University of Connecticut

Department of Economics Working Paper Series

Mortgage Lending in Chicago and Los Angeles: A Paired Testing Study of the Pre-Application Process

Stephen L. Ross
University of Connecticut

Margery Austin Turner
Metropolitan Housing and Communities Policy Center, Urban Institute

Erin Godfrey
New York University

Robin R. Smith
Metropolitan Housing and Communities Policy Center, Urban Institute

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

Abstract

This paper analyzes data from a recently completed study of discrimination against African-American 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 isolated significant levels of differential treatment on the basis of race and ethnicity in Chicago with African Americans and Hispanics receiving less information and assistance than comparable whites. Adverse treatment of African-Americans and Hispanics is also observed in Los Angeles for specific treatments, but the overall pattern of treatment observed did not differ statistically from equal treatment. Multivariate analyses for Chicago indicate that large lenders treat minorities more favorably than small lenders and that lenders with substantial numbers of applications from African-Americans treat African Americans more favorably than lenders with predominantly white application pools.

Journal of Economic Literature Classification: G21, J15, L85, R30

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

Margery Austin Turner, Director, Metropolitan Housing and Communities Policy Center,
    Urban Institute

Erin Godfrey, Ph.D. Candidate, New York University

Stephen L. Ross, Associate Professor, University of Connecticut

Robin R. Smith, Research Associate, Metropolitan Housing and Communities Policy
    Center, Urban Institute

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 amount to the property's assessed value, or the ratio of housing expenses to the borrower's income. The most recent and high profile 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 powerful tool for investigating adverse and differential treatment of racial and ethnic minorities.[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 provides 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. This key advantage of paired testing has been cited in a recent panel and earlier workshop sponsored by the National Resource Council, Blank, Dabady, and Citro (2004) and Foster, Mitchell, and Fienberg (2002).[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 potentially leading to greater racial disparities in underwriting. Temkin, Levy, and Levine (1999) examine a medium-sized mortgage lender that has an explicit policy of including the loan officer, who knows the race of the applicant, in the underwriting decision for problem applications.[7] Squires and Kim (1999) merge employment information from the Equal Employment Opportunity Commission with HMDA data and find that the approval rates of African-American applicants rise as the share of African-American employees at the lender increases.[8] Finally, Yezer, Phillips, and Trost (1994) and Ross and Yinger (1999) find evidence of a simultaneity between loan application attributes and a lender's underwriting standards. As in the case study, a likely way for race to influence how loan applications respond to lender underwriting standards is through the loan officer since underwriters rarely meet applicants.[9]

In addition, disparate treatment by loan officers may discourage minority applicants who as a result never submit an application or seek higher cost financing in the subprime market. Specifically, discrimination may weaken minority borrowers' 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).

While the paired testing methodology has been widely applied to study housing market (Ross and Turner, In press; Ondrich, Ross, and Yinger, 2003; Yinger, 1987) and employment discrimination (Bertrand and Mullainathan, 2004; Neumark, 1996; Kenny and Wissoker, 1994), only a few, enforcement-oriented testing efforts have been conducted in the mortgage market.[10] 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 found that 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.[11] This study recorded the treatment of minority testers posing as 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. 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 homeownership.

The paper is organized as follows: section 2 describes the paired testing methodology and the data generated by this study, section 3 describes the empirical methodologies used to measure and test for differential treatment of minorities and to

6

examine how such treatment varies across lenders, section 4 presents empirical

evidence concerning the extent and pattern of racial and ethnic differences in treatment,

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

differential treatment of African Americans and Hispanics relative to whites 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.[12] 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.[13] Cash assets required were calculated as the sum of the 5 percent downpayment and closing costs, which were based on loan amount and evidence from a pre-test phase respectively. Total housing expenses were calculated as the sum of interest and principle payments plus private mortgage insurance, homeowners insurance, and property taxes.[14] 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.[15]

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),[16] 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.[17] 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).[18]

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 racial/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. 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. These differences suggest that the walk-in protocol used in this study may not represent the market wide incidence of differential treatment, but nonetheless captures the experiences of a sizable set of potential minority borrowers.

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. [19] 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 following detailed and uniform instructions.

- *Step #2 - Make the Initial Request.* At their appointments, testers were instructed to very clearly request (up to three times, if necessary) 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 detailed information on income, debts, assets, credit history, as well as other personal and financial characteristics when requested by a lender.[20]

- *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.[21] 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 notes taken and materials obtained during the test.[22]

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:[23]

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 or quoted a higher loan amount provides a benefit to the borrower by increasing his or her financial options.

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).[24]

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, which is consistent with imposing no assumptions concerning whether a type of treatment is favorable or unfavorable.

Statistical Analysis Methodology

Each paired test typically generates data on a series of treatments *t* that are experienced by the majority and minority testers and recorded as W_{it} and M_{it}. Incidence measures are developed by comparing the treatment of the two testers in each test *i* and classifying the test as being majority favored, equal treatment, or minority favored. Gross majority or minority favored treatment is defined as the fraction of tests classified as majority or minority favored, respectively, and the net measure of adverse treatment is defined as gross majority favored treatment minus gross minority favored treatment. Specifically,

$$N_t = \Pr[W_{it} - M_{it} > 0] - \Pr[W_{it} - M_{it} < 0] \tag{1}$$

where the probability (*Pr*) is captured by the sample frequency and treatments are initialized so that a large value is interpreted as positive treatment. Similarly, a severity measure for a treatment is defined as the difference in the treatment experienced by the two testers or

$$S_t = E[W_{it} - M_{it}] \tag{2}$$

where the expected value is captured by the sample mean of the difference or a weighted mean if pooling tests across a stratified sample.

Net incidence and severity 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.[25]

In the analysis presented here, the gross incidence and the mean treatment

experienced for continuous variables are reported by both majority and minority testers.

Statistical tests for differential treatment are then conducted to determine whether the

net incidence or severity arising from the differences of majority and minority treatment

systematically differ from zero using a two sided test.[26] Even though earlier language

refers to favorable or adverse treatment, the two-sided test provides a more

conservative test for differential treatment, which does not rely on specific interpretations

of favorable versus unfavorable treatment.

Due to small sample sizes, conventional tests based on the normal distribution or

asymptotic relationships may be biased, and we follow the suggestion of Heckman and

Siegelman (1993) to use Fisher's exact (permutation) tests. In the case of incidence

data, the null hypothesis may be written as

$$H_{0t}: \Pr[W_{it} - M_{it} > 0 \,|\, W_{it} - M_{it} \neq 0] = 0.5 \tag{3}$$

The null hypothesis in equation (3) leads to the standard sign test, which can be formally

written as

$$\sum_{j=D_t^C}^{D_t^M} \Pr[D_t = j \,|\, D_t^M] \leq 0.025 \tag{4}$$

where D_t is the difference in the number of white and minority favored tests, D_t^M is the

total number of tests with differential treatment of some form, D_t^C is the critical statistic

for a 5 percent chance of a type I error where the null hypothesis of differential treatment

is rejected if D_t or D_t^M minus D_t is greater than or equal to that statistic, and

$$\Pr[D_t \,|\, D_t^M] = \frac{D_t^M! / D_t!(D_t^M - D_t)!}{2^{D_{Max}}} \tag{5}$$

which is simply the number of permutations where D_t ones are observed in D_t^M binary

variables divided by the total number of permutations for D_t^M binary variables.

In the case of severity measures, the null hypothesis may be written as

$$H_{0t}: E[W_{it} - M_{it}] = 0 \qquad\qquad (6)$$

The null hypothesis in equation (6) is tested using the Wilcoxon signed rank test. As

with the sign test, the Wilcoxon signed rank test is simply the application of a standard

permutation test for ranks after conditioning on differential treatment. Specifically, the

difference between white and minority treatment is calculated for tests where this

difference is zero and absolute differences are ordered from largest to smallest in order

to assign ranks. The signed rank statistic is the sum of ranks from all pairs that are

consistent with differential treatment, and the critical statistic is obtained as in equation

(4) by summing permutation based probabilities of observing rank sums above some

threshold. See Agresti (1990) and Ramsey and Schafer (1997) for detailed

presentations of the sign and Wilcoxon signed rank tests.

Multiple Testing Environment

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.

A common approach to control for multiple tests is a Bonferroni correction where

the simple version involves multiplying the likelihood of a type I error by number of

hypothesis tests conducted (Shaffer, 1995). This correction is appropriate for

considering the type I error associated with a single statistically significant finding that has been taken from a larger set of tests. In this case, however, the analysis of a sample of tests may yield multiple findings that are consistent with differential treatment by race or ethnicity. The overall pattern of differential treatment observed may have been highly unlikely to arise by chance even if individual findings are statistically insignificant after applying the Bonferroni correction.

In our opinion, the appropriate null hypothesis for differential treatment is that differential treatment was not identified on any treatment variables considered or

$$H_{0s}: \quad T_{1s}=0 \text{ and } T_{2s}=0 \text{ and } T_{3s}=0 \text{ and } \ldots T_{Ks}=0 \tag{7}$$

where T_{ks} is the *kth* test for differential treatment that is zero when the individual null is maintained and K is the total number of tests conducted for a sample s. Under the null hypothesis, T_{ks} is a binary variable (Bernoulli) that takes the value one with a probability of 0.05, the likelihood of a type I error. The K hypotheses form a multinomial variable, and the cumulative probability distribution of the multinomial distribution can be used to calculate the likelihood of rejecting the null of equal treatment in the sample.[27] Specifically,

$$\sum_{t=0}^{T_s^C} \Pr\left[\sum_{k=1}^{K} T_k = t\right] > 0.95 \tag{8}$$

where T_s^C is the number of individual findings of differential treatment necessary to reject the null with a type I error rate of 0.05, and

$$\Pr[t] = \frac{K!}{t!(K-t)!}(0.05)^t (0.95)^{K-t} \tag{9}$$

Of course, a full set of hypothesis tests are conducted for four samples of tests, two groups in each of two sites. In this case, a Bonferroni correction is completely appropriate since a finding of differential treatment for one group in either site will be highlighted as evidence that either race or ethnicity affects the process of mortgage

lending in that market regardless of the findings for the other group or in the other market.   Since these four samples are independent, the simple Bonferroni correction can be applied where the likelihood of a type I error for each sample is multiplied by four (Shaffer, 1995), and the 0.95 in equation (8) must be replaced with 0.9875.

In order to focus on a more parsimonious set of tests, the paper examines the six overall forms of treatment: provision of information, loan amount, number of products, coaching, recommendation of FHA, and follow-up contact. Due to the presentation of both incidence and severity measures, two tests were conducted for many of these treatments, and in order to be conservative we only count a treatment as a finding of differential treatment when equal treatment was rejected with no more than a 0.05 type I error rate for both the severity and the incidence measure. The resulting 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. After applying the Bonferroni correction, we conclude that a significant finding of differential treatment for three and four of the six null hypotheses for a given sample of tests allows us to reject the null of no differential treatment with a 0.01 and a 0.0005 type I error rate, respectively.

Tester Heterogeneity

Any paired testing effort involves taking a naturally heterogenous 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. For one of the four samples, Heckman and Siegelman rejected the null hypothesis of equal differential treatment across tester pairs suggesting that tester identity influences

17

treatment.[28] Alternatively, Turner, Ross, Galster, and Yinger (2002) examined the effect

of actual tester attributes, such as education and income,[29] on differential treatment and

found that these attributes 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 on or in some cases even increased the estimated

incidence of adverse treatment.

This paper assesses tester heterogeneity following a method similar to Heckman

and Siegelman (1993) where contingency tables were created to test whether treatment

differed systematically across tester pairs. Unlike Heckman, however, this study used a

relatively large number of testers, Chicago tests used between 9 and 12 testers for each

group and the Los Angeles tests used between 6 and 7,[30] and the testers were not

paired sometimes conducting tests with multiple partners. Therefore, contingency tables

are created that describe the pattern of treatment outcomes for individual testers, and for

each sample twelve tables are created representing the pattern of treatment for majority

and minority testers for each of the six overall treatments.[31]

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. The resulting contingency tables are *2xJ* where *J* is the total number of

testers of a given race or ethnicity in a sample of tests. The detailed variable

construction is described prior to presenting the results of the heterogeneity tests. In

spite of these provisions, many cells have small sizes, and a Fisher's exact test is used

to test for table homogeneity. The specific test used here is a test for homogeneity

across the rows of the table, which is typically conducted using a Chi-square test with

larger cell populations, and the distribution of permutations is described by the multiple

hypergeometric distribution, see Agresti (1990).

The specific null hypothesis tested is

$$H_0: \quad \Pr[W_{i_{jt}} \mid j,s] = \Pr[W_{it} \mid s] \text{ and } \Pr[M_{i_{jt}} \mid j,s] = \Pr[M_{it} \mid s] \quad (10)$$

for all testers *j* and all treatment variables *t*. Specifically, the hypotheses that treatment probabilities are the same for all majority or for all minority testers in a sample is tested via the group and treatment specific homogeneity tests, and the application of this test to both majority and minority testers for six treatment variables gives rise to the twelve tables referred to above.

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, see equations (8) and (9). The resulting likelihood of a type I error is then multiplied by four due to the four samples considered in order 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 overall null hypothesis of homogeneity can be rejected based on the number of individual hypothesis test rejection for any of following the significance thresholds: 0.10, 0.05, 0.01, 0.005, or 0.001.

Multivariate Analysis

While the test scenario was the same for all tests, the lenders visited exhibited considerable heterogeneity. This section describes our attempts to examine how differential treatment varied across lenders. In order to increase the efficiency of estimates, we restrict the treatment of majority testers to be the same in the same site *m* regardless of the minority group involved in testing, i.e. the specific sample *s*. We also start with a linear model for all treatments whether the underlying variable is discrete, ordinal, or continuous. Specifically,

19

$$W_{itsm} = \alpha_{Wtm} X_{Wism} + \beta_{Wtm} Z_{ism} + \xi_{j_{Wi}sm} + \delta_{itsm} + \varepsilon_{Witsm} \qquad (11)$$

where X_{Wism} represent test attributes that are unique to the majority tester's visit, such as the race of the agent encountered; Z_{ism} represents attributes of test i in sample s in site m that are invariant across visits, primarily attributes of the lender visited; α_{Wtm} and β_{Wtm} describe the relationship between visit and test attributes and the treatment of the majority tester on treatment variable t, which vary only across sites; δ_{itsm} represents a test fixed effect; ξ_{jsm} is the fixed effect associated with tester j, and ε_{Witsm} captures random events that arose during a tester's visit. Similarly, minority treatment is modeled as

$$M_{itsm} = \alpha_{Mtsm} X_{Mism} + \beta_{Mtsm} Z_{ism} + \xi_{j_{Mi}sm} + \delta_{itsm} + \varepsilon_{Mitsm} \qquad (12)$$

where the key difference between equations (11) and (12) is that the parameters α and β are indexed by the minority group tested and the tester j is identified as the minority tester for test i in equation (12).

Tester fixed effects are eliminated by mean differencing equations (11) and (12) by the identity of the majority or minority tester, respectively. This model is estimated using Ordinary Least Squares (OLS). MacKinnon and White (1985) show, however, that the traditional heteroskedasticity consistent covariance matrix estimator may provide seriously biased standard error estimates in small samples. Horowitz (2000) raises similar concerns about pairwise bootstrap estimation in small samples. Mammen (1993) and Flachaire (1999, In Press) all suggest either applying the HC_2 or HC_3 correction to the traditional pairwise bootstrap and/or shifting to a wild bootstrap procedure.

Accordingly, consistent standard errors and F-statistics are generated using a wild bootstrap with the HC_2 correction.[32] Specifically, 10,000 bootstrap samples are generated using the following specification for the dependent variable in each bootstrap sample b:

$$\widetilde{Y}_{itsmb} = \hat{Y}_{ltsm} + \varpi_{itsmb}\left(\hat{\varepsilon}_{ltsm}\Big/(1 - h_{lsm})^{0.5}\right) \qquad (13)$$

where i is a randomly chosen observation from the sample s in site m, $\hat{Y}_{ltsm}$ is the

predicted value of $(W_{ltsm} - M_{ltsm})$ based on the OLS estimates obtained from mean

differenced versions of equations (11) and (12), ϖ_{itsmb} is a random normal variable with

mean zero and variance one drawn separately for each treatment, observation, and

bootstrap sample. $\hat{\varepsilon}_{ltsm}$ is the predicted residual or the difference between

$(W_{ltsm} - M_{ltsm})$ and $\hat{Y}_{ltsm}$, and

$$h_{lsm} = \mathrm{X}_{lsm}(\mathrm{X}_m'\mathrm{X}_m)^{-1}\mathrm{X}_{lsm}' \text{ with } \mathrm{X}_m = [(X_{Wism} - X_{Mism}), X_{Mism}, Z_{ism}] \qquad (14)$$

which is often referred to as the leverage of the observation. Formally, one minus h_{lsm} is

the variance of the predicted residual for observation i.

An additional analysis is conducted to mitigate the impact of test fixed effects.

Specifically, majority and minority mean differenced equations are differenced to yield

$$\overline{W}_{itsm} - \overline{M}_{itsm} = \alpha_{Wtm}(\overline{X}_{Wism} - \overline{X}_{Mism}) + (\alpha_{Wtm} - \alpha_{Mtsm})\overline{X}_{Mism} +$$
$$(\beta_{Wtm} - \beta_{Mtsm})\overline{Z}_{ism} + (\overline{\delta}_{itsm}^{W} - \overline{\delta}_{tism}^{M}) + (\overline{\varepsilon}_{Witsm} - \overline{\varepsilon}_{Mitsm}) \qquad (15)$$

where the bar above a variable indicates a tester mean differenced variable and the

superscipt on the mean differenced test fixed effect indicates whether the effect was

mean differenced based on the white or minority tester identity. This differencing does

not completely eliminate the influence of test fixed effects. Specifically,

$$(\overline{\delta}_{itsm}^{W} - \overline{\delta}_{tism}^{M}) = Mean_{i|j=j_{Wi}}[\delta_{itsm}] - Mean_{i|j=j_{Mi}}[\delta_{itsm}] \qquad (16)$$

Nonetheless, there are a number of reasons to believe that any bias arising from

test unobservables will be mitigated by the use of equation (15). First, since pairs of

testers are assigned to randomly selected advertisements, the mean test effects in

equation (16) should limit to zero as the number of tests conducted by each tester

increases. The average number of tests per tester is about seven and ten in Chicago

and Los Angeles, respectively. Second, equation (16) equals zero if majority and

minority testers are always paired with the same teammate. Even though this project did

not use fixed pairs of testers, testers were often assigned to the same partner due to

schedules, matching on age and gender, and habit, which should further reduce the

influence of test fixed effects in equation (15).

Before concluding this section, we believe that it is important to discuss our

choice of the linear model over non-linear maximum likelihood estimators like the logit,

probit or ordered probit. The linear model offers two important advantages. First, the

model allows for a simple approach for eliminating test fixed effects that is consistent

across all dependent variables considered in our analysis. Second, our bootstrap

estimating procedure is both supported by existing research on the linear model and

facilitated by the computational advantages of that model. On the other hand, the

consistency advantage typically offered by non-linear models is lost in the context of

both the small sample size and the potentially complex error structure in our data.

Empirical Results

The first three treatment variables considered refer directly to the information

requested in the testing protocols, and the results are shown in Table 3.   The table

presents both incidence indicators that represent the frequency with which the white or

majority tester is 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 # for significance at

the 10% level, * for significance at the 5% level and ** for significance at the 1% level or

better. This discussion focuses primarily on findings significant at the 5% level or better.

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 (0-4) 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.[33] 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.[34]

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.[35] 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

The pattern of findings above are summarized in Table 5, which shows the significance level for estimated differential treatment for each of the six broad treatment categories with the levels for specific variables listed in the order presented in Tables 3 and 4. At the bottom of the table, the number of findings at the 0.05 significance level is summarized for each sample where findings are not counted if any variables for a given treatment category are not statistically significant. In Chicago, Significant differential treatment is observed for three and four of the treatment categories for Anglo-Hispanic and black-white tests, respectively, leading to a rejection of the null of equal treatment with only a 0.01 and 0.001 chance of type I error. Differential treatment is rejected for only one of the six treatments for both samples in Los Angeles, and the overall pattern of treatment in Los Angeles is consistent with the null hypothesis of equal treatment of majority and minority testers.

The reader should also note that all findings of differential treatment in Chicago are in the direction that had an a priori interpretion as differential treatment of minority borrowers. Hispanics were quoted lower loan amounts, provided less information on

specific loan products, and received less assistance or coaching than white testers. Similarly, African-Americans were less likely to receive the specific information requested, provided less information on specific loan products, received less assistance or coaching than white testers, and were less likely to experience follow-up contact from the lender. This pattern reinforces our conclusion that systematic differences exist between the treatment experienced by minorities and whites during the pre-application phase of applying for a mortgage.[36]

Some readers may feel that we are being too conservative in interpreting the results in Los Angeles. For example, in the Anglo-Hispanic sample, differential treatment on number of products shown is statistically significant for the severity measure, which contains more information than the incidence variable. Even if we include this finding, however, there are still only two findings of differential treatment for the sample, and such a pattern still cannot be distinguished statistically from equal treatment.[37]

In our opinion, the key finding in Los Angeles that deserves additional consideration is differential treatment on the coaching variable for black-white tests. The coaching differences 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. Moreover, a simple Bonferroni correction to the significance level of downpayment assistance (multiplying by 92 for the total number of tests in Tables 2 and 3) yields a significance level of 0.01. Therefore, even after considering the multiple testing environment, statistically significant differences in the provision of 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.

26

Tester Heterogeneity

As discussed earlier, this paper assesses tester heterogeneity following a method similar to Heckman and Siegelman (1993). The six specific treatment 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,[38] the loan amount variable is 1 if the loan amount provided by the loan officer is above the average value for all testers in a sample of tests, 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.[39]

The results of the heterogeneity analysis are shown in Table 6. For each treatment 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.[40] The last row in the top panel of the table shows the number of testers in each sample. 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, tester homogeneity is rejected at the 5% significance level using both the 0.01 and 0.001 significance level thresholds, and for Hispanic-Anglo tests in Chicago tester homogeneity is rejected based on the 0.1 threshold. For black-white tests in Chicago and Hispanic-Anglo tests in Los Angeles, there is no evidence of tester heterogeneity. 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, if behavioral differences between testers of the same race affect relative treatment in those samples, similar differences between the white and minority testers may affect differential treatment, and

there is no way to determine whether these behavioral differences vary systematically by race or ethnicity in the samples of testers.

In terms of interpreting the previous findings of differential treatment, this analysis supports the finding of differential treatment of African-Americans in Chicago, but raises some questions about the finding for Hispanics in Chicago. In spite of these concerncs, however, we believe that the finding of differential treatment of Hispanics in Chicago is still quite important. The only study that has quantified the influence of tester attributes on treatment (Turner, Ross, Galster, and Yinger, 2002) found that in most cases controlling for tester attributes has no influence on differential treatment and when attribute differences affected differential treatment they tended to improve outcomes for minorities biasing analyses away from finding discrimination. Moreover, 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 minor behavioral differences that have no connection to the lender's business goals.[41]

Exploring Variation across Lenders

A multivariate analysis is conducted for each of the six major treatment variables presented in Tables 3 and 4. In terms of detailed loan information we examine the determinants of the information provided index, the loan amount quoted by the loan officer, and the number of products discussed. Additional treatments considered include whether follow-up contact was made, whether the tester was encouraged to consider FHA financing, and number of types of coaching provided, as well as whether coaching on downpayment assistance was provided since this specific form of coaching appeared important for three of the four samples.

Table 7 presents the mean and standard errors for visit and lender attributes. The four columns represent the four samples of tests over two sites and two minority groups. The first four rows present the frequency with which the majority and minority testers in each sample encountered African-American or Hispanic loan officers. As expected, the representation of Hispanics among loan officers is substantially higher in Los Angeles than Chicago, but there is no evidence to suggest that black or Hispanic testers were more likely to encounter minority agents. The last four rows present the average lender share of African-American and Hispanic applications, average the number of loans in the market made by each lender in 1999, and the representation of depository lenders in each sample. The share of applications from Hispanic borrowers is obviously higher in Los Angeles. In addition, the lenders are larger on average and are less likely to be depository lenders. The final row contains the sample sizes.

Table 8 contains the estimates for the treatment models in Chicago controlling for tester fixed effects. Each column contains the estimates for a specific treatment variable. The first six rows present the coefficients that describe the treatment of white testers. The second and third sets of rows present the racial and ethnic differences in treatment, respectively. The last four rows present the F-statistics for each model (white, racial difference, and ethnic difference) and the sample size.

In Chicago, three of the seven white models and three of the fourteen racial and ethnic difference models are statistically significant at varying levels of confidence. However, given the large number of hypothesis tests conducted, we focus on behavioral relationships that are consistent across a variety of treatments. In terms of the treatment of white testers, African-American loan officers provide applicants more coaching, are more likely to make follow-up contact, and more likely to encourage FHA financing, which is viewed as a high cost form of financing. Lenders with a higher share of black applicants are less likely to provide the requested information, provide information about

fewer products, less coaching, and less likely to initiate follow-up contact, but are more likely to encourage the applicant to consider FHA financing. Finally, larger lenders (in terms of number of loans originated) provide less coaching concerning down payment assistance, are less likely to make follow-up contact, but again more likely to encourage FHA financing.

These same variables appear important in terms of the relative treatment of black and Hispanic testers. In fact, the pattern of findings often mirrors the white coefficients. Hispanic testers encountering African-American loan officers receive less coaching including less downpayment assistance and are less likely to receive follow-up contact relative to the treatment experienced by white testers. Black testers visiting lenders with a high share of black applicants are provided information on more products, more coaching including downpayment assistance, more likely to receive follow-up contact, and less likely to be encouraged to apply of FHA financing. In addition, Black and Hispanic testers visiting large lenders receive higher loan amounts (Hispanic only), more coaching including downpayment assistance, more follow-up contact, and are less likely to be encouraged to apply for FHA Financing (Hispanic only). In summary, Black loan officers appear to treat Hispanics worse than white and Hispanic loan officers, but large lenders and lenders with a large share of Black applications appear to treat one or more of the minority groups relatively better and are less likely to steer those minority groups to high cost FHA financing.

Table 9 shows comparable estimates for the model that controls for tester and test fixed effects. The precision of the estimates clearly declines with the degrees of freedom, and the magnitude of the parameter estimates varies somewhat, but the overall pattern of results is quite robust. The one exception is the findings concerning African-American loan officers. Specifically, the tendancies of African-American loan officers to provide more coaching overall and to be more likely to initiate follow-up

contact with Hispanic testers both drop in magnitude substantially. The sensitivity of the estimates associated with loan officer race should not be surprising. The tester fixed effect models cannot distinguish between the behavior of minority loan officers and the behavior of loan officers at lenders who hire a substantial number of minority loan officers. The test fixed effects specification mitigates the influence of lender unobservables such as the tendency to hire minority loan officers. The reader should also note that the basic pattern of results observed in Tables 8 and 9 is found in simple ordinary least squares estimates of treatment, i.e. no tester or test fixed effects.

Table 10 presents the tester fixed effect estimates for Los Angeles. The variables that are important for determining treatment are similar, but the pattern of treatment appears less consistent and the statistical evidence for differential treatment is relatively weaker. First, in terms of the treatment of white testers in Los Angeles, African-American loan officers provide information about fewer products, less downpayment assistance, and are less likely to initiate follow-up contact, and similarly Hispanic loan officers provide less downpayment assistance. On the other hand, both African-American and Hispanic loan officers are more likely to provide the basic information requested by the white tester. Furthermore, the results for number of products and follow-up contact are not robust to controlling for test fixed effects (table 11). A larger share of Hispanic loans, i.e. loans to the dominant minority group, is consistent with a higher likelihood of both follow-up contact and encouraging FHA financing. Finally, larger lenders appear less likely to provide coaching assistance including downpayment assistance.

In terms of differential treatment, only one of the fourteen models of racial and ethnic differences in treatment is statistically significant at the 5% level or better. In addition, none of the significant coefficient estimates in the black model are robust across more than one treatment variable. In terms of differential treatment of Hispanic

testers, African-American loan officers appear to provide Hispanic testers with information about a relatively larger number of products and are relatively more likely to provide downpayment assistance. As in Chicago, larger lenders appear to provide more favorable treatment to Hispanics in terms of providing the information requested and providing more coaching including downpayment assistance. The effect of lender size is robust to controlling for test effects, but the observed tendency of black loan officers to provide information about fewer products is not robust, see Table 11. In summary, the finding that larger lenders treat Hispanic testers more favorably is robust across alternative specifications and multiple treatments and consistent with findings for Chicago. On the other hand, the relationship between loan officer race and the treatment of Hispanic testers is less robust and not consistent with findings in Chicago.

Conclusion

This paper finds strong evidence of adverse treatment of African Americans and Hispanics in the Chicago metropolitan area across a wide variety of pre-application treatments. African Americans were less likely to be provided the requested information during their visit, received detailed information about fewer products, and were provided with less coaching and were less likely to experience follow-up contact. Hispanics in Chicago were quoted lower loan amounts and house prices, received information about fewer products, and received less coaching. Furthermore, the finding of adverse treatment of Hispanic and African-Americans is supported by empirical regularities in the pattern of treatment. Larger lenders provide more favorable treatment to black and Hispanic testers relative to their white teammates in terms of loan amount, coaching, follow-up contact, and not being steered towards FHA financing. In addition, lenders with a larger fraction of African-Americans in their applicant pool provide more favorable treatment to African-Americans in terms of number of products discussed, amount of coaching, follow-up contact, and not steering toward FHA financing.

The evidence of adverse treatment in Los Angeles is substantially weaker. Statistically significant racial and ethnic differences in treatment are identified for only two of the six major treatment variables for Hispanic testers and only one of the six for black testers. After accounting for the large number of tests conducted, the overall pattern of differential treatment is not statistically significant. Similarly, the multivariate analysis of treatment in Los Angeles provides less support for the notion that minority testers receive different levels of pre-application information and assistance. The multivariate models exhibit less statistical significance overall relative to the models for Chicago. In fact, the multivariate analysis finds almost no evidence of systematic variation in the treatment of black testers in Los Angeles other than on the FHA encouragement variable.

Two key findings for Los Angeles are worth noting. Racial differences in the provision of downpayment assistance are large and statistically significant even after controlling for the multiple testing environment. Lower levels of downpayment assistance for minority testers are identified for three of the four samples suggesting that adverse treatment of minorities on this particular treatment may be widespread. In terms of the multivariate analysis, Hispanics appear to experience relatively more favorable treatment from large lenders, which is consistent with the findings for black and Hispanic testers in Chicago. Accordingly, the study offers fairly strong evidence that large lenders are less likely to discriminate during the pre-application process than small lenders.

The paper, however, finds 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 determining treatment for these two samples. While the findings for Hispanic-Anglo tests in Chicago clearly constitute adverse treatment by ethnicity that is unrelated to a borrower's financial qualifications, the findings 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.

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 (2004), "Are Emily and Gred more
Employable than Lakisha and Jamal ? A Field Experiment on Labor Market
Discrimination." _American Economic Review_ 94, 991-1013.

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).

Blank, Rebecca M., Marilyn Dabady, and Constance F. Citro Eds. (2004), _Measuring
Racial Discrimination_ (Washington, D.C.: The National Academies Press).

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 Researve 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.

Charles, Kerwin and Erik Hurst. 2002. "The transition to home ownership and the black/white wealth gap." *Review of Economics and Statistics* 84, 281-297.

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).

Flachaire, E. (Forthcoming), "Bootstrapping Heteroskedastic Regression Models: Wild Bootstrap versus Pairs Bootstrap," *Computational Statistics and Data Analysis*.

Flachaire, E. (1999), "A Better Way to Bootstrap Pairs," *Economics Letters*, 64, pp. 257-262.

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. In press. "The causes and consequences of distinctively black names." *Quarterly Journal of Economics*.

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.

Horowitz, J.L. (2000) "The Bootstrap" in J. Heckman and E. Leamer (Eds), *Handbook of Econometrics, Vol 5*. Elsevier Science.

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.

MacKinnon, J.G and H.L. White. (1985), "Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties. *Journal of Econometrics*, 21, pp. 53-70.

Mammen E. (1993), "Bootstrap and Wild Bootstrap for High Dimensional Linear Models,"
Annals of Statistics, 21, pp. 255-285.

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).

Neumark, David. (1996). "Sex discrimination in restaurant hiring: An audit study."
Quarterly Journal of Economic 111, pp.915-41.

Ondrich, Jan, Stephen L. Ross, and John Yinger. 2003. "Now you see it, Now you don't:
Why do real estate agents withhold available houses from black customers."
Review of Economics and Statistics 85, pp. 854-873.

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.

Rachlis, Mitchell B. and Anthony M. Yezer. (1993). "Serious flaws in statistical tests for
discrimination in mortgage markets." _Journal of Housing Research_ 4, pp. 315-336.

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 L. and Margery A. Turner. In press. "Housing discrimination in metropolitan American: Explaining changes between 1989 and 2000." *Social Problems*.

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.

White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a

      Direct Test for Heteroskedasticity." *Econometrica*, 48, pp. 817-838.

Yezer, Anthony H., Robert F. Phillips, and Robert P. Trost. 1994. "Bias in estimates of

      discrimination and default in mortgage lending." *Journal of Real Estate Finance*

      and Economics, 9, pp. 197-215.

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).

Yinger, John. 1987. "Measuring Discrimination with Fair Housing Audits: Caught in the

Act." *American Economic Review,* 76 (December): pp. 881-893.

Table 1: Population of Eligible Lending Institutions

	# of Institutions		Application Volume		% of HMDA Activity	
Lenders Tested	_LA_	_Chicago_	_LA_	_Chicago_	_LA_	_Chicago_
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	

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	

Sites	Chicago		Los Angeles	
Group	**Hispanic-Anglo**	**Black-White**	**Hispanic-Anglo**	**Black-White**
Information Provided	0.647	0.037*	0.027*	0.424
Loan Amount or Price	0.010/0.002/.0003/.0001**	0.136/0.473/0.138/0.461	0.766/0.542/0.602/0.900	0.864/0.618/0.978/0.864
Products Discussed	0.036/0.050*	0.024/0.009*	0.233/0.032	0.736/0.437
Coaching Provided	0.004/0.004**	0.043/0.010*	0.897/0.666	0.011/0.002*
Follow-Up Contact	0.607/0.209	0.012/0.009*	0.388/0.062	1.0000.672
FHA Encouraged	0.824/1.000	0.302/0.754	1.000/1.000	0.180/0.375
Statistical Thresholds	**Number of Rejections**	**Number of Rejections**	**Number of Rejections**	**Number of Rejections**
0.05 or better	3**	4**	1	1

Table 5: Summary of Findings and Significance Levels

Table 6: 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	

Table 7: Means and Standard Errors of Lender Variables				
	Chicago		Los Angeles	
Control Variables	**Hispanic-Anglo**	**Black-White**	**Hispanic-Anglo**	**Black-White**
Majority tester encounters Af-Am. officer	0.127 (0.335)	0.117 (0.323)	0.089 (0.286)	0.068 (0.253)
Majority tester encounters Hispanic officer	0.063 (0.245)	0.117 (0.323)	0.228 (0.422)	0.324 (0.471)
Minority tester encounters Af-Am. officer	0.063 (0.245)	0.143 (0.352)	0.101 (0.304)	0.068 (0.253)
Minority tester encounters Hispanic officer	0.076 (0.267)	0.117 (0.323)	0.215 (0.413)	0.311 (0.466)
Share of African-American Applications	0.128 (0.159)	0.125 (0.144)	0.070 (0.058)	0.071 (0.060)
Share of Hispanic Applications	0.120 (0.117)	0.118 (0.101)	0.300 (0.209)	0.269 (0.178)
Loan volume in 1000's	2.279 (2.332)	2.264 (2.315)	5.172 (4.820)	5.298 (4.921)
Whether lender is a depository Lender	0.671 (0.473)	0.740 (0.441)	0.595 (0.494)	0.621 (0.488)
Sample Size	79	77	79	74

Table 8: Regression Analysis of Differential Treatment in Chicago with Tester Fixed Effects

Variables	Information Index	Loan Amount in $1,000's	Number of Products	Amount of Coaching	Downpayment Assistance	Follow-up Contact	FHA Encouraged
Loan Officer African-American	-0.082	1.436	0.028	0.490**	0.156	0.278**	0.197#
	(0.31)	(0.32)	(0.05)	(2.90)	(1.32)	(2.41)	(1.73)
Loan Officer Hispanic	-0.338	10.883	-0.512	-0.215	-0.118	0.096	-0.080
	(1.00)	(1.45)	(1.03)	(0.94)	(0.79)	(1.04)	(0.67)
Share African-American Applications	-1.313**	3.248	-2.762**	-0.973*	-0.415	-0.323**	0.448#
	(2.35)	(0.23)	(3.27)	(2.13)	(1.49)	(2.36)	(1.82)
Share Hispanic Applications	0.211	5.857	1.887*	-0.023	0.126	-0.123	0.999**
	(0.33)	(0.23)	(2.02)	(0.03)	(0.24)	(0.54)	(2.44)
Loan Volume	-0.049	-0.297	0.006	-0.034	-0.032*	-0.018#	0.032#
	(1.22)	(0.35)	(0.09)	(1.20)	(2.00)	(1.75)	(1.89)
Depository Lender	-0.269	-5.113	0.126	-0.086	-0.107	-0.028	-0.141
	(1.52)	(0.73)	(0.37)	(0.49)	(1.04)	(0.54)	(1.46)
Black Tester: Loan Officer Af-Am.	-0.309	16.039	-0.678	-0.188	-0.161	-0.132	0.365#
	(0.67)	(0.96)	(0.80)	(0.37)	(0.67)	(0.83)	(1.74)
Black Tester: Loan Officer Hispanic	0.077	2.337	-0.019	0.306	0.279	-0.084	0.265
	(0.16)	(0.26)	(0.02)	(0.74)	(1.40)	(0.86)	(1.61)
Black Tester: Share Af-Am. Applications	-0.172	4.869	1.575	1.381	1.221**	0.281	-0.762#
	(0.12)	(0.20)	(1.19)	(1.15)	(2.41)	(1.63)	(1.69)
Black Tester: Share Hisp Applications	0.033	14.329	-0.471	-0.143	0.255	0.245	1.150**
	(0.03)	(0.28)	(0.29)	(0.10)	(0.37)	(0.83)	(2.35)
Black Tester: Loan Volume	0.035	0.274	0.029	0.104*	0.069**	0.027#	-0.017
	(0.54)	(0.19)	(0.33)	(1.97)	(2.51)	(1.94)	(0.75)
Black Tester: Depository Lender	-0.011	-3.079	0.284	-0.122	0.089	0.085	-0.214
	(0.04)	(0.29)	(0.56)	(0.29)	(0.50)	(1.23)	(1.40)
Hisp Tester: Loan Officer Af-Am.	0.278	15.301	1.381	-1.081**	-0.432#	-0.404**	0.042
	(0.44)	(1.35)	(1.02)	(2.37)	(1.91)	(2.46)	(0.15)
Hisp Tester: Loan Officer Hispanic	-0.258	-6.154	-0.358	0.600	-0.032	0.023	0.190
	(0.31)	(0.43)	(0.29)	(1.48)	(0.15)	(0.12)	(0.74)
Hisp Tester: Share Af-Am. Applications	-1.204	13.824	0.224	-0.025	0.436	0.127	-0.439
	(1.26)	(0.50)	(0.14)	(0.04)	(1.38)	(0.46)	(0.87)
Hisp Tester: Share Hisp Applications	2.101	4.574	-1.197	-1.131	-0.455	-0.367	0.572
	(0.13)	(0.12)	(0.43)	(1.09)	(0.76)	(0.76)	(0.97)
Hisp Tester: Loan Volume	0.006	2.916*	-0.060	0.078*	0.037*	0.028	-0.041#
	(0.08)	(1.97)	(0.63)	(1.96)	(2.03)	(1.23)	(1.77)
Hisp Tester: Depository Lender	0.723#	4.257	-0.161	-0.133	0.010	-0.037	-0.077
	(1.88)	(0.42)	(0.31)	(0.50)	(0.08)	(0.30)	(0.61)
F-Test: White	1.29 [0.262]	0.72 [0.634]	3.30 [0.0004]**	2.67 [0.015]*	1.49 [0.181]	1.43 [0.203]	7.15 [0.0000]**
F-Test: Black	0.14 [0.991]	0.29 [0.941]	0.30 [0.936]	1.44 [0.185]	2.61 [0.018]*	1.10 [0.362]	3.51 [0.002]**
F-Test: Hispanic	1.06 [0.387]	1.34 [0.239]	0.31 [0.931]	2.26 [0.038]*	1.51 [0.175]	1.22 [0.296]	0.96 [0.453]
Sample Size	312	297	261	312	312	312	312

Table 9: Regression Analysis of Differential Treatment in Chicago with Tester and Test Fixed Effects

Variables	Information Index	Loan Amount in $1,000's	Number of Products	Amount of Coaching	Downpayment Assistance	Follow-up Contact	FHA Encouraged
Loan Officer African-American	0.107 (0.19)	9.939 (1.14)	-0.164 (0.40)	0.208 (0.73)	0.141 (0.76)	0.362** (2.30)	0.342# (1.76)
Loan Officer Hispanic	-0.203 (0.38)	24.557# (1.78)	-0.715 (1.31)	0.159 (0.48)	-0.026 (0.13)	0.125 (0.89)	0.095 (0.56)
Black Tester: Loan Officer Af-Am.	-0.691 (1.19)	11.890 (0.64)	0.268 (0.44)	0.075 (0.11)	-0.179 (0.61)	-0.274 (1.21)	0.256 (0.84)
Black Tester: Loan Officer Hispanic	0.253 (0.46)	-10.844 (0.94)	0.530 (1.06)	0.667 (1.32)	0.365 (1.41)	-0.046 (0.38)	0.273 (1.28)
Black Tester: Share Af-Am. Applications	0.047 (0.04)	20.047 (0.53)	1.661# (1.78)	2.795* (2.04)	1.457# (1.83)	0.520# (1.83)	-0.557 (0.93)
Black Tester: Share Hisp Applications	0.348 (0.31)	53.534 (0.90)	1.113 (0.63)	0.175 (0.11)	0.354 (0.45)	0.535 (1.28)	1.195* (2.03)
Black Tester: Loan Volume	-0.002 (0.04)	0.195 (0.11)	0.043 (0.51)	0.083 (1.18)	0.061# (1.77)	0.041* (2.13)	-0.016 (0.66)
Black Tester: Depository Lender	-0.056 (0.16)	1.693 (0.12)	0.250 (0.61)	0.039 (0.09)	0.236 (1.17)	0.069 (0.69)	-0.116 (0.75)
Hisp Tester: Loan Officer Af-Am.	0.204 (0.39)	6.567 (0.44)	0.331 (0.27)	-1.084* (2.09)	-0.323 (1.40)	-0.169 (1.11)	-0.023 (0.08)
Hisp Tester: Loan Officer Hispanic	0.481 (0.31)	-3.415 (0.11)	1.229 (1.34)	0.419 (0.67)	-0.373 (1.08)	0.028 (0.15)	0.343 (1.23)
Hisp Tester: Share Af-Am. Applications	-1.608 (0.90)	-10.239 (0.30)	-0.905 (0.74)	-1.063 (1.14)	0.402 (0.94)	0.014 (0.04)	-0.448 (0.72)
Hisp Tester: Share Hisp Applications	1.445 (0.81)	-12.025 (0.28)	-1.564 (0.57)	-1.718 (1.22)	-0.757 (1.12)	-0.823 (1.23)	0.802 (1.24)
Hisp Tester: Loan Volume	0.058 (0.63)	4.185** (2.37)	-0.091 (1.14)	0.081# (1.79)	0.061** (2.61)	0.041# (1.80)	-0.029 (1.35)
Hisp Tester: Depository Lender	0.575 (1.41)	-0.117 (0.01)	-0.597 (1.34)	-0.397 (1.35)	-0.146 (1.07)	-0.090 (0.68)	-0.092 (0.69)
F-Test: White	0.10 [0.905]	1.61 [0.204]	0.84 [0.434]	0.34 [0.712]	0.35 [0.705]	2.63 [0.076]#	1.64 [0.198]
F-Test: Black	0.42 [0.865]	0.63 [0.706]	0.84 [0.542]	1.65 [0.138]	1.45 [0.200]	1.37 [0.231]	1.94 [0.078]#
F-Test: Hispanic	0.65 [0.690]	1.36 [0.236]	0.86 [0.527]	2.02 [0.067]#	1.80 [0.103]	0.80 [0.571]	0.87 [0.519]
Sample Size	156	141	130	156	156	156	156

Table 10: Regression Analysis of Differential Treatment in Los Angeles with Tester Fixed Effects

Variables	Information Index	Loan Amount in $1,000's	Number of Products	Amount of Coaching	Downpayment Assistance	Follow-up Contact	FHA Encouraged
Loan Officer African-American	0.370#	-1.97	-0.990*	-0.274	-0.298**	-0.156**	-0.038
	(1.67)	(0.13)	(2.31)	(0.68)	(2.50)	(2.79)	(0.39)
Loan Officer Hispanic	0.354#	-4.81	-0.008	-0.235	-0.199**	0.004	0.005
	(1.83)	(0.45)	(0.02)	(1.49)	(2.34)	(0.08)	(0.09)
Share African-American Applications	-0.365	-77.70	0.553	-0.852	-0.294	0.014	1.338#
	(0.20)	(1.03)	(0.20)	(0.50)	(0.42)	(0.03)	(1.88)
Share Hispanic Applications	0.267	-0.33	0.520	0.274	0.033	0.280#	0.603**
	(0.59)	(0.02)	(0.79)	(0.67)	(0.18)	(1.81)	(3.17)
Loan Volume	-0.006	-0.19	0.043	-0.030*	-0.013#	-0.003	0.003
	(0.37)	(0.20)	(1.53)	(2.10)	(1.84)	(0.07)	(0.47)
Depository Lender	0.161	-13.81	0.308	-0.030	0.140	0.051	-0.061
	(0.78)	(1.45)	(0.97)	(0.18)	(0.18)	(0.92)	(0.95)
Black Tester: Loan Officer Af-Am.	0.269	-19.58	0.627	-0.160	0.078	0.099	0.192
	(0.65)	(0.51)	(0.80)	(0.31)	(0.44)	(0.12)	(0.81)
Black Tester: Loan Officer Hispanic	-0.229	-13.54	0.745	0.247	0.285**	-0.102	0.116
	(0.57)	(0.64)	(1.45)	(0.96)	(2.49)	(1.30)	(0.94)
Black Tester: Share Af-Am. Applications	0.631	215.26	-2.145	2.077	1.175	-0.445	-2.751**
	(0.22)	(0.86)	(0.56)	(0.83)	(1.13)	(0.57)	(2.74)
Black Tester: Share Hisp Applications	-0.027	-37.90	1.398	-0.473	-0.251	-0.030	0.479
	(0.03)	(0.92)	(1.24)	(0.66)	(0.78)	(0.13)	(1.36)
Black Tester: Loan Volume	-0.002	-0.43	-0.002	-0.006	0.001	0.004	-0.024*
	(0.09)	(0.25)	(0.06)	(0.26)	(0.09)	(0.47)	(2.29)
Black Tester: Depository Lender	-0.286	-34.10#	-0.228	0.186	-0.006	-0.118	0.044
	(0.85)	(1.73)	(0.51)	(0.56)	(0.04)	(1.21)	(0.35)
Hisp Tester: Loan Officer Af-Am.	0.297	16.10	1.411**	0.482	0.492*	-0.147	0.091
	(0.82)	(0.77)	(2.34)	(0.85)	(2.07)	(1.33)	(0.48)
Hisp Tester: Loan Officer Hispanic	-0.910#	9.79	-0.706	0.104	0.056	-0.107	-0.003
	(1.80)	(0.41)	(1.44)	(0.33)	(0.44)	(1.00)	(0.03)
Hisp Tester: Share Af-Am. Applications	-1.722	14.40	-3.262	-1.261	0.944	1.153	0.249
	(0.42)	(0.09)	(0.75)	(0.44)	(0.72)	(0.90)	(0.22)
Hisp Tester: Share Hisp Applications	0.290	9.98	0.374	0.269	-0.344	-0.306	-0.191
	(0.32)	(0.22)	(0.38)	(0.40)	(1.24)	(1.14)	(0.61)
Hisp Tester: Loan Volume	0.052	-0.75	-0.037	0.052#	0.026*	-0.008	-0.010
	(1.62)	(0.43)	(1.00)	(1.88)	(2.10)	(1.06)	(0.97)
Hisp Tester: Depository Lender	-0.243	-0.05	-0.485	-0.080	-0.057	0.056	-0.001
	(0.60)	(0.00)	(1.10)	(0.29)	(0.45)	(0.05)	(0.01)
F-Test: White	1.16 [0.328]	0.57 [0.754]	1.45 [0.196]	1.85 [0.089]#	4.41 [0.000]**	1.75 [0.109]	5.91 [0.000]**
F-Test: Black	0.33 [0.921]	0.90 [0.495]	0.88 [0.510]	0.35 [0.910]	1.39 [0.218]	0.90 [0.495]	2.86 [0.010]**
F-Test: Hispanic	1.17 [0.322]	0.20 [0.977]	2.01 [0.065]#	0.78 [0.586]	1.77 [0.105]	1.01 [0.419]	0.35 [0.910]
Sample Size	306	281	267	306	306	306	306

Table 11: Regression Analysis of Differential Treatment in Los Angeles with Tester and Test Fixed Effects

Variables	Information Index	Loan Amount in $1,000's	Number of Products	Amount of Coaching	Downpayment Assistance	Follow-up Contact	FHA Encouraged
Loan Officer African-American	2.189	18.606	0.110	-0.108	-0.303	0.091	0.169
	(4.35)	(0.42)	(0.22)	(0.14)	(1.28)	(1.31)	(1.14)
Loan Officer Hispanic	0.293	39.233	-0.046	-0.241	-0.266	-0.037	-0.140
	(0.68)	(1.65)	(0.11)	(0.84)	(2.01)	(0.39)	(0.91)
Black Tester: Loan Officer Af-Am.	-0.153	-15.936	0.419	-0.014	0.222	-0.043	0.135
	(0.19)	(0.32)	(0.71)	(0.02)	(0.73)	(0.50)	(0.58)
Black Tester: Loan Officer Hispanic	-0.519	-24.615	0.587	0.266	0.293	-0.216	0.166
	(1.15)	(1.03)	(1.39)	(0.94)	(2.31)	(2.28)	(1.12)
Black Tester: Share Af-Am. Applications	-2.850	225.493	-4.024	-2.159	-0.377	-0.459	-2.694
	(0.69)	(0.73)	(0.95)	(0.77)	(0.30)	(0.49)	(1.98)
Black Tester: Share Hisp Applications	-0.908	-59.186	0.908	-0.769	-0.382	0.181	0.437
	(0.79)	(1.07)	(0.89)	(0.87)	(1.00)	(0.75)	(0.98)
Black Tester: Loan Volume	-0.017	-1.638	-0.011	-0.030	0.002	0.000	-0.029
	(0.45)	(0.96)	(0.33)	(0.94)	(0.20)	(0.03)	(2.42)
Black Tester: Depository Lender	-0.628	-49.119	-0.668	-0.370	-0.164	-0.041	-0.097
	(1.53)	(2.19)	(1.88)	(0.91)	(1.00)	(0.41)	(0.70)
Hisp Tester: Loan Officer Af-Am.	-0.151	16.084	0.724	0.948	0.704	-0.071	0.065
	(0.36)	(0.62)	(1.45)	(1.44)	(2.92)	(0.54)	(0.32)
Hisp Tester: Loan Officer Hispanic	-0.784	-10.419	-0.126	0.151	0.052	-0.090	-0.101
	(1.35)	(0.42)	(0.24)	(0.43)	(0.34)	(0.82)	(0.68)
Hisp Tester: Share Af-Am. Applications	-0.490	-16.588	-0.139	1.766	2.020	1.267	-0.267
	(0.14)	(0.08)	(0.04)	(0.72)	(1.81)	(0.91)	(0.22)
Hisp Tester: Share Hisp Applications	0.557	24.930	1.161	0.401	-0.246	-0.366	-0.052
	(0.61)	(0.46)	(1.09)	(0.58)	(0.74)	(1.16)	(0.16)
Hisp Tester: Loan Volume	0.059	0.726	-0.016	0.059	0.021	-0.001	-0.006
	(1.79)	(0.39)	(0.42)	(1.86)	(1.46)	(0.23)	(0.46)
Hisp Tester: Depository Lender	-0.140	7.350	0.109	0.267	0.004	-0.007	0.106
	(0.35)	(0.31)	(0.25)	(0.91)	(0.03)	(0.07)	(0.81)
F-Test: White	9.48 [0.000]**	1.36 [0.261]	0.04 [0.961]	0.36 [0.698]	2.57 [0.080]#	1.24 [0.293]	1.10 [0.336]
F-Test: Black	0.94 [0.469]	1.32 [0.254]	0.99 [0.375]	0.48 [0.822]	1.16 [0.331]	0.91 [0.490]	1.38 [0.227]
F-Test: Hispanic	0.89 [0.504]	0.21 [0.973]	0.94 [0.394]	1.15 [0.337]	2.23 [0.040]*	0.72 [0.634]	0.44 [0.851]
Sample Size	153	131	133	153	153	153	153

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

[2] For more information on paired testing and its role in both measurement and enforcement, see Blank, Dabady, and Citro (2004), Foster, Mitchell, and Fienberg (2002), Ross (2002) and Fix and Turner (1999).

[3] Pair testing attempts to capture disparate treatment as defined by law, which is unequal treatment of equals based on membership in a protected class while economists often focus on market discrimination, which is 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 if minorities overcompensate for discrimination placing them at a disadvantage in negotiations or in high cost market segments, such as the subprime mortgage market (Ross and Yinger, 2002).

[4] Ross and Yinger (2002) and Blackburn and Vermilyea (2003) 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).

[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] Temkin, Levy, and Levine's (1999) case study concluded that the underwriting system was race neutral in their assessment of the process, but in terms of outcomes they found large racial disparities in lending, numerous discrimination complaints, and one active law suit under the ECOA.

[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 racial 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] These empirical tests are based on a theoretical model proposed by Rachlis and Yezer (1993) where 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 application terms and underwriting.

[10] 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, or to record closed form variables that describe specific treatments experienced by testers.

[11] 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

[12] 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) and Charles and Hurst (2002).

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

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

[15] 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.

[16] 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 characteristics and location of all applications and loans.

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

[18] 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.

[19] 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.

[20] 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. 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.

[21] The tester report form and other test instruments are available upon request.

[22] 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.

[23] 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.

[24] 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).

[25] This is important in interpretating 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.

[26] 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.

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

[28] 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.

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

[30] While the large number of testers presents some problems in testing for heterogeneity, Murphy (2002) argues that pair testing efforts should use as large a number of testers as is feasible in order to base inference on a broad sample of testers.

[31] 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.

[32] As is standard practice, the OLS parameter estimates are presented and the F-statistics are based on the OLS estimates and the bootstrapped variance-covariance matrix. The bootstrap required an unusually large number of bootstrap samples because the F-statistics converged more slowly than the standard errors. Presumably, the off-diagonal elements of the variance-covariance matrix converged more slowly than the variance terms.

[33] 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.

[34] 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.

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

[36] In fact, a sign test can be implemented to test whether the significant findings of adverse treatment of minority testers arise more frequently than findings that are consistent with favorable treatment. Results have been presented for six classes of treatment and 23 individual variables representing a total of 24 major and 92 minor hypothesis tests over both groups and sites. At a 5 percent significance threshold, 26 of the 92 null hypotheses are rejected in the direction of finding adverse treatment of minorities, and none are rejected as favorable treatment. The likelihood of such an event arising by chance is less than 0.0001. Focusing only on the six

46

major treatment categories, 9 of the 24 null hypotheses are rejected in favor of adverse treatment, none for favorable treatment, and the likelihood of this result arising by chance is less than 0.005.

[37] If we conduct the sign test described in the previous endnote separately for Chicago and Los Angeles, the likelihood of the Chicago findings arising by chance is less than 0.02 while the likelihood of the Los Angeles findings arising by chance is 0.50 for the set of major hypotheses.

[38] This definition is equivalent to distinguishing between whether testers were "pre-qualified" or not.

[39] 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.

[40] 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.

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