Comparing Loan Performance Between Races as a Test for Discrimination

George Galster
The Urban Institute

Abstract

This critique argues that analysis of loan performance data cannot provide a clear indicator of prejudice-motivated or profit-motivated mortgage market discrimination. It evaluates the Berkovec, Canner, Gabriel, and Hannan model by considering various possible assumptions that can be made by the lender and by the researcher regarding the dimensions of a loan applicant's creditworthiness.

The issue of discrimination directed toward minority applicants for mortgages has received a marked increase in public attention during the 1990s. A variety of evidence, including disparities observed in Home Mortgage Disclosure Act (HMDA) data, stories of blatant discrimination, a smattering of court cases, pilot studies using paired testers, and multivariate statistical analyses of lenders' loan application files, has combined to substantiate the view that discrimination in mortgage lending exists. (See Galster, 1992; Yinger, 1993; and Cloud and Galster, 1993.)

More recently, this view has been challenged by an entirely different analytical perspective. First articulated by Peterson (1981), and later expounded by Becker (1993) and Brimelow and Spencer (1993), this position argues that discrimination in the mortgage underwriting process can be deduced by examining the racial and ethnic differences in the performance of loans granted. In its simplest form, the argument states that underwriters who are motivated by an irrational prejudice against minority applicants will accept their applications only if they exceed the typical minimal standards of loan approval applied to white applicants. Thus their prejudice leads them to reject qualified prospective minority borrowers. Assuming that qualified minority applicants who are rejected cannot secure mortgages elsewhere, the result of such prejudice-based discrimination should be the creation of a pool of minority borrowers who are more qualified than they would be in the absence of such discrimination. Logically, then, minorities should evince better loan performance (for example, lower default rates) than whites if discrimination is present. Observations to the contrary have led the aforementioned analysts to conclude that there is no discrimination in mortgage lending.

Previous critiques of this line of reasoning have already discredited it (Yinger, 1993; Galster, 1993; and Tootell, 1993). To recapitulate only the central criticism: Although the *marginal* minority borrower will be better qualified (and thus have a better performing loan) than the marginal white borrower in the presence of prejudice-based discrimination, the relationship between the *average* minority and white borrowers cannot be accurately ascertained without resorting to arbitrary assumptions about the cross-race distributions of creditworthiness among borrowers.

In a 1995 paper, Berkovec, Canner, Gabriel, and Hannan (BCGH) attempt to provide a more sophisticated defense of the default approach for testing lender discrimination. Their contribution is to model the creditworthiness (C) of an applicant as the sum of two sets of factors, the first observable both to the lender and to the research analyst and the second observable (or measurable) only to the lender. BCGH proceed to discuss prejudiced lenders' behavior along the lines articulated above, except that they further classify the underwriting decision into those applicants qualified for conventional mortgages, those qualified only for FHA mortgages, and those unqualified for any mortgage. Higher standards for C are established for minorities at both conventional/FHA and FHA/no-loan thresholds by prejudiced lenders, leading to a somewhat better qualified pool of minority holders of both conventional and FHA mortgages.

As illustrated with hypothetical data in table 1, white applicants require a minimum C of 60 to obtain a conventional loan and 20 to obtain an FHA loan. Prejudice-based discrimination might lead to higher C thresholds for minorities, such as 65 and 25, respectively. Thus the minority applicant with C=61 is rejected for a conventional loan and relegated to the FHA borrower pool. Similarly, the minority applicant with C=22 is rejected for all loans. Both acts raise the average C among minority FHA borrowers.

These data are insufficient to produce the result that the average C value for minority FHA borrowers is higher (that is, default rates are lower, on average) than the average C for whites. As the data in table 1 make plain, the average C value in the minority FHA pool is improved as a result of discrimination, but not necessarily to the point that it exceeds the average C for whites. To their credit BCGH note that such a comparison of average loan performance across races creates an ambiguous indicator of discrimination.

In its place, they offer a subtle new argument that attempts to compare the treatment of the *marginal* applicants. Recall that they assume that some of the elements comprising C are not measured (or statistically controlled) by the researcher who is analyzing the lenders' loan files. Discrimination should systematically result in these unobserved factors favoring minorities, due to the higher loan approval standards to which minorities are held. Thus, BCGH argue, if one statistically controls for observable characteristics of the applicant, a finding that minority status—which serves as a proxy for the unobserved characteristics—is associated with better loan performance (for example, a lower rate of default) would constitute evidence of discrimination. Inasmuch as BCGH attain the opposite finding in their empirical work, they conclude that prejudice-based discrimination is absent.

Whether loan performance provides a valid indicator of lender discrimination depends on the way one specifies the factors that constitute the index of creditworthiness (C) that lenders observe. C is assumed by BCGH to be directly related to loan performance (or profitability) and inversely related to defaults. As argued below, regardless of the way the components of C are specified the BCGH procedure does not yield a valid estimate of discrimination, although the precise reasoning that illustrates this conclusion varies in each of the four cases considered.

Table 1

Hypothetical Illustration of Berkovec, Canner, Gabriel, and Hannan Framework

Creditworthiness Scores of Applicants	Minority	Majority
	10	
	16	16
		18
20 = Majority threshold - - - - - - - - - - -		
for FHA loan	22	
25 = Minority threshold - - - - - - - - - - -		
for FHA loan	32	32
	42	
		50
		59
60 = Majority threshold - - - - - - - - - - -		
for conventional loan	61	
65 = Minority threshold - - - - - - - - - - -		
for conventional loan	71	
		74
	81	
		84
		94

Average Creditworthiness Scores of Applicants	Minority	Majority
No discrimination		
Rejected pool	14	17
FHA pool	32	47
Conventional pool	71	84
Discrimination		
Rejected pool	16	17
FHA pool	45	47
Conventional pool	76	84

Source: Author

Following the BCGH hypothesis, one assumes that research analysts (not lenders) can observe a set of characteristics (X) of the loan and the applicant that are related to C through a vector of coefficients (β) plus an error term indicating characteristics observable only to the lender (ε), such that $C = X\beta + \varepsilon$.

Case 1: All components of C are observable ($\varepsilon = 0$), and race of applicant (R)[1] is not a component of X, and is uncorrelated with all components of X.

In this case the research analyst has completely specified a model that reflects the thinking of the lender. Although the distribution of X (and thus C) may well differ according to R, a loan performance regression that *controls* for X cannot help but yield a statistically insignificant coefficient for R. In other words if everything that affects C is controlled and R does not, by assumption, affect C, there can only be a zero coefficient for R in the BCGH regressions, *regardless of whether or not the lender was discriminating.* Comparisons of the hypothetical data in table 1 that have the same C value could not reveal any independent racial differential. If the lender were discriminating in the way hypothesized by BCGH, one would likely observe lower C values for *some* of the whites accepted for loans than one would observe for *any* minorities accepted for similar loans. Recall, however, that this is not the BCGH test.

Case 2: Not all components of C are observable to the researcher, and R is not a component of X and is not correlated with X or ε.

In this case there are omitted variables in the loan performance equation (such as components of ε), but there is no bias in the R coefficient because R and ε are assumed to be independent. In a nondiscriminatory world, there would be no reason for the R coefficient to be statistically significant. But if there were prejudice-based discrimination by lenders, forcing approved minority applicants to possess higher values of unobserved characteristics for any level of observable ones, R would tend to be significant. Yet such a theory would violate the assumption of this case. That is, we cannot begin by making the assumptions of case 2, because a particular finding of the test would violate them.

Case 3: Not all components of C are observable, and R is not a component of X but is correlated with ε.

In this case the BCGH regressions will suffer from omitted-variables bias (that is, the components of ε), the degree of which will be proportional to the correlation between R and ε. In such a case, the coefficient of R will be biased and any interpretation of it misleading. Unfortunately, it is precisely such a correlation between R and ε that is the essence of the BCGH test, inasmuch as they assume that the *only* source of such a correlation is lender discrimination that reputedly forces minorities to have superior unobserved characteristics (ε). As Galster (1993) and Yinger (1993) have argued, however, there are persuasive reasons for a high correlation between R and ε, even in the absence of prejudice-based lending discrimination. Thus the BCGH test cannot be conclusive if one admits these alternative sources. In other words BCGH assume that discrimination makes the unobservables of minorities superior to those of whites, but if other factors skew the unobservables in the opposite direction, the net result will be an empirical situation in which the reputed discrimination effect will be obscured.[2]

Deeper analysis of the preceding cases reveals a more fundamental conundrum for BCGH. The key feature of the BCGH test is that it *necessitates* the existence of applicant variables observed by the lender that are not observed (or controlled) by the research analyst. These would appear as higher hurdles erected for minority applicants by the prejudiced, discriminatory lender, producing an observed correlation between R and ε.

Thus the BCGH method depends on omitted variables to bias the coefficient of R in a discriminatory setting. Unfortunately, if too many variables are omitted, the model loses probity for other reasons.

To ascertain why this key methodological feature is flawed, consider two extremes of unobserved variables (that is, variables omitted from the regression). Case 1, which has none, showed that the coefficient of R was statistically insignificant regardless of the absence or presence of discrimination. The other extreme is embodied in a simple comparison of the average loan performance of white borrowers and minority borrowers, equivalent to estimating a regression with R as the only independent variable. For reasons explained above in the discussion of table 1, such a test is similarly incapable of providing an indicator of discrimination, as demonstrated by Galster (1993), Yinger (1993), and Tootell (1993).

Now let us return to the completely specified model of case 1 and add one unobserved variable (Z). According to BCGH logic, the R coefficient in a loan performance regression should provide an indicator of discrimination if it proves correlated with Z and thus proves statistically significant through omitted-variables bias (assuming Z is strongly correlated with C). Whether this indicator will manifest itself is doubtful, depending not only on the correlation between R and Z but also on the correlation between R and X (potential multicollinearity). As one continues this process by positing an increased number of unobserved Z variables that are more strongly related to C and fewer observed (controlled) X variables, it becomes more plausible that R will be more highly correlated with Z but less plausible that such a correlation can be attributed to discrimination. The most extreme case, in which all variables are unobserved (and uncontrolled), makes this point clear. Even with discrimination, minorities accepted for loans could, as a group, have lower C values than whites (see table 1). Thus the BCGH method is deeply flawed, because it cannot extricate itself from a dilemma that renders the test ambiguous. If there are too few[3] unobserved (uncontrolled) variables, the omitted-variables bias is too weak to be due to discrimination. If there are too many[4] unobserved variables, the signal is strong but meaningless, because it cannot credibly be attributed to discrimination.

Case 4: All components of C are observable, and R is a component of X.

In this case it is assumed that race itself is a strong and unique predictor of creditworthiness, due perhaps to less stable incomes or the greater likelihood of declining property values associated with minority neighborhoods. Given such assumptions, many lenders will assess two otherwise identically qualified applicants differently on the basis of R, because they have different C values. When that occurs, discrimination due to prejudice becomes indistinguishable from discrimination motivated by profit; neither is clearly observable by the BCGH method.

Should an assessment such as that described in case 4 produce differential treatment, the BCGH loan performance regression that controls for X will observe a statistically significant coefficient for R. One notes, however, that such a result is not produced because lenders are creating higher C standards for minorities; rather, for the same X variable, minorities have a lower C factor. If a lender does not discriminate, the statistical result will nevertheless be the same. More less-qualified minorities may be granted loans than under a discriminatory regime because, although their X values are identical to those of whites, they have lower C values. Thus the BCGH regression would show an identical coefficient for R, regardless of whether or not the lender is discriminating. Only the pool of marginally accepted borrowers would distinguish the two behaviors.

To their credit BCGH admit that their method is not designed to test for discrimination motivated by profit maximization (statistical discrimination), such as that posited in case 4. The point to bear in mind is that profit-motivated discrimination obscures the search for discrimination motivated by prejudice.

BCGH have conducted a careful study of racial-ethnic differentials in the performance of FHA mortgages. Unfortunately, they exaggerate when they assert that their findings suggest an absence of lending discrimination based on prejudice. Examination of loan performance, even in the sophisticated BCGH manner, cannot yield unambiguous indications of the absence or presence of discrimination, whether it be motivated by prejudice or by profit.

Author

Dr. Galster is Director of Housing Research at The Urban Institute's Center for Public Finance and Housing. He earned his Ph.D. in Economics from the Massachusetts Institute of Technology and received undergraduate degrees from Wittenberg and Case Western Reserve Universities. Dr. Galster has had broad experience with academic, governmental, nonprofit, and for-profit research and policymaking organizations, including HUD and numerous municipalities, community organizations, and civil rights groups.

Notes

1. In BCGH, R represents the conditional probability of default.
2. Note that this situation would occur if there were large numbers of variables correlated with C that were unobserved by both researchers and mortgage lenders.
3. Few means the number and degree of correlation with C.
4. Many means the number and degree of correlation with C.

References

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

Berkovec, James, Glenn Canner, Stuart Gabriel, and Timothy Hannan. January 1995. "Discrimination, Default, and Loss in FHA Mortgage Lending." Paper presented at the American Real Estate and Urban Economics Association Annual Meeting, Washington, D.C.

Brimelow, Peter, and Leslie Spencer. January 1993. "The Hidden Clue," *Forbes* 4:48.

Cloud, Cathy, and George Galster. 1993. "What Do We Know About Racial Discrimination in Mortgage Markets?" *Review of Black Political Economy* 22(1):101–120.

Galster, George. 1992. "Research on Discrimination in Housing and Mortgage Markets: Assessment and Future Directions," *Housing Policy Debate* 3(2):639–683.

______. 1993. "The Facts of Lending Discrimination Cannot Be Argued Away by Examining Default Rates," *Housing Policy Debate* 4(1):141–146.

Peterson, Richard. Autumn 1981. "An Investigation of Sex Discrimination in Commercial Banks' Direct Consumer Lending," *Bell Journal of Economics* 12:547–561.

Tootell, Geoffrey. September/October 1993. "Defaults, Denials, and Discrimination in Mortgage Lending," *New England Economic Review*, pp. 45–51.

Yinger, John. May 18, 1993. "Discrimination in Mortgage Lending: A Literature Review." Paper presented at the Conference on Discrimination and Mortgage Lending: Research and Enforcement, Washington, D.C.