Trost. 1994. "Bias in Estimates of Discrimination and Default in Mortgage Lending: The Effects of Simultaneity and Self-Selection." *The Journal of Real Estate Finance and Economics 9'* 197-216.

Yinger, John. 1998. "Evidence on Discrimination in Consumer Markets**.**" *Journal of Economic Perspectives 12* (Spring): 23-40.

_____. 1996. "Discrimination in Mortgage Lending: A Literature Review." In *Mortgage Lending, Racial Discrimination, and Federal Policy,* edited by John Goering and Ron Wienk (Washington, D.C.: The Urban Institute Press), pp. 29-74.

_____. 1998. “Why Default Rates Cannot Shed Light on Mortgage Discrimination.” *Cityscape: A Journal of Policy Development and Research 2 (1): 25-32.*

APPENDIX

LAGS BETWEEN DEFAULT AND COMPLETION OF THE CLAIM PROCESS

Section 4.2 noted that there are racial differences in the lag between the occurrence of default and the completion of the resulting claim process. To see the basis for this argument, first consider Table A-1 below, which records, for each calendar year of default, the cumulative rate at which claims were completed in each month after default.[42] Thus, for example, we see that among defaults occurring in 1992, 59.9 percent had gone to claim by the end of 1993. Because our recording of default and claim activity is limited to claims that are seen in our data, cumulative claim activity always reaches 100% by the time the observation window closes in July 2000. Nonetheless, cumulative claim activity on defaults occurring in 1992 is 98.5 percent by the end of 1996, and thus it seems reasonable to believe that we have observed all or nearly all of the defaults that occurred in 1992 by the time the observation window closes in July 2000.[43] Similarly, the defaults occurring in 1993, 1994, and 1995 seem likely to be fairly free of censoring associated with the close of the observation window. As we proceed through later default years, however, it seems very likely that we are encountering more serious censoring of defaults because (a) we observe that claim activity for these defaults remained high even in 1999, and (b) we observe that rates of claim processing for a given number of years following the year of default differ markedly from what is seen in, say, 1992 defaults. For example, we see that among 1998 defaults, 68.55 percent (= 81.55 – 13.12) of the claims were completed during 1999, one calendar year after the calendar year of default. For defaults occurring in 1992, 1993, 1994, 1995, and 1996, less than 60 percent of the claims were completed in the calendar year following the calendar year of default.

Next consider Table A-2, which presents, for each calendar year of default, the mean elapsed time (in days) from the date of default to the completion of the claim process. The table gives the means for whites and blacks separately, as well as the difference between the two means.[44] Counts are the sample sizes over which the means are calculated. To allow for the role of judicial foreclosure, separate panels consider states with and states without judicial foreclosure. Notice that within each panel, the mean elapsed time for blacks invariably exceeds that for whites, often by 100 days or more. Not surprisingly, black-white differences tend to be particularly large for defaults occurring in the early to mid-1990s when censoring of defaults plays a smaller role. The means within race and the racial differential tend to be much smaller for 1999 defaults; censoring is presumably a much more important phenomenon in the 1999 defaults than in defaults from early years.

Table A-3 provides a bit of additional information. Panels A (for whites) and B (for blacks) show, for each racial group, the cumulative percentage of claims that were completed at each 3-month interval following the date of default. The data are classified by the calendar year in which the default occurred; once again, only defaults associated with claims that were

[42] This exercise utilizes the full universe of applications in 1992, 1994, and 1996, not simply the samples used in logit estimation.

[43] Clearly, we cannot be positive that we have observed all defaults even in 1992 since we are, by assumption, contending with censored default data. Yet, it is hard to imagine that censoring is a serious problem when one permits at least seven years for a claim to be paid following the occurrence of the default.

[44] Median elapsed times tend to differ from the means, but patterns are similar to those seen in Table A-2.

processed by July 2000 are available for analysis. Thus, for example, we see that among borrowers who defaulted in 1992 (and for which claim processing was completed by July 2000), 52.99 percent of the white claims were processed by 12 months after default, but only 43.10 percent of black claims were processed within the same 12-month interval following default. Comparisons between corresponding entries in Panels A and B show that this pattern is typical: the percentage of white claims completed at a given interval following default almost always exceeds that of black claims.

Tables A-1, A-2, and A-3 suggest strongly that, first, defaults occurring near the end of the observation period are censored and, second, that measured default rates are likely to be more strongly affected for blacks than for whites. There are at least three possible responses to this difficulty in analyzing default probabilities by race. First, we could ignore the problem and continue to use virtually all defaults recorded in the data. This procedure seems inadvisable. Although differential censoring may not in fact have a major influence on the estimates in this paper, it seems unwise to ignore the potential effect on measured black-white differentials, particularly when a relatively simple fix-up can go a long way towards removing the problem.

A second potential solution is to modify the estimation framework so that it properly models the observed outcomes. That is, the observed default outcomes are only those that go to claim within the observation window. Thus, the probability that we should be modeling in these data is the probability of observing a loan that both defaults AND is observed completing the claim process within the observation interval. This joint probability may be expressed as the product of a marginal and a conditional probability:

Prob (default and claim completed) = Prob(default) * Prob(claim completed| default).

The first of the terms on the right-hand side is the main subject of this inquiry, and this is the probability about which we have hypotheses based on the presence or absence of racial discrimination. The second term on the right-hand side may be estimable, but we have no particular insight as to its form, nor is it involved in any of the hypotheses at issue in this paper. Attempts to model the latter probability would take us well beyond the scope of this paper, are likely beside the main point, and, if done incorrectly, may even compromise our ability to extract the information we do seek. Hence, we opt for a third alternative: setting a threshold time for recording defaults that is sufficiently early relative to the close of the observation window that we are comfortable that any remaining censoring of defaults will not contaminate the empirical findings.

The setting of this threshold date is a clearly a judgment call. For this purpose, we have chosen to close the default window on April 30, 1998, thus counting as defaults only those occurring on or before that date.[45] This choice implies that any default would have a minimum of 27 months to show up as a claim before the end of July 2000. As shown in Table A-3, experience with the defaults occurring from 1992 through 1995 suggests that 80 to 87 percent of black defaults and 91 to 95 percent of white defaults in April 1998 may have been captured in the observed claims data; these figures will of course be even higher for defaults in earlier months when the bulk of defaults occurred.

To see how estimates presented in Tables 2 and 3 would be affected by eliminating late

[45] We are thus ignoring information on defaults occurring beyond April 30, 1998, but as noted, proper utilization of this information would require a substantial expansion of the scope of this paper.

defaults (but without restricting defaults to those occurring within four years of amortization start), we first recalculated sample default rates by counting only those defaults that occurred by April 30, 1998. These default rates are presented in Panel A of Table A-4 and may be contrasted with those presented in Panel B of Table 2. As might be expected, the percentage reductions in the default rates for whites exceed the percentage reductions in the default rates for blacks. Finally, we reran the Table 3 logits to see how the estimates would be affected if only the limitation on default dates were imposed. The results for the estimated black coefficient are given in Panel B of Table A-4. Note that the pattern of findings regarding racial differentials is similar to what was seen in Table 3. In particular, in all cases the introduction of credit history data yields a dramatic reduction in the estimated black coefficient.

Table 1

Definitions of Variables

BCGH Variables

RMISSING		1 if borrower race is unknown, 0 if known
BLACK		1 if black borrower, 0 if any other race
AMIND		1 if American-Indian borrower, 0 if any other race
ASIAN		1 if Asian borrower, 0 if any other race
HISPANIC		1 if Hispanic borrower, 0 if any other race
LTV		Loan-to-value ratio
INVEST		1 if investment property, 0 if noninvestment property
REFIN		1 if loan is a refinance, 0 if initial financing
CONDO		1 if property is a condominium, 0 if not a condominium
DIRECT		1 if insurance approved under direct endorsement, 0 if not approved under direct
URBAN		1 if property located in an urban area, 0 if nonurban
RURAL		1 if property located in rural area, 0 if nonrural (1992 Census MSA code is
SUBURBAN		1 if property located in a sururban area, 0 if nonsuburban
COMP	(NA)	1 if application indicates compensating factors, 0 if no compensating factors
FIRSTBUY		1 if borrower is a first-time homebuyer, 0 if not a first-time buyer
REPEATBUY		1 if borrower is not a first-time homebuyer, 0 if a first-time homebuyer
NEW		1 if property is a new house, 0 if not a new house
CBUNMARD		1 if borrower is not married to coborrower, 0 if borrower and coborrower are
DEPNUM		Number of dependents (excluding borrower and coborrower)
SELFEMP	(NA)	1 if borrower is self employed, 0 if otherwise employed
LQASS		assets available after closing
NOCBINC		1 if no coborrower or coborrower income is zero, 0 if coborrower's income is greater than 0
PCBINC	(NA)	Percent of household income earned by coborrower
LQASS2		Square of liquid assets
AGE < 25		1 if borrower is under 25 years of age, 0 if older than 25 years
AGE 25-35		1 if borrower is between 25 and 35 years of age, 0 if younger than 25 or older than
AGE 35-45		1 if borrower is between 25 and 45 years of age, 0 if younger than 35 or older than
BUYDOWN		1 if mortagage interest rate has been bought down by seller, 0 if interest rate has not been bought down
INCOME		Total annual effective family income
INCOME2		Square of income
SHRTMOR		1 if mortgage term is less than 30 years, 0 if term is greater or equal to 30 years
SINGLEM		1 if borrower is male and there is no coborrower, 0 if there is a coborrower
SINGLEF		1 if borrower is female and there is no coborrower, 0 if there is a coborrower
HVAL		Appraised value of the property at time of purchase
HVAL2		HVAL squared
POTHINC	(NA)	Percent of borrower income that is from other (nonsalary) sources

Table 1

Definitions of Variables

BCGH Variables

HEI 20-38		1 if housing expense to income ratio is between .20 and .38, 0 otherwise
HEI 38-50		1 if housing expense to income ratio is between .38 and .50, 0 otherwise
HEI > 50		1 if hosuing expense to income ration is above .50, 0 otherwise
DTI 20-40		1 if total debt to income ratio is between .20 and .41, 0 otherwise
DTI 41-53		1 if total debt to income ratio is between .52 and .65, 0 otherwise
DTI 53-65		1 if total debt to income ratio is between .53 and .65, 0 otherwise
DTI > 65		1 if total debt to income ratio is above .65, 0 otherwise
CTBLACK		Black percentage of census tract population
CTAMIND		American Indian/Alaskan native percentage of census tract population
CTASIAN		Asian percentage of census tract population
CTHISPANIC		Hispanic percentage of census tract population
CTMISS		Percentage of census tract population with race or ethnicity unknown
CTINCOME		Median family income of the census tract as a proportion of the median family income of the metropolitan area as a whole (multiplied by 100)
CTHVAL	(NA)	Median value of owner-occupied homes in the census tract
CTVCRAT		Percentage of one-to-four family housing units vacant in the census tract
CTMEDAGE	(NA)	Median age of residential properties in the census tract
CTUNEMP		Unemployment rate of the census tract
CTRENTRATE		Proportion of housing units in the census tract that are rental
CHGMEDV	(NA)	The change between 1980 and 1990 in the median value of owner-occupied homes in the census tract
HERF		The Hirschmann-Herfindahl index of market concentration, defined as the sum of market shares of the number of home purchase loans of lenders in each MSA

Additional Variables (See Text for Additional Details)

armflag	1 if mortgage is ARM, 0 otherwise
HPrelPW	House price relative to PricewaterhouseCoopers reference price
HPrelLL	House price relative to FHA loan limit /0.95
LLmin	1 if area loan limit at legislative minimum, 0 otherwise
LLmax	1 if area loan limit at legislative maximum, 0 otherwise
HPcMSA00	Proportional post-origination change in quarterly MSA house price index (if available), 0 otherwise
HPcST00	Proportional post-origination change in quarterly state-level house prices index when MSA house price index not available, 0 otherwise
STdata00	1 if MSA house price index unavailable, 0 otherwise
FICOsb	Operational FICO for lone borrower (no coborrower present), 0 otherwise
FICO2b	Average of operational FICOs for borrower/coborrower pair when both have scores, 0 otherwise
FICO2b1	Operational FICO for borrower/coborrower pair in which FICO available for only one individual, 0 otherwise

Table 2
Panel A

Means and Standard Deviations of Explanatory Variables Used in the Logit Models

Variable	1992 Sample Mean	1992 Sample Std. Dev.	1994 Sample Mean	1994 Sample Std. Dev.	1996 Sample Mean	1996 Sample Std. Dev.
armflag	0.2008	0.4006	0.3006	0.4585	0.2803	0.4492
FICOsb	686.00	64.03	672.43	64.83	657.90	65.87
FICO2b	687.00	53.00	675.90	54.06	665.38	55.41
FICO2b1	681.00	60.07	669.10	60.93	654.75	62.13
HPrelPW	0.9108	0.3105	0.8962	0.3004	0.8847	0.2983
HPrelLL	0.6345	0.1806	0.6348	0.1801	0.6568	0.1852
LLmin	0.0365	0.1875	0.0000	0.0000	0.0089	0.0942
LLmax	0.1342	0.3409	0.1638	0.3701	0.1008	0.3011
HPcMSA00	0.3917	0.1968	0.3054	0.1218	0.2273	0.0894
HPcST00	0.4035	0.1249	0.3063	0.0694	0.2170	0.0629
STdata00	0.1443	0.3514	0.1049	0.3064	0.2671	0.4425
black	0.0992	0.2989	0.1232	0.3287	0.1108	0.3139
amind	0.0035	0.0593	0.0045	0.0670	0.0052	0.0723
asian	0.0168	0.1284	0.0177	0.1317	0.0167	0.1281
hispanic	0.0785	0.2690	0.1083	0.3108	0.1262	0.3321
ltv	0.9336	0.0639	0.9430	0.0570	0.9414	0.0569
invest	0.0003	0.0178	0.0002	0.0152	0.0003	0.0175
refi	0.0626	0.2422	0.0273	0.1628	0.0392	0.1941
condo	0.0229	0.1495	0.0299	0.1703	0.0235	0.1516
direndor	0.9977	0.0481	0.9999	0.0073	1	0
urban	0.9990	0.0313	0.9989	0.0334	0.9221	0.2680
suburban	0.5203	0.4996	0.5486	0.4976	0.4951	0.5000
firstime	0.6358	0.4812	0.6688	0.4706	0.7171	0.4504
new	0.0918	0.2887	0.0941	0.2920	0.0612	0.2397
cbunmard	0.1068	0.3089	0.1256	0.3314	0.1257	0.3315
depnum	0.7957	1.1144	0.7992	1.1281	0.7281	1.0798
lqass	$3,654	$10,503	$3,654	$11,317	$3,732	$11,436
nocbinc	0.9130	0.2818	0.9999	0.0073	1.0000	0.0057
ageles25	0.1120	0.3154	0.1220	0.3272	0.1250	0.3307
age25_35	0.5323	0.4990	0.5045	0.5000	0.4939	0.5000
age35_45	0.2456	0.4304	0.2507	0.4334	0.2528	0.4346
buydown	0.0000	0.0021	0	0	0.0000	0
income	$39,295	$15,552	$40,694	$16,457	$41,582	$17,169
shrtmor	0.0418	0.2002	0.0233	0.1508	0.0212	0.1442
singlem	0.1360	0.3428	0.1356	0.3424	0.1437	0.3508
singlef	0.1524	0.3594	0.1639	0.3702	0.1689	0.3747
hval	$77,345	$27,885	$84,280	$30,871	$87,815	$31,841
hei20_38	0.6123	0.4872	0.6698	0.4703	0.6637	0.4725
hei38_50	0.0070	0.0836	0.0086	0.0921	0.0135	0.1153
hei50_	0.0017	0.0411	0.0014	0.0369	0.0013	0.0363
dti20_40	0.8440	0.3629	0.8003	0.3998	0.7562	0.4294
dti41_53	0.1184	0.3230	0.1691	0.3748	0.2131	0.4095
dti53_65	0.0035	0.0588	0.0032	0.0569	0.0040	0.0631
dti65_	0.0046	0.0675	0.0048	0.0693	0.0039	0.0621
ctblack	0.1021	0.1892	0.1022	0.1841	0.0988	0.1813
ctamind	0.0051	0.0100	0.0052	0.0101	0.0058	0.0142
ctasian	0.0210	0.0419	0.0231	0.0390	0.0219	0.0370
cthisp	0.0735	0.1334	0.0843	0.1422	0.0855	0.1458
ctincome	103.00	27.00	103.00	26.00	103.00	26.00
ctunemp	0.0564	0.0342	0.0560	0.0334	0.0577	0.0344
ctrent	0.3168	0.1768	0.3165	0.1788	0.3167	0.1764
herf2	547.64	338.64	358.58	215.07	347.68	204.44

Notes: The numbers of observations were as follows: 61,630 in 1992, 59,245 in 1994, and 90,894 in 1996. Averages for state indicators are not reported but are available upon request.

Table 2
Panel B

Means and Standard Deviations of FHA Default Rates and FICO Scores by Race

	1992 Sample		1994 Sample		1996 Sample	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
1. Default Rates by Race						
White	4.1%	19.9%	4.0%	19.6%	2.9%	16.7%
Black	8.1	27.3	7.6	26.6	4.8	21.5
Indian	5.6	22.9	5.0	21.9	4.4	20.5
Asian	6.3	24.3	6.0	23.8	3.7	18.8
Hispanic	11.0	31.3	8.5	27.8	5.4	22.6
Other	9.0	28.7	6.8	25.1	4.5	20.1
2. FICO Scores by Race						
White	690.4	57.6	679.3	58.8	667.6	59.9
Black	651.9	59.8	641.6	58.6	627.1	57.8
Indian	675.8	62.8	670.3	56.5	648.5	59.7
Asian	696.6	51.2	685.1	56.5	671.8	57.2
Hispanic	675.1	55.1	666.4	55.1	654.6	57.1
Other	676.4	61.1	663.9	64.0	654.3	61.6
Total	686.3	58.7	673.2	59.7	661.2	60.7

Table 3
Panel A

Logit Estimates of BCGH and Adjusted Models of FHA Claim Defaults: 1992 Sample

	BCGH Model		BCGH Enhanced		BCGH with Credit History		BCGH Enhanced with Credit History	
	Coefficient	z	Coefficient	z	Coefficient	z	Coefficient	z
armflag	0.08072	1.552	0.0875685	1.678	0.066676	1.267	0.0727927	1.376
FICOsb					-0.0083138	-24.324	-0.0083492	-24.362
FICO2b					-0.0089634	-26.951	-0.0089881	-26.945
FICO2b1					-0.008213	-24.488	-0.0082475	-24.517
HPrelPW			-0.9572414	-7.221			-0.9304016	-6.885
HPrelLL			-1.506888	-8.55			-1.473327	-8.2
LLmin			-0.7679523	-3.551			-0.7627812	-3.494
LLmax			-0.7410507	-5.771			-0.7133787	-5.44
HPcMSA00			-1.704427	-7.06			-1.759988	-7.169
HPcST00			-2.329031	-3.718			-2.400617	-3.794
STdata00			0.3888208	1.91			0.4321039	2.098
black	0.3224573	4.587	0.3109092	4.42	0.0707147	0.993	0.0617895	0.867
amind	-0.0519595	-0.167	-0.0840598	-0.269	-0.2517346	-0.794	-0.2777618	-0.873
asian	-0.1432994	-0.967	-0.123351	-0.834	-0.1548943	-1.032	-0.1381015	-0.921
hispanic	0.0525084	0.769	0.0250994	0.366	-0.0267851	-0.386	-0.0547474	-0.785
ltv	3.54214	8.61	3.937408	9.423	3.764907	9.051	4.133673	9.76
invest								
refi	0.0863526	0.85	-0.0426021	-0.413	0.1319719	1.273	-0.001876	-0.018
condo	0.2990165	2.827	0.2827292	2.629	0.346806	3.22	0.3296132	3.01
direndor	-0.3894606	-1.063	-0.4324207	-1.181	-0.2695191	-0.728	-0.3023873	-0.816
urban	-0.4703953	-0.927	-0.4564977	-0.896	-0.4286699	-0.839	-0.4288131	-0.837
suburban	-0.0587542	-1.353	-0.0780688	-1.782	-0.0492747	-1.119	-0.0668034	-1.502
firstime	0.0019149	0.042	-0.0075293	-0.165	-0.0277655	-0.602	-0.0369167	-0.797
new	-0.0256274	-0.358	0.0149611	0.207	-0.0250578	-0.346	0.0156417	0.214
cbunmard	-0.0992338	-1.521	-0.1205803	-1.843	-0.1022361	-1.524	-0.121451	-1.804
depnum	0.1885125	11.489	0.1898761	11.546	0.1467208	8.66	0.1480457	8.725
lqass	-0.0000221	-5.476	-0.0000226	-5.583	-0.0000111	-2.741	-0.0000117	-2.871
nocbinc	0.0147058	0.205	-0.0016697	-0.023	-0.0672604	-0.915	-0.0842205	-1.144
lqass2	1.67E-10	5.098	1.70E-10	5.149	1.03E-10	2.989	1.07E-10	3.095
ageles25	0.1078924	1.336	0.1061011	1.313	0.029607	0.362	0.0269101	0.328
age25_35	-0.1451351	-2.276	-0.1393326	-2.181	-0.1703411	-2.643	-0.1670303	-2.587
age35_45	-0.1134455	-1.657	-0.1036078	-1.511	-0.1144357	-1.653	-0.1052914	-1.518
buydown								
income	-0.0000173	-2.761	-0.0000144	-2.293	-0.0000179	-2.798	-0.0000153	-2.353
income2	7.56E-11	1.572	5.73E-11	1.177	7.46E-11	1.506	5.80E-11	1.15
shrtmor	-0.8201856	-5.098	-0.852407	-5.296	-0.7190738	-4.447	-0.7503761	-4.634
singlem	0.1166651	1.928	0.0989489	1.631	-0.03487	-0.412	-0.0463222	-0.547
singlef	-0.2446332	-3.864	-0.255573	-4.033	-0.3971834	-4.703	-0.4029814	-4.768
hval	-0.0000121	-3.998	1.81E-07	0.056	-9.19E-06	-3.096	2.99E-06	0.822
hval2	4.63E-11	3.937	1.44E-11	1.245	3.74E-11	3.333	5.78E-12	0.413
hei20_38	0.14355	2.194	0.164296	2.508	0.1359207	2.062	0.1576731	2.378
hei38_50	0.0927448	0.443	0.1509835	0.721	0.1519342	0.712	0.2064636	0.963
hei50_	-0.2084653	-0.438	-0.0960694	-0.202	-0.0477461	-0.098	0.073143	0.151
dti20_40	0.137317	0.973	0.1559844	1.104	0.1117225	0.778	0.1287434	0.896
dti41_53	0.196464	1.312	0.2156757	1.439	0.1544641	1.016	0.1703682	1.119
dti53_65	0.4053825	1.232	0.4617756	1.394	0.3111832	0.931	0.3637032	1.082
dti65_	0.0468467	0.144	0.1538885	0.477	-0.0145578	-0.044	0.0889913	0.272
ctblack	0.3993883	3.098	0.2873451	2.193	0.2500117	1.923	0.1411481	1.068
ctamind	0.4637573	0.165	0.6610806	0.24	-0.2786151	-0.094	-0.1446443	-0.049
ctasian	-0.8963245	-1.747	-0.6446346	-1.24	-0.8753817	-1.689	-0.595557	-1.133
cthisp	-0.0588109	-0.338	-0.1939855	-1.102	-0.1480691	-0.834	-0.2823061	-1.574
ctincome	-0.0054084	-4.698	-0.0045618	-3.848	-0.0049687	-4.282	-0.0041588	-3.477
ctunemp	0.3391881	0.449	0.8415873	1.095	0.2189529	0.284	0.7367948	0.942
ctrent	-0.2634155	-1.921	-0.2052842	-1.49	-0.1883689	-1.358	-0.1358308	-0.974
herf2	-0.0004694	-5.212	-0.0003784	-3.434	-0.0004147	-4.597	-0.0003521	-3.175
cons	-1.908644	-2.45	-2.553566	-3.24	3.521999	4.325	2.942388	3.565
No. of Obs	61,603		61,603		61,603		61,603	
Log Likelihood	-10,926.16		-10,869.98		-10,526.10		-10,471.23	
LR Chi2	3,112.42		3,224.78		9,912.54		4,022.28	
Prob > Chi2	0.0000		0.0000		0.0000		0.0000	
Pseudo R2	0.1247		0.1292		0.1567		0.1611	

Table 3
Panel B

Logit Estimates of BCGH and Adjusted Models of FHA Claim Defaults: 1994 Sample

	BCGH Model		BCGH Enhanced		BCGH with Credit History		BCGH Enhanced with Credit History	
	Coefficient	z	Coefficient	z	Coefficient	z	Coefficient	z
armflag	0.0827577	1.756	0.0827787	1.753	0.0230916	0.484	0.0236776	0.495
FICOsb					-0.0088783	-25.11	-0.0088969	-25.149
FICO2b					-0.009432	-27.31	-0.0094486	-27.349
FICO2b1					-0.0088829	-25.345	-0.0089035	-25.396
HPrelPW			-0.4331701	-3.099			-0.4692289	-3.317
HPrelLL			-0.7071988	-3.65			-0.7624727	-3.88
LLmin								
LLmax			-0.5212447	-3.652			-0.5304069	-3.663
HPcMSA00			-0.699967	-2.509			-0.7829256	-2.755
HPcST00			0.1524409	0.13			0.0832467	0.071
STdata00			-0.038127	-0.106			-0.0273194	-0.075
black	0.3358931	5.208	0.3340883	5.176	0.1001598	1.536	0.0978527	1.499
amind	0.0645044	0.226	0.0635792	0.222	0.0012715	0.004	-0.0015899	-0.006
asian	0.0719489	0.507	0.0805622	0.567	0.0898484	0.622	0.0983515	0.682
hispanic	0.0013117	0.019	-0.0080259	-0.119	-0.0606527	-0.891	-0.0705556	-1.034
ltv	4.205059	8.501	4.457156	8.893	4.440717	8.921	4.704489	9.331
invest								
refi	0.2344101	1.743	0.1868464	1.378	0.1918351	1.404	0.1374159	0.996
condo	0.0821435	0.704	0.0848459	0.72	0.1602912	1.351	0.1619226	1.351
direndor								
urban	0.708419	0.717	0.6927622	0.701	0.6318244	0.637	0.619896	0.624
suburban	-0.0274637	-0.622	-0.0293242	-0.658	-0.0268647	-0.602	-0.0295728	-0.656
firstime	-0.0407383	-0.9	-0.0437287	-0.964	-0.0821783	-1.797	-0.0863205	-1.883
new	-0.0215249	-0.306	-0.0039167	-0.055	-0.0137932	-0.194	0.0050631	0.07
cbunmard	-0.0196407	-0.314	-0.0257656	-0.411	0.0340465	0.525	0.0262171	0.404
depnum	0.1538412	9.443	0.1536083	9.407	0.108291	6.428	0.1075897	6.367
lqass	-0.000034	-7.168	-0.0000339	-7.142	-0.00002	-4.31	-0.0000199	-4.291
nocbinc								
lqass2	1.91E-10	5.268	1.90E-10	5.234	1.11E-10	2.853	1.10E-10	2.831
ageles25	0.2554362	3.272	0.2568115	3.288	0.2626815	3.321	0.2640342	3.336
age25_35	-0.1589884	-2.52	-0.1561576	-2.474	-0.1606282	-2.524	-0.157255	-2.47
age35_45	-0.032892	-0.495	-0.0309845	-0.466	-0.0325177	-0.485	-0.0296402	-0.442
buydown								
income	-0.0000112	-1.872	-0.0000108	-1.796	-0.0000135	-2.197	-0.0000131	-2.124
income2	5.11E-11	1.12	4.86E-11	1.066	6.12E-11	1.311	5.92E-11	1.271
shrtmor	-0.5658676	-3.063	-0.5867709	-3.172	-0.4795528	-2.575	-0.5027366	-2.696
singlem	0.027329	0.433	0.0241693	0.382	-0.1265183	-1.547	-0.1307572	-1.597
singlef	-0.0885532	-1.467	-0.0895292	-1.482	-0.2431574	-3.143	-0.2442204	-3.155
hval	-0.0000125	-3.496	-4.79E-06	-1.142	-8.35E-06	-2.283	4.47E-08	0.01
hval2	4.36E-11	2.89	2.14E-11	1.277	3.20E-11	2.082	7.77E-12	0.456
hei20_38	0.0669599	1.004	0.0709642	1.062	0.0114304	0.17	0.015505	0.23
hei38_50	0.0491805	0.232	0.0667016	0.314	0.0407369	0.19	0.0607125	0.282
hei50_	-0.5470641	-0.869	-0.5333124	-0.846	-0.6321233	-0.99	-0.6109106	-0.953
dti20_40	0.2779929	1.652	0.2822919	1.678	0.2744249	1.613	0.2806196	1.649
dti41_53	0.3641695	2.098	0.3728249	2.148	0.3318135	1.891	0.3420207	1.949
dti53_65	-0.0278049	-0.067	-0.0092657	-0.022	0.0386996	0.093	0.0494331	0.118
dti65_	0.3220259	0.931	0.3232689	0.933	0.3378314	0.961	0.3363656	0.954
ctblack	0.263196	2.032	0.2178909	1.657	0.1438168	1.099	0.0975193	0.734
ctamind	3.61516	1.521	3.648806	1.55	3.069006	1.298	3.035818	1.293
ctasian	-1.950908	-3.511	-1.829809	-3.247	-2.064359	-3.666	-1.964404	-3.438
cthisp	0.2303252	1.349	0.1081524	0.626	0.2145731	1.241	0.0838917	0.48
ctincome	-0.0036202	-3.215	-0.0038265	-3.27	-0.0029355	-2.59	-0.0030937	-2.624
ctunemp	1.982453	2.644	2.108219	2.759	1.965876	2.576	2.122514	2.732
ctrent	-0.5404016	-3.872	-0.5305703	-3.789	-0.4571435	-3.244	-0.4438396	-3.138
herf2	-0.0003071	-2.369	-0.0003837	-2.278	-0.0002229	-1.709	-0.0003079	-1.82
cons	-5.189573	-4.597	-5.43881	-4.779	0.5772399	0.5	0.3197331	0.275
No. of Obs	59,226		59,226		59,226		59,226	
Log Likelihood	-10,851.34		-10,839.96		-10,453.92		-10,441.69	
LR Chi2	1,834.08		1,856.84		2,628.91		2,653.38	
Prob > Chi2	0.0000		0.0000		0.0000		0.0000	
Pseudo R2	0.0779		0.0789		0.1117		0.1127	

Table 3
Panel C

Logit Estimates of BCGH and Adjusted Models of FHA Claim Defaults: 1996 Sample

	BCGH Model		BCGH Enhanced		BCGH with Credit History		BCGH Enhanced with Credit History	
	Coefficient	z	Coefficient	z	Coefficient	z	Coefficient	z
armflag	0.2011691	4.073	0.191247	3.861	0.1310401	2.614	0.1219421	2.425
FICOsb					-0.0114592	-29.711	-0.0114152	-29.564
FICO2b					-0.0118848	-31.618	-0.0118645	-31.528
FICO2b1					-0.0114261	-29.43	-0.0113974	-29.327
HPrelPW			-0.0837988	-0.66			-0.1111164	-0.857
HPrelLL			-0.241792	-1.486			-0.2274903	-1.367
LLmin			-0.0921466	-0.275			-0.0610332	-0.18
LLmax			-0.3153615	-2.062			-0.3187293	-2.043
HPcMSA00			-2.578488	-6.726			-2.482473	-6.379
HPcST00			-2.040813	-1.383			-1.987823	-1.329
STdata00			0.1196307	0.352			0.1763981	0.513
black	0.1595602	2.34	0.1474926	2.157	-0.1238323	-1.81	-0.1329035	-1.936
amind	0.0771149	0.303	0.0554078	0.217	-0.0875756	-0.339	-0.1024644	-0.395
asian	0.088847	0.566	0.0892765	0.568	0.1306225	0.819	0.1376053	0.863
hispanic	-0.0143165	-0.215	-0.0286093	-0.428	-0.0964563	-1.429	-0.1093161	-1.615
ltv	4.527842	8.132	4.698851	8.38	4.673305	8.384	4.836874	8.62
invest								
refi	0.136983	1.005	0.1126227	0.821	0.044693	0.324	0.0241283	0.174
condo	-0.028608	-0.198	0.0658314	0.452	0.0376787	0.257	0.1279239	0.865
direndor								
urban	-1.346815	-3.307	-1.313388	-3.204	-1.273879	-3.029	-1.217744	-2.86
suburban	-0.0090233	-0.194	0.0000592	0.001	-0.0087424	-0.186	0.0041985	0.088
firstime	0.0080221	0.156	0.0209153	0.404	-0.0471749	-0.903	-0.0344261	-0.657
new	-0.1833195	-1.981	-0.2013596	-2.167	-0.1656049	-1.77	-0.1821782	-1.939
cbunmard	0.0224938	0.337	0.0238034	0.357	0.0839694	1.204	0.0875274	1.253
depnum	0.1538376	8.506	0.1601073	8.846	0.0878766	4.697	0.0931301	4.972
lqass	-0.0000494	-8.596	-0.0000489	-8.522	-0.0000297	-5.45	-0.0000292	-5.355
nocbinc								
lqass2	2.73E-10	6.818	2.73E-10	6.801	1.75E-10	4.162	1.73E-10	4.09
ageles25	0.3346697	4.067	0.332585	4.041	0.3388534	4.062	0.3382265	4.054
age25_35	-0.0368918	-0.553	-0.0477331	-0.714	-0.05347	-0.793	-0.0619167	-0.917
age35_45	-0.0997416	-1.387	-0.1047402	-1.455	-0.1139356	-1.567	-0.1175474	-1.615
buydown								
income	-0.0000107	-1.671	-0.0000115	-1.792	-0.0000127	-1.921	-0.0000133	-1.999
income2	3.79E-11	0.765	4.05E-11	0.815	4.41E-11	0.857	4.49E-11	0.87
shrtmor	-0.8427525	-3.399	-0.8455805	-3.408	-0.682371	-2.723	-0.6850647	-2.733
singlem	0.3092211	5.199	0.3164211	5.316	0.1430783	1.848	0.1390323	1.794
singlef	-0.0683311	-1.097	-0.0609071	-0.977	-0.2414367	-3.136	-0.2431185	-3.156
hval	-0.0000128	-3.557	-9.49E-06	-2.412	-8.62E-06	-2.291	-5.41E-06	-1.306
hval2	4.93E-11	3.33	4.71E-11	3.1	3.89E-11	2.482	3.70E-11	2.28
hei20_38	0.1689063	2.338	0.1555566	2.149	0.114242	1.568	0.1028407	1.408
hei38_50	-0.0849734	-0.431	-0.0732675	-0.371	-0.0941506	-0.47	-0.0798518	-0.397
hei50_	-0.6862035	-0.895	-0.7369204	-0.96	-0.5747746	-0.74	-0.6088938	-0.783
dti20_40	0.0641242	0.38	0.062994	0.373	0.0357835	0.21	0.0307181	0.18
dti41_53	0.2036417	1.176	0.2045158	1.18	0.1350734	0.771	0.1299048	0.741
dti53_65	0.0066805	0.017	-0.0128794	-0.033	-0.1550978	-0.393	-0.1876732	-0.475
dti65_	0.1813858	0.469	0.1579312	0.408	0.1375459	0.348	0.1041318	0.263
ctblack	0.4845262	3.737	0.548403	4.167	0.2619351	2.001	0.3285567	2.476
ctamind	2.487196	1.484	2.121003	1.269	2.574359	1.549	2.232032	1.327
ctasian	-3.327441	-4.919	-2.616636	-3.885	-3.369549	-4.911	-2.70106	-3.952
cthisp	0.2050799	1.218	0.0863055	0.503	0.1517323	0.884	0.0385798	0.221
ctincome	-0.0045997	-3.766	-0.0065333	-5.14	-0.0038134	-3.073	-0.0057085	-4.419
ctunemp	2.71195	3.635	1.983699	2.602	2.536113	3.3	1.811842	2.316
ctrent	-0.4489919	-3.094	-0.4992824	-3.415	-0.2945133	-2.006	-0.3453635	-2.335
herf2	0.0000334	0.235	-0.000388	-2.033	0.0001164	0.808	-0.0003784	-1.956
cons	-4.113015	-5.742	-3.466299	-4.74	3.080995	4.045	3.691434	4.744
No. of Obs	74,538		74,538		74,538		74,538	
Log Likelihood	-10,503.53		-10,474.68		-9,967.94		-9,941.65	
LR Chi2	1,624.32		1,682.02		2,695.51		2,748.07	
Prob > Chi2	0.0000		0.0000		0.0000		0.0000	
Pseudo R2	0.0718		0.0743		0.1191		0.1214	

Table 4

Logit Estimates of BCGH and Adjusted Models of FHA Claim Defaults: 4 Year Defaults for 1992 and 1994 Samples

	1992 Sample				1994 Sample			
	BCGH Enhanced		BCGH Enhanced with Credit History		BCGH Enhanced		BCGH Enhanced with Credit History	
	Coefficient	z	Coefficient	z	Coefficient	z	Coefficient	z
armflag	0.036169	0.588	0.0221217	0.355	0.0910603	1.795	0.0309542	0.603
FICOsb			-0.0091346	-22.937			-0.0091483	-23.972
FICO2b			-0.0099061	-25.557			-0.0097655	-26.185
FICO2b1			-0.0089935	-23.037			-0.0092324	-24.384
HPrelPW	-0.6188917	-4.003	-0.5774896	-3.657	-0.4371838	-2.927	-0.4980386	-3.293
HPrelLL	-0.9753462	-4.673	-0.9085549	-4.266	-0.6452119	-3.134	-0.7540992	-3.607
LLmin	-0.5981494	-2.276	-0.5770626	-2.175				
LLmax	-0.520139	-3.384	-0.4796811	-3.051	-0.5174012	-3.369	-0.5664553	-3.618
HPcMSA00	-4.910533	-7.617	-5.056197	-7.725	-0.2550955	-0.496	-1.045493	-1.984
HPcST00	-5.527539	-5.562	-5.748493	-5.732	0.6935159	0.41	-0.600782	-0.35
STdata00	0.1630615	1.491	0.2069278	1.865	0.0064677	0.022	0.126172	0.425
black	0.2693133	3.308	-0.0055536	-0.067	0.3947461	5.735	0.149903	2.155
amind	0.0158123	0.046	-0.1952624	-0.554	-0.0144751	-0.045	-0.078488	-0.241
asian	-0.1802164	-0.994	-0.19259	-1.049	0.057125	0.369	0.0785626	0.5
hispanic	-0.1285034	-1.558	-0.2191477	-2.615	0.010753	0.148	-0.0476058	-0.652
ltv	3.589746	7.357	3.801077	7.689	4.317657	8.017	4.554346	8.412
refi	0.0363594	0.306	0.1008668	0.83	0.05936	0.388	-0.0068564	-0.044
condo	0.3205293	2.494	0.3705329	2.835	0.0735185	0.586	0.1603643	1.259
direndor	-0.533152	-1.252	-0.3921957	-0.911				
urban	-0.1473396	-0.211	-0.1029639	-0.147				
suburban	-0.0806591	-1.57	-0.0713016	-1.367	-0.0442499	-0.92	-0.0462516	-0.951
firstime	-0.0257431	-0.481	-0.0577111	-1.065	-0.030405	-0.618	-0.0716113	-1.442
new	0.0140881	0.166	0.0080284	0.093	0.0029231	0.038	0.012751	0.164
cbunmard	-0.1172503	-1.514	-0.1229295	-1.544	-0.0557024	-0.821	0.0082638	0.118
depnum	0.2119811	11.405	0.1697475	8.805	0.1498392	8.561	0.1038976	5.72
lqass	-0.000029	-5.701	-0.0000165	-3.269	-0.0000337	-6.602	-0.0000196	-3.947
nocbinc	0.1139253	1.308	0.0196188	0.221				
lqass2	1.94E-10	4.734	1.22E-10	2.796	1.94E-10	5.105	1.15E-10	2.853
ageles25	0.1648969	1.758	0.0788183	0.829	0.2608176	3.087	0.2719948	3.179
age25_35	-0.1605012	-2.146	-0.1906118	-2.522	-0.1505465	-2.205	-0.1525073	-2.217
age35_45	-0.1271554	-1.586	-0.1317755	-1.625	-0.026338	-0.366	-0.0258801	-0.357
income	-0.000021	-2.877	-0.0000211	-2.787	-0.0000137	-2.132	-0.0000152	-2.297
income2	1.03E-10	1.842	1.00E-10	1.72	6.68E-11	1.378	7.27E-11	1.464
shrtmor	-0.9105307	-4.62	-0.7871886	-3.976	-0.6302049	-2.986	-0.5503835	-2.591
singlem	0.0301411	0.418	-0.1309232	-1.337	0.0576939	0.854	-0.1288073	-1.483
singlef	-0.267226	-3.612	-0.4314022	-4.457	-0.1233637	-1.873	-0.310752	-3.742
hval	-7.63E-07	-0.208	1.98E-06	0.466	-3.86E-06	-0.87	1.32E-06	0.291
hval2	1.53E-11	1.191	6.77E-12	0.417	2.24E-11	1.268	8.51E-12	0.473
hei20_38	0.1417078	1.844	0.1334559	1.717	0.0536167	0.738	0.0006345	0.009
hei38_50	0.2264818	0.936	0.2905005	1.17	0.0039649	0.017	0.0100314	0.043
hei50_	-0.3383237	-0.531	-0.1649096	-0.256	-0.8736978	-1.171	-0.9627368	-1.27
dti20_40	0.1246548	0.754	0.1023999	0.61	0.2854078	1.543	0.2915035	1.56
dti41_53	0.1302617	0.741	0.0961285	0.539	0.3668131	1.924	0.3466558	1.8
dti53_65	0.0350533	0.081	-0.0894404	-0.205	-0.1215336	-0.256	-0.0558704	-0.117
dti65_	-0.0443592	-0.114	-0.0920461	-0.234	0.3895867	1.047	0.4196191	1.111
ctblack	0.5306705	3.542	0.376392	2.493	0.2255741	1.608	0.1088056	0.768
ctamind	-0.1480464	-0.044	-1.153688	-0.327	3.239166	1.186	2.324492	0.844
ctasian	-1.227249	-1.852	-1.186441	-1.772	-1.842516	-3.056	-1.96613	-3.223
cthisp	0.2400084	1.167	0.1522251	0.726	0.1085086	0.584	0.0877736	0.467
ctincome	-0.0047429	-3.391	-0.0043112	-3.06	-0.0040527	-3.211	-0.0034271	-2.696
ctunemp	0.2129366	0.236	0.1024997	0.112	1.768722	2.132	1.736843	2.057
ctrent	-0.3305859	-2.042	-0.2579821	-1.575	-0.6144279	-4.056	-0.5307187	-3.472
herf2	-0.0002991	-2.275	-0.0002673	-2.023	-0.0004942	-2.638	-0.0004175	-2.218
cons	-3.666093	-3.696	2.266834	2.198	-4.837761	-7.926	1.06865	1.63
No. of Obs	61,612		61,612		59,167		59,167	
Log Likelihood	-8,487.45		-8,122.27		-9,588.13		-9,224.34	
LR Chi2	2,038.44		2,768.81		1,590.35		2,317.98	
Prob > Chi2	0.0000		0.0000		0.0000		0.0000	
Pseudo R2	0.1072		0.1456		0.0766		0.1116	

Table 5

Black/White Differentials by Risk Class

1992 Sample

Risk Class	Actual Default Rate (All Races)	Logit Estimates of Black Coefficient		Ratios of Black to White Estimated Default Probabilities	Difference Between Black and White Estimated Default Probabilities
Based on FICO Score		**Coefficient**	**z**		
score<=595	11.451	-0.0589	-0.407	0.9522	-0.0050
score> 595 & score<= 625	8.495	0.0403	0.252	1.0353	0.0027
score> 625 & score<= 660	7.174	-0.0322	-0.227	0.9722	-0.0017
score> 660	3.491	0.1547	1.246	1.1494	0.0042
Based on Estimated Log Odds of Default					
pctile>85	17.857	-0.0707	-0.7507	0.9448	-0.0095
pctile> 70 & pctile<=85	7.472	0.0286	0.2127	1.0266	0.0020
pctile> 50 & pctile<=70	3.914	0.1762	0.9895	1.1820	0.0070
pctile<=50	1.116	0.5122	2.1591	1.6453	0.0068

1994 Sample

Risk Class	Actual Default Rate (All Races)	Logit Estimates of Black Coefficient		Ratios of Black to White Estimated Default Probabilities	Difference Between Black and White Estimated Default Probabilities
Based on FICO Score		**Coefficient**	**z**		
score<=595	10.562	0.0389	0.3232	1.0340	0.0032
score> 595 & score<= 625	8.044	0.1823	1.3083	1.1745	0.0120
score> 625 & score<= 660	6.385	-0.0094	-0.0699	0.9915	-0.0005
score> 660	2.944	0.2035	1.6019	1.2110	0.0050
Based on Estimated Log Odds of Default					
pctile>85	14.887	0.0129	0.1388	1.0108	0.0016
pctile> 70 & pctile<=85	7.625	0.197	1.5439	1.1971	0.0144
pctile> 50 & pctile<=70	4.435	-0.0662	-0.3982	0.9392	-0.0027
pctile<=50	1.485	0.2965	1.4759	1.3346	0.0047

1996 Sample

Risk Class	Actual Default Rate (All Races)	Logit Estimates of Black Coefficient		Ratios of Black to White Estimated Default Probabilities	Difference Between Black and White Estimated Default Probabilities
Based on FICO Score		**Coefficient**	**z**		
score<=595	7.941	-0.2038	-1.9419	0.8349	-0.0123
score> 595 & score<= 625	5.419	-0.0949	-0.685	0.9158	-0.0041
score> 625 & score<= 660	3.964	-0.2314	-1.4235	0.8049	-0.0072
score> 660	1.423	0.1023	0.5616	1.1043	0.0013
Based on Estimated Log Odds of Default					
pctile>85	11.27	-0.2415	-2.4495	0.8096	-0.0211
pctile> 70 & pctile<=85	4.991	0.008	0.0545	1.0076	0.0004
pctile> 50 & pctile<=70	3.052	-0.118	-0.705	0.8924	-0.0032
pctile<=50	0.907	0.0785	0.3801	1.0802	0.0007

Note: All default probabilities are calculated using characteristics of whites.

Table A-1

Cumulative Percentage of Claims Completed in Each Calendar Year, by Year of Default

	Year of Default								
Year Claim Completed	1992	1993	1994	1995	1996	1997	1998	1999	2000
1992	1.07								
1993	59.87	12.36							
1994	90.13	72.01	14.78						
1995	95.92	93.75	71.10	11.09					
1996	98.50	97.62	92.06	61.22	10.49				
1997	99.57	99.13	98.08	89.92	65.62	10.27			
1998	99.57	99.66	99.28	97.50	92.26	71.01	13.12		
1999	100.00	99.97	99.81	99.43	98.67	95.98	81.55	26.96	
2000	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Note: Data consist of all defaults that resulted from endorsed applications (other than streamline refinaces) in 1992, 1994 and 1996, and which appeared as claims by July 2000

Table A-2

Mean Number of Days from Default to Completion of Claim Process, by Year of Default and Race[1]

In States Without Judicial Foreclosure

Year of Default	Mean Number of Days		Counts		Difference in Means
	White[2]	Black	White	Black	
1992	409	452	159	64	-43
1993	378	475	1252	400	-97
1994	388	468	2175	662	-80
1995	437	539	4711	1578	-102
1996	409	518	5515	1685	-109
1997	394	491	7668	2316	-97
1998	354	413	6411	1825	-59
1999	288	300	2669	660	-12

In States With Judicial Foreclosure

Year of Default	Mean Number of Days		Counts		Difference in Means
	White	Black	White	Black	
1992	510	612	109	52	-102
1993	444	594	1098	287	-150
1994	473	578	1767	534	-105
1995	526	654	3586	1247	-128
1996	548	667	3717	1265	-119
1997	512	596	5479	1590	-84
1998	447	497	4634	1029	-50
1999	334	336	1446	228	-2

[1]Data consist of all defaults that resulted from endorsed applications (other than streamline refinances) in 1992, 1994 and 1996, and which appear as claims by July 2000.
[2]White includes races/ethnicities other than Black, Hispanic, Asian, and American Indian.

Table A-3

Cumulative Percentage of Claims Completed at 3-Month Intervals After Default, by Year of Default and Race[1]

Panel A
White[2] Borrowers

Months After Default	Year of Default 1992	1993	1994	1995	1996	1997	1998	1999
3	1.12	1.45	1.55	1.70	1.91	2.14	2.20	2.67
6	14.55	17.23	17.55	11.57	9.04	8.25	8.95	15.29
9	35.45	39.96	38.86	29.12	26.44	26.32	30.93	52.30
12	52.99	58.51	57.48	46.84	46.35	47.27	55.41	82.36
15	67.54	71.79	70.04	60.62	62.63	65.22	74.21	96.01
18	78.36	82.68	79.60	72.10	75.23	77.96	87.24	100.00
21	85.07	88.51	86.28	80.75	83.31	86.41	94.92	100.00
24	89.93	92.89	90.21	87.03	89.06	91.80	98.37	100.00
27	92.16	95.40	93.51	91.13	92.60	95.22	99.68	100.00
30	94.78	96.38	95.41	94.06	94.95	97.60	100.00	100.00
33	95.90	97.53	96.37	95.89	96.74	99.00	100.00	100.00
36	96.27	98.17	97.34	97.20	97.82	99.68	100.00	100.00
39	96.27	98.55	98.27	97.89	98.62	99.91	100.00	100.00
42	97.01	98.85	98.71	98.51	99.24	100.00	100.00	100.00
45	97.76	99.23	99.01	98.93	99.66	100.00	100.00	100.00
48	98.13	99.49	99.21	99.33	99.88	100.00	100.00	100.00
51	98.13	99.53	99.47	99.63	99.99	100.00	100.00	100.00
54	98.88	99.66	99.54	99.75	100.00	100.00	100.00	100.00
57	100.00	99.70	99.67	99.84	100.00	100.00	100.00	100.00
60	100.00	99.74	99.72	99.96	100.00	100.00	100.00	100.00
63	100.00	99.79	99.82	99.99	100.00	100.00	100.00	100.00
66	100.00	99.91	99.85	100.00	100.00	100.00	100.00	100.00
72	100.00	99.96	100.00	100.00	100.00	100.00	100.00	100.00

[1]Data consist of all defaults that resulted from endorsed applications (other than streamline refinances) in 1992, 1994 and 1996, and which appear as claims by July 2000.
[2]White includes races/ethnicities other than Black, Hispanic, Asian, and American Indian

Table A-3

Cumulative Percentage of Claims Completed at 3-Month Intervals After Default, by Year of Default and Race[1]

Panel B
Black Borrowers

Months After Default	Year of Default							
	1992	1993	1994	1995	1996	1997	1998	1999
3	0.86	0.44	0.50	0.50	0.75	0.49	0.74	2.25
6	8.62	14.41	15.89	7.40	4.68	3.51	5.92	13.29
9	28.45	28.68	35.62	21.17	17.05	15.54	22.07	52.59
12	43.10	43.52	46.82	33.73	28.92	32.13	42.50	80.41
15	56.03	57.50	59.36	45.73	44.44	48.26	61.28	95.83
18	64.66	68.41	67.73	58.19	57.15	62.88	78.28	100.00
21	77.59	77.00	75.08	68.25	67.86	74.12	90.29	100.00
24	83.62	83.41	80.02	74.73	76.85	82.82	96.22	100.00
27	87.07	86.90	84.28	80.53	83.08	89.02	99.47	100.00
30	89.66	89.08	87.37	85.38	87.36	93.65	100.00	100.00
33	93.10	91.85	90.55	89.27	91.36	97.21	100.00	100.00
36	94.83	92.58	92.47	92.07	94.10	99.08	100.00	100.00
39	94.83	94.03	94.57	94.23	96.00	99.85	100.00	100.00
42	97.41	95.49	96.32	95.50	97.69	100.00	100.00	100.00
45	98.28	96.65	97.24	96.57	98.92	100.00	100.00	100.00
48	98.28	96.94	97.91	97.52	99.63	100.00	100.00	100.00
51	99.14	97.38	98.16	98.58	99.93	100.00	100.00	100.00
54	99.14	98.25	98.66	99.33	100.00	100.00	100.00	100.00
57	99.14	98.54	99.00	99.65	100.00	100.00	100.00	100.00
60	99.14	98.98	99.50	99.82	100.00	100.00	100.00	100.00
63	99.14	99.13	99.67	100.00	100.00	100.00	100.00	100.00
66	99.14	99.56	99.92	100.00	100.00	100.00	100.00	100.00
72	99.14	99.56	100.00	100.00	100.00	100.00	100.00	100.00
78	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

[1]Data consist of all defaults that resulted from endorsed applications (other than streamline refinances) in 1992, 1994 and 1996, and which appear as claims by July 2000.

Table A-4

Default Rates and Logit Estimates Based on Defaults Through April 30, 1998

Panel A
Default Rates

Race	1992 Sample Mean	1992 Sample Std. Dev.	1994 Sample Mean	1994 Sample Std. Dev.	1996 Sample Mean	1996 Sample Std. Dev.
White	3.7%	19.0%	3.3%	18.0%	1.6%	12.7%
Black	7.6	26.5	6.8	25.1	3.2	17.6
Indian	4.6	20.9	3.9	19.3	2.7	16.3
Asian	5.4	22.6	5.0	21.8	1.9	13.7
Hispanic	10.2	30.3	7.3	26.0	3.3	17.9
Other	8.1	27.3	6.5	24.6	3.6	18.8

Panel B
Logit Estimates of Black Coefficient

	BCGH Model		BCGH Enhanced		BCGH with Credit History		BCGH Enhanced with Credit History	
	Coefficient	z	Coefficient	z	Coefficient	z	Coefficient	z
1992 Logit	0.3347559	4.601	0.321087	4.411	0.0783777	1.064	0.0675679	0.0917
1994 Logit	0.3932796	5.716	0.3928847	5.707	0.1501085	2.159	0.1494924	2.149
1996 Logit	0.2182836	2.613	0.2071689	2.474	-0.0846944	-1.015	-0.0922498	-1.103