3-1-2016

Discrimination in Mortgage Lending: Evidence from a Correspondence Experiment

Andrew Hanson
Marquette University, andrew.r.hanson@marquette.edu

Zackary Hawley
Texas Christian University

Hal Martin
Georgia State University

Bo Liu
Georgia State University

Accepted version. *Journal of Urban Economics,* Vol. 92 (March 2016): 48-65. DOI.

Discrimination In Mortgage Lending: Evidence From A Correspondence Experiment

Andrew Hanson
Department of Economics, Marquette University, Milwaukee, WI
Zackary Hawley
Department of Economics, Texas Christian University, Fort Worth, TX
Hal Martin
Department of Economics, Georgia State University, Atlanta, GA
Bo Liu
Department of Economics, Georgia State University, Atlanta, GA

Abstract: We design and implement an experimental test for differential response by mortgage loan originators (MLOs) to requests for information about loans. Our e-mail correspondence experiment is designed to analyze differential treatment by client race and credit score. Our results show net discrimination by 1.8% of MLOs through non-response. We also find that MLOs offer more details about loans and are more likely to send follow up correspondence to whites. The effect of being African American on MLO response is equivalent to the effect of having a credit score that is 71 points lower.

Keywords: Discrimination; Field experiment; Mortgage lending; Race; JEL; J15; C93

1. Introduction

There are substantial documented differences between African Americans and whites in the price paid for credit. During the 2004–2008 housing boom, Home Mortgage

Disclosure Act (HMDA) data released by the Federal Financial Institutions Examination Council[1] shows a 27 basis point difference (favoring whites) in contract mortgage rates.[2] Conditioning on borrower characteristics, Ghent et al. (2014) find that for 30 year adjustable rate mortgages African Americans borrowers face interest rates 12 basis points higher than non-Hispanic borrowers. Bayer, Ross, and Ferreira (2014) find that even after conditioning on previously unavailable credit characteristics, African American borrowers have a 7.7 percentage point higher likelihood of being in a high cost mortgage (relative to a market-wide incidence of 14.8%). Bayer et al. (2014) also show that lender fixed effects reduce the unexplained differences across race in being a high-cost borrower by 60–70%, suggesting that a large portion of market-wide differences in outcomes may be driven by sorting across (or differential access to) lenders, rather than differential treatment by lenders.

We examine the incidence of differential treatment by mortgage lenders by testing for racial discrimination using a matched-pair correspondence experiment on Mortgage Loan Originators (MLOs). MLOs are essentially licensed mortgage salespeople who assist customers with loan applications and have the ability to offer and negotiate the terms of a mortgage with applicants. The role of information provider and advisor in the lending process, and the discretion MLOs have in dealing with customers makes them an integral part of the borrowing process from a client's perspective. Discrimination by MLOs could result in different lending outcomes between minority and majority borrowers, and also influence outcomes as the home buying process proceeds. For example, a borrower who is delayed or who is pre-approved for a smaller loan amount may be treated differently by a real estate agent in terms of search effort, neighborhood choice, or expediency of service. If differences in initial treatment by an MLO are severe (offering different interest rates, fees, or suggesting credit repair services), this could conceivably affect a home buyer in all aspects of the home purchase, even if they are successful in obtaining a loan.[3]

Our matched-pair experiment examines the response MLOs offer to initial contact from a potential client interested in obtaining information about a mortgage loan. We design the experiment to test for differential treatment by client race (white or African American) and by credit score. We randomly assign pairs of e-mail inquiries to MLOs according to our design to test for the effects of a borrower's race, credit score, and the interaction between these two. We reveal client race to MLOs using selected client names within each e-mail inquiry. We use only names that have a high likelihood of being given to only one race in a sample of birth certificate for male babies born in New York City in 1990. We examine the propensity for MLOs to respond to our inquiries, the propensity to follow up, and the content of the response to test for differential treatment.

To our knowledge, this is the first experimental test of discrimination by MLOs that uses e-mail correspondence and a nationally representative sample.[4] This is in contrast to an earlier study by Ross et al. (2008), which relies on in-person interaction between MLOs and actors and uses only select metropolitan area samples.[5] Heckman (1998) and Heckman and Siegelman (1993) critique the use of actors when testing for discrimination because actors may bias results if they are not identical along all dimensions except race. While the Heckman critique is valid in theory, tester heterogeneity is something that can be examined in practice. Modern in-person tests such as HUD's 2000 Housing Discrimination Study (HDS) collect actual characteristics of testers and allow for an examination of how these characteristics may affect results (Ross, 2002). Ross et al. (2008) does a formal examination of tester heterogeneity in an in-person study and demonstrates that at conventional levels of statistical significance, tester homogeneity cannot be rejected.

While we believe there is value in using in-person studies, and they offer ways to examine discrimination by MLOs that our study cannot, our work provides some advantages over in-person studies.[6] Most importantly, we avoid the threat of actor bias by relying solely on electronic communication with MLOs that are identical in presentation in all ways except the indication of race. This also allows us to dramatically increase the scope of the experiment and the geographic area covered relative to in-person studies. Using electronic communication provides a detailed record of correspondence that allows us to examine the timing and content of MLO responses to our inquiries. The use of the internet in general is becoming a standard part of the home search and borrowing process which has yet to receive much attention in the academic literature. Bricker et al. (2010) report that 41.7% of borrowers use the internet for information about borrowing,[7] and over 90% of home buyers in 2012 reported using the internet in some capacity during their home search (NAR, 2012).

Our results show that MLOs discriminate on the basis of race and treat clients differently by their reported credit score. We find that on net, 1.8% of MLOs discriminate by not responding to inquiries from African Americans while responding to inquiries from white clients.[8] We find larger net response differences across credit score types, with 8.5% of MLOs responding to clients in our high credit score group while not responding to clients who do not report a credit score. We also find that credit score differences exacerbate differences in response between races. Overall, the effect of being African American on MLO response is roughly equivalent to the effect of having a credit score that is 71 points lower. We also find that MLOs are more likely to send follow-up correspondence to whites than African Americans.

The remainder of the paper is organized as follows: Section 2 describes an MLO's role in the lending process. Section 3 outlines the design of our experiment, while Section 4 details implementation and sample characteristics. Section 5 presents our primary results while Section 6 offers several robustness checks. The final section of the paper offers concluding comments.

2. MLOs in the lending process

MLOs are typically the initial and primary contact person for borrowers seeking a mortgage, and have discretion over how they respond to customer inquiries. MLOs may, for example suggest that a borrower attempt to improve their credit score before completing a loan application, or may encourage a borrower to act quickly to take advantage of low interest rates. They may also present different fees or interest rates to borrowers, offer encouragement or discourage the borrower from moving forward with the loan, or offer other financial advice related to obtaining a mortgage.[9]

MLOs typically have contact with the client throughout the entire lending process, from initial inquiry through loan closing, but they are particularly important in the application process. Clients who have marginal credit depend on MLOs to give advice on what products to apply for, what steps to take to improve their credit and whether their application will ultimately be successful. MLOs may communicate with an underwriter, but do not directly make decisions about accepting or denying a loan. Ross et al. (2008) point out that while minorities are less likely than whites (controlling for observable factors) to obtain a loan,[10] this fact alone does not indicate that differential treatment by the same lender is occurring. Differences in lending outcomes may be explained by differences in minority selection of lenders or access to a different set of lenders. Ross et al. also point out that differences in outcomes could be consistent with differential treatment by MLOs, but evidence on this aspect is limited.

3. Experiment design

To test for discrimination among MLOs we design a matched pair correspondence experiment using e-mail to inquire about assistance with a home mortgage.[11] The matched pairs are structured to test outcome differences due to race and credit score differences among potential borrowers. Each MLO receives two e-mails in the experiment. This design, along with the use of names to identify race, follows the Bertrand and Mullainathan (2004) resume experiment.[12] The audit methodology has a long history in the housing discrimination literature starting with Yinger's (1986) real estate agent experiment.

We use three credit score groups in our experiment: no credit score, low credit score, and high credit score. The low credit score group reports a randomly assigned credit score between 600 and 650; the high credit score group reports a randomly assigned credit score between 700 and 750.[13] As a precaution against exposing the experiment, we randomly assign a credit score for each e-mail (rather than each pair) from a uniform distribution within each category (low or high). Although there is a chance the credit scores within a matched pair are exactly the same, most often the scores will be different within a small range. We use the randomly assigned differences in credit scores to test how MLOs respond to credit score and race differences. For the portion of experiments where a credit score is reported, the average credit score is 675. For the high score group the average is 725 for the low score group it is 625.

We also divide our experiment into groups by the content of correspondence. We chose this design to guard against exposing the experiment. Each group includes one of two types of questions to be asked of the MLO. Using different questions across MLOs may make our inquiries less suspicious for company spam filters or for co-workers who discuss client e-mails. E-mails to one group contain a question about interest rates and a question about mortgage fees (all questions for this set are listed in Appendix 1 in the boxes labeled Question #1a and Question #1b). E-mails to the other group contain a question about loan availability and a question about what information is necessary to proceed in the process of obtaining a loan (all questions for this set are listed in Appendix 1 in the boxes labeled Question #2a and Question #2b). To further guard against exposing the experiment to MLOs, we randomly assign the phrases within the structure of our e-mail inquiries. For example, we randomly assign each e-mail one of five possible greetings (Hello, Hi, Hi There, Hey, or Dear), and ensure that the other e-mail sent to the same MLO does not use the exact greeting. We view the benefits from not matching the text exactly (reducing the risk of exposing the experiment) as exceeding the cost that any of our greetings (or other text elements) might influence outcomes in a meaningful way.[14]Appendix 1 details the exact layout of our correspondence and the randomly assigned text that populates each e-mail.

Our experiment includes 30 different matched pair types, representing all of the combinations between cells in Fig. 1.[15] This allows us to examine the marginal effect of race and credit score (on the extensive and intensive margin), as well as to examine if there is a different marginal effect of credit score across races. We randomly assign each MLO to a matched pair type, and randomly vary the credit score within the range for that type. The matched pair, or within-subjects, design means that each MLO in our experiment receives two e-mail inquiries.

	Question set #1 Black	Question set #1 White	Question set #2 Black	Question set #2 White
No CS	1	2	7	8
Low CS	3	4	9	10
High CS	5	6	11	12

Fig. 1. Experiment design.

We reveal borrower race to MLOs through the name associated with each e-mail inquiry. The source of first names is the New York City Department of Health and Human Hygiene (DHHH) records for babies born in 1990. The DHHH birth records provide counts of babies born by gender, race, and first name. We begin by calculating the probability a baby is born either white or African American for each name in the sample. We use only male baby names for this calculation. The DHHH data do not report a count for names with fewer than 10 babies born in a given race-gender match. This makes our probabilities for names that are very likely to be associated with only one race equal to one, when in fact they could be less than one. Because of this censoring, and the primary concern of signaling race, we also consider the raw number of occurrences each name has within a given race. After compiling a list based on probabilities and counts, we eliminate most names that have a Muslim or Jewish origin from our list as we want to minimize any confounding effects these characteristics would bring to the experiment.

The source of surnames is Word et al.'s analysis of 2000 Census data. This analysis reports counts of surnames for the general population, and by race/ethnicity of respondents to the census. For African American surnames we use the same criteria as first names, choosing those with the highest probability of belonging to African Americans. We choose the surnames with the largest probability of belonging to African Americans regardless of total count, as the data shows a large number of African Americans with these surnames in all cases. We use slightly different criteria for white surnames, as many of the names with the highest probabilities of belonging to whites have a strong ethnic component (for example the highest probabilities are Yoder, Mueller, Koch, all are from a

German origin). For white surnames, we choose three names (Miller, Nelson, Baker) from the most common (by count) names that have greater than a 0.8 probability of being white and less than a 0.15 probability of being African American. We choose the other two names (Krueger and Schmitt) from the list with the highest probabilities of being white, regardless of their ethnic attachment.

Table 1. Names identifying race.

	First name			Last name		
	P(Race|Name)	Count	Rank	P(Race|Name)	Count	Rank
White						
Zachary Miller	1	164	1	0.86	969910	NA
Brendan Nelson	1	55	5	0.8	329788	NA
Jake Krueger	1	43	9	0.97	36694	2
Ethan Schmitt	1	38	10	0.97	35326	6
Maxwell Baker	1	36	15	0.82	343081	NA
Spencer Miller	1	31	17	0.86	969910	NA
Brett Nelson	1	28	20	0.8	329788	NA
Conor Schmitt	1	21	33 (tie)	0.97	35326	6
Luke Krueger	1	22	31	0.97	36694	2
Seth Baker	1	21	33 (tie)	0.82	343081	NA
African American						
Jamal Washington	1	96	1	0.9	163036	1
Jerome Jefferson	1	38	27	0.53	666125	2
DaQuan Booker	1	68	10	0.66	35101	3
Terrell Banks	1	66	12	0.54	99294	4
Darnell Jackson	1	65	13	0.53	666125	5
Tyrone Washington	1	56	14	0.9	163036	1
Kadeem Jefferson	1	84	2	0.75	51361	2
Reginald Jackson	1	51	18	0.75	51361	5
Jermaine Booker	1	49	22	0.66	35101	3
DaShawn Banks	1	39	26	0.54	99294	4

Notes: The source of first names is the New York City Department of Health and Human Hygiene (DHHH) records for babies born in 1990. The DHHH data do not report a count for names with fewer than 10 babies born in a given race-gender match. This makes our probabilities for names that are very likely to be associated with only one race equal to one, when in fact they could be less than one. The first name count is the number of babies born with that name for each race. The first name rank is where each name ranks in the count distribution. The source of surnames is Word et al.'s analysis of 2000 Census data. For African Americans we choose the surnames with the largest probability of belonging to African Americans regardless of total count. The white surnames Krueger and Schmitt were chosen with the same criteria. Because we are concerned that many of the highest probability white surnames have a German origin, we choose three white surnames (Miller, Nelson, and Baker) using alternative criteria. The alternative criteria is to use the most common (by count) names that have greater than a 0.8

probability of being white and less than a 0.15 probability of being African American. White names chosen with the alternative criteria are not ranked in the top ten for all white names, thus we report their value for rank as NA.

Table 1 shows the list of names used to signal race in the experiment. The first three columns of Table 1 show the probability a baby is African American or white given they are born with that name, the count of babies born with that name in 1990, and the rank (by count) for each name. The last three columns of Table 1 show the probability a person is African American or white given the surname, the count of persons with that name in 2000, and the rank (by count) for each name. White names chosen with the alternative criteria are not ranked in the top ten for all white names, thus we report their value for rank as NA. MLOs are exposed to the name associated with each inquiry in three ways: the actual e-mail address,[16] the signature at the bottom of each e-mail (styled as "First name Surname"), and the name plate in the MLOs inbox (styled as "First name Surname").

Table 2 shows the frequency each name occurs in the experiment. Each name represents approximately 5% of the sample of e-mails sent, or about 520 e-mails. The least frequent name in our sample is Kadeem Jefferson, with 459 e-mails or about 4.4% of the total e-mails in our experiment. The most frequent name in our sample is Tyrone Washington, with 577 e-mails or about 5.6% of the total e-mails in our experiment. All differences in name frequency are due to the random assignment of names to matched-pair types, and random assignment of matched-pair types to MLOs in our experiment.

Table 2. Frequency of names in experiment.

	Frequency of occurrence	Percentage of e-mails
White		
Zachary Miller	509	4.91%
Brendan Nelson	535	5.16%
Jake Krueger	526	5.08%
Ethan Schmitt	528	5.10%
Maxwell Baker	515	4.97%
Spencer Miller	489	4.72%
Brett Nelson	502	4.84%
Conor Schmitt	508	4.90%
Luke Krueger	547	5.28%
Seth Baker	523	5.05%
African American		
Jamal Washington	513	4.95%
Jerome Jefferson	571	5.51%
DaQuan Booker	543	5.24%

	Frequency of occurrence	Percentage of e-mails
Terrell Banks	518	5.00%
Darnell Jackson	485	4.68%
Tyrone Washington	577	5.57%
Kadeem Jefferson	459	4.43%
Reginald Jackson	490	4.73%
Jermaine Booker	497	4.80%
DaShawn Banks	527	5.09%
Total	10362	

Notes: Names are randomly assigned to an audit type after the audit type is randomly determined for each MLO. Random assignment is done without replacement for each MLO so that names are not repeated within an audit. Difference in frequency of names in the experiment is due to random assignment.

4. Experiment implementation and sample characteristics

We identify a set of MLOs with their e-mail addresses as subjects for the experiment through internet search. Collection of MLO contact information occurred from February through April of 2012. We used multiple styles of internet search including Google Maps, Google.com, Yellow Pages (YP.com) and Better Business Bureau (bbb.org). For each MLO we identify the following information: name (first and last), state, website, email address, title, physical address and company affiliation. When a photograph is available, we also identify their presumed gender and race.

To ensure a broad sample and limit the potential for the experiment to be exposed, we limit sampling of MLOs operating in the same workplace. We categorize MLO workplace according to their place of employment on two levels: the company and the branch. We consider MLOs to work for the same company if they work for an employer with the same company name (for example, Bank of America or Wells Fargo). We consider MLOs to work for the same branch if they advertise the same physical address on their website. We limit our sample to 8 MLOs per branch, but do not restrict the number of MLOs at the company level. The average number of MLOs per branch in our experiment is 2.38, with an average per company of 27.

We balance the number of MLOs by state level geography, using the proportion of the US population in 2010. For instance, Mississippi has 0.96% of the 2010 US population and we target 0.96% (50) of MLOs in our sample to come from that state. Our goal is not perfect geographic representativeness, but rather a broad geographic sample of the target population of MLOs. Table 3 shows a state-by-state count of the MLOs in our experiment.[17]

The state with the largest number of subjects is California with 423 MLOs, whereas Alaska and West Virginia have the smallest number, two each. The difference between the proportion of MLOs in our sample and 2010 sample population ranges from under sampling by 3.9 percentage points in California to over sampling by 1.71 percentage points in Illinois. Most states are within 0.50 percentage points of the population proportion.

Table 3. Number of audits and response rate across states.

	Number of audits	Overall response rate (%)	Responded to at least one inquiry (%)	Percent of audit population	Percent of US population
Full sample	5181	68.50%	84.93%	–	–
Alabama	70	70.71%	88.57%	1.35%	1.55%
Alaska	2	75.00%	100.00%	0.04%	0.23%
Arizona	125	64.00%	81.60%	2.41%	2.07%
Arkansas	49	57.14%	89.80%	0.95%	0.94%
California	423	69.27%	83.69%	8.16%	12.07%
Colorado	99	70.20%	86.87%	1.91%	1.63%
Connecticut	69	57.97%	73.91%	1.33%	1.16%
Delaware	15	63.33%	80.00%	0.29%	0.29%
Florida	226	73.01%	86.28%	4.36%	6.09%
Georgia	224	71.43%	86.61%	4.32%	3.14%
Hawaii	28	60.71%	71.43%	0.54%	0.44%
Idaho	34	69.12%	88.24%	0.66%	0.51%
Illinois	304	70.89%	83.88%	5.87%	4.16%
Indiana	143	75.87%	92.31%	2.76%	2.10%
Iowa	60	69.17%	85.00%	1.16%	0.99%
Kansas	37	74.32%	89.19%	0.71%	0.92%
Kentucky	93	70.97%	83.87%	1.80%	1.41%
Louisiana	70	72.14%	84.29%	1.35%	1.47%
Maine	44	80.68%	95.45%	0.85%	0.43%
Maryland	158	70.25%	85.44%	3.05%	1.87%
Massachusetts	106	51.89%	68.87%	2.05%	2.12%
Michigan	231	73.81%	91.77%	4.46%	3.20%
Minnesota	95	72.63%	86.32%	1.83%	1.72%
Mississippi	57	71.05%	85.96%	1.10%	0.96%
Missouri	101	65.35%	87.13%	1.95%	1.94%
Montana	16	87.50%	100.00%	0.31%	0.32%
Nebraska	51	82.35%	94.12%	0.98%	0.59%
Nevada	57	73.68%	94.74%	1.10%	0.87%

	Number of audits	Overall response rate (%)	Responded to at least one inquiry (%)	Percent of audit population	Percent of US population
New Hampshire	54	68.52%	79.63%	1.04%	0.43%
New Jersey	111	73.87%	82.88%	2.14%	2.85%
New Mexico	48	65.63%	83.33%	0.93%	0.67%
New York	328	68.75%	83.23%	6.33%	6.28%
North Carolina	64	74.22%	90.63%	1.24%	3.09%
North Dakota	27	61.11%	74.07%	0.52%	0.22%
Ohio	261	72.22%	91.19%	5.04%	3.74%
Oklahoma	64	47.66%	87.50%	1.24%	1.22%
Oregon	100	68.50%	83.00%	1.93%	1.24%
Pennsylvania	177	77.97%	89.27%	3.42%	4.11%
Rhode Island	34	73.53%	82.35%	0.66%	0.34%
South Carolina	103	63.59%	83.50%	1.99%	1.50%
South Dakota	12	45.83%	66.67%	0.23%	0.26%
Tennessee	58	68.97%	87.93%	1.12%	2.06%
Texas	225	52.44%	80.44%	4.34%	8.14%
Utah	85	46.47%	62.35%	1.64%	0.90%
Vermont	15	83.33%	100.00%	0.29%	0.20%
Virginia	144	58.33%	75.69%	2.78%	2.59%
Washington	159	68.87%	86.16%	3.07%	2.18%
West Virginia	2	75.00%	100.00%	0.04%	0.60%
Wisconsin	110	76.82%	90.91%	2.12%	1.84%
Wyoming	13	50.00%	76.92%	0.25%	0.18%

Notes: Sampling is intended to follow US state population proportions. The exact representativeness of state populations in our data depends on availability of MLO contact information. Collection of MLO contact information occurred from February through April of 2012. We used multiple styles of internet search including Google Maps, Google.com, Yellow Pages (YP.com) and Better Business Bureau (bbb.org) to locate MLO contact information.

We have 5,181 MLOs as subjects in our experiment. A notable feature of MLOs is their relative demographic homogeneity. We are able to identify the race of MLO using photographs for about 75% of our sample. 93% of the race-identified MLOs are white. Gender is more evenly split, where 53% of gender-identified MLOs are male. We know of no existing demographic statistics to corroborate our sample as representative of the industry. Nonetheless, we did not specifically seek MLOs based on race or gender, and believe these statistics to be representative of the industry participants who list information on the internet. Table 4 shows complete demographic characteristics for the sample of MLOs in the experiment.

Our experiment began on Monday, April 30, 2012 at 2:00 pm with a set of 200 pilot emails to a sub-sample of MLOs. The full experiment commenced 1 week later, on Tuesday, May 8 at 1:00 pm. An MLO was randomly selected to participate in one of five rounds (one pilot and four regular).[18] A round consists of a first and second e-mail going out to the same group of MLOs separated by 1 week. For instance, Round 3 recipients received their first email on Thursday, May 10; they received their second email on Thursday, May 17. All emails following the pilot were sent beginning at 1 pm Eastern Daylight Time (EDT) and ending by 2 pm EDT. Regular rounds were conducted on Tuesdays, Wednesdays and Thursdays between May 8 and May 29. The schedule is designed to minimize possible confounding effects. Mondays and Fridays are avoided in the experiment to minimize the impact of weekend lag times or end-of-week effects. The start time of each half round is held constant to minimize unobserved time-of-day effects.

Keeping days of the week and time of e-mail sending similar within rounds is designed to reduce noise in the experiment. Bias is eliminated through randomization of treatment assignment and treatment order. Whichever audit type an MLO is assigned to, the order in which the MLO receives the treatment is randomized. Should our efforts to eliminate confounding effects be inadequate, randomizing the order of treatment causes such effects to attenuate the outcome to zero rather than bias it. This is a key strength of our experimental design.

Given the volume of mail to be delivered (two emails to each MLO from one of twenty different client email accounts), we used an automated email-sending program to minimize human error and speed processing. The program was set to initiate individual Simple Mail Transfer Protocol (SMTP) sessions for each email sent with the Gmail servers, and to wait 1–2 s between sending emails. The resulting emails are indistinguishable from messages that would be sent directly from the Gmail web interface.

Of the 10,362 e-mail inquiries sent in our experiment, about 1.3% received an automated response or out-of-office reply. We do not count these as a response unless a follow up e-mail was sent by the MLO.

5. Results

The experiment provides a wealth of information on MLOs response to standard inquiries for assistance with obtaining a home mortgage. We maintain all content for each MLO response to use in the analysis, in addition to the date/time stamp indicating when the e-mail was sent to examine speed of response.

5.1. Response versus non-response by race

The most basic indication of equal treatment across race and credit score types is whether or not an MLO responds to our inquiry for assistance with a mortgage loan. We consider a response to be one that we deem genuinely written by a human; out-of-office or other automatic replies are not considered a response. Our measure of response is if the MLO ever sent a genuine e-mail response within 2 weeks of our inquiry. This measure means that we count a genuinely human response to have occurred even if it is not the first correspondence we received from the MLO.[19] In several instances, we received out of office replies, and subsequent genuinely human responses. In other cases we received automatically generated commercial replies followed by genuinely human responses.

Table 4. Mortgage loan originator characteristics.

	Number of audits	Frequency	Overall response rate
Gender			
Female	1916	36.98%	87.37%
Male	2202	42.50%	84.92%
Not identified	1063	20.52%	80.53%
Race			
White	3619	69.85%	86.57%
Non-White	273	5.27%	85.71%
Arabic	1	0.02%	100.00%
Asian	57	1.10%	80.70%
Black	90	1.74%	91.11%
Hispanic	115	2.22%	84.35%
Indian	7	0.14%	71.43%
Native American	3	0.06%	100.00%
Not identified	1289	24.88%	80.14%

Notes: We identified race and gender of MLOs by visually inspecting photographs when available on lender webpages. Our approach was conservative in identifying both gender and race: if we felt there was any room for argument about either, we categorized the demographic information as not identified. The not identified category includes all instances where there was ambiguity in assigning race and/or gender and when a photograph was not available. For the summary statistics shown, the Non-White race category includes Arabic, Asian, Black, Hispanic, Indian, and Native American, the breakdown of each is shown within the non-white category.

Table 5. Response rate and mortgage loan originator (MLO) level response.

	Overall response rate			Response at MLO level				
	(1) White	(2) African American	(3) (1)-(2)	(4) Respond to neither	(5) Respond to both	(6) White only	(7) African American only	(8) (6)-(7)
All audits	68.31%	65.68%	2.63%	16.28%	49.77%	17.88%	16.07%	1.81%
	[3540]	[3402]	$p = 0.0022^{***}$	[632]	[1932]	[694]	[624]	$p = 0.0573^{*}$
Depository Lender	73.35%	70.51%	2.84%	11.51%	55.49%	17.69%	15.31%	2.37%
	[1663]	[1614]	$p = 0.0163^{**}$	[194]	[935]	[298]	[258]	$p = 0.0980^{*}$
Mortgage Bank	64.69%	61.23%	3.46%	20.12%	46.28%	17.97%	15.63%	2.34%
	[1411]	[1305]	$p = 0.0094^{***}$	[327]	[752]	[292]	[254]	$p = 0.1039$
White MLOs	70.67%	67.41%	3.26%	14.49%	51.96%	17.85%	15.71%	2.14%
	[2554]	[2443]	$p = 0.0014^{***}$	[392]	[1406]	[483]	[425]	$p = 0.0585^{*}$
Non-White MLOs	68.50%	65.57%	2.93%	13.04%	49.28%	20.77%	16.91%	3.86%
	[187]	[179]	$p = 0.2337$	[27]	[102]	[43]	[35]	$p = 0.4282$
Missing Race MLOs	61.70%	60.80%	0.90%	21.98%	43.76%	17.34%	16.92%	0.42%
	[799]	[780]	$p = 0.3189$	[213]	[424]	[168]	[164]	$p = 0.8693$
Male MLOs	69.16%	66.00%	3.16%	16.66%	50.75%	17.74%	14.85%	2.89%
	[1496]	[1479]	$p = 0.0124^{**}$	[277]	[844]	[295]	[247]	$p = 0.0434^{**}$
Female MLOs	70.80%	67.23%	3.57%	12.98%	51.48%	18.83%	16.71%	2.12%
	[1382]	[1264]	$p = 0.0085^{***}$	[184]	[730]	[267]	[237]	$p = 0.1964$
Missing Gender MLOs	62.04%	62.23%	−0.19%	21.35%	44.69%	16.48%	17.48%	−1.00%
	[662]	[659]	$p = 0.5351$	[171]	[358]	[132]	[140]	$p = 0.6713$

Notes: The p-value represented in column (3) is from a one-sided t-test (alternative hypothesis of a positive difference) with a null hypothesis that the difference in average response rate is zero. The p-value reported in column (8) is from McNemar paired difference in proportions test. This test is designed for testing the difference in proportion of respondents for paired subjects, the test statistic is $\chi^2 = (N\_(\text{Only } W)-N\_(\text{Only AA}))^2/(N\_(\text{Only } W)+N\_(\text{Only AA}))$, where N represents the number of MLOs only responding to one group. The test statistic has a chi-squared distribution, and we calculate all p-values accordingly. Number of MLOs shown in [], *=0.10 significance, **= 0.05 significance, ***= 0.01 significance.

Table 5 shows the difference in genuine response by MLOs to our inquiries across race groups. The first and second columns of Table 5 show the overall response rate, while columns 4–7 offer a breakdown at the MLO level for the sub-sample of the experiment that includes MLOs who received separate e-mails from both an African American and a white borrower.[20] The response rate to white borrowers for the full sample is 68.31%. For

African American borrowers, the response rate is 2.63 percentage points lower, or 65.68%. As shown in the third column of Table 5, this difference is statistically significant at the 1% level, indicating that MLOs are less likely to respond to inquiries from African American borrowers than they are from whites.

We find a slightly higher response rate difference (3.46 percentage points in favor of white borrowers) for the subsample of lending institutions classified as mortgage banks.[21] This result is precisely estimated at conventional levels using response rate differences that include across lender comparisons, but statistical precision is strained when using the within-MLO sample. For institutions defined as depository lenders, which are more like traditional banks, we find about the same level of discrimination as the full sample (2.84 percentage points in favor of white borrowers). The result for depository lenders is statistically significant for the response rate tests at the 5% level, and at the 10% level for the within-MLO tests.

We find a slightly higher response rate difference (3.26 percentage points in favor of white borrowers) for the subsample of MLOs whose race is white. The magnitude of our results is similar for the sub-sample of non-white MLOs, but the small sample size of this group strains statistical precision. We find a smaller difference in response rate for the sample of MLOs where we are not able to identify race (a 0.9 percentage point difference, favoring whites), although this result is also imprecise. The level of discrimination measured by response rate differences between male and female MLOs is similar: 3.16 percentage points for male MLOs, 3.57 percentage points for females, both favoring white borrowers and both precisely estimated. Among the sample of MLOs where we do not identify gender, we find only a small, imprecisely estimated difference in response rate (0.19 percentage points, favoring African Americans).

Although we do find a statistically significant difference in response rates between African American and white borrowers, the MLO level results shown in columns 4–7 of Table 5 show that 66.05% of the MLOs in our sample treat e-mail inquiries the same—either by responding to both (49.77) or responding to neither (16.28) inquiry. We measure discrimination at the MLO level by the net amount of discrimination, or the difference in the proportion of MLOs who respond only to whites, and those that only respond to African Americans. Using this measure, we find a smaller level of discrimination: MLOs responding only to whites outnumber MLOs responding only to African Americans by 1.81 percentage points, a difference which is close to statistically precise at conventional levels (a p-value of 0.0573) despite the smaller sample size and more rigorous significance test.[22]

The MLO-level results across race and gender of MLOs are similar to the response rate results, with smaller magnitudes and some loss of statistical precision. We still find that white MLOs discriminate, but only by 2.14 percentage points on net, with statistical significance at the 10% level. Our results for male MLOs also remain largely the same, showing a slightly higher level of discrimination than the full sample (2.89 percentage points on net), and maintaining statistical precision; however, we lose statistical precision on the results for female MLOs.

Table 6. Response rate and mortgage loan originator (MLO) level response.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Response rate differences						
	High credit	Low credit	No credit	(1)–(2)	(1)–(3)	(2)–(3)
Response rate (includes all audits)	69.46%	65.76%	65.77%	3.70%	3.69%	–0.01%
	[2397]	[2276]	[2269]	$p = 0.0005^{***}$	$p = 0.0005^{***}$	$p = 0.5024$
Panel B: MLO level differences						
	Respond to neither	Respond to both	Respond to Higher only	Respond to Low/No only	(3)-(4)	
High versus low credit	15.21%	49.88%	19.47%	15.44%	4.03%	
	[132]	[433]	[169]	[134]	$p = 0.0506^{*}$	
High versus no credit	15.17%	53.00%	20.14%	11.66%	8.48%	
	[129]	[450]	[171]	[99]	$p = 0.0000^{***}$	
Low versus no credit	17.80%	50.17%	17.22%	14.81%	2.41%	
	[155]	[437]	[150]	[129]	$p = 0.2311$	

Notes: The p-value represented in columns (4), (5), and (6) is from a one-sided t-test (alternative hypothesis of a positive difference) with a null hypothesis that the difference in average response rate is zero. The p-value reported in column (5) of Panel B is from McNemar paired difference in proportions test. This test is designed for testing the difference in proportion of respondents for paired subjects. The test statistic is $\chi^2 = (N_{(\text{Only } W)} - N_{(\text{Only AA})})^2 / (N_{(\text{Only } W)} + N_{(\text{Only AA})})$, where N represents the number of MLOs only responding to one group. The test statistic has a chi-squared distribution, and we calculate all p-values accordingly. Number of MLOs shown in [], *,0.10 significance; **, 0.05 significance; ***, 0.01 significance.

5.2. Response versus non-response by credit score

Table 6 shows the difference in genuine response by MLOs to our inquiries across credit score groups. Panel A shows the overall response rate difference across the high, low, and no credit score groups. Panel B shows the difference in response at the MLO level for the sub-sample of MLOs who received separate e-mails from borrowers with different credit scores (or one including a credit score and the other excluding). The high credit score group received the highest response rate, 69.46%. The response rate for the low credit score group is 3.7 percentage points lower, or 65.76%. The difference in response rate between the high and low credit score group is statistically significant at the 1% level, indicating that MLOs are more likely to respond to inquiries from borrowers with a higher credit score.

We find a similar gap in response rates between the high and no credit score groups of 3.69 percentage points, with the response rate of the no credit score group at 65.77%. The difference in response rate between the high and no credit score group is also statistically significant at the 1% level. We find only a small, statistically imprecise gap between the response rate for the low and no credit score groups that slightly favors the no credit score group.

The MLO level results show that the majority of MLOs respond (between 49.88 and 53%) or do not respond (between 15.17 and 17.8%) to both credit score groups. As with the race results, we measure equal treatment by examining the net proportion of MLOs that respond differently across credit score groups. The MLO level analysis reveals about the same level of differential treatment between the high and low credit score groups: 4.03 percentage points on net, which is statistically significant at the 10% level (p value of 0.0506). The biggest difference is the differential treatment between the high and no credit score group, where the high credit score group is favored by 8.48% of MLOs on net. This result is statistically meaningful at the 1% level.

Comparing the race and credit score differences shows that MLOs are relatively more sensitive to differences in credit scores when deciding whether or not to respond to a borrower inquiry for assistance with a mortgage loan. The relative difference between race and credit score groups depends on which response measure is used. At the MLO level, going from a low to high credit score roughly doubles the difference in net unequal treatment in response/non-response compared to the difference between African American and white borrowers. The mean difference between credit score groups (100 points), assuming a linear relationship between credit score and response, suggests that the effect of having an African American name on MLO response is roughly equivalent to having a credit score that is 71 points lower.

5.3. Response versus non-response by race and credit score

The design of our experiment allows us to look at differential treatment across several possible race-credit score combinations. Table 7 shows results for audits where whites and African Americans are assigned to the same credit score category, where whites are in a higher credit score category, and where African Americans are in a higher credit score category. The magnitude of the response rate differences and MLO level differences will be exactly the same for these results, as they use the same set of audits; however, the statistical tests at the MLO level count only differential response by individual MLOs, so statistical significance may vary.

Table 7. Response rate and mortgage loan originator (MLO) level response.

	Overall response rate			Response at MLO level				
	(1) Group 1	(2) Group 2	(3) (1)-(2)	(4) Respond to neither	(5) Respond to both	(6) White only	(7) African American only	(8) (6)-(7)
Equal credit within audit								
Credit categories combined	67.34%	65.72%	1.62%	16.31%	49.36%	17.97%	16.35%	1.62%
	[1746]	[1704]	$p = 0.1083$	[423]	[1280]	[466]	[424]	$p = 0.1693$
Both high credit	69.79%	69.13%	0.66%	14.76%	50.17%	18.34%	16.72%	1.62%
	[1199]	[1198]	$p = 0.3366$	[128]	[435]	[159]	[145]	$p = 0.4560$
Both low credit	68.12%	63.37%	4.75%	18.12%	48.43%	17.77%	15.68%	2.09%
	[1186]	[1090]	$p = 0.0016^{**}$	[156]	[417]	[153]	[135]	$p = 0.3165$
Both no credit	67.03%	64.50%	2.53%	16.07%	49.48%	17.80%	16.65%	1.15%
	[1155]	[1114]	$p = 0.0588^{*}$	[139]	[428]	[154]	[144]	$p = 0.6022$
White higher credit within au dit	71.26%	61.05%	10.21%	15.91%	48.22%	23.04%	12.83%	10.21%
	[300]	[257]	$p = 0.0009^{***}$	[67]	[203]	[97]	[54]	$p = 0.0006^{***}$
African American Higher credit within audit	65.26%	72.37%	−7.11%	15.00%	52.63%	12.63%	19.74%	−7.11%
	[248]	[275]	$p = 0.0173^{**}$	[57]	[200]	[48]	[75]	$p = 0.0187^{**}$

Notes: Equal credit within audit implies credit category is the same (high, low, no) for a given audit. Higher credit score comparisons exclude low versus no credit audits. Including no versus low credit score audits makes differences between race even larger than the differences shown here. The *p*-value represented in column (3) is from a one-sided *t*-test (alternative hypothesis of a positive (or negative for African American Higher Credit) difference) with a null hypothesis that the difference in average response rate is zero. The *p*-value reported in column (8) is from McNemar paired difference in proportions test. This test is designed for testing the difference in proportion of respondents for paired subjects. The test statistic is χ^2 = (N_(Only W)-N_(Only AA))^2/ (N_(Only W)+N_(Only AA)), where N represents the number of MLOs only responding to one group. The test statistic has a chi-squared distribution, and we calculate all *p*-values accordingly. Number of MLOs shown in [], *,0.10 significance; **, 0.05 significance; ***, 0.01 significance.

Table 8. Intensity of mortgage loan originator (MLO) response.

	(1) White	(2) African American	(3) (1)-(2)
Panel A: Time elapsed until respo nse			
Time until response (h:mm)	8:20	9:23	1:03
	(29:50)	(34:04)	$p = 0.0879^{*}$
Panel B: Length ofr esponse			
Character count, all audits	426.00	431.84	2.52
	(356.35)	(388.34)	$p = 0.3890$
Character count, dropping audits with equal (within 10 characters) length replies	448.74	431.84	16.90
	(347.55)	(361.01)	$p = 0.0986^{*}$
Panel C: Follow-upr esponse			
Follow-up e-mail received	6.91%	5.17%	1.74%
	[358]	[268]	$p = 0.0001^{***}$
Number of follow-up e-mails received	2.19	2.25	−0.06
	(0.51)	(0.71)	$p = 0.9014$

Notes: Row 1 shows the average time elapsed between when an inquiry is sent and when a MLO reply is received, reported in h:mm format, these averages do not include e-mails where no reply was made. Rows 2 and 3 examine the character count for MLO responses. Row 2 examines all audits and includes counting non-response as zero characters, Row 3 excludes replies that were of equal length and does not count non-responses. Row 4 and 5 examine additional genuine e-mail responses after the first genuine e-mail response. Row 4 shows the percentage of MLOs who sent a follow-up e-mail, Row 5 shows the average number of follow-up e-mails received. In all cases, *p*-values are from standard difference in means *z* tests. Standard deviations are reported in (), number of MLOs shown in [],*=0.10 significance; **, 0.05 significance; ***, 0.01 significance.

The magnitude of discrimination using only audits in equal credit score categories is slightly smaller than the level for the entire sample; we find net discrimination by 1.62% of MLOs, as opposed to 1.9 for the full sample. The statistical significance of this relationship is strained, as the MLO level result has a *p*-value of 0.1693, outside of traditional significance levels, while the response rate test is close to statistical precision at the 10% level. Given that the magnitude of these results is similar to the full sample, the loss of statistical precision does not seem alarming, especially considering that these results rely on a smaller sample.[23]

The results across race where clients report different credit scores show that the higher credit score group is favored regardless of whether they are white or African American, but that the degree to which they are favored is larger for whites with higher credit scores. The middle row of Table 7 shows that whites are favored by 10.21% (statistically significant at the 1% level) of MLOs in audits where the white client reports a higher credit score and the African American reports either a low or no credit score. The bottom row of Table 7 shows that while African Americans with higher credit scores are favored over whites with low or no credit score, the difference is only 7.11% of MLOs on net (statistically significant at the 5% level). The MLO level results show that more MLOs choose to reply to both clients when the African American has a higher credit score, as opposed to replying to only the white client when African Americans have a lower credit score. These results suggest that while there is a level of discrimination that exists regardless of credit score, discrimination increases with credit score differences between races. A non-constant level of discrimination may mean that the source of discrimination is not taste-based, but instead a form of statistical discrimination based on perceived group differences.

5.4. Content of response

E-mail communication with MLOs allows us to examine not only the propensity to respond, but also the nature of responses in our experiment. Table 8 shows how MLOs responded in terms of the timing, length, and propensity to send a follow-up response in the experiment.

Table 9. Side-by-side comparison of mortgage loan originator (MLO) response content.

	(1) Neutral	(2) Prefer White	(3) Prefer African American	(4) (2)–(3)
Panel A: Author blind revie w				
MLOs responding to both races	57.25%	22.67%	20.03%	2.64%
(1932 matched pairs)	[1106]	[438]	[387]	$p = 0.0817^{*}$
Reason for Preference				
More favorable terms		5.94%	4.13%	$p = 0.1200$
Friendliness		33.56%	32.82%	$p = 0.4103$
Included more details		46.58%	40.83%	$p = 0.0484^{**}$
Explained the process		5.48%	7.24%	$p = 0.8498$
Un-preferred e-mail was negative		5.25%	8.01%	$p = 0.9451$
Facilitated the transaction		27.85%	30.75%	$p = 0.8193$
Un-preferred e-mail steered or was pushy		1.37%	0.78%	$p = 0.2059$
Other		9.82%	10.85%	$p = 0.6873$
Panel B: Outside reviewer blind re view				
MLOs responding to both races	43.94%	29.19%	26.40%	2.79%
(1932 Matched pairs)	[849]	[564]	[510]	$p = 0.1058$
Reason forp reference				
More favorable terms		5.67%	5.10%	$p = 0.3384$
Friendliness		47.16%	45.29%	$p = 0.2698$
Included more details		54.26%	51.57%	$p = 0.1892$
Explained the process		20.04%	18.82%	$p = 0.3082$

	(1) Neutral	(2) Prefer White	(3) Prefer African American	(4) (2)–(3)
Un-preferred e-mail was negative		7.98%	7.45%	$p = 0.6268$
Facilitated the transaction		26.24%	26.47%	$p = 0.5340$
Un-preferred e-mail steered or was pushy		2.13%	1.37%	$p = 0.1743$
Other		3.90%	5.29%	$p = 0.8628$

Notes: Side-by-Side comparison uses the visual basic interface shown in Appendix 2 for all MLOs responding to both e-mails in matched pairs with clients of different race. Column (2) totals includes all instances where whites were preferred or strongly preferred, Column (3) includes all instances where African Americans were preferred or strongly preferred. All indications of preference in panel A are judged by the authors in a blind review where information about clients and MLOs is masked. All indications of preference in panel B are indicated by outside reviewers in a blind review where information about clients and MLOs is masked. The p-value represented in column (4) is from a one-sided t-test (alternative hypothesis of a positive difference) with a null hypothesis that the difference in average response rate is zero. Percentages for reasons indicated for preference do not sum to one because graders were allowed to indicate multiple reasons for preference. See Appendix 3 for instructions given to graders and definitions of reasons for preference. Number of MLOs shown in [], *,0.10 significance; **, 0.05 significance; ***, 0.01 significance.

Individual results show that MLOs are substantially slower to respond to African American clients than they are to whites. Among MLOs that responded to both inquiries, whites received a response in 8 h and 20 mins on average. MLOs took an average of an hour and 3 mins longer to respond to African Americans, a result that is statistically precise at the 10% level. In terms of the length of response, a measure of response intensity, we find no difference in the number of characters in a response when examining all e-mails. Examining a sub-set of responses that differ by at least 10 characters (to eliminate short generic, or form responses), we do find that whites are favored and receive a response that is about 4% longer.

The propensity to attempt a follow up with white potential clients was much greater than for African American clients. We find that 6.9% of MLOs sent at least one follow-up reply to whites, while only 5.2% sent a follow-up reply to African Americans. Statistically, this is one of our strongest results, as it is statistically meaningful at the 1% level. We find, however, that the average number of follow-up replies sent to the two groups is not different.

5.5. Content of response: side by side comparisons

We designed a side-by-side analysis tool that allows us to make a direct comparison of the response from a single MLO to clients with a different race. Appendix 2 shows a screen shot of the side-by-side analysis tool used to grade the difference between two e-mails sent by the same MLO to different clients. To ensure an unbiased grading, all identifying information from both MLOs and clients was masked when using the side-by-side analysis tool.

We conducted both an internal (author examined) and external (a team that did not include any authors)[24] reviews of e-mail pairs using this tool. We designed the analysis tool to randomly assign a left-side and right-side e-mail, to guard against any ordering effects in grading responses. Graders were instructed to indicate if they thought the responses were "Neutral", meaning they were similar in content, language, and nature; they felt one was "Preferred" in some way over the other, or they thought that one was "Strongly Preferred" over the other. Graders were instructed to strictly record their opinion about the response, and that they did not have to justify their feelings. In addition to an opinion about how favorable the responses were relative to one another, we also offered a series of check boxes for the reason. Graders were instructed that they were not required to use a reason, and there was also a place to write in other reasons. Reasons for favoring included: offering more favorable terms, friendliness, and facilitation of the transaction. Appendix 3 provides the instructions given to graders and the full list of reasons for preference.

We graded all responses from MLOs replying to both inquires for mortgage assistance. For white/African American matched pairs, this was 1932 pairs of responses or 3864 e-mails. Panel A of Table 9 shows the results of the side-by-side analysis for author graded responses. We find that even among MLOs that respond to both inquiries for mortgage assistance, some discrimination exists. Blind grading shows that about 57% of e-mails were perceived as being neutral between the different inquiries. Blind grading also shows that among e-mails where some preference was indicated, we were more likely to perceive e-mails sent to whites (22.6%) as being favorable to those sent to African Americans (20%). The magnitude of this difference suggests that another 2.6% of MLOs discriminate by responding more favorably to whites—this is in addition to the 1.8% of MLOs that discriminate through non-response. The side-by-side comparison results are statistically precise at the 10% level with a *p*-value of 0.0817.

Table 10. Adjusted *p*-values for primary results.

	(1)	(2)	(3)
Outcome	**Result**	**Unadjusted *p*-value**	**Adjusted *p*-value**
Panel A: Table 5			
Response difference: all audits	2.63%	0.0022	0.0308
Panel B: Table 8			
Time until response (h:mm)	1:03	0.0879	0.9804
Character count, all audits	2.52	0.3890	1

Outcome	(1) Result	(2) Unadjusted *p*-value	(3) Adjusted *p*-value
Character count, dropping audits with equal (within 10 characters) length replies	16.9	0.0986	0.986
Follow-up e-mail received	1.74%	0.0001	0.0015
Number of follow-up e-mails received	−0.06	0.9014	1
Panel C: Table 9			
MLOs responding to both races	57.25%	0.0817	0.9804
Reason forp reference(Author blind revie w)			
More favorable terms		0.1200	1.0000
Friendliness		0.4103	1.0000
Included more details		0.0484	0.6292
Explained the process		0.8498	1.0000
Un-preferred e-mail was negative		0.9451	1.0000
Facilitated the transaction		0.8193	1.0000
Un-preferred e-mail steered or was pushy		0.2059	1.0000
Other		0.6873	1.0000

Significance levels of primary results are adjusted according to Holm (1979)'s procedure to adjust for family-wise error, as implemented in Newson (2010). Columns (1) and (2) repeat the results for outcomes and *p*-values in earlier tables for reference. The adjusted *p*-value in column (3) describes a conservative (upper bound) probability that rejecting the null hypothesis will result in one or more Type I errors across the group of 15 results.

The most common reason an e-mail was graded as preferred for white clients was that they were given more details (46.6% of white preferred e-mails); this reason was also the most common when preference was given to African Americans, but occurred only 40.8% of the time. The second most common reason an e-mail was graded as preferred for white clients was that the tone of the e-mail was more friendly (33.56%), which was also the second most common reason African Americans were preferred (32.8%). It was also fairly common to choose that preference was given to whites because the e-mail facilitated the loan transaction (27.8%), although among MLOs that gave preference to African Americans this occurred more often (30.7%). It was less common for graders to indicate that preference was given to whites because of strong overt measures of discrimination like offering more favorable terms (5.9%) [25] or steering into a product or being pushy (1.4%).

The side-by-side comparison done by outside reviewers is remarkably consistent with our internal grading. Although outside reviewers graded e-mails as neutral less often than the authors (43.9%, as opposed to 57.2), the net level of discrimination for favoring whites is only 0.15 percentage points higher. Panel B of Table 8 shows tests for discrimination in the content of response using the outside reviewer's opinion of MLO replies. The outside reviewers perceived that e-mails to whites were more favorable 29.2% of the time, while perceiving favorable e-mails for African Americans 26.4% of the time. The net incidence of unequal treatment is 2.8% of MLOs, but this result is not quite statistically significant with a *p*-value of 0.106. Outside reviewers were more likely to use friendliness and the inclusion of details as reasons why whites were favored and less likely to use the "other" category.

5.6. Type I error and multiple hypotheses

With any testing that involves multiple hypotheses, type I error, or rejecting a null hypothesis that is true, may be a concern. In the context of discrimination experiments, Ross et al. (2008) discuss the risk of type I error extensively and correct for it across multiple hypotheses. Despite the strength of random assignment in experimental design in protecting against bias, readers must remain cautious when interpreting the significance of individual tests. We guard against type I error in two ways: first, we report *p*-values for the range of single-hypothesis tests we set out to conduct, whether or not the results are significant. This allows the reader to evaluate the significance of individual results against the pattern of tests conducted. Second, we present a conservative adjustment to the *p*-values for our primary results that control for type I error across multiple tests.

There are several strategies available to adjust hypothesis tests in order to yield information about the probability of type I error for a group of tests. Anderson (2008) provides an overview of two types of adjustments: one controls for the familywise error rate, identifying the likelihood of at least one type I error occurring across a family of tests. Another approach is to control the false discovery rate, or the rate at which one expects type I errors to occur. We control for the familywise error rate using a conservative adjustment procedure proposed in Holm (1979), which combines a Bonferroni correction with a step-down testing approach. The method produces adjusted *p*-values for each test that reflect the probability of at least one type I error across the family of tests should the null be rejected for that particular test, as implemented in Newson (2010).

Table 10 presents a summary of the 15 outcomes in our primary results. Column (1) shows the point estimate of the difference between whites and African Americans for the

15 tests and column (2) repeats the *p*-value under single hypothesis testing. Column (3) shows the adjusted *p*-value for each test in the context of the group. This value is the upper bound of the probability that rejecting the null hypothesis will result in at least one type I error among the 15 outcomes. Two of our primary results are impervious to this correction at the 5% threshold: the response difference across all audits and whether a follow-up email was received. The marginally significant results in Table 9, however, fail to reject the null hypothesis at any acceptable level of confidence for detecting a single type I error across the family of 15 tests.[26]

5.7. Why do MLOs discriminate?

Although our experiment is not specifically designed to test for the reasons why MLOs discriminate against African American clients, the data generated from our experiment allows us to explore this question to some degree. The standard theories behind why discrimination occurs fall into two basic categories: taste-based (Becker, 1957), and statistical (Phelps, 1972). In statistical discrimination, an agent treats a minority client differently because they are making some assumption about their characteristics (typically this is described as assigning group averages), and find interacting with them to be less profitable in expectation. Taste based discrimination reflects an agent's preferences toward majority and minority groups. List (2004) differentiates between statistical and taste based discrimination by carefully designing a series of experiments in the sports card marketplace that involves a dictator game with market participants, a design which is not possible in our setting.[27]

As an alternative to a formal test, we consider how the interaction between credit score and race should generally affect response between whites and African Americans. If MLOs are practicing taste-based discrimination, we should find a constant level of discrimination across all credit score categories (compared to an equivalent white credit score). If MLOs practice statistical discrimination, are agnostic about the quality of credit reporting between groups, and generally assume that African Americans have lower credit scores, then reporting higher credit scores should confirm what they believe about whites (not changing response) and raise the likelihood they respond to African Americans. This would result in mitigating discrimination among the high credit score group, relative to the no credit score reported group. Reporting lower credit scores should confirm what MLOs believe about African Americans (not changing response), but lower the likelihood they respond to whites, resulting in mitigating discrimination relative to the no credit score group.

If MLOs hold the beliefs described above, then the results in Table 7 broken down by credit category generally support that MLOs are not practicing taste-based discrimination. The difference in response by race between credit score groups is not constant, ranging from as much as 4.75 percentage points in the lower credit score group to as small as 0.66 percentage points in the high credit score group. The difference between African Americans and whites is much larger when both report no credit score than when both report high credit scores, suggesting that statistical discrimination is what is driving the results. In fact, when both report high credit scores, the response rate and MLO level tests cannot reject equal treatment. The strongest result within credit score groups is that discrimination exists when both groups report low credit scores, which does not match our intuition for statistical discrimination. The larger level (4.75 percentage points) of discrimination relative to the no credit score group is driven by an increase in white response rates, and a decrease in African American response rates. While this does not fit neatly with our explanation of standard statistical discrimination, one possibility for this finding is that MLOs differentially weigh credit score information from whites and African Americans. It could be that when whites report a low score it is viewed as something that can be improved, but when African Americans report a low credit score it is viewed as revealing a problem borrower.

Bertrand et al. (2005) suggests an alternative to the standard taste-based and statistical explanations for discrimination. They suggest that agents may not make a conscious choice when discriminating, but instead discriminate unintentionally or implicitly because of an unconscious association between a person of a certain type and some identified attribute. This type of discrimination is tested in the laboratory by showing subjects photographs of people in a rapid manner and requiring that they assign them to some category (e.g. good or bad). Requiring quick reaction attempts to identify the subconscious thoughts of the subject. Our experiment does not lend itself to a formal test of implicit discrimination, but the nature of the timing in MLO responses is suggestive. Table 8 shows a large gap between the average time it takes an MLO to respond to our subjects—they respond faster to whites. This is suggestive that on the margin, more MLOs are making the quick decision to respond to whites and to set aside the e-mail from the African American until later, which suggests the motivation for discrimination may be implicit rather than conscious discrimination.

6. Robustness of findings

We explore the sensitivity of our findings to three particular choices in the original experiment: choosing a sample of MLOs based on state population weights, the particular names used in correspondence with MLOs, and the primary statistical tests for inference.

To create a national sample of MLOs, we decided to use the population of each state relative to the country as a whole to guide our selection across geography. This choice means that our sample necessarily includes more MLOs from highly populous states like California, Texas, and Florida, regardless of the African American population living in those states. While we are confident that our results represent the level of disparate treatment in our sample, we recognize that choosing a sample with different criteria may affect the overall level of reported discrimination.

To explore the possibility that the level of discrimination we find would be different if we would have initially chosen a different sample of MLOs, we re-weight our results by various population characteristics that may affect the level of discrimination. The characteristics we consider as alternative weights are: homeownership, African American homeownership, household income, and African American household income. These weights represent characteristics that influence participation in the mortgage market, which may vary for African Americans in certain areas. For each weighting scheme, we apply weights equal to the percentage of national totals in each state and calculated a weighted response/non-response based on these weights.

Table 11 shows our primary results after re-weighting our sample by the characteristics alongside our population weighted standard results. We generally find that re-weighting increases the level of discrimination relative to our primary findings. Weighting by household income produces the largest differences, with net discrimination rates as high as 3.38 percentage points. In all cases our results remain qualitatively similar, but the different weighting schemes do highlight that sample choice may influence the level of reported discrimination.

Table 11. Robustness of primary results to alternative sample weights.

	Overall response rate			Response at MLO level				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	White	**African American**	(1)–(2)	**Respond to neither**	**Respond to both**	**White only**	**African American only**	(6)–(7)
All audits	68.31%	65.68%	2.63%	16.28%	49.77%	17.88%	16.07%	1.81%
	[3540]	[3402]	$p = 0.0022^{***}$	[632]	[1932]	[694]	[624]	$p = 0.0573^{*}$

	Overall response rate			Response at MLO level				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	White	African American	(1)–(2)	Respond to neither	Respond to both	White only	African American only	(6)–(7)
Homeownership weights	67.99%	65.03%	2.96%	16.05%	48.75%	18.02%	16.45%	1.57%
			$p = 0.0000^{***}$					$p = 0.0964^{*}$
African American homeownership weights	68.00%	65.48%	2.52%	16.17%	48.53%	18.51%	16.63%	1.88%
			$p = 0.0000^{***}$					$p = 0.0483^{**}$
Median household income weights	68.00%	64.63%	3.37%	15.70%	50.25%	17.38%	15.15%	2.23%
			$p = 0.0000^{***}$					$p = 0.0147^{**}$
African American household income weights	68.12%	64.74%	3.38%	15.53%	50.60%	17.35%	14.99%	2.36%
			$p = 0.0000^{***}$					$p = 0.0096^{***}$

Notes: The p-value represented in column (3) is from a one-sided t-test (alternative hypothesis of a positive difference) with a null hypothesis that the difference in average response rate is zero. The p-value reported in column (8) is from McNemar paired difference in proportions test. This test is designed for testing the difference in proportion of respondents for paired subjects, the test statistic is χ^2 = (N_(Only W)-N_(Only AA))^2/ (N_(Only W)+N_(Only AA)), where N represents the number of MLOs only responding to one group. The test statistic has a chi-squared distribution, and we calculate all p-values accordingly. Number of MLOs shown in [], *,0.10 significance; **, 0.05 significance; ***, 0.01 significance.

Table 12. Name robustness and national popularity.

Name	Response rate	Different than own race	National popularity
Jake Krueger	72.62%	$p = 0.0422$	140
Brett Nelson	70.92%	$p = 0.2304$	84
Brendan Nelson	70.09%	$p = 0.3988$	155
Ethan Schmitt	69.70%	$p = 0.5146$	58
Luke Krueger	69.65%	$p = 0.5215$	118
Tyrone Washington	68.80%	$p = 0.1324$	260
Jermaine Booker	68.01%	$p = 0.2948$	306
Zachary Miller	67.98%	$p = 0.8761$	22
Reginald Jackson	67.96%	$p = 0.3082$	230
Darnell Jackson	67.42%	$p = 0.4379$	353
Conor Schmitt	67.13%	$p = 0.5834$	458
Spencer Miller	66.26%	$p = 0.3510$	105
Jamal Washington	66.08%	$p = 0.8533$	247
Seth Baker	65.77%	$p = 0.2352$	102

Name	Response rate	Different than own race	National popularity
Kadeem Jefferson	64.71%	$p = 0.6751$	538
Maxwell Baker	64.66%	$p = 0.0901$	188
DaQuan Booker	63.54%	$p = 0.3184$	709
Jerome Jefferson	63.05%	$p = 0.2101$	264
Terrell Banks	62.93%	$p = 0.2110$	282
DaShawn Banks	62.62%	$p = 0.1599$	732

Notes: P-value is for a difference in means *t*-test between the response rate for each name and names of the same race. National popularity ranking comes from the Social Security Administration website using counts of baby names from 1990 at: http://www.ssa.gov/cgi-bin/popularnames.cgi

Identifying discrimination in correspondence experiments relies on the choice of names being representative of race groups. The birth certificate data we use demonstrates that the names in our experiment are highly correlated with the race we assign, but they leave open the possibility that particular names are treated differently for other reasons. While we do not have the ability to infer why this might be, we can examine how our choice of names may affect our results.

We consider two sets of robustness checks with the names in the experiment. First, we examine response rate differences between each name and other names of the same race, and exclude names that are treated statistically different. Next we examine national data on name popularity and exclude the most and least popular names in each race.

Table 12 shows the response rate by name in the experiment. Since our findings show that African Americans experience lower response rates, most of the top response rates come from white names. We preform *t*-tests of the response rate differences between each name and the average within-race response rate; these results are shown in the second column of Table 12. There are only two statistically distinguishable (among own race) names in our experiment: Jake Krueger has a higher response rate than other whites ($p = 0.0422$) and Maxwell Baker has a lower response rate than other withes ($p = 0.0901$).

Excluding white names with a higher than average response rate (Jake Krueger) narrows the gap between white and African American response rates, as shown in column 1 of Table 13. We are somewhat confident in these results as both tests show a similar (albeit smaller) gap between whites and African Americans, and the response rate tests maintain statistical significance. The within MLO results do not maintain statistical significance, but this is likely due to sample size restriction that comes with excluding 10% of the already smaller sample. Results excluding the other name that showed statistically different outcomes, Maxwell Baker, are in line with our primary results, and in fact show a slightly larger response rate difference. The third row shows that if we exclude white name with the highest and African American name with the lowest response rate then the magnitude of our results is smaller but still statistically meaningful for the response rate difference, but it loses marginal statistical significance for the MLO level test.

Table 13. Results excluding popular or unique names.

	Overall response rate			Response at MLO level				
	(1) White	(2) African American	(3) (1)-(2)	(4) Respond to neither	(5) Respond to both	(6) White only	(7) African American only	(8) (6)–(7)
Exclude Jake Krueger (5% level different than own race)	67.83%	65.68%	2.15%	16.41%	49.71%	17.53%	16.35%	1.18%
	[3158]	[3402]	$p = 0.0119^{**}$	[571]	[1730]	[610]	[569]	$p = 0.2440$
Exclude Jake Krueger and Maxwell Baker (10% different than own race)	68.46%	65.68%	2.78%	15.87%	50.31%	17.75%	16.07%	1.68%
	[2835]	[3402]	$p = 0.0023^{***}$	[490]	[1553]	[548]	[496]	$p = 0.1144$
Exclude DaShawn Banks and Jake Krueger (highest white, lowest AA)	67.83%	66.02%	1.81%	16.27%	49.98%	17.61%	16.14%	1.47%
	[3158]	[3072]	$p = 0.0321^{**}$	[508]	[1561]	[550]	[504]	$p = 0.1657$
Exclude Zachary Miller and DaShawn Banks (most popular white, least popular AA)	68.35%	66.02%	2.33%	16.02%	50.94%	17.43%	15.61%	1.82%
	[3194]	[3072]	$p = 0.0083^{***}$	[503]	[1599]	[547]	[490]	$p = 0.0820^{*}$
Exclude Zachary Miller and Ethan Schmitt (2 most popular)	68.18%	65.68%	2.50%	16.29%	50.47%	17.27%	15.97%	1.30%
	[2826]	[3402]	$p = 0.0054^{***}$	[503]	[1558]	[533]	[493]	$p = 0.2234$
Exclude DaShawn Banks and DaQuan Booker (2 least popular)	68.31%	66.35%	1.96%	15.62%	50.36%	17.85%	16.17%	1.68%
	[3540]	[2727]	$p = 0.0224^{**}$	[482]	[1554]	[551]	[499]	$p = 0.1155$

Notes: The p-value represented in column (3) is from a one-sided t-test (alternative hypothesis of a positive difference) with a null hypothesis that the difference in average response rate is zero. The p-value reported in column (8) is from McNemar paired difference in proportions test. This test is designed for testing the difference in proportion of respondents for paired subjects, the test statistic is χ^2 = (N_(Only W)-N_(Only AA))^2/ (N_(Only W)+N_(Only AA)), where N represents the number of MLOs only responding to one group. The test statistic has a chi-squared distribution, and we calculate all p-values accordingly. Number of MLOs shown in [], *,0.10 significance; **, 0.05 significance; ***, 0.01 significance.

Table 14. Robustness of primary results to alternative inference method.

	(1)	(2)	(3)	(4)	(5)
	White only	**African American only**	**McNemar**	**Liddell**	**Sign test**
All audits	17.88%	16.07%	1.81%	1.11	
	[694]	[624]	$p = 0.0573^{*}$	$p = 0.0286^{**}$	$p = 0.0287^{**}$
Equal credit	17.97%	16.35%	1.62%	1.09	
	[466]	[424]	$p = 0.1693$	$p = 0.0847^{*}$	$p = 0.0846^{*}$
White higher credit	23.04%	12.83%	10.21%	1.71	
	[97]	[54]	$p = 0.0006^{***}$	$p = 0.0003^{***}$	$p = 0.0002^{***}$
African American higher credit	12.63%	19.74%	–7.11%	1.53	
	[48]	[75]	$p = 0.0187^{**}$	$p = 0.0009^{***}$	$p = 0.0097^{***}$

Notes: The p-value represented in column (3) is from the McNemar test described in all other tables. The p-value reported in column (4) is from the Liddell Test. This test is designed for testing the ratio of subjects only responding to one group over those only responding to the other; the statistic is F= (N_(Only W))/ (1+N_(Only AA)), where N represents the number of MLOs only responding to one group. The test statistic has a chi-squared distribution, and we calculate all p-values accordingly. The sign test compares the number of times a MLO favors a white (African American) relative to the expectation that this would occur in half of observations; the test statistic is from a binomial distribution with the number of observations equal to the number of times when one group was favored (1318 in our data). Number of MLOs shown in [], *,0.10 significance; **, 0.05 significance; ***, 0.01 significance.

We also examine differences in general popularity of the names in our experiment using Social Security Administrative data. This data reports counts of name for babies born nationally each year (unconditional on race) for the 1000 most common names. All names in our experiment are among the 1000 most common in popularity for the year of our birth certificate data (1990). As a result of population shares, white names are necessarily more popular among the general population, but one of the white names (Conor) is quite unpopular nationally relative to other white names. The third column of Table 12 shows popularity ranks for all names in our data.

Excluding the most popular white name (Zackary Miller) and the least popular African American name (DaShawn Banks) does not affect either the response rate or MLO level results, as shown in Table 13. In each case, the magnitude of the difference is extremely similar and we maintain the same level of significance as our primary results.

We also explore excluding the two most popular white names, and (separately) excluding the two least popular African American names. In either case the response rate results are similar in magnitude to our primary results and maintain statistical significance. The MLO level results also show a similar magnitude to the primary results, but lose statistical significance.

Our primary means of inference at the MLO level is the McNemar test, which is designed to account for an experiment that is run in pairs, as it considers that some proportion of the subjects in the experiment will respond only to whites or only to African Americans. The McNemar test uses all of this information, where simple difference-in-proportions tests do not. There are several alternatives to the McNemar test for paired data, the most common being the sign test and the exact Liddell test (Liddell, 1983). The Liddell test is preferred in medical applications as it offers a confidence interval on the ratio of subjects that only respond to the treated and only to the untreated (in our application this is the proportion responding to whites only, divided by the proportion responding to African Americans only). Ross et al. (2008) use a sign test for inference in an in-person mortgage discrimination experiment with matched pairs. The sign test, like the McNemar test relies on the difference in occurrence of unequal treatment from the same MLO across the sample.

We re-examine inference for our key results using both the sign test and the exact Liddell test. The results of these tests, reported in Table 14, all show increased levels of statistical precision over the McNemar test. We are therefore confident that the means of inference are not driving our results.

7. Conclusion

We find evidence of discrimination against African Americans in the market for mortgage loans. The discrimination we find occurs at the initial information gathering stage for borrowers in response to a simple e-mail inquiry about assistance with obtaining a mortgage. We find that MLOs, the primary contact person for a borrower looking to obtain a mortgage, are less likely to respond to inquiries from clients with African American names than they are to clients with white names. We also find that MLOs responding to inquiries from both races are more likely to write a preferential e-mail to white clients. The level of discrimination we find is large for a characteristic that should not matter (race) relative to one that should matter (credit score).

Finding discrimination in the information gathering stage is likely to influence outcomes for minority borrowers throughout the lending and home buying process. If African American borrowers are less likely to receive communication from an MLO and the MLO treats them differently when communication does occur, it makes submitting a loan application more difficult, and the remainder of the home purchase more arduous. In addition, our work shows that the growing importance of e-mail communication between clients and lenders, where in-person meetings are less and less common, does not mean that discrimination on the basis of race will not occur.

The magnitude of discrimination we find is smaller than the most recent in-person study (Ross et al., 2008); however, the standard for compliance is much lower in our most basic test: we only examine if MLOs are willing to respond to an e-mail. Several of our correspondence based results are similar in nature to the in-person results in Ross et al. (focusing on their differential treatment tests for the Chicago sample). Both the Ross et al. work and our work show that that African Americans are less likely than whites to receive follow up communication, and less likely than whites to be given details about loan products. Ross et al. find additional differences showing that whites were more likely to be given the information they requested and more likely to be given advice or "coaching" that may help in qualifying for a loan.

Our findings confirm that discrimination still exists in the lending industry, and that it exists across a larger sample, and geographic scope than previous studies have examined. We are also able to compare the difference in treatment between whites and African Americans with the difference in treatment across credit score groups. Our average differences suggest an African American name reduces the probability that an MLO responds by the same magnitude as does reporting a credit score that is 71 points lower.

Our results suggest that examining lending outcomes is not sufficient to uncover the level of discrimination that minorities face in the lending process. Our work also suggests that to uncover the full extent of discrimination in this market, multiple types of communication should be used in addition to in-person audits, and that enforcement of Fair Lending Laws would be more robust if audits included other means of communication.

Appendix 1. Correspondence construction

[GREETING] [FIRST NAME][27],
I'm interested in [PRODUCT].
[SOURCE]
[PLEASANTRY]
[*If score known, then* CREDIT SCORE] [RANDOMLY ASSIGNED SCORE]
[QUESTION #1a or QUESTION #2a, depending on question set type]
[QUESTION #1b or QUESTION #2b, depending on question set type]
[VALEDICTION]
[RANDOMLY ASSIGNED NAME from Table 1]

<u>GREETING</u>
Hello
Hi
Hi There
Hey
Dear

<u>PRODUCT</u>
a home loan.
a mortgage.
getting a home loan.
getting a loan.
information on mortgages.

<u>SOURCE</u>
I found your information on-line, and thought you could help.
I got your contact information on-line, and hope you can help me.
I looked you up on-line, hopefully you can help.
I found you on-line, and think you can help.
I got your information on the web, and thought you might be able to help.

<u>PLEASANTRY</u>
I just have a few questions.
I have a few questions for you.
I'm curious about a few things.
I'd like to ask you a couple of questions.
I'm wondering about a few things.

<u>CREDIT SCORE</u>
I know that my credit score is
My credit score is
I have a credit score of
I already know my credit score is

<u>VALEDICTION</u>
Thank you for your time,
Thanks in advance,
I look forward to hearing from you,
I look forward to your reply,
Thanks for taking the time,

<u>QUESTION #1a</u>
How are interest rates looking?
What interest rate can I expect?
Can you tell me about current interest rates?
What do interest rates look like right now?
How should I expect interest rates to look?

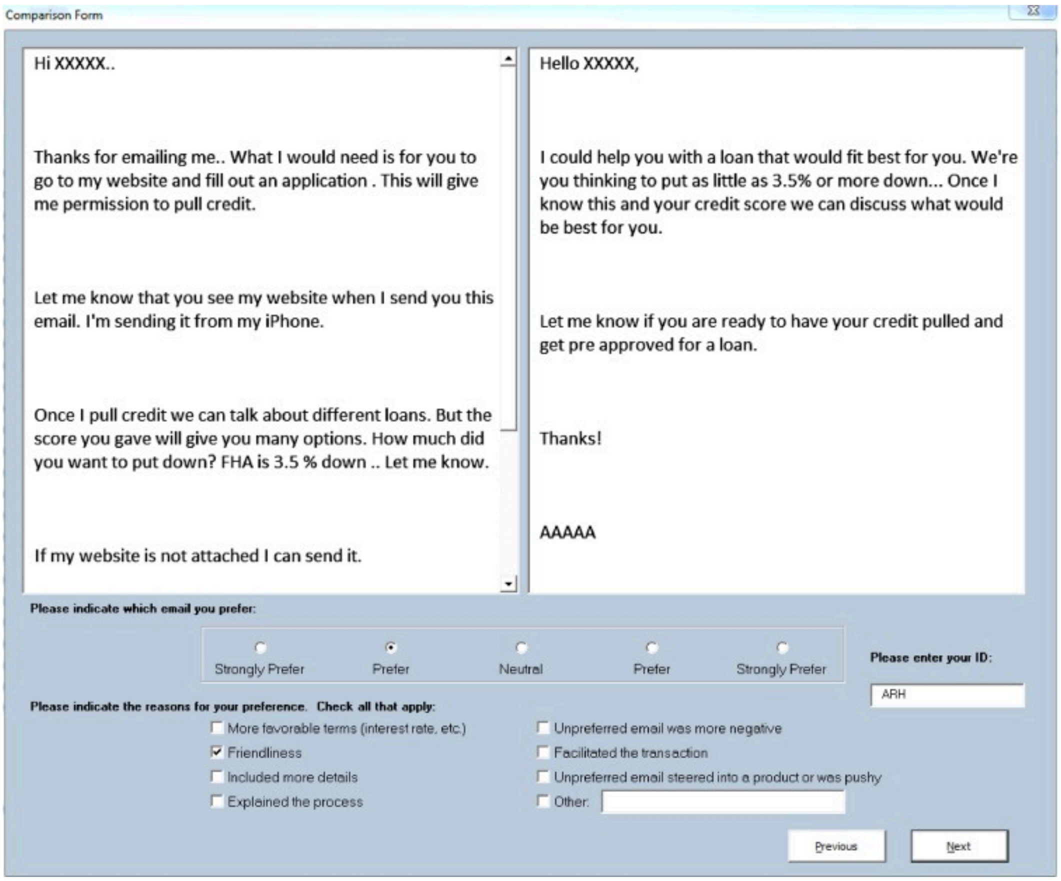
<u>QUESTION #2a</u>
What types of loans might be available for me?
Can you tell me about the types of loans you have?
What sort of loans are available for someone like me?
Can you offer advice on what type of loans are available?
What kind of loans do you have available?

<u>QUESTION #1b</u>
What sort of fees are involved?
What fees should I expect?
How do the fees work?
Are there fees that I need to worry about?
What are the typical fees?

<u>QUESTION #2b</u>
What other information do you need from me?
Do you need any other information to start the process?
What sort of information do you need to move forward?
What more do you need from me to proceed?
Do you know what else I need to begin the process?

[27] We use actual MLO first names given on the webpage where we found contact information.

Appendix 2. Side-by-side comparison tool



Comparison Form

Hi XXXXX..

Thanks for emailing me.. What I would need is for you to go to my website and fill out an application . This will give me permission to pull credit.

Let me know that you see my website when I send you this email. I'm sending it from my iPhone.

Once I pull credit we can talk about different loans. But the score you gave will give you many options. How much did you want to put down? FHA is 3.5 % down .. Let me know.

If my website is not attached I can send it.

Hello XXXXX,

I could help you with a loan that would fit best for you. We're you thinking to put as little as 3.5% or more down... Once I know this and your credit score we can discuss what would be best for you.

Let me know if you are ready to have your credit pulled and get pre approved for a loan.

Thanks!

AAAAA

Please indicate which email you prefer:

Strongly Prefer | Prefer | Neutral | Prefer | Strongly Prefer

Please enter your ID: ARH

Please indicate the reasons for your preference. Check all that apply:

- [ ] More favorable terms (interest rate, etc.)
- [x] Friendliness
- [ ] Included more details
- [ ] Explained the process
- [ ] Unpreferred email was more negative
- [ ] Facilitated the transaction
- [ ] Unpreferred email steered into a product or was pushy
- [ ] Other:

Previous | Next

Appendix 3. Grader instructions for side-by-side comparison

Thank you for agreeing to help out with this research project. Your task is simple-to review a set of e-mail responses we received from mortgage lenders and compare them. We are interested to know if you feel the responses you read tend to be more favorable toward one set of recipients than another, or if they are treated approximately the same. Essentially, we want to know your opinion. Please use the format we've supplied by enabling external content in excel and simply clicking on the "Open Form" button. After entering the form, please start by typing your name in the ID box. Next, read both e-mails **carefully**. After reading both e-mails, please indicate if you thought the mortgage lender strongly preferred one, preferred the other, or treated the recipients neutrally using the check box indicator.

If you thought that the mortgage lender expressed a preference, please use the next section of check boxes to indicate why you felt this way. Feel free to use the "other" check box in the event that your opinion does not match a reason listed, or if you can't quite describe why you feel that way. You can also fill in a reason for "other" to describe your reasoning. Please use the following as guidance when checking boxes for your reasons:

More favorable terms (interest rate, etc.): Check this box if the lender replied with more favorable loan terms to one recipient than the other. This could be in a quoted or suggested interest rate, length of loan, type of loan, fees, or anything else that has to do with costs to the borrower.

Friendliness: Check this box if you feel the lender was more 'friendly' to one recipient than the other. Again, this is your opinion, we will not hold it against you.

Included more details: Check this box if you feel the lender gave a more detailed description of the products, application materials needed, or generally gave answers with more depth to one recipient than the other.

Explained the process: Check this box if you feel the lender offered more guidance on the lending process, the application process, or the home purchase process to one recipient than the other. This might include offers on how to improve credit, or necessary paper work to complete an application.

Un-preferred email was more negative: Check this box if you thought that one of the e-mails was negative, even if the other e-mail was neutral. This might include negative language, unusually short replies (relative to the other), or a rude tone in writing.

Facilitated the transaction: Check this box if you feel the lender attempted to facilitate a successful transaction more with one recipient than the other. This might include offers for future communication, providing application materials, encouraging an application, or offering help with credit issues or home search.

Un-preferred email steered into a product or was pushy: Check this box if you feel the lender was being pushy about selling a loan, or suggested a specific product that was "right" for the recipient and not the other. Differentiating between this box and facilitating a transaction will largely depend on your interpretation of the language of the e-mail. Remember, this is your opinion.

References

Anderson, 2008 Anderson M. **Multiple inference and gender differences in the effects of early intervention: a reevaluation of the Abecedarian, Perry preschool, and early training projects** *Journal of the American Statistical Association*, 103 (484) (2008), pp. 1481-1495

Bayer et al., 2014 Bayer, P., Ferreira F. and Ross S. (2014) "Race, ethnicity, and high cost mortgage lending" *National Bureau of Economic Research Working Paper* 20762.

Becker, 1957 Becker Gary *The Economics of Discrimination* University of Chicago Press, Chicago (1957)

Bertrand and Mullainathan, 2004 Bertrand Marianne, Mullainathan Sendhil **Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination** *American Economic Review*, 94 (4) (2004), pp. 991-1013

Bifulco et al., 2011 Bifulco Robert, Jason Fletcher, Stephen Ross **The effect of classmate characteristics on post-secondary outcomes: evidence from the add health** *American Economic Journal: Economic Policy*, 3 (1) (2011), pp. 25-53

Bricker et al., 2010 Bricker Jesse, Arthur Kennickell, Kevin Moore, John Sabelhaus **Changes in U.S. Family Finances from 2007 to 2010: Evidence from the Survey of Consumer Finances** *Federal Reserve Bulletin*, 98 (2) (2010), pp. 1-80

Doleac and Stein, 2013 Doleac Jennifer, Luke Stein **The visible hand: Race and online market outcomes** *Economic Journal*, 123 (2013), pp. 469-492

Ewens et al., 2014 Ewens Michael, Bryan Tomlin, Choon Wang **Statistical discrimination or prejudice? A large sample field experiment** *Review of Economics and Statistics*, 96 (1) (2014), pp. 119-134

Ghent et al., 2014 Ghent Andra, Ruben Hernández-Murillo, Michael Owang **Differences in subprime loan pricing across races and neighborhoods** *Regional Science and Urban Economics*, 48 (2014), pp. 199-215

Gruenstein Bocian et al., 2008 Gruenstein Bocian Debbie, Keith Ernst, Wei Li**Race, ethnicity and subprime home loan pricing** *Journal of Urban Economics*, 60 (1) (2008), pp. 110-124

Hanson and Hawley, 2011 Hanson Andrew, Zackary Hawley **Do landlords discriminate in the rental housing market? Evidence from an internet field experiment in US cities** *Journal of Urban Economics*, 70 (2-3) (2011), pp. 99-114

Heckman, 1998 Heckman James **Detecting discrimination** *Journal of Economic Perspectives*, 12 (2) (1998), pp. 101-116

Heckman and Siegelman, 1993 Heckman James, Peter Siegelman **The urban institute audit studies: their methods and findings** Michael E. Fix, Raymond J. Struyk (Eds.), In *Clear and Convincing Evidence: Measurement of Discrimination in America,* Urban Institute Press, Washington (1993), pp. 187-258

Holm, 1979 Holm Sture **A simple sequentially rejective multiple test procedure** *Scandanavian Journal of Statistics*, 6 (2) (1979), pp. 65-70

Ladd, 1998 Ladd Helen **Evidence on discrimination in mortgage lending** *Journal of Economic Perspectives*, 12 (2) (1998), pp. 41-62

Liddell, 1983 Liddell Francis Douglass Kelly **Simplified exact analysis of case-referent studies: matched pairs; dichotomous exposure** *Journal of Epidemiology and Community Health*, 37 (1983), pp. 82-84

List, 2004 List John **The nature and extent of discrimination in the marketplace: evidence from the field** *Quarterly Journal of Economics*, 119 (1) (2004), pp. 49-89

Munnell et al., 1996 Munnell Alicia, Geoffrey Tootell, Lynn Browne, James McEneaney **Mortgage lending in Boston: interpreting HMDA data** *American Economic Review*, 86 (1) (1996), pp. 25-53

Neumark, 2012 Neumark David **Detecting discrimination in audit and correspondence studies** *Journal of Human Resources*, 47 (4) (2012), pp. 1128-1157

Newson, 2010 Newson Roger **Frequentist *q*-values for multiple-test procedures** *Stata Journal*, 10 (4) (2010), pp. 568-584

Phelps, 1972 Phelps Edmund **The statistical theory of racism and sexism** *American Economic Review*, 62 (4) (1972), pp. 659-661

Pope and Sydnor, 2011 Pope Devin, Justin Sydnor **What's in a picture? Evidence of discrimination from Prosper.com** *Journal of Human Resources*, 46 (1) (2011), pp. 53-92

Ravina, 2012 Ravina, Enrichetta (2012) "The effect of beauty and personal characteristics in credit markets" *Social Science Research Network Working Paper*. Available at SSRN: http://ssrn.com/abstract=1101647.

Ross, 2002 Ross Stephen **Paired testing and the 2000 housing discrimination survey** Williams Foster, Angela Faith Mitchell, Stephen E. Fienberg (Eds.), *Measuring Housing Discrimination in a National Study: Report of a Workshop,* National Academies Press, Washington, D.C (2002)

Ross et al., 2008 Ross Stephen L., MargeryAustin Turner, Erin Godfrey, RobinR. Smith **Mortgage lending in Chicago and Los Angeles: a paired testing study of the pre-application process** *Journal of Urban Economics*, 63 (3) (2008), pp. 902-919

Ross and Yinger, 2002 Ross Stephen L., John Yinger *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair-Lending Enforcement* MIT Press, Cambridge (2002)

Smith and Delair, 1999 Smith Robin, Michelle Delair **New evidence from lender testing: discrimination at the pre-application stage** Turner Margery, Felicity Skidmore (Eds.), *Mortgage Lending Discrimination: A Review of Existing Evidence,* Urban Institute Press, Washington, D.C (1999)

Turner and Skidmore, 1999 Turner Margery, Felicity Skidmore **Introduction, summary, and recommendations** Turner Margery, Felicity Skidmore (Eds.), *Mortgage Lending Discrimination: A Review of Existing Evidence,* Urban Institute Press, Washington, D.C (1999)

Word et al., 2000 Word, David L., Charles D. Coleman, Robert Nunziata, and Robert Kominski. (not dated) "Demographic aspects of surnames from census 2000. *Technical Report for the U.S. Census Bureau.*

Yinger, 1986 Yinger John **Measuring racial discrimination with fair housing audits: caught in the act** *American Economic Review*, 76 (5) (1986), pp. 881-893

We thank seminar participants at Georgia State University, Marquette University, Texas Christian University, University of Texas at Arlington, the Lincoln Institute, the Urban Economics Association Annual Meetings, the American Real Estate and Urban Economics Annual Meetings, the Royal Economic Society Annual Conference, the Race, Ethnicity and Place Bi-annual Conference, the Western Economic Association Annual Conference, and the Federal Reserve Bank of Cleveland Policy Summit for offering feedback on this work. We also thank Mike Eriksen, Elliot Anenberg, and Kaitlyn Harger for detailed feedback. All errors are our own. The authors obtained Institutional Review Board approval for this project.

[1]The FFIEC maintains summary statistics of HMDA data on its website at: http://www.ffiec.gov/hmdaadwebreport/NatAggWelcome.aspx.

[2]This difference is not conditional on borrower characteristics and is the difference in the mean interest rate reported for loans where the interest rate is known on conventional, 1-4 family home purchase loans (excluding manufactured homes) between 2004 and 2008. See Gruenstein-Bocian et al. (2008) for a study that examines interest rate differences across race groups conditional on borrower characteristics.

[3]See Ross and Yinger (2002) for a particularly lucid explanation of discrimination in the lending process, including an explanation and critique of research methodology.

[4]There are several recent studies that use e-mail correspondence to test for discrimination in the market for rental housing. See Hanson and Hawley (2011) for a recent example and a review of this literature. Also see Ladd (1998), Yinger and Ross (2002), and Ross et al. (2008) for a review of the literature on discrimination in mortgage markets in particular.

[5]See Smith and Delair (1999) for a summary of early evidence on discrimination by mortgage lenders from a sample of enforcement-based in-person audits which covers five US cities.

[6]See Doleac and Stein (2013) for a novel approach to avoiding the use of actors by studying discrimination using pictures in an on-line market. This work varies the skin color of the seller to test for discrimination among buyers of iPods.

[7]39.5% of borrowers report using sellers of financial services as a method of obtaining information about borrowing. The most commonly used source of information about borrowing is "friends, relatives, and associates" with 43.9% of borrowers using that channel (Bricker et al., 2010).

[8]The net level of discrimination measures the difference in the percentage of MLOs that only reply to an inquiry from a white client against the percentage of MLOs that only reply to an inquiry from an African American client. The gross level of discrimination or the percentage of MLOs that only reply to an inquiry from a white client is 17.8% of MLOs. The overall difference in response rates is 2.6 percentage points favoring whites—this difference does not match the net discrimination level because some of our experiments involved sending inquiries from same race clients to the same MLO.

[9]The Secure and Fair Enforcement for Mortgage Licensing Act (SAFE), part of the larger Housing and Economic Recovery Act of 2008, included several provisions to tighten regulations of MLOs. These provisions included requiring licensing of MLOs, creating a Nationwide Mortgage Licensing System (NMLS), issuing uniform licensing applications and reporting requirements across states, and creating a national clearing house for collecting consumer complaints.

[10]See Munnell et al. (1996) for a study that identifies denial rate differences between African American and white clients controlling for credit differences.

[11]The experiment attempts to uncover discrimination from differential treatment based on minority status. Scholars (and governments) have also recognized that disparate impact, or having a policy that disproportionately impacts minorities while lacking business purpose is discrimination. See Turner and Skidmore (1999) for a discussion about the difference between differential treatment and disparate impact in mortgage lending.

[12]In general, the use of names to identify race may be concerning as names may also reveal something else about the characteristics of a client, such as social class (which may be important if it is correlated with ability to repay a loan). We believe the inclusion of credit scores in our experiment helps minimize the concern that important unobservables besides race are inferred through the client name. Doleac and Stein (2013) use pictures to identify race in an experimental setting to avoid this criticism. Research by Pope and Sydnor (2011) and Ravina (2012) use pictures to identify race in observational studies of credit markets.

[13]The low credit score range approximates the 15–30th percentile of the national distribution of credit scores according to the Fair Isaac Company (FICO). The high credit score range approximates

the 40–60th percentile of the national FICO score distribution (FRB, 2007). Most MLOs seem to operate using a rule on an acceptable credit score (like a minimum of 620); we noticed that the reported rule varied across the responses we received. Our low score sample seems to straddle the rule in all areas.

[14]There is almost no difference in the response rate across types of e-mails. The response rate for the question set #1 group is 67.2. The response rate for the question set #2 group is 66.8. This difference is not statistically meaningful.

[15]We also randomly vary the order in which e-mails are sent. For each matched pair type the order of the treatment difference (e.g. high versus low credit score) is randomly reversed in exactly half the emails. We do this to ensure that order effects do not drive any results.

[16]We use Gmail addresses exclusively, and all take the form firstname.surnameXXX@gmail.com, where XXX is a random three-digit number.

[17]The actual sample in our experiment is not exactly proportional with 2010 state populations. This is partly due to randomly selecting MLOs from this sample to be subjects and partly due to the fact that availability of MLO information is not uniform across geography.

[18]Random selection for a round depends on the number of MLOs at a branch. Spreading a branch's MLOs through multiple rounds is done to ensure no more than two emails from any of our clients arrive at an office on any given day. The structure of this selection, however, is independent of treatment assignment.

[19]Of the responses we consider genuinely human, 99.8% of them were received within 2 weeks of our original inquiry. We count the other 0.02% as non-responses.

[20]Recall that the full sample includes some audits where inquiries had the same race, but different credit characteristics.

[21]We identified different types of institutions using the institution name in our data.

[22]For all statistical significance tests at the MLO level, we use the McNemar test. This test is designed for testing the difference in proportion of respondents for paired subjects. The test statistic is $\chi 2=(\mathrm{NOnlyW}-\mathrm{NOnlyAA})2/(\mathrm{NOnlyW}+\mathrm{NOnlyAA})$ where N represents the number of MLOs only responding to one group. The McNemar test statistic has a chi-squared distribution, and we calculate all p-values accordingly.

[23]We also test for differences when both scores are high, both low, and both do not include credit. These results show a similar level of discrimination as the results that combine these groups, but statistical precision is even more strained due to smaller sample sizes in these sub groups.

[24]We used a team of 10 different reviewers.

[25]Graders were asked to indicate that more favorable terms were offered if the lender replied with more favorable loan terms to one recipient than the other. This could be in a quoted or suggested interest rate, length of loan, type of loan, fees, or anything else that has to do with costs to the borrower.

[26]We also examined the results of e-mail content for joint significance, using a similar method to Bifulco et al. (2011). These tests are intended to uncover an overall pattern of discrimination that would not have occurred by chance, even if individual tests are not significant. The results of the joint tests do not produce any evidence of discrimination that was not apparent in our standard or corrected individual tests.

[27]Also see Ewens et al. (2014) and Neumark (2012) for work that explicitly tests for the cause of racial discrimination.

