



University of
Connecticut

Department of Economics Working Paper Series

Other Things Being Equal: A Paired Testing Study of Discrimination in Mortgage Lending.

Margery Austin Turner
Urban Institute

Erin Godfrey
New York University

Stephen L. Ross
University of Connecticut

Robin R. Smith
Urban Institute

Working Paper 2003-09R

March 2003, revised May 2004

341 Mansfield Road, Unit 1063
Storrs, CT 06269-1063
Phone: (860) 486-3022
Fax: (860) 486-4463
<http://www.econ.uconn.edu/>

Abstract

This article summarizes a recently completed study, funded by the U.S. Department of Housing and Urban Development (HUD) and conducted by the Urban Institute, of discrimination against black and Hispanic homebuyers when they visit mortgage lending institutions in two major metropolitan markets to make pre-application inquiries. It represents the first application of paired testing to rigorously measure discrimination in the mortgage lending process. The paired tests disclosed significant levels of adverse treatment on the basis of race and ethnicity, with African Americans and Hispanics receiving less information and assistance than comparable whites, even at this very early stage in the application process.

Other Things Being Equal: A Paired Testing Study of Discrimination in Mortgage Lending

Introduction

Considerable evidence indicates that minority homebuyers are less likely than whites to obtain mortgage loans and, if they are successful, receive less favorable loan amounts and terms. Racial and ethnic disparities in loan denial rates are consistently found in data collected as part of the Home Mortgage Disclosure Act (HMDA) across income categories, loan types, and markets (Avery, Beeson, and Sniderman, 1996; Holloway and Wyly, 2002). More formal analyses of lender underwriting behavior, such as Munnell, Tootell, Browne, and McEneaney (1996), Schafer and Ladd (1981), and King (1980), all find evidence of racial and/or ethnic differences in mortgage underwriting. Finally, the few studies that have considered the price of credit, such as Crawford and Rosenblatt (1999) and Courchane and Nickerson (1997), also find significant racial differences in lending practices.

However, considerable disagreement persists about the extent to which discrimination is the cause of these unequal outcomes, or whether the differences result primarily from unequal qualifications and creditworthiness. The vast majority of empirical work on racial and ethnic differences in underwriting is based on HMDA data, which do not contain many critical lender and loan attributes such as credit history, the ratio of the loan to the assessed value, or the ratio of housing expenses to borrower income. The most recent and high profile underwriting study that controls for detailed underwriting variables, Munnell et. al. for the Boston market in 1990, was highly controversial, with some follow-up studies confirming its findings (Ross and Yinger, 1999; Carr and

Megbolugbe, 1993) while others raised doubts about the conclusions (Horne, 1997; Day and Liebowitz, 1998). Finally, recent analyses by the Office of the Comptroller of the Currency (Courchane, Nebhut, and Nickerson, 2000; Stengel and Glennon, 1999) find less widespread evidence of discrimination when estimating underwriting models for individual lenders.¹

Paired testing provides a uniquely powerful tool for investigating both the incidence and forms of adverse treatment based on race or ethnicity.² Following the paired test methodology, two individuals, one white and one minority, can pose as homebuyers and inquire about the availability and terms for home mortgage loans. Because the two members of a tester team present themselves as equally qualified borrowers in every respect except their race or ethnicity, systematic differences in the treatment they receive provide direct evidence of disparate treatment by mortgage lenders.³ Furthermore, evidence provided by paired testing can be viewed as complementary to traditional regression studies because it can capture forms of discrimination that are not accessible through administrative data (Foster, Mitchell, and Fienberg, 2002), such as discouragement of minority applicants who as a result never submit an application or seek higher cost financing in the subprime market.⁴

It is important to note that paired testing by design is required to focus on early stages of the market transaction. In the case of the mortgage market, this requirement implies that testing focus on the behavior of loan officers during the pre-application stage of the mortgage transaction. Disparate treatment during this stage is illegal under the Equal Credit Opportunity Act (ECOA),⁵ but such tests cannot directly test for disparate treatment in the approval of mortgage applications, which may be of greater policy significance than illegal behavior during the pre-application stage.⁶

Nonetheless, a number of existing studies suggest that loan officers can influence the eventual outcome of a mortgage application and that differential treatment

by loan officers can lead to greater racial disparities in underwriting. Temkin, Levy, and Levine (1999) perform a case study of a medium-sized mortgage lender. They document an apparently race neutral underwriting system where only the loan officer knows the race of the applicant and yet that system has led to large racial disparities in lending, numerous discrimination complaints, and one active law suit under the ECOA.⁷ Squires and Kim (1999) merge employment information from the Equal Employment Opportunity Commission with HMDA data and find that the approval rates of black applicants rise as the share of black employees at the lender increases.⁸ As in the case study, the most likely way for race to influence outcomes is through the loan officers since underwriters rarely meet applicants. Finally, Rachlis and Yezer (1993) propose that the loan officer provides information to borrowers about the lender's specific underwriting standards during the pre-application and application stages of the mortgage transaction, which affects both the borrower's application and the likelihood of approval creating a simultaneity between loan terms and application approval. Yezer, Phillips, and Trost (1994), Phillips and Yezer (1995), and Ross and Yinger (1999) all find evidence of such a simultaneity.⁹

The paired testing methodology has been widely used to detect and measure adverse treatment by rental and sales agents in the housing market and in entry level job searches,¹⁰ but only a few, enforcement-oriented testing efforts have been conducted in the mortgage market.¹¹ Specifically, in the early 1990s, the National Fair Housing Alliance (NFHA) conducted tests in seven cities: Atlanta, Chicago, Dallas, Denver, Detroit, Oakland, and Richmond (Smith and Cloud, 1996) and a local fair housing group conducted tests in Philadelphia (Lawton, 1996). In both cases, testers posed as first-time homebuyers and refinancers inquiring about financing terms and conditions at the pre-application stage. A reanalysis of the NFHA testing data by the Urban Institute concluded that differential treatment occurred at significant levels in at

least some cities. Minorities were less likely to receive information about loan products, received less time and information from loan officers, and were quoted higher interest rates in most of the cities where tests were conducted (Delair and Smith, 1999).

This paper analyzes data from a pilot paired testing study of mortgage lending institutions that was funded by the U.S. Department of Housing and Urban Development (HUD) and conducted by the Urban Institute in the spring and summer of 2000.¹² This study recorded the treatment of testers posing as black and Hispanic first-time homebuyers with limited assets and their comparable white teammates when they visited mortgage lending institutions in Chicago and Los Angeles to make pre-application inquiries. It represents the first application of paired testing to rigorously measure discrimination in the mortgage lending process. Moreover, while the results cannot be generalized to the entire population of minority homebuyers, the study's focus on first time, downpayment constrained homebuyers targets a group that is likely to be dependent upon pre-application assistance, and as such this study captures behaviors that may pose a significant barrier to minority households in their attempts to transition from renting to homeownership.

The paper is organized as follows: section 2 describes the methodology used and data generated by this study, section 3 describes the empirical methodology used to measure systematic adverse treatment of minorities, section 4 presents empirical evidence of adverse treatment, section 5 discusses the implications of this research for future testing efforts, and the final section briefly summarizes the paper's contributions and findings.

Paired Testing Methodology and Data

This study involves approximately 250 paired tests of a representative sample of mortgage lending institutions in Los Angeles, California and Chicago, Illinois (approximately 75 tests per group in each site). These tests followed a single,

standardized set of protocols in order to yield statistically rigorous measures of adverse treatment against African Americans and Hispanics in the two metropolitan housing markets. Specifically, testers posing as first-time homebuyers with limited assets visited mortgage lending institutions to make a general, uninformed request for information about how much house they could afford and what loan products might be available to them.

All of the testers were assigned financial profiles that qualified them for products targeted to borrowers with A- credit quality in their respective housing markets. Each tester was randomly assigned one of six fictitious credit history profiles containing one or two minor credit blemishes, usually a late payment of some kind. The testers were assigned income and asset levels sufficient to purchase a median-priced house in their metropolitan area, assuming a 30-year conventional fixed-rate loan at a market specific interest rate and a 5 percent downpayment. The two members of each tester pair were given virtually identical financial and household characteristics, with the minority partner always slightly better qualified than the white.

The financial profiles were designed to make the testers downpayment constrained in terms of the loan amount for which they could qualify rather than being constrained by their income or debt levels.¹³ For a specific tester pair, the target house price was chosen randomly from an interval around the market median house price, \$175,000 to \$190,000 for Chicago and \$275,000 to \$300,000 for Los Angeles.¹⁴ Cash assets required were calculated as the sum of the 5 percent downpayment plus closing costs, which were based on loan amount and evidence from a pre-test phase. Total housing expenses were calculated as the sum of interest and principle payments plus private mortgage insurance, homeowners insurance,¹⁵ and property taxes.¹⁵ The other expenses were estimated based on standard, publicly available data sources. Financial

variables were set to match key financial ratios, and included small amounts of random variation in order to avoid detection by loan officers.¹⁶

Testers were matched by local testing staff on gender, age, and general appearance. These matches were not permanent, but rather testers could be paired with multiple partners if multiple testers were available with the same gender, comparable age, and broadly similar appearance. Finally, in the case of Hispanic-Anglo tests, Hispanic testers were allowed to have an accent, but were required to be fluent in both spoken and written English. Previous paired testing studies that have imposed this standard have found little evidence that accent leads to treatment differences between Hispanic testers (Yinger, 1995; Kenny and Wissoker, 1994).

In both Chicago and Los Angeles, testers visited a representative sample of mortgage lending institutions in the metropolitan area that reported under the Home Mortgage Disclosure Act (HMDA),¹⁷ accepted at least 90 mortgage loan applications in the market in 1998, and had offices in the region that a first-time homebuyer could realistically find and visit.¹⁸ Based on these criteria, the population of qualifying lenders included 67 institutions in Chicago and 106 institutions in Los Angeles. The population of eligible lenders represented 56% and 62% of the total HMDA-reported application volume in the Chicago and Los Angeles markets, respectively (see Table 1).¹⁹

In order to draw a market representative sample, lenders were selected with replacement and with a probability of selection based on loan volume. Lending institutions with very large application volumes not only had a high probability of selection, but were likely to appear in the sample more than once. This sampling strategy allows us to draw conclusions about the incidence of differential treatment by large lending institutions in Chicago and Los Angeles that are representative of the pattern of lending in the marketplace. In the Los Angeles metropolitan area, 35 lenders were selected for black/white testing and 34 were selected for Hispanic-Anglo testing,

and in Chicago 49 lenders were selected for black/white testing and 51 were selected for Hispanic-Anglo testing. For both markets and race/ethnic groups, the lenders selected accounted for approximately half of the application activity captured in HMDA data.

Table 2 shows the effect of the lender selection process on market representativeness in terms of the share of loan applications made to depository lenders, as well as the racial and ethnic composition of loan applications. In Los Angeles, the sample of eligible lenders is drawn more heavily from depository lenders rather than mortgage banks with the percent of applications to depository lenders increasing from 43 to 59 percent, but the selection process did not have a large effect on racial and ethnic composition. If minority treatment differed between depository and non-depository lenders, the findings presented in this paper may not represent the overall level of discrimination in the marketplace, but as is noted later in the paper no evidence is found to suggest that the pattern of differential treatment varies by lender. In Chicago, the sample of eligible lenders receives a somewhat smaller share of African-American loan applications than the full sample, 11 rather than 15 percent. This difference suggests that the walk-in protocol used in this study may not be the ideal strategy for testing lenders who serve minority homebuyers. The difference does not imply, however, that the results in the paper are biased because as with depository lenders there is no evidence to suggest that differential treatment varies between lenders who serve predominantly white applicants and lenders who have large minority applicant pools.

While the sample was drawn based on lending institutions, each test was conducted by visiting an individual branch office, which was randomly selected from all of the institution's local offices.²⁰ To select a branch office for a test, a list of the target lender's local branches was compiled. Urban institute staff made calls to local offices to verify addresses and determine which branches potential borrowers could visit to receive

information on mortgage loans. Once the list of local branches was reduced to those providing mortgage information, the branch to be tested was selected randomly from the list.

The testing protocols that testers followed when conducting a test can be summarized in five basic steps.

- *Step #1 - Obtain an Appointment.* All testers called to arrange “in person” visits with lenders. Testers were provided detailed instructions on how to complete such calls and how to avoid protracted conversations with loan officers over the phone.
- *Step #2 - Make the Initial Request.* When testers arrived for their appointments, their first step was to very clearly state (up to three times, if necessary) that the purpose of their visit was to obtain help in figuring out a price range of housing that they might be able to afford and an estimated loan amount for which they might qualify.
- *Step #3 - Exchange Personal/Financial Information.* Testers were trained to be forthcoming and provide income, debts, assets, credit information and other personal and financial characteristics when requested by a lender. Testers were instructed to be precise when providing their financial information and to refer to their “cheat sheets” if necessary. Under no circumstances, however, did testers provide a social security number or date of birth or authorize a credit check.²¹
- *Step #4 - Record Information on Financing Options Recommended.* Testers were required to take notes and record information provided by the lender such as suggested home price range, an estimated loan amount, and details about any financing options recommended.

- *Step #5 - End the Visit.* Testers were instructed to thank the lender for any assistance and allow the lender to suggest any follow-up contact.

Following every test visit, testers completed a *Test Report Form* that recorded their responses to specific questions about the test experience and the information that was provided by the lender.²² Testers were instructed to complete all forms as soon as possible following contact with a lender and were not allowed to discuss their experiences with their testing partner. Testers completed the forms based on their recollection of what occurred during the test, and on their review of notes taken and materials obtained during the test.²³

Even at the pre-application stage, inquiries about mortgage products and terms are complicated interactions and differences in treatment can take many forms. As a result, the study gather treatment information on six major questions about the information and assistance that lending institutions provided:²⁴

- 1) Did testers receive the information they requested about loan amounts and house prices they could afford?
- 2) How much were testers told they could afford to borrow and/or buy?
- 3) How many specific products were discussed with the tester?
- 4) How much “coaching”, such as offers of advice on paying down debts, downpayment assistance, or a prequalification letter, did testers receive to help them qualify for a loan?
- 5) Did testers receive follow-up calls from lenders?
- 6) Were testers encouraged to consider FHA loans as an option?

During the study design, researchers defined specific hypotheses concerning findings that would be interpreted as adverse treatment of minority testers. For the first

three questions, these hypotheses are straightforward. Specifically, minorities are assumed to have experienced adverse treatment if they receive less information overall, are told that they are qualified for a smaller loan, or are told about fewer products than their white partners. The test scenario instructs testers to be persistent in requesting information on obtaining a loan, and failure to satisfy the minority tester's request is naturally interpreted as negative treatment. Similarly, being provided detailed information on multiple loan products provides a benefit to the borrower by increasing his or her financial options. Finally, quoting minority homeseekers lower house prices than comparable whites unfairly constrains their housing options and may cause them to conclude that they cannot afford the house they want or to limit their search to lower-priced neighborhoods than they can afford.

Although some researchers have questions about what constitutes unfavorable treatment on the remaining three questions, these hypotheses reflect current thinking in the mortgage lending literature. While some forms of coaching (such as being told to pay down debt) might be interpreted negatively by minority homebuyers, the standard view in the fair lending literature is that assistance in paying down debts or obtaining a downpayment increases the likelihood of an application being approved and therefore constitutes favorable treatment (Yinger, 1996; Siskin and Cupingood, 1996). Similarly, even though excessive follow-up by a loan officer might be viewed negatively by some homebuyers, receiving follow-up contact reflects marketing effort and an eagerness to do business, and therefore is considered favorable treatment, as has been the case in every previous paired testing study. Finally, being encouraged to pursue FHA financing is viewed as negative treatment because FHA financing is substantially more expensive than conventional financing and an on-going policy question has been whether minority borrowers are disproportionately steered to FHA financing (Bradford and Shlay, 1996).²⁵

Even if a reader is somewhat skeptical about these specific interpretations, all of the results presented below can be interpreted as differential treatment of equally qualified individuals seeking assistance and advice on obtaining a home mortgage. Given the careful pairing of testers on assigned financial characteristics and the fact that the testers approach the same lender following a common protocol, differential treatment can reasonably be attributed to the race or ethnicity of the testers. In order to be conservative, all hypothesis tests presented are conducted as two sided tests so that if a result is significant with a 0.05 chance of falsely rejecting the null hypothesis then there is only a 0.025 chance that the hypothesis was improperly rejected in the direction observed.

Statistical Analysis Methodology

Data from paired testing can be used to construct three different types of measures: 1) measures of the *gross incidence* of differential treatment; 2) measures of the *net incidence* of differential treatment; and 3) measures of the *severity* of differential treatment. Each of these measures is briefly explained in turn. A *gross incidence* measure is defined as the share of all tests in which the minority receives less favorable treatment than his or her white Anglo partner. Gross incidence measures provide very simple and understandable indicators of how often minorities are treated less favorably than equally qualified white Anglos. However, there are also cases in which minority testers receive better treatment than their white Anglo partners. Thus, we constructed measures of the gross incidence of minority-favored treatment as well as measures of the gross incidence of white-favored treatment.

Net incidence measures focus on the difference between these two gross incidence measures by subtracting the gross incidence of minority-favored treatment from the gross incidence of white-favored treatment for a given indicator.

A large net incidence measure would suggest that unfavorable treatment of minorities is substantially more prevalent than unfavorable treatment of white Anglos. A small net incidence measure, on the other hand, suggests that while lenders do not always provide comparable treatment to similar customers, they are just as likely to treat minority customers favorably as white Anglo customers.

Net incidence measures are often interpreted as estimates of systematic discrimination against minorities. If one assumes that white Anglos rarely experience systematic adverse treatment, then all cases of minority-favored treatment can be interpreted as random differences in treatment, unrelated to race or ethnicity. If this assumption is correct, then by subtracting cases of minority-favored treatment from the cases of white-favored treatment, the net incidence measure removes the element of random error and reflects the true incidence of discrimination against minorities. However, if the assumption is incorrect (and systematic discrimination against white Anglos does sometimes occur), then the net measure may actually understate the incidence of discrimination against minorities.²⁶

In the analysis presented here, gross incidence measures are reported for both white-favored and minority-favored treatment. When these two gross measures are significantly different from one another (the net measure is significantly greater than zero), then we conclude that a systematic pattern of differential treatment based on race or ethnicity has occurred. If, on the other hand, the incidence of minority-favored treatment is essentially the same as the incidence of white-favored treatment, we cannot conclude that these differences are systematically based on race or ethnicity.²⁷ Because our sample sizes are relatively small, and the data may not be normally distributed, conventional tests of statistical significance may provide a biased test for systematic differences in treatment. Therefore, we used the Sign Test as suggested by Heckman

and Siegelman (1993) to determine whether the incidence of white-favored treatment was significantly different than the incidence of minority-favored treatment.²⁸

Gross and net incidence measures are complemented by measures of the *severity* of differential treatment, which reflect the size or magnitude of differences in treatment between minority testers and their white Anglo partners.²⁹ Severity measures can only be constructed for forms of treatment that yield continuous differences, such as dollars of loan amount or number of loan products. Again because of our small sample sizes and the potential that the data are not normally distributed, we employed the more robust Wilcoxon Rank Sum Test to measure the statistical significance of these severity measures.³⁰

By its nature, paired testing represents an attempt to assess a complex set of personal interactions and so typically surveys a wide variety of treatment variables. In this study, results are presented for six classes of treatment and a large number of individual variables covering four different samples of tests, two groups tested in each of two sites. This large number of tests raises the question of whether some or all of the statistically significant results presented below represent an incorrect rejection of the null hypothesis of equal treatment, a type I error.

As discussed above, this study was designed and implemented to test specific hypotheses regarding adverse treatment of minority homebuyers. Using these hypotheses, a sign test can be implemented to test whether the number of significant findings at the 0.05 level that are consistent with adverse treatment of minority testers differs statistically from the number of findings that are consistent with favorable or equal treatment.³¹ If a substantial number of significant findings exist and most are associated with adverse treatment based on criteria that were defined before the study was conducted, this statistical test, as well as common sense, will dictate that the findings are not attributable to Type I error.

Moreover, additional tests were conducted without relying on the presumed direction of adverse treatment.³² Under the null hypothesis of no differential treatment, the results of each of the six major hypothesis tests may be considered a 1/0 variable (Bernoulli) that takes the value one with a probability of 0.05, the likelihood of a type I error. This series of six Bernoulli variables form a multinomial variable, and the cumulative probability distribution of this multinomial distribution can be used to calculate the likelihood of rejecting the null of equal treatment for any number of the six hypothesis tests. In this case, the likelihood of rejecting one or more, two or more, three or more, or four or more of the six hypotheses under the null hypothesis is 0.2649, 0.0328, 0.0022, and 0.0001.³³

Of course, these six hypothesis tests were conducted for four different samples of tests, two groups in each of two sites. Since these four samples are independent, a simple Bonferroni correction can be applied where the likelihoods of a type I error are multiplied by four since there are four hypotheses concerning differential treatment by race or ethnicity and site (Shaffer, 1995). After applying the Bonferroni correction, we conclude that a significant finding at the 0.05 confidence level for four of the six null hypotheses for a given sample of tests allows us to reject the null of no differential treatment with a confidence level of 0.001, and a significant finding for three of the six allows us to reject the null of no differential treatment at the 0.01 confidence level.

Finally, any paired testing effort involves taking a naturally heterogeneous group of people (testers) and training them to follow a common protocol and act in very similar ways during their test visits. Insufficient training or supervision, or a vague protocol may allow individual testers to behave differently during their visits, potentially creating differences in treatment that are unrelated to race. Heckman and Siegelman (1993) examined the tester pairs used in an employment discrimination study by the Urban Institute in order to determine if the individual pairs experienced different patterns of

treatment. For one of the four samples of tests, Heckman and Siegelman rejected the null hypothesis of similar patterns across tester pairs. They concluded that unobservable tester characteristics had influenced the treatment of testers and that the results from these tests should be interpreted with care.³⁴ Alternatively, Turner, Ross, Galster, and Yinger (2002) examined the results of the newest national sample of paired tests. They found that actual tester attributes, such as education and income,³⁵ sometimes influenced treatment even though this information was not directly available to the real estate agent. They also found, however, that controlling for these differences had little effect or in some cases even increased the incidence of adverse treatment reported in the study.

This paper assesses tester heterogeneity following a method similar to Heckman and Siegelman (1993). Frequency or contingency tables are created that describe the pattern for treatment outcomes by individual tester. For each treatment variable considered, eight tables are set up: white testers from black-white tests in Los Angeles, African-American testers from black-white tests in Los Angeles, white testers from black-white tests in Chicago, etc.³⁶

This study used a relatively large number of testers when compared to the employment tests examined by Heckman and Siegelman.³⁷ The Chicago tests used between 9 and 12 testers for each racial and ethnic group, and the Los Angeles tests used between 6 and 7. In order to limit the problems associated with empty and near-empty cells, all testers who conducted only one or two tests are dropped from the sample, and the treatment variables considered are constructed to have only two possible outcomes (1/0) for each tester. In spite of these provisions, many cells have small sizes, and a Fisher's exact test is used to test for table homogeneity.³⁸

As with the core analysis, there are multiple tests for each sample, and four samples are considered. In this case, there are 12 tests, 6 for the majority and 6 for the

minority testers, for each sample of tests. Accordingly, a multinomial distribution of 12 Bernoulli variables is used to calculate the likelihood of observing a specific number of heterogeneity findings or more for any sample of tests under the null hypothesis of homogeneous testers. The resulting likelihood of a type I error is then multiplied by four to obtain a significance level. In order to be aggressive in identifying heterogeneity (conservative in interpreting the core findings), multiple significance level thresholds are examined, and a pool of testers are presumed heterogenous if the null hypothesis of homogeneity can be rejected for any of the significance thresholds: 0.10, 0.05, 0.01, 0.005, or 0.001.

Empirical Results

The first three questions refer directly to the information requested in the testing protocols, and the results for the associated treatment variables are shown in Table 3. The table presents both incidence indicators that represent the frequency with which the white tester was favored over their minority partner or visa versa, as well as severity indicators that present the average experience of both the white and minority tester. The significance tests for whether these frequencies or averages differ between the white and minority tester are recorded in the minority favored column as a * for significance at the 5% level and a ** for significance at the 1% level or better. Note that findings that are significant at the 10% level are labeled with a # in order to provide additional information to the reader, but these results are not discussed in the text because they do not meet our pre-determined threshold for significance.

The results for whether testers received the information that they requested are shown in the first five rows under incidence indicators. These contain the frequency of white or minority favored tests for whether the tester was provided with a loan amount, a house price, specific product options, or exchanged financial details with the loan

officers, as well as an index that is based on the four treatment variables. The index scores a tester's treatment as lowest possible treatment if no house price or loan amount is provided reflecting the design of the testing scenario where the tester repeatedly requested a loan amount or house price that would be affordable. The next lowest score arises if only a loan amount and/or house price is provided, and higher scores reflect the provision of additional information on products or the exchange of additional financial details.³⁹ Focusing on row 5, significant differences in treatment based on the information provided index are found for black-white tests in Chicago and Hispanic-Anglo tests in Los Angeles with 16 and 13 percentage point differences between white and minority favored treatment, respectively. The row that follows shows samples sizes for each minority group tested in each site.

The next four rows contain the incidence of differential treatment on the loan amount or house price provided to the tester, as well as sample sizes for these comparisons. A test is considered white-favored if the white tester was provided an estimated loan amount or house price that is five percent above the value provided to the minority tester, and similarly minority-favored tests arise when the value provided to the white tester is five percent below the minority tester's value. Significant differences are found for both loan amount and house price for Hispanic-Anglo tests in Chicago with 31 and 35 percentage point differences, respectively. Severity tests are also significant for this same set of tests, and the differences in values quoted white and Hispanic testers are \$16,600 for loan amount and \$16,100 for house price. Note that the sample sizes are smaller than in the information provided rows because comparisons were only made if both testers received a loan amount or house price based on the testers' financial characteristics.⁴⁰

The final treatment variable considered in this table is the provision of information on specific loan products. Specifically, the incidence indicator captures whether the

white or minority tester was provided information on more products than his or her partner, and the severity indicator compares the number of products discussed with each tester. For the incidence measure, significant differences of 24 percentage points are found for black-white and Hispanic-Anglo tests in Chicago. For the severity measures, significant differences are found for black-white tests in Chicago and for Hispanic-Anglo tests in both Los Angeles and Chicago with differences between 0.3 and 0.4 products shown on average. Note that the definition for whether a product is discussed is fairly rigid, requiring that the product be considered with reference to the individual's own financial details. Naturally, products cannot be discussed at this level of detail unless financial details are exchanged. Therefore, tests are only considered if both testers exchanged financial details; to do otherwise would double count differential treatment that was found in the received information variables.

The first seven rows of table 4 (under incidence indicators) contains the results for the six coaching variables considered: discussions on paying down debts, discussions on debt consolidation, downpayment assistance, seller assistance, pre-qualification letter, and offer of a homebuying seminar; as well as an indicator for whether the white or minority tester received coaching on a larger number of items than his or her partner. Focusing on the final row (more coaching), significant differences of 25, 21, and 24 percentage points were found for Hispanic-Anglo and black-white tests in Chicago, and black-white tests in Los Angeles, respectively. The provision of downpayment assistance appears central to this finding for all three sites while pre-qualification and seller assistance may play some role for tests in Chicago. The severity indicators are also statistically significant for these three sets of tests, and the differences in coaching fell between 0.3 and 0.5 types of coaching being provided.

Table 4 also presents the results for whether testers received follow-up contact and whether FHA was encouraged or discouraged. The incidence measure for follow-up

captures whether one tester received at least one follow-up contact while his or her partner did not, and the severity measure compares the average number contacts received by white and minority testers.⁴¹ Significant differences in follow-up contact are found only for black-white tests in Chicago where incidence differences are 11 percentage points and severity differences are 0.2 contacts. No significance differences in treatment are found for the FHA variables.

Statistical Significance with Multiple Tests

Results have been presented for six classes of treatment and 23 individual variables representing a total of 24 major and 92 minor hypothesis tests for the two groups tested in two sites. At a 5 percent significance threshold, 26 of the 92 null hypotheses are rejected in the direction of finding discrimination against minorities, and none are rejected in the direction of finding reverse discrimination. The likelihood of such an event arising by chance can be tested using a sign test, and the event has a probability of less than 0.0001. This confidence level may be artificially precise, however, because many of the treatment variables reflect similar or related forms of behavior by loan officers. If we focus only on the six classes of treatment, 9 of the 24 null hypotheses are rejected in favor of discrimination, none for reverse discrimination, and the likelihood of this result arising by chance is less than 0.005.⁴²

As discussed above, additional tests are conducted without relying on the presumed direction of discriminatory treatment. The significance and insignificance of specific tests for differential treatment is modeled as a multinomial variable, and a Bonferroni correction factor is applied to control for the fact that separate hypothesis tests are conducted for the four samples of paired tests. A significant finding at the 0.05 confidence level for four of the six null hypotheses for a given sample of tests allows us to reject the null of no differential treatment with a confidence level of 0.001, and a significant finding for three of the six allows us to reject the null of no differential

treatment at the 0.01 confidence level. These criteria result in strong evidence of the differential treatment of both African Americans and Hispanics in Chicago, but suggest that the findings for Los Angeles should be interpreted with caution.⁴³

While differential treatment in Los Angeles only arises for a relatively small subset of treatment variables, the finding of differential treatment on the coaching variables for Anglo-Hispanics in Chicago deserves additional consideration. The differences in coaching in terms of severity, as well as the incidence on downpayment assistance individually, are highly significant for this sample of tests, better than 0.002 and 0.0001 significance levels. A simple Bonferroni correction to both of these statistics (multiplying by 24 or 92 for the individual measure) yields significance levels of 0.05 and 0.01, respectively.⁴⁴ Therefore, even after considering the multiple testing environment, statistically significant differences in the extent of coaching especially concerning downpayment assistance have been found in three of the four samples of tests, suggesting that racial and ethnic differences on this form of loan officer behavior may be fairly widespread.

Tester Heterogeneity

As discussed earlier, this paper assesses tester heterogeneity following a method similar to Heckman and Siegelman (1993). The specific variables are constructed as follows: the information provided variable is 1 if either the tester is provided a loan amount or house price and financial details are exchanged, the loan amount variable is 1 if loan amount provided by the loan officer is above the average value for all testers of this race in this sample, the product information variable is 1 if the tester is told about two or more products, the coaching variable is 1 if downpayment assistance was offered, the follow-up variable is 1 if any follow-up contact is made, the FHA variable is 1 if FHA was encouraged, and all variables are set to zero otherwise.⁴⁵

As with the core analysis, there are multiple tests for each sample, and four samples being considered. Accordingly, a multinomial distribution is used to calculate the likelihood of observing a specific number of heterogeneity findings or more for any sample of tests under the null hypothesis of homogeneous testers. The results of these analyses are shown in Table 5. For each variable, the row shows the confidence with which the null hypothesis of homogeneous testers can be rejected, and the row below shows the sample size.⁴⁶ The second panel of rows shows the number of heterogeneity findings for each sample of tests for each standard of significance considered.

For black-white tests in Los Angeles, homogeneity is rejected at the 5% significance level using both the 0.01 and 0.001 thresholds, and for Hispanic-Anglo tests in Chicago homogeneity is rejected based on the 0.1 significance level threshold. For black-white tests in Chicago and Hispanic-Anglo tests in Los Angeles, there is no evidence of tester heterogeneity. For the last two samples mentioned, the tests for heterogeneity do not approach statistical significance even when the Bonferroni correction factor is not applied. These findings imply that behavioral differences between testers of the same race lead to systematic differences in treatment between those testers in two of the samples. More crucially, these findings suggest that behavioral differences between the white and minority samples of testers might have contributed to or mitigated the observed levels of adverse treatment.⁴⁷

Such a possibility should not, however, lead the reader to discount the policy significance of statistically significant findings of adverse treatment. While tester heterogeneity raises questions about whether the treatment differences identified represent disparate treatment discrimination, these findings represent real differences in treatment between testers who are equally qualified, visit the same firm, and make the same request. Differences in treatment that have an adverse impact on minorities must

by law satisfy a business necessity standard that is unlikely to include responses to behavioral differences that have no connection to the lender's business goals.⁴⁸

Implications for Future Research and Enforcement Testing

Based upon the results of this pilot testing effort, we conclude that paired testing at the pre-application stage of the mortgage lending process is both feasible and effective for both enforcement and research purposes. The results demonstrate convincingly that unequal treatment of minority homebuyers at the pre-application stage of the mortgage lending process remains a significant problem for African Americans and Hispanics in some U.S. metropolitan areas. Adverse treatment of both African Americans and Hispanics occurs on a substantial number of treatment variables in Chicago, and for a smaller number in Los Angeles. The results for Los Angeles, however, must be interpreted with some care because of the small number of statistically significant findings relative to the number of statistical tests. Finally, the interpretation of findings for Hispanic-Anglo tests in Chicago and black-white tests in Los Angeles also requires some care because unobserved tester characteristics (other than race or ethnicity) may have influenced the testers' treatment experience.

Additional testing clearly should be conducted. Due to the complexity of this kind of testing and the differences between metro area results, however, we do not recommend a large-scale national study of adverse treatment at the pre-application stage of the mortgage lending process. Instead, systematic studies should be conducted on a site-by-site basis, with ample time and resources for effective training, test coordination, and quality control. Paired testing in the mortgage market obviously requires considerable attention to the individuals who are used as testers. Tester training should be a priority and may require substantially more time and resources than training for paired testing in other markets, like housing. Also, future studies should

record the actual characteristics of the testers and make those characteristics available for analysis. Alternatively, when the number of tests is small, testers should be matched on key attributes, such as their education level or previous experience in the mortgage market.⁴⁹

The testing protocols implemented for the pilot study focused on only one source of pre-application information, one set of borrower profiles, and one type of pre-application inquiry. They do not tell us how often adverse treatment occurs for other types of borrowers, making different types of pre-application inquiries to other information providers. Therefore, future testing efforts can and should focus on other information sources, including mortgage brokers, new home sales agents, mobile home dealers, and referrals from real estate agents. They should include testers posing as less well-qualified borrowers as well as homebuyers with a particular house price in mind. And they should employ different protocols, including scenarios in which testers approach lending institutions with a specific house price in mind, and request detailed information about available loan products.⁵⁰

Some of the tests conducted for this pilot provided anecdotal evidence that lenders may encourage or discourage customers from considering particular neighborhoods. This may include comments that steer minority customers away from predominantly white areas or that discourage whites from investing in minority neighborhoods. However, our testing protocols were not designed to systematically elicit or record comments of this kind or to analyze the characteristics of communities discussed by loan officers. Future testing efforts should consider strategies for systematically capturing and analyzing this type of differential treatment.

Finally, our experience demonstrates that research and enforcement testing can be linked without compromising the integrity of either. It is important to distinguish testing for research from testing conducted primarily for law enforcement purposes.

Testing for research generally seeks to produce generalizable results regarding the prevalence of adverse treatment for a market area or for the nation as a whole. To achieve these generalizable results, tests are randomized using an accepted sampling frame, and quite large numbers of tests are conducted in order to support statistically significant comparisons. To generate reliable and objective comparisons of minority and white experiences across a large number of tests, researchers usually use highly structured reporting forms, with closed-ended, “check the box”-type items.

By contrast, the purpose of an enforcement test is to establish legal violations and to correct them either through settlement or litigation. Testing for enforcement is often complaint driven, and typically is targeted to a single firm or a selected set of firms. Enforcement testing often requires multiple tests of a single firm, but generally does not involve the large number of tests typical of research testing. As a consequence, enforcement testing report forms tend to be much more open-ended, requiring test partners to provide greater narrative detail, rather than check boxes. These forms are generally analyzed pair-by-pair by a knowledgeable analyst who compares the treatment of test partners across all aspects of the encounter, including subjective as well as objective information.

Although research and enforcement testing differ in significant ways, the distinctions between the two should not be overdrawn. Both are based on the same core methodology and protocols, differing primarily in the way test results are recorded and analyzed. Thus, innovative testing methodologies developed for one application can be adapted for the other. Moreover, randomized testing of large numbers of market transactions need not be limited to research. They can and should be applied in targeting for enforcement. Furthermore, research and enforcement testing can be conducted in tandem, yielding *both* market-wide estimates of the incidence of adverse

treatment *and* case-specific evidence of individual violations (evidence that might be followed-up with enforcement oriented testing).

Conclusion

This paper finds strong evidence of differential treatment of African Americans and Hispanics in the Chicago metropolitan area, along with more limited indications of differential treatment in Los Angeles using data from a rigorously designed and implemented paired testing study. Differential treatment takes a variety of forms. In Chicago, African Americans were less likely to be provided any information in response to their inquiries, received detailed information about fewer products, and were provided with less coaching and follow-up. Hispanics in Chicago were quoted lower loan amounts and house prices, received information about fewer products, and received less coaching. In Los Angeles, African Americans received less coaching than comparable whites, while Hispanics were less likely to be provided any information in response to their inquiries and received information about fewer products. Results for Los Angeles, especially for Hispanics, should be interpreted with some caution, given the smaller number of treatment variables for which statistically significant differences occurred.

The paper also finds, however, some evidence of tester heterogeneity in two of the four samples: black-white tests in Los Angeles and Hispanic-Anglo tests in Chicago. These findings imply that a tester's unobserved characteristics may influence his or her treatment, which suggests that actual differences between testers (other than their race or ethnicity) can affect the outcomes of tests. While testers were trained to behave in similar ways and to follow common protocols, individual differences appear to have played some role in these two samples. While findings for these samples clearly constitute adverse treatment by race and ethnicity that is completely unrelated to a borrower's financial qualifications, they cannot be unambiguously interpreted as disparate treatment discrimination.

Nonetheless, this study demonstrates that African American and Hispanic homebuyers in Chicago face a significant risk of receiving less favorable treatment than comparable whites across a broad set of loan officer behaviors when they visit mortgage lending institutions to inquire about financing options. Discriminatory treatment at this early stage in the mortgage lending process, though subtle, has the potential to influence minority borrowers' ability to obtain credit through a number of important mechanisms. Minority homeseekers may be discouraged from applying for a mortgage, either abandoning their housing search or seeking mortgage credit in the subprime market at a much higher cost. To the extent that loan officers can influence the underwriting process, differential treatment at the pre-application stage by the loan officer may provide an indication of their treatment during underwriting stage. Finally, loan officers may provide more support and information to white applicants during the pre-application and application stages allowing white applicants to prepare an application that has a better chance of acceptance than the application of a similarly qualified minority applicant.

Readers might reasonably ask about the relevance of these results at a time when mortgage transactions are increasingly conducted over the internet without any need to walk into a lender's office or talk to a loan officer. While internet lending is growing, however, it represented only three percent of U.S. mortgages in 2001 and is forecast to grow to 12 percent by 2005 (Pastore, 2001). Moreover, access to the internet is not evenly distributed across potential homebuyers. The U.S. Department of Commerce (2000) documented substantial racial, ethnic, and poverty gaps in access to the internet with only 29 percent of black households, 24 percent of Hispanic households, and 19 percent of households making less than \$15,000 in annual income having access. These numbers suggest that the in-person, pre-application contacts examined in this study are likely to continue well into this decade and beyond.⁵¹

References

- Agresti, Alan (1990), Categorical Data Analysis (New York: John Wiley).
- Avery, Robert B., Patricia E. Beeson, and Mark S. Sniderman (1996), "Accounting for Racial Differences in Housing Credit Markets," in John Goering and Ron Wienk's (Eds.), Mortgage Lending, Racial Discrimination, and Federal Policy (Washington, DC: Urban Institute Press).
- Bertrand, Marianne and Sendhil Mullainathan (2003), "Are Emily and Gred more Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." NBER Working Paper #9873.
- Blackburn, McKinley, and Todd Vermilyea (2003), "Racial Discrimination in Home Purchase Mortgage Lending Among Large National Banks," in Proceedings of the 2003 Joint Statistical Meetings (Alexandria, VA: American Statistical Association).
- Bostic, Rapheal (2003), "A Test of Cultural Affinity in Home Mortgage Lending." Journal of Financial Services Research 23, pp. 89-112.
- Bradford, Calvin and Anne Shlay (1996), "Assuming a Can Opener: Economic Theory's Failure to Explain Discrimination in FHA Lending Markets." Cityscape: A Journal of Policy Development and Research 2, pp. 77-87.
- Carr, James H. and Isaac F. Megbolugbe (1993), "The Federal Reserve Bank of Boston Study on Mortgage Lending Revisited." Journal of Housing Research 4, pp. 277-314.
- Carr, James H. and Jenny Schuetz (2001), "Financial services in distressed communities: Framing the issue, finding solutions" (Washington, DC: Fannie Mae Foundation).

- Courchane, Marsha, Amos Golan, and David Nickerson (2000), "Estimation and Evaluation of Loan Discrimination: An Informational Approach." *Journal of Housing Research* 11, pp. 67-90.
- Courchane, Marsha, David Nebhut, and David Nickerson (2000), "Lessons Learned: Application of Statistical Techniques to Fair Lending Compliance Examinations." *Journal of Housing Research* 11, pp. 277-295.
- Crawford, Gordon W., and Eric Rosenblatt (1999), "Differences in the Cost of Mortgage Credit: Implications for Discrimination." *Journal of Real Estate Finance and Economics* 19, pp. 147-159.
- Day, Todd, and Stanley J. Liebowitz (1998), "Mortgage discrimination in Boston: Where's the bias." *Economic Inquiry* 36, pp. 1-27.
- Figlio, David (2003), "Names, Expectations, and Black Children's Achievement." Unpublished manuscript.
- Fix, Michael and Margery Austin Turner (1999), *A National Report Card on Discrimination in America: The Role of Testing* (Washington, DC: The Urban Institute).
- Foster, A.W., F. Mitchell and S. E. Feinberg (2002), *Measuring Housing Discrimination in a National Study* (Washington, DC: National Academy Press).
- FreddieMac (2000), *The Automated Underwriting Report. Chapter 5 Expanding Markets* (McLean, VA: FreddieMac).
- Freiberg, Fred and Carla Herbig (2002), *Guide to Enforcement Tools for Fair Lending Testing* (Washington, DC: The Urban Institute).
- Fryer, Roland and Steven Levitt (2003), "The Causes and Consequences of Distinctively Black Names." NBER Working Paper #9938.

- Goering, John M. and Ron Wienk (1996), *Mortgage Lending, Racial Discrimination, and Federal Policy* (Washington, DC: The Urban Institute Press).
- Heckman, James J. and Peter Siegelman (1993), "The Urban Institute Audit Studies: Their Methods and Findings" in Michael Fix and Raymond J. Struyk (Eds.), *Clear and Convincing Evidence: Testing for Discrimination in America* (Washington, DC: The Urban Institute Press).
- Holloway, Steven R. and Elvin K. Wyly (2002), "The Disappearance of Race in Mortgage Lending." *Economic Geography* 78, pp. 129-169.
- Horne, David K (1997), "Mortgage Lending, Race, and Model Specification." *Journal of Financial Services Research* 11, pp. 43-68.
- Johnson, Norman L., Samuel Kotz, N. Balakrishnan (1997), *Discrete Multivariate Distributions* (New York: John Wiley).
- Kenny, Genevieve M. and Douglas A. Wissoker (1994), "An Analysis of the Correlates of Discrimination Facing Young Hispanic Job-Seekers." *American Economic Review* 84, pp. 674-683.
- Kim, Sunwoong and Greg Squires (1999), "The Color of Money and the People Who Lend It." *Journal of Housing Research* 9, pp. 271-284.
- King, Thomas A (1980), "Discrimination in Mortgage Lending: A Study of Three Cities" (Washington, DC: Federal Home Loan Bank Board).
- Ladd, Helen (1998), "Evidence on Discrimination in Credit Markets." *Journal of Economic Perspectives* 12, pp. 41-62.
- Lax, Howard, Michael Manti, Paul Raca, and Peter Zorn (2000), *Subprime lending: An investigation of economic efficiency* (McLean, VA: Freddie Mac).
- Lawton, Rachel (1996), "Preapplication Mortgage Lending Testing Program: Lender Testing by a Local Agency," in John Goering and Ron Wienk (Eds.), *Mortgage*

- Lending, Racial Discrimination, and Federal Policy* (Washington, DC: Urban Institute Press).
- Linneman, Peter, and Susan M. Wachter (1989), "The Impacts of Borrowing Constraints on Homeownership" *AREUEA Journal* 17, pp. 389-402.
- Listokin, David and Elvin Wyly (1998), *Successful Lending Industry Strategies, Volume 1: Successful Strategies* (New Brunswick, NJ: Center for Urban Policy Research).
- Longhofer, Stanley D. (1996), "Cultural Affinity and Mortgage Discrimination." *Economic Review, Federal Reserve Bank of Cleveland* 3rd Quarter, pp. 12-24.
- Massey, Douglas S. and Garvey Lundy (2001), "Use of Black English and Racial Discrimination in Urban Housing Markets: New Methods and Findings." *Urban Affairs Review* 36, pp. 452-469.
- Munnell, Alicia, Lynne E. Browne, James McEneaney, and Geoffrey M.B. Tootell (1996), "Mortgage Lending in Boston: Interpreting HMDA Data." *American Economic Review* 86, pp. 25-53.
- Murphey, S.A. (2001), "Audit Studies and the Assessment of Discrimination," in A.W. Foster, F. Mitchell and S.E. Fienberg (Eds.), *Measuring Housing Discrimination in a National Study: A Report of a Workshop* (Washington, DC: National Academy Press).
- Pastore, Michael (2001), "Net Plays Increasing Role in Mortgage, Home-Buying Process." *Cyberatlas.com* July 24, 2001.
- Pennington-Cross, Anthony and Anthony M. Yezer (2000), "The Federal Housing Administration in the New Millennium." *Journal of Housing Research* 11, pp. 357-372.

- Ramsey, Fred L. and Daniel W. Schafer (1997). *The Statistical Sleuth: A Course in Methods of Data Analysis* (Belmont, CA: Wadsworth Publishing Company).
- Ross, Stephen L. (2002), "Paired Testing and the 2000 Housing Discrimination Study," in A.W. Foster, F. Mitchell and S. E. Feinberg (Eds.), *Measuring Housing Discrimination in a National Study* (Washington, DC: National Academy Press).
- Ross, Stephen and John Yinger (2002), *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair-Lending Enforcement* (Cambridge, MA: The MIT Press).
- Ross, Stephen L., and John Yinger (1999), "Does Discrimination in Mortgage Lending Exist? The Boston Fed Study and Its Critics," in Margery Austin Turner and Felicity Skidmore (Eds.), *Mortgage Lending Discrimination: A Review of Existing Evidence* (Washington, DC: The Urban Institute).
- Schafer, Robert, and Helen F. Ladd (1981), *Discrimination in Mortgage Lending* (Cambridge, MA: MIT Press).
- Shaffer, Juliet Popper (1995) "Multiple Hypothesis Testing." *Annual Review of Psychology* 46, pp. 561-584.
- Siskin, Bernard and Leonard Cupingood (1996), "Use of Statistical Models to Provide Statistical Evidence of Discrimination in the Treatment of Mortgage Loan Applicants: A Study of One Lending Institution," in John Goering and Ron Wienk (Eds.) *Mortgage Lending, Racial Discrimination, and Federal Policy* (Washington, DC: Urban Institute Press).
- Smith, Shanna, and Cathy Cloud (1996), "The Role of Private, Nonprofit Fair Housing Enforcement Organizations in Lending Testing," in John Goering and Ron Wienk (Eds.) *Mortgage Lending, Racial Discrimination, and Federal Policy* (Washington, DC: Urban Institute Press).

- Stengel, Mitchell and Dennis Glennon (1999), "Evaluating Statistical Models of Mortgage Lending Discrimination: A Bank-Specific Analysis." *Real Estate Economics* 27, pp. 299-334.
- Tempkin, Kenneth, Diane K. Levy, and David Levine (1999), "Inside a Lender: A Case Study of the Mortgage Application Process," in Margery Austin Turner and Felicity Skidmore (Eds.), *Mortgage Lending Discrimination: A Review of Existing Evidence* (Washington, DC: The Urban Institute).
- Turner, Margery A., Stephen Ross, George Galster, and John Yinger (2002), *Discrimination in Metropolitan Housing Markets: Results from Phase I of HDS2000* (Washington, DC: U.S. Department of Housing and Urban Development).
- Turner, Margery A. and Felicity Skidmore (1999), *Mortgage Lending Discrimination: A Review of Existing Evidence* (Washington, D.C.: The Urban Institute).
- U.S. Dept of Commerce (2000), "Falling Through the Net: Toward Digital Inclusion", October 2000.
- Yezer, Anthony M. (1995), *Fair Lending Analysis: A Compendium of Essays on the Use Of Statistics* (Washington, D.C: American Bankers Association).
- Yinger, John (1996), "Discrimination in Mortgage Lending: A Literature Review," in John Goering and Ron Wienk (Eds.) *Mortgage Lending, Racial Discrimination, and Federal Policy* (Washington, DC: Urban Institute Press).
- Yinger, John (1995), *Closed doors, opportunities lost : the continuing costs of housing discrimination* (New York : Russell Sage Foundation).

Table 1: Population of Eligible Lending Institutions						
	# of Institutions		Application Volume		% of HMDA Activity	
Lenders Tested	<i>LA</i>	<i>Chicago</i>	<i>LA</i>	<i>Chicago</i>	<i>LA</i>	<i>Chicago</i>
Lender Population	67	106	89,788	103,017	56	62
Black-White Sample	35	49	81,031	78,655	50	47*
Hispanic-Anglo Sample	34	51	80,447	85,214	50	51*

*During the course of testing, one institution in our sample merged with another large lender. HMDA application volume for this lender is difficult to determine and is not represented in these totals.

Table 2: Share of Loans for Lender Samples				
Site	LA		Chicago	
Full	Full	Testing	Full	Testing
Percent depository	0.429	0.591*	0.642	0.689
Percent African-American	0.072	0.066	0.149	0.108*
Percent Hispanic	0.246	0.283	0.106	0.118

*Represents differences that are significant at the 5% level. Note that no other differences were significant even at the 10% level.

Table 3: Detailed Loan Information

Sites	Chicago				Los Angeles			
Group	Hispanic-Anglo		Black-White		Hispanic-Anglo		Black-White	
Incidence Indicators	% White Favored	% Hisp. Favored	% White Favored	% Black Favored	% White Favored	% Hisp. Favored	% White Favored	% Black Favored
1. Information Provided								
Loan amount	5.1	3.8	6.3	1.3	11.4	1.3*	6.8	4.1
House price	6.3	6.3	5.1	3.8	10.1	1.3*	6.8	6.8
Specific options	5.1	7.6	6.3	0.0#	11.4	1.3*	6.8	5.5
Financial details	10.1	13.9	17.7	8.9	17.7	10.1	13.7	11.0
Information Index	15.2	22.8	26.6	10.1*	22.8	10.1*	20.3	13.5
Sample Size	79		79		79		74	
2. Amount of Credit								
Loan amount	50.9	20.0**	30.7	16.1	42.9	37.5	32.1	28.6
Sample Size	62		55		57		56	
House price	49.1	14.6**	29.5	21.3	42.1	33.3	35.7	28.6
Sample Size	61		55		56		56	
3. Detailed Product Information								
Shown More Products	47.6	23.8*	39.1	17.4*	40.3	26.9	28.8	24.2
Sample Size	69		63		67		66	
4. Severity Indicators								
Severity Indicators	White	Black	White	Black	White	Hispanic	White	Hispanic
2. Amount of Credit								
Loan Amount (\$1,000's)	192.4	175.8**	183.0	180.4	271.4	265.1	264.4	269.3
House Price (\$1,000's)	202.3	186.2**	194.1	194.0	286.0	286.6	285.9	288.6
3. Detailed Product Information								
Number of Products	2.43	2.09*	2.39	1.97**	2.27	1.81*	1.88	1.79

Table 4: Additional Loan Officer Treatments

Sites	Chicago				Los Angeles			
Group	Hispanic-Anglo		Black-White		Hispanic-Anglo		Black-White	
Incidence Indicators	% White Favored	% Hisp. Favored	% White Favored	% Black Favored	% White Favored	% Hisp. Favored	% White Favored	% Black Favored
4. Coaching and Related Assistance								
Paying down debts	5.1	0.0	6.3	2.5	6.3	1.3	1.4	2.7
Debt consolidation	0.0	0.0	0.0	0.0	1.3	0.0	1.4	0.0
Downpayment	24.1	6.3**	32.9	16.5*	16.5	17.7	29.7	2.7**
Seller assistance	10.1	1.3*	12.7	10.1	17.7	22.8	28.4	16.2
Pre-qualification	24.1	21.5	30.4	15.2#	22.8	15.2	18.9	14.9
Homebuyer's Seminar	6.3	3.8	5.1	5.1	1.3	1.3	2.7	0.0
More Coaching	40.5	15.2**	50.6	29.1*	39.2	36.7	43.2	18.9**
5. Follow-Up Contact								
Received Contact	7.6	11.4	12.7	1.3**	5.1	10.1	5.4	5.4
6. Consideration of FHA Loan Products								
FHA Encouraged	13.9	11.4	6.3	12.7	11.4	11.4	5.4	13.5
FHA Discouraged	2.5	1.3	7.6	5.1	0.0	0.0	5.4	1.4
Severity Indicators								
	White	Black	White	Black	White	Hispanic	White	Hispanic
4. Coaching and Related Assistance								
Amount of Coaching	0.87	0.51**	1.37	0.99**	1.20	1.13	1.22	0.76**
5. Follow-Up Contact								
Number of Contacts	0.16	0.28	0.22	0.03**	0.19	0.38#	0.11	0.05
Sample Size								
	79		79		79		74	

Table 5: Testing for Tester Heterogeneity

Sites	Chicago				Los Angeles			
Group	Hispanic-Anglo		Black-White		Hispanic-Anglo		Black-White	
Incidence Indicators	% White Favored	% Hisp. Favored	% White Favored	% Black Favored	% White Favored	% Hisp. Favored	% White Favored	% Black Favored
Information Provided	0.130	0.470	0.619	0.096#	0.285	0.231	0.826	0.079#
Sample Size	77	74	73	67	77	78	69	70
Amount of Credit	0.377	0.063#	0.326	0.265	0.645	0.270	0.673	0.973
Sample Size	54	52	59	56	55	55	52	54
Product Information	0.010**	0.313	0.306	0.336	0.229	0.411	0.053#	0.062#
Sample Size	62	59	64	61	65	66	61	64
Coaching	0.818	0.001**	0.214	0.014*	0.023*	0.311	0.002**	0.280
Sample Size	77	74	73	67	77	78	69	70
Follow-up Contact	0.908	0.198	0.542	0.716	0.693	0.389	0.146	0.077#
Sample Size	77	74	73	67	77	78	69	70
FHA Loan Products	0.372	0.139	0.797	0.855	0.611	0.357	0.784	0.077#
Sample Size	77	74	73	67	77	78	69	70
Number of Testers	5	5	9	8	5	5	4	4
Statistical Thresholds	Number of Rejections		Number of Rejections		Number of Rejections		Number of Rejections	
0.10 or better	3		2		1		6**	
0.05 or better	2		1		1		1	
0.01 or better	2*		0		0		1	
0.005 or better	1#		0		0		1#	
0.001 or better	1*		0		0		0	

Endnotes

¹ See Ross and Yinger (2002), Turner and Skidmore (1999), Ladd (1998), Goering and Wienk (1996), Yezer (1995), and Turner and Skidmore (1999) for more extended discussions of these issues.

² For more information on paired testing and its role in both measurement and enforcement, see Ross (2002) and Fix and Turner (1999).

³ Of note, Ross and Yinger (2002) and Blackburn and Vermilyea (In Press) both find evidence that market level studies overstate adverse treatment against minorities in lending due to variations in underwriting standards across lenders. Paired testing involves sending testers with nearly identical requests to the same lender, and evidence of disparate treatment arises from direct comparisons of behavior by the same lender rather than comparisons across lenders as in market level studies like Munnell et. al. (1996) and Shafer and Ladd (1981).

⁴ Discrimination may weaken minority borrower's attachment to formal financial markets, and borrowers with low levels of such attachment are the primary users of the subprime market (Carr and Schuetz, 2001). In fact, many users of the subprime market are qualified for financing in the primary market (FreddieMac, 2000), and these borrowers on average pay a substantial premium above what they would have paid in the prime market (Lax, Manti, Raca and Zorn, 2000).

⁵ See Ross and Yinger (2002) for a detailed discussion of the ECOA. Also, see Listokin and Wyly's (1998) discussion of industry practices to support fair lending. Many of these practices emphasize the training of and the incentives facing loan officers rather than focusing solely on the loan underwriting system. Finally, see Lawton (1996) and Smith and Cloud (1993) for discussions of paired testing in the mortgage market by fair housing groups and the resulting legal actions brought by those groups.

⁶ In principle, paired testing might be used to measure discrimination at the application and underwriting stage of the mortgage lending process as well. However, most experts believe that federal laws prohibiting the submission of false credit applications make it impossible to extend the paired testing methodology into the application stage. For further discussion of this issue, see Turner and Skidmore (1999).

⁷ In fact, this lender has a policy that creates a loan team to assess problem applications and the loan officer and the underwriter are both on this team. This team creates an explicit mechanism for the loan officer to influence underwriting outcomes.

⁸ This evidence is consistent with the cultural affinity hypothesis where racial similarities increase the quality of communication between the borrower and the loan officer leading to superior outcomes for borrowers who share the same race and cultural background with loan officers. See Bostic (2003) and Longhofer (1996) for additional tests of the cultural affinity hypothesis in the mortgage market.

⁹ The above discussion relates to a larger issue in the debate of paired testing concerning the distinction between disparate treatment as defined by law, which is unequal treatment of equals based on membership in a protected class, and market discrimination as defined by economists, which is often defined as the overall effect of discrimination on market outcomes. Estimates of disparate treatment may be larger than market discrimination because minorities can take steps to avoid discriminatory firms. On the other hand, estimates of market discrimination can exceed the level of disparate treatment either if minorities overcompensate for discrimination placing them at a disadvantage in negotiations or in high cost market segments, such as the subprime mortgage market, or if heterogeneous firm behavior biases estimates of

market discrimination upwards, as has been found in the mortgage market (Ross and Yinger, 2002).

¹⁰ Heckman and Siegelman (1993) have argued that the pairing of testers is likely to overstate discrimination because it leaves no legitimate mechanism for the firm to use in distinguishing between applicants and so even very weak levels of prejudice may lead to large differences in treatment during the test. This argument depends upon the notion that there is one job or apartment that the firm must allocate between the two testers, and so the argument does not apply to testing in the mortgage market where presumably testers are not in direct competition for a scarce pool of mortgage credit.

¹¹ These efforts were intended to monitor local lenders and to support legal complaints against lenders if evidence of discrimination was detected. Enforcement oriented testing programs usually place considerably more emphasis on the narrative description of each tester's experience, and less effort is expended to precisely match the situation experienced by testers, to control the behavior of testers during a test, and to record closed form variables that describe specific treatments experienced by testers.

¹² Because of the complexity of the mortgage application process and the challenges it presents for paired testing, this study was divided into two basic stages: a *pre-test* stage and a *pilot* stage. The pre-test stage was used to experiment with a fairly wide variety of paired testing scenarios and sources of mortgage financing information. A total of 78 tests were conducted in Orange County, California and New Orleans, Louisiana in 1999 targeting six different information sources including conventional mortgage lenders, sub-prime mortgage lenders, mortgage brokers, real estate agents, new home sales offices, and mobile home dealers. The pre-tests also experimented by varying the testing scenario over factors such as whether a specific house had been identified, the borrower's qualifications, and visits by couples versus a single tester posing as a married individual

¹³ The downpayment constraint is thought to be the most common constraint faced by first-time homebuyers and is especially relevant to minority homebuyers because black and Hispanics tend to have substantially lower levels of accumulated assets than whites even after controlling for income, see Linneman and Wachter (1989).

¹⁴ Although minority and white tester financial characteristics were closely matched, testing protocols did call for the minority tester to be slightly more qualified than his or her white partner.

¹⁵ Principle and interest were calculated using interest rates of 7.625% for Chicago and 8.25% for Los Angeles and standard annuity tables.

¹⁶ Specifically, income and debt were set so that housing expense and total debt expense to income ratios fell between 0.33 and 0.35 and between 0.255 and 0.265, respectively, which are well within secondary market guidelines. Also note that co-borrower income share was set to a random number between 0.375 and 0.45.

¹⁷ HMDA requires all independent mortgage companies and mortgage lenders owned by depository institutions that make at least 100 home purchase and/or refinancing loans in a given year to report on the demographic and locational characteristics of all applications and loans.

¹⁸ This last criteria was verified by pre-test phone calls and where necessary visits to the lender's local office.

¹⁹ In 1998, HMDA contained 515 and 793 lenders with loan applications for properties located in Chicago and Los Angeles. Only 171 and 198 institutions remained in Chicago and Los Angeles after deleting institutions that did not have a substantial presence in the Chicago or Los Angeles markets, i.e. at least 90 applications in 1998. The vast majority of the remaining institutions that were excluded from the population were not eligible because they did not have

local offices, were credit unions with restricted membership, or were no longer in business at the time of the study. Note that the levels and patterns of discrimination may be different for smaller lending institutions or when inquiries are made by telephone or internet rather than in person.

²⁰ Home Mortgage Disclosure Act data reports applications at the lender level rather than the branch, and therefore Information on the volume of activity conducted by each branch is not available.

²¹ Testers were instructed to explain that they did not want the lender to run a credit check because they were just beginning their credit search and did not want their credit history to show a lender's inquiry. Since the time of this study, the major credit reporting agencies have changed policies concerning when credit inquiries are recorded on the credit history and how they enter a borrower's credit score in order to account for the increased use of credit checks and increased volume of refinancing. As a result, this explanation for not authorizing a credit check may not work in future pre-application testing studies.

²² The tester report form and other test instruments are available upon request.

²³ Testers also completed a *Test Narrative Form* providing a detailed, chronological, account of the test experience in the tester's own words. Additional narrative forms were completed by testers following any phone contact by a lender. Because of the complexity of lender testing, the detailed narratives played a particularly important role in quality control.

²⁴ Testers also recorded the terms and conditions of specific loan product they were offered. Ideally, one would compare terms and conditions for comparable loan products offered to both white and minority customers. However, because HTP protocols called for testers to approach lending institutions with a very general request for information, the product-specific information they received was very diverse. Similar products were listed in different order and given different names, and testers did not always receive a complete set of terms and conditions for every product discussed. Therefore, it is not possible to match products and compare terms and conditions.

²⁵ This question is in part driven by the well-documented fact that minority households are disproportionately represented in the FHA pool of loans after controlling for the financial characteristics of those loans (Pennington-Cross and Yezer, 2000).

²⁶ This is important in interpreting paired tests of housing markets because minorities may be sometimes favored because white testers are being steered away from certain housing units or neighborhoods, see Turner, Ross, Galster, and Yinger (2002) and Ross (2000). In principle, some mortgage lenders could practice reverse discrimination while other discriminate against minorities.

²⁷ It is important to note that even when no statistical pattern of race-based differential treatment is observed, individual cases of discrimination may have occurred. A qualitative review of the entire test file might be needed to assess the overall outcome across multiple measures.

²⁸ The sign test provides a non-parametric statistical test for whether a binary variable is systematically more likely to take one of the two possible values. The test involves the calculation of all data permutations under the null hypothesis that the variable is distributed evenly between the two outcomes. The fraction of permutations that yield patterns as or more unevenly distributed than the data is an estimate for the likelihood that the data arose by chance or the likelihood that the null hypothesis is true, see Ramsey and Schafer (1997) for details. This test is applied to paired testing data by dropping cases in which the two testers are treated the same.

²⁹ Note that severity measures are inherently net measures, since they reflect the average of differences between minority and white outcomes. Thus, cases where minorities were favored effectively cancel out cases where whites were favored.

³⁰ This test is also an exact permutation-based test. A signed-rank statistic is calculated by ordering the tests based on racial differences in treatment, assigning ranks to all tests, and summing the ranks of all tests for which the racial difference is consistent with minority adverse treatment. The statistical test is constructed by calculating the fraction of data permutations that yield the same or higher rank sum, see Ramsey and Schafer (1997) for details

³¹ As above, the hypothesis tests where equal treatment is not rejected are dropped, and the test determines the likelihood that the frequency of findings of adverse treatment differs systematically from the frequency of findings of favorable treatment.

³² It should be noted that philosophical differences in the interpretation of treatment variables actually does not make the sign test inappropriate. Statistical validity simply requires that the hypotheses be established prior to the study being conducted or at least the data being analyzed. These hypotheses are well documented in the original proposal and the research design plan that were submitted to the Department of Housing and Urban Development.

³³ See Johnson, Kotz, and Balakrishnan (1997) for details on the multinomial distribution.

³⁴ Their results are quite compelling based on the frequency tables. For one pair, the minority tester appears is treated worse than their white partner for most tests, and for the other the minority tester is usually treated better than the white tester. Given that each tester pair conducted a random sample of tests, the best explanation is that one minority performed better than the white tester in terms of obtaining better treatment, and one performed worse. Note that there was no evidence of differential experiences across tester pairs for the other three samples of tests. Those frequency tables showed that minority testers were consistently treated worse than their white partners.

³⁵ These data are collected directly from the tester's employment application at the fair housing group.

³⁶ Heckman and Siegelman (1993) examine tester pairs rather than individual testers because the testing procedure paired testers for all tests. In this study, testers were not paired and actually conducted tests with many partners. A second issue that arises is the decision to separate white testers from the black-white and Hispanic-Anglo tests in the same site. This exactly follows the approach of Heckman and Siegelman because as with the employment audits the black-white and Hispanic-Anglo tests documented in our paper were conducted as separate efforts with their own pool of testers and independent training sessions.

³⁷ See Murphy (2002) who argues that pair testing efforts should use as large a number of testers as is feasible.

³⁸ The Fisher's exact test is a permutation test in which the permutations associated with the observed outcome is compared to total number of permutations possible. The specific test used here is a test for homogeneity across the rows of the table, and the distribution of permutations is described by the multiple hypergeometric distribution. See Agresti (1990).

³⁹ Specifically, no information on house price or loan amount is scored as a zero, provision of only house price and/or loan amount is a one, additional provision of specific financing options or product options is a two, additional discussion of financial details is a three, and both discussion of products and financial details is a four. Discussion of financial details is given a higher priority than discussion of products because these details are required for

substantive discussions on financing options, but a reversal of this priority yields very similar incidences of white and minority favored tests.

⁴⁰ Specifically, a house price or loan amount is only considered in this comparison if financial details were exchanged. Results are very similar when all tests in which both testers received a loan amount or house price are included in the analysis.

⁴¹ Mail and telephone follow-up were given equal weight, with each piece of mail or telephone call counted as one follow-up contact.

⁴² Differences in number of products is significant for the severity, but not the incidence measure. In order to be conservative, this finding is not included among the nine. Reverse discrimination is found with a 10% significance level for the follow-up severity measure. The sign test is recalculated using a 10% threshold and an either/or criteria when the tests for incidence and severity measures do not agree. Based on these criteria, the likelihood of the findings arising by chance is less than 0.02.

⁴³ In fact, if we conduct the earlier sign test for adverse versus favorable treatment separately for black-white and Hispanic-Anglo hypothesis tests, the likelihood of black-white findings arising by chance is less than 0.02 while the likelihood of the Hispanic-Anglo findings arising by chance is 0.50.

⁴⁴ In fact, we can also apply the technique that we have throughout this paper and simply vary the confidence level. Specifically, we calculate the likelihood of one rejection out of six at the 0.002 level of significance and apply the Bonferroni correction. This results in a likelihood of a type I error that is less than 5 percent or coaching findings is significant at the 0.05 level.

⁴⁵ The threshold that financial details be exchanged is consistent with our emphasis on detailed discussions of options between the loan officer and tester. The threshold of two products is chosen because almost all testers in the relevant subsample were told about at least one product. Downpayment assistance is used for coaching because this variable is the primary driver of all empirical findings concerning adverse treatment in coaching, and finally FHA encouraged is used for the FHA variable because it is rare that FHA is explicitly brought up by the agent and then discouraged.

⁴⁶ The sample sizes are less than the sample sizes in Tables 2 and 3 because testers who conducted a small number of tests have been deleted and the final number of testers in the analysis are shown in the last row of the first panel in the table.

⁴⁷ Similar techniques were used to test whether the pattern of treatment differed across different types of lenders or test circumstances, such as depository versus non-depository lender, lender size, share of lenders applications that are from African Americans or Hispanics, testers encountering minority agents, and testers encountering the same agent. No statistically significant differences were found in the patterns of treatment except for the same agent variable. Furthermore, the differences identified for the same agent are associated with the frequency of equal treatment. The racial differences in treatment appear unaffected by whether the tester saw the same agent.

⁴⁸ See Ross and Yinger (2002) for a more complete discussion of adverse impact discrimination.

⁴⁹ A package of tools developed and lessons learned during the project have been made available to fair housing groups and other organizations to facilitate and improve enforcement testing (see Freiberg and Herbig, 2002).

⁵⁰ The pre-test phase of the project experimented with multiple scenarios, and the lessons and experiences from this phase are also documented in Freiberg and Herbig (2002).

⁵¹ Finally, even a dramatic increase in internet lending for all groups is unlikely to eliminate racial and ethnic differences in the provision of pre-application assistance and information. Lenders will still need to provide new homebuyers information on the overall mortgage process and the mortgage products available. Even if this information is provided by phone or email, loan officers or other customer service representatives may be able to identify race or ethnicity by name or voice. Some recent studies of discrimination by name include Bertrand and Mullainathan (2002), Fryer and Levitt (2003), and Figlio (2003) and by voice Massey and Lundy (2001).