CONSUMER FINANCIAL PROTECTION BUREAU | NOVEMBER 2024

Matched-Pair Testing in Small Business Lending Markets

Consumer Financial Protection Bureau

Consumer Financial Protection Bureau

Table of contents

Executive Summary **2**
1. Introduction **3**
2. Background **5**
2.1 Background on Small Business Credit 5
2.2 Background on Matched-Pair Testing 9
3. Research Methodology **13**
3.1 Overview of Testing Methodology and Hypotheses 13
3.2 Bank and Branch Selection for Testing 14
3.3 Tester Recruitment, Training, and Protocols 17
3.4 Outcome Measures and Analytic Approach 21
4. Statistical Analyses and Results **26**
4.1 Descriptive Overview of Test Visits 26
4.2 Primary Statistical Analyses 28
4.3 Aggregate Analysis of All Relevant Measures 43
5. Conclusion **48**
5.1 Summary of Results 48
5.2 Limitations 49
5.3 Implications for Research and Industry 50
Appendix A: Supplementary analyses and results **53**

Executive Summary

The Consumer Financial Protection Bureau ("CFPB") and the U.S. Department of Justice ("DOJ") conducted matched-pair testing to test for and detect differential treatment of fictitious well-qualified Black and white small business owners seeking credit at large bank lenders in two counties. This report describes a research study analyzing the matched-pair testing data.

Pairs of testers visited 25 bank branches located in Fairfax County, Virginia and 25 branches located in Nassau County, New York—consisting of 100 total visits across 23 financial institutions—over several months in 2023. Testers were trained to present themselves based on a pre-developed small business profile and inquire about available business financing options. Each visit was audio recorded and testers completed a post-visit survey documenting their visit to generate the data analyzed in this report.

Our aggregate-level analyses reveal that Black testers received less favorable treatment than paired white testers in two of the four treatment domains we considered. First, Black testers received less favorable treatment on measures of encouragement/discouragement to apply for financing than white testers. Second, in the domain of small business loan products discussed and potential steering, bank representatives were more likely to discuss non-requested credit products—such as business credit cards or real estate-secured loans—with Black testers than with white testers. These differences in treatment are statistically significant.

Given the design and scope of this pilot research, these findings should not be generalized to the broader small business lending market or to specific financial institutions. These findings do, however, highlight the existence of differential treatment in small business lending. This research reveals evidence of—and provides a framework for detecting—differential treatment of well-qualified Black and white small business owners seeking credit at large bank lenders in select counties.

1. Introduction

Small business owners seeking to build or expand their businesses can often benefit from obtaining financing. While small business lending can provide a pathway for growth, small business owners may face barriers to accessing credit throughout the loan application process. One such barrier may include differential treatment based on a small business owner's race.[1]

This report describes a study analyzing the outcomes of investigatory research conducted by the Consumer Financial Protection Bureau ("CFPB") and the U.S. Department of Justice ("DOJ"). The research analyzed the results of matched-pair testing to test for and detect differential treatment of fictitious well-qualified Black and white small business owners seeking credit in person. It focused on large bank lenders in two counties in separate metropolitan statistical areas.

The study aggregates the results of 50 matched-pair tests (comprising 100 tester visits) of small business lending institutions. Each test consisted of one Black tester and one white tester separately visiting a single bank branch to inquire about financing options for their small business. The 50 tests were conducted at fifty different bank branches—25 in Nassau County, NY and 25 in Fairfax County, VA—across 23 different financial institutions. Bank branches were selected at random from among a prespecified set of large banks offering in-person small business services in one or both counties. Testers received training on testing protocols before conducting any tests. To capture tester experiences, visits were audio recorded and testers completed a detailed post-visit survey about their interaction with bank representatives. These experiences were analyzed as distinct outcomes across four domains of treatment: (1) level of encouragement/discouragement to apply for financing;[2] (2) information provided to the tester about available credit products and potential steering toward product types; (3) overall quality of treatment or customer service; and (4) business and credit information requested of the tester.

We find evidence of less favorable treatment of Black testers relative to paired white testers for two out of the four treatment domains we considered. First, in the domain of encouragement/discouragement

---

[*] Lynn Conell-Price, Patrick Heck, and Elle Tibbitts (CFPB Office of Research) are the corresponding authors for this report.
[1] Race-based gaps in business financing have been demonstrated in representative field surveys. See "2022 Report on Firms Owned by People of Color: Based on the 2021 Small Business Credit Survey." 2022. Small Business Credit Survey. Federal Reserve Banks. Available at https://doi.org/10.55350/sbcs-20220629; Scott, M. L., Bone, S. A., Christensen, G. L., Lederer, A., Mende, M., Christensen, B. G., and Cozac, M., "Revealing and Mitigating Racial Bias and Discrimination in Financial Services" (2023). *Journal of Marketing Research* 61(4): 598-618. Available at https://doi.org/10.1177/00222437231176470.
[2] We use the term "encouragement" as including factors that increase an individual's motivation to apply for credit at a given bank. We use the term "discouragement" as including factors that decrease an individual's motivation to apply for credit at a given bank. Throughout this report, our use of "discouragement" does not necessarily correspond to the usage of this term for regulatory purposes in ECOA/Regulation B. (see 12 C.F.R. § 1002.4(b)). Our use of the words "encouragement" and "discouragement" also corresponds with the wording in surveys that testers completed.

we find that Black testers received less favorable treatment, reflecting less encouragement and more discouragement, than white testers. In the domain of products discussed and potential steering, we find that bank representatives were more likely to initiate discussions of products that were not explicitly requested when interacting with a Black (versus white) tester. These differences in treatment are statistically significant.[3]

This research reveals evidence of—and provides a framework for detecting—differential treatment of well-qualified Black and white small business owners seeking credit at large bank lenders in select counties. Although the scope of this pilot research was limited, it offers novel preliminary insights into the treatment of well-qualified Black and white small business owners and provides a framework that can help inform larger and more representative studies.

In the next section of this report, we provide background on small business lending and the matched-pair testing methodology, including its application in prior government and research testing. We then provide greater detail on our research methodology and analytic approach in this study, before reporting our findings and conclusions.

[3] See 4. Statistical Analyses and Results for details on levels of statistical significance and the size of the effects we observed.

2. Background

2.1 Background on Small Business Credit

2.1.1. Small Business Market and Credit Needs

This research focuses on credit access within the small business market, where consumers may face unique challenges compared to other markets.

Although there is no universal definition of what constitutes a small business, definitions applied by financial institutions, researchers, and the federal government typically specify thresholds of business revenues, number of employees, or loan amounts obtained by a commercial borrower to delineate a "small business" from businesses in general.[4] For example, the U.S. Small Business Administration ("SBA") defines small businesses according to a set of size standards that vary based on the specific industry in which an entity principally operates.[5] These standards generally specify a threshold for either number of employees or average annual business receipts as the largest size at which a business is classified as a small business. Organizationally, a small business can be a sole proprietorship, cooperative, partnership, or corporate structure with few employees and a limited amount of revenue and assets. Due to their more limited revenues, small businesses will generally have more restricted access to credit than larger commercial entities.[6]

According to 2020 data from the SBA there were 33.3 million businesses with fewer than 500 employees in the country employing more than 61.6 million people in total, or about half of the total number of private sector workers.[7] Many factors that shape the trajectory of a small business fall outside of those businesses' control, including demand, costs, and the prices they can charge consumers. Small businesses' limited financial reserves make them more vulnerable to economic shocks than larger businesses. For example, both the Great Recession beginning in 2007 and the COVID-19 pandemic period beginning in 2020 disproportionately impacted employment for smaller

[4] In CFPB's final Small Business Lending Rule, which implements section 1071 of the Dodd-Frank Act, a business is defined as a small business "if its gross annual revenue, as defined in § 1002.107(a)(14), for its preceding fiscal year is $5 million or less," see § 1002.106 Business and small business.
[5] See SBA's online resources for size standards, e.g., https://www.sba.gov/document/support-table-size-standards
[6] See Nanda, R., and Phillips, G., "Small Firm Financing: Sources, Frictions, and Policy Implications." In *Handbook of the Economics of Corporate Finance*, edited by B. Eckbo, G. Phillips, and M. Sorenson, 1:107-35. North-Holland, 2023. Available at https://doi.org/10.1016/bs.hecf.2023.02.003.
[7] "2023 Small Business Profile." U.S. Small Business Administration, 2023. Available at https://advocacy.sba.gov/wp-content/uploads/2023/11/2023-Small-Business-Economic-Profile-US.pdf

relative to larger businesses.[8,9] Such challenges may be mitigated if small business owners have existing credit lines or access to new credit sufficient to cover unexpected cash flow deficiencies.

Of the total number of American small businesses, nearly one third are owned or primarily owned by individuals identified as any race and national origin other than white and non-Hispanic (i.e., "minority-owned," as defined in the Annual Business Survey of the U.S. Census Bureau).[10] However, these entrepreneurs own a smaller proportion of employer businesses. The Census Bureau estimated that, as of 2023, 1.2 million businesses with employees were minority-owned, constituting 21 percent of all employer enterprises. Examining businesses without employees, generally sole proprietorships, the Minority Business Development Agency of the U.S. Department of Commerce found that, in 2018, approximately 8.7 million non-employer businesses (33.6 percent of all non-employer businesses) were minority-owned.[11]

As of 2021, 87 percent of Black-owned businesses were located in urban areas,[12] where populations reflect increasingly more racial and ethnic diversity relative to the larger U.S. population.[13] Recent research focused on small businesses operating in more racially diverse communities shows how they may be particularly vulnerable to negative demand and cost shocks because they tend to be smaller, less experienced, and have fewer resources than those operating in other communities. For example, a 2019 report from the J.P. Morgan Chase Institute found that in 90 percent of majority Black and Hispanic communities, most small businesses operated with less than 14 days of cash as a buffer against

---

[8] Between 2007 and 2009, employment at businesses with under 50 employees declined by 10.4%, compared to 7.5% at larger employers. *p.* 1, Şahin, A., Kitao, S., Cororaton, A., and Laiu, S. "Why Small Businesses were Hit Harder by the Recent Recession," (2011). Current Issues in Economics and Finance, 17(4). Federal Reserve Bank of New York. Available at https://www.newyorkfed.org/medialibrary/media/research/current_issues/ci17-4.pdf.

[9] An estimate of the increase in small business permanent closures that resulted during the first year of the pandemic compared to the normal closure baseline has been calculated to be as many as 200,000, estimating excess establishment exits and analyzing other estimates of small business exits during the pandemic. See Crane, L., Decker, A., Flaaen, A., Hamins-Puertolas, A, and Kurz, C. "Business Exit During the COVID-19 Pandemic: Non-Traditional Measures in Historical Context," (April 2021). Finance and Economic Discussion Series. Washington: Board of Governors of the Federal Reserve System. Available at https://www.federalreserve.gov/econres/feds/files/2020089r1pap.pdf. The paper defines "exit" as permanent shutdown and calculates "excess" exits by comparing the number of exits during March 2020-February 2021 with previous years.

[10] See press release, "Census Bureau Releases New Data on Minority-Owned, Veteran-Owned and Women-Owned Businesses," (Press Release Number CB23-112). (October 2023). Available at https://www.census.gov/newsroom/press-releases/2023/annual-business-survey-employer-business-characteristics.html.

[11] Nearly 8.7 million minority non-employer firms in the U.S. generated $306.1 billion in revenues in 2018. See "All Minority-Owned Firms: Fact Sheet," (June 2022). Minority Business Development Agency, U.S. Department of Commerce. Available at https://www.mbda.gov/sites/default/files/2022-06/All%20Minority%20Owned%20Firms%20Fact%20Sheet%20-%20FINAL%206.10.2022.pdf.

[12] See Leppert, R., "A Look at Black-owned Business in the U.S.," (February 2024). Pew Research Center. Available at https://www.pewresearch.org/short-reads/2024/02/16/a-look-at-black-owned-businesses-in-the-us/.

[13] See Frey, W., "2020 Census: Big cities grew and became more diverse, especially among their youth.," (October 2021). Brookings Institute. Available at https://www.brookings.edu/articles/2020-census-big-cities-grew-and-became-more-diverse-especially-among-their-youth/

unanticipated events.[14] Meanwhile, this was the case for only 35 percent of small businesses in majority white communities. These findings indicate that small businesses operating in communities of color—where small businesses with non-white, non-Hispanic owners frequently operate—may have a greater need for financing to address cash flow insufficiencies than better-resourced small businesses in white communities. Businesses with greater access to cash and financing are less likely to be vulnerable to business closures and more likely to expect future revenue growth.

Black and Hispanic small business owners may also experience more difficulty accessing affordable credit. For example, an analysis of the Federal Reserve Banks' 2017 Small Business Credit Survey[15] found that both Black and Hispanic small business owners were more likely to be denied credit or discouraged from applying for credit based on their credit scores.[16]

2.1.2. The Credit Process and Potential Barriers in Small Business Lending

To obtain financing, small business owners must navigate the credit process while interacting with potential lenders. The process can include identifying a need for credit, obtaining information about credit alternatives, preparing to apply, submitting a formal application along with requested supplemental information, and understanding a credit decision and any associated contingencies established by the lender (see "Navigating the credit process" below).

[14] See "Place Matters: Small Business Financial Health in Urban Communities" (Sept. 2019). JP Morgan Chase and Co. Institute. Available at https://www.jpmorganchase.com/content/dam/jpmc/jpmorgan-chase-and-co/institute/pdf/institute-place-matters.pdf; Farrell, D., Wheat, C., and Mac, C. "Small Business Owner Race, Liquidity, and Survival" (July 2020). JP Morgan Chase and Co. Institute. Available at https://www.jpmorganchase.com/content/dam/jpmc/jpmorgan-chase-and-co/institute/pdf/institute-small-business-owner-race-report.pdf.

[15] Robb, A. "Minority-Owned Employer Businesses and their Credit Market Experiences in 2017" (July 2020). Office of Advocacy, Small Business Administration. Available at https://advocacy.sba.gov/wp-content/uploads/2020/07/Minority-Owned-Employer-Businesses-and-their-Credit-Market-Experiences-in-2017.pdf.

[16] The Equal Credit Opportunity Act (ECOA) makes it "unlawful for any creditor to discriminate against any applicant, with respect to any aspect of a credit transaction," on a prohibited basis such as race or national origin. 15 U.S.C. § 1691(a). Creditors that discourage applicants and prospective applicants on a prohibited basis violate ECOA and Regulation B. *See Consumer Fin. Prot. Bureau v. Townstone Fin., Inc.*, --- F.4th ----, 2024 WL 3370023 (7th Cir. July 11, 2024).

Navigating the Credit Process:
Borrowers and lenders must complete many steps during the credit-seeking process. Barriers to access can occur during any of the steps required of the borrower, lender, or both. Some examples of these steps include:

Borrower:
- Determining credit amount needed
- Choosing product (e.g., line of credit or loan)
- Gathering financial information
- Identifying lender(s)
- Completing application
- Providing supplemental information

Financial Institution/Lender:
- Review for completeness
- Approval/denial
- Pricing
- Loan or line amount assignment
- Required collateral determination
- Loan structure elements

Although barriers to credit access can occur at any point,[17] a great deal of regulation has focused on potential differential credit decisioning outcomes. For example, CFPB's Small Business Lending Rule, which implements section 1071 of the Dodd-Frank Act, requires financial institutions to collect and report to the CFPB data regarding applications for credit by small businesses.[18] Information to be collected and made public—subject to consideration of privacy interests—includes information on the race, sex, and ethnicity of the principal owner(s) of the business, as well as the characteristics and disposition of the credit request.

Though usable data from the Small Business Lending Rule will not be available for some time, other federal regulatory data collection efforts currently provide a limited window into credit availability in small business lending. These efforts, restricted to originated loans, include data collected under the guidelines of the Community Reinvestment Act of 1977 (CRA), call reports by covered banks and credit unions to their federal prudential regulators (Federal Reserve System, Federal Deposit Insurance Corporation (FDIC), National Credit Union Administration (NCUA) and Office of the Comptroller of the Currency (OCC)), and information provided to the Community Development Financial Institutions (CDFI) Fund by its participants.

Despite regulatory improvements around monitoring credit application outcomes, small business owners may also face substantial and difficult challenges early in the credit process—well before a formal application has been submitted. Prior experiences or expectations of discrimination, for example, may discourage a small business owner from even submitting an application.[19] These

[17] Ibid.
[18] See Small Business Lending Under the Equal Credit Opportunity Act (Regulation B), available at https://www.federalregister.gov/d/2023-07230
[19] For an example of how anticipated discrimination can affect homeownership rates, see Pol, L., Guy, R., Ryker, R., and Chan, W. (1981). "Anticipated Discrimination in the Home Lending Market." *Housing and Society* 8: 3-11.

experiences and expectations may also influence the specific products they seek or the lenders they choose to engage.

Critically, and as one motivating aspect of the present research, barriers that occur prior to the submission of a formal application may not be visible in administrative data, even to lenders themselves. This obscures important pieces of the small business lending process, making them difficult to monitor, understand, and address.

We seek to partially address this gap by reporting the results of a research study, conducted on investigative research data, that directly assesses differential treatment of small business applicants before a formal credit application has been submitted. This research analyzes data from matched-pair testing, a field-testing methodology that, in the lending context, enables comparison of two similarly situated fictitious credit seekers who differ on one dimension of interest.

2.2 Background on Matched-Pair Testing

Matched-pair testing is a long-standing field-testing methodology that allows investigators to observe how lenders, sellers, or other supply-side agents behave in actual markets. This method tests for differential treatment in a market by matching pairs of individuals ("testers") on all relevant characteristics except the one being studied. These testers then interact with an agent and may, depending on the test design, inquire about, request, negotiate, or apply for a good or service. The outcomes of these interactions are documented and analyzed to determine if they indicate differential treatment.[20] Matched-pair testing methodology has long been used by government, academic researchers, and advocacy groups as a tool to detect and measure differential treatment of consumer groups in domains other than small business lending, including housing, residential and automobile lending, access to public accommodations and government services, and employment.[21,22]

[20] See Fix, M., and Struyk, R. "Clear and Convincing Evidence: Measurement of Discrimination in America." Urban Institute Press, (Oct. 1993). https://webarchive.urban.org/publications/105136.html; Yinger, J. (1998). "Evidence on Discrimination in Consumer Markets." *Journal of Economic Perspectives* 12(2): 23-40. Available at https://www.aeaweb.org/articles?id=10.1257/jep.12.2.23.

[21] For example, both the CFPB and DOJ have used evidence from matched-pair testing to establish ECOA violations (see, e.g., *United States and Consumer Financial Protection Bureau v. BancorpSouth Bank,* No. 1:16cv118-GHD-DAS (N.D. Miss. filed June 29, 2016), files.consumerfinance.gov/f/documents/201606_cfpb_bancorpsouth-joint-complaint.pdf; *United States v. Guaranteed Auto Sales* (D. Md.)). Neither the CFPB nor DOJ has filed a small business lending lawsuit under ECOA based on matched-pair testing evidence.

[22] Matched-pair testing and similar methodologies have also been used by industry groups, who may conduct "mystery shopping" to assess customer service and provide training.

There is a body of academic research on "paired testing," which includes matched-pair testing.[23] In economics and other social sciences, research conducted using paired testing methodologies is often referred to as "audit studies." This research typically tests for discrimination in consumer markets via controlled in-person or otherwise real-world engagements. Researchers in economics recognize audit study methodology (which can include matched-pair testing) as one of two complementary empirical strategies to identify discrimination, the other being regression studies of existing data (see e.g., Yinger, 1998; Bertrand and Duflo 2017).[24,25] Audit study methodology is an important complement to regression studies because field studies with proper randomization allow researchers to draw causal conclusions (e.g., that observed differences in treatment were likely caused by a controlled tester characteristic). In contrast, regression analyses tend to leverage larger sample sizes and can address an important limit to audit studies, namely, that an individual audit study should not be interpreted as capturing the discrimination faced by average consumers in the market, because the results are specific to the research design choices and sample for that study.

Consistent with limitations of the audit study methodology, this study was not designed to produce generalizable findings. Matched-pair tests typically prioritize reducing differences in testers' characteristics other than the factor of interest. They also often focus on unambiguously well-qualified applicants to ensure that any unfavorable treatment is driven by the factor of interest, rather than economic incentives conveyed by differences or noise in the testers' financial profiles. However, results of studies where all tester profiles are designed to be well-qualified do not reflect the experience of all consumers in the market. For example, prior research shows that marginally qualified testers experience greater discrimination than well-qualified testers (Hunter and Walker 1996),[26] indicating

[23] The terms "paired testing" and "matched-pair testing" are often used interchangeably but are not necessarily synonymous. We refer to "matched-pair testing" as a specific case of "paired testing" where testers are carefully matched on characteristics other than the variable of interest. Some prior studies that use "paired testing" methodologies have also carefully matched testers into pairs, while others have not.

[24] Yinger, 1998, see footnote 20; Bertrand, M., and Duflo, E. "Field Experiments on Discrimination." In *Handbook of Economic Field Experiments*, edited by A. Banerjee and E. Duflo, 1:309-393. North Holland, 2017. Available at https://www.sciencedirect.com/science/article/pii/S2214658X1630006X?via%3Dihub.

[25] Yet another relevant research methodology, "correspondence studies," relies on correspondence (e.g., via mail or email) to infer bias or discrimination. Bertrand and Duflo (2017) discuss the proliferation over the last two decades of these correspondence studies, a particular type of paired testing in which fictitious applicants correspond via mail or email. Correspondence studies often use identical applications but for variation in names selected to convey race. This approach has gained popularity because of the low marginal cost of additional tests and ability to match more observable characteristics by using nearly identical applications. Given that the small business credit process typically involves more borrower-lender interaction than mere correspondence, we did not use this methodology for our research.

[26] Hunter, W., and Walker, M. B., (1996). "The Cultural Affinity Hypothesis and Mortgage Lending Decisions." *Journal of Real Estate Finance and Economics* 13(1): 57-70. Available at https://link.springer.com/article/10.1007/BF00174551.

that results of testing conducted with well-qualified testers may underestimate discriminatory treatment (HUD/Urban 2012).[27]

Prior audit and matched-pair testing methodology has revealed discrimination in other contexts, including in the housing market. The Department of Housing and Urban Development (HUD) and the Urban Institute (Urban) have together conducted matched-pair studies monitoring racial and ethnic discrimination in rental and sales housing markets since the late 1970s. Their 2012 housing discrimination report presented evidence from over 8,000 paired tests conducted in a nationally representative sample of 28 metropolitan areas in which a white and matched Black, Hispanic, or Asian tester contacted a housing provider to inquire about a housing unit.[28] The study found that when contacting housing providers, Black, Hispanic, and Asian testers were generally just as likely as similarly qualified white testers to get an appointment and hear about at least one available housing unit. However, when differences in treatment occurred within individual pairs, the white tester tended to receive more favorable treatment than Black, Hispanic, and Asian testers. In particular, these testers were told about and shown fewer homes than otherwise similar white testers who inquired around the same time. An important advantage of the study's large sample is that it enabled rigorous statistical analysis within specific regions and testers' races and ethnicities. For example, the study found that Black, Hispanic, and Asian testers whose races and ethnicities were more identifiable experienced greater discrimination than Black, Hispanic, and Asian testers who were more likely to be mistaken as white. Testing at this scale also allowed the researchers to conclude that their findings were not isolated within specific regions. Since the 1970s, this research has advanced the methodology of matched-pair studies and prompted productive discussion of the evidentiary value of data generated using this methodology. Heckman and Siegelman (1993), for example, discuss the appropriateness of various statistical approaches to inferring discrimination from paired test data—a discussion that informs the analyses we present.[29]

Turning to the small business credit market, two peer-reviewed research studies have used matched-pair testing methodology to investigate potential discrimination. Both studies were conducted with the National Community Reinvestment Coalition (NCRC), an advocacy organization. Bone et al. described their 2019 research as "a case study to demonstrate how a matched-pair mystery shopping methodology can be used to investigate potential differences in lending treatment to minority small business

[27] Turner, M., Levy, D., Wissoker, D., Aranda, C., Pitingolo, R., and Santos, R. "Housing Discrimination Against Racial and Ethnic Minorities 2012." (June 2013) Report prepared for U.S. Department of Housing and Urban Development. Available at https://www.huduser.gov/portal/publications/fairhsg/hsg_discrimination_2012.html.
[28] Ibid.
[29] Heckman, J., and Siegelman P., (1993). "The Urban Institute Audit Studies: Their Methods and Findings." In *Clear and Convincing Evidence: Measurement of Discrimination in America*, edited by Fix, M., and Struyk, M., 187-258. The Urban Institute Press, 1993.

consumers".[30] This research reported results from 26 matched-pair tests (52 test visits) of Black and white male borrowers seeking small business credit in two major United States metropolitan areas. Bone et al. (2019) report that Black testers were asked to provide more financial information (statistically significant at the five percent level) and were less likely to receive encouragement and assistance in completing applications at the bank (significant at the ten percent level) compared to matched white testers. This study, published before we conducted any testing, informed our hypothesis and research design development. A second paper by some of the same coauthors, Scott et al. (2023), further developed the paired testing methodology using a combination of field and laboratory studies to test hypotheses about the mechanisms underlying discriminatory treatment in the small business lending market. This second paper was released after our data collection and did not inform our design, though it offers complementary evidence to our research.

[30] *p*. 394, Bone, S., Glenn L. Christensen, J., Williams, S., Lederer, A., and Lubin, P., (2019). "Shaping Small Business Lending Policy Through Matched-Pair Mystery Shopping." *Journal of Public Policy & Marketing* 38(3): 391-399. Available at https://journals.sagepub.com/doi/full/10.1177/0743915618820561.

3. Research Methodology

This section describes the methodology for conducting testing and our analytic approach. We review the research hypotheses and testing approach before detailing the selection of tested banks and branches, tester recruitment and testing protocols, and our analytic approach.

3.1 Overview of Testing Methodology and Hypotheses

Testers were recruited and trained to represent fictitious well-qualified small business owners inquiring about credit for their businesses at in-person appointments. Each complete matched-pair test included one Black and one white tester assigned to visit the same bank branch location on different dates to inquire about financing to expand their business. Testers within a pair were either both male or both female to minimize any differences that might emerge between sex or gender categories. All testers were assigned a test profile describing the business and financial circumstances they were to portray and were instructed to follow a standardized test protocol to minimize the possibility that any differences in tester treatment would be driven by variation in tester behavior. Ensuring that visits were as standardized as possible was also a goal of the training testers received, described below. Consistent with prior research, each Black tester was assigned a slightly more favorable financial profile compared to their matched white counterpart. Immediately after each visit, testers completed a detailed online survey including closed- and open-ended questions designed to capture their treatment and experience. Unknown to bank representatives, test visits were audio recorded[31] to validate survey responses, ensure that testers followed the testing protocol, and allow for detailed review of individual tests. Testers provided email and phone contact information during their appointments, and phone and email accounts were monitored for follow-up contact from bank employees.

Using the data generated from these tests, we examined whether Black and white testers received equivalent treatment in the following domains: (1) level of **encouragement/discouragement** to apply for financing; (2) information provided to the tester about available credit **products and potential steering** toward product types; (3) overall quality of treatment or **customer service**; and (4) **business and credit information** requested of the tester.

[31] All testing was conducted in single-party consent states and testers audio recorded each visit without informing bank staff.

Based on related prior work, we hypothesized that white testers would receive favorable outcomes relative to Black testers who were at least equivalently qualified. Broadly, we predicted that any differences in treatment would tend to favor white relative to Black testers, who were each slightly more qualified than their white counterparts. We specifically predicted that, relative to white testers, Black testers would tend to (1) receive less favorable encouragement/discouragement to apply for credit at the institution, (2) receive information about products that the tester did not request (which might not meet presented business needs), (3) receive lower quality customer service, and (4) have more business and credit information requested of them by bank representatives.

Testing resulted in a prespecified sample size of 50 complete matched-pair tests (i.e., 100 tester visits). This provided sufficient sample size to conduct statistical analysis with statistical power of 80 percent or greater (using effect sizes from prior research as inputs to a power analysis).[32]

3.2 Bank and Branch Selection for Testing

3.2.1. Geographic Regions Tested

Testing focused on two counties in the New York City and Washington, DC metropolitan areas—Nassau County, NY and Fairfax County, VA—selected because of their geographic convenience, demographic characteristics, and the CFPB and DOJ's shared preference to consider more than one region. The CFPB and DOJ divided responsibility for administering tests, with CFPB conducting 25 matched-pair tests in Nassau County, NY through a contractor operating under supervision of the CFPB, the Fair Housing Justice Center ("FHJC"), and DOJ conducting the other 25 matched-pair tests in Fairfax County, VA, through its in-house Fair Housing Testing Program.

Table 1 summarizes the demographic characteristics of each county. Importantly for our study, both counties have sizable Black populations, although there is considerable geographic variation in density of Black and white populations within each county, reflecting a potential history of segregation. Both counties are also population centers near major cities. According to the Annual Business Survey conducted by the Census Bureau, Nassau County had a larger total number of employer businesses (those with employees in addition to the owner) than Fairfax County in 2017 (see Table 2). Both

[32] Prior to viewing any data, we planned the outcome measures, experimental design, hypotheses, and analyses. Statistical power is the probability with which the combination of an experiment and hypothesis test can detect an effect (here, a difference in treatment between groups) if such an effect exists in the population of interest. To determine requisite sample size, we conducted a power analysis assuming expected effect sizes, one-tailed tests, an alpha level of 0.05, and statistical power of 80% or greater. Expected effect sizes were informed in part by the one published article on matched-pair testing in the small business credit context (at the time): Bone et al. (2019, *Journal of Public Policy & Marketing*), where differences in treatment ranged from 11 to 30 percentage points between the two tester categories.

counties had a similar number of non-white, non-Hispanic-owned employer businesses (20.7 percent in Nassau and 33.3 percent in Fairfax). Table 2 shows employer business numbers for each county.

TABLE 1: DEMOGRAPHIC CHARACTERISTICS BY COUNTY (2022)

Metric	Nassau County, NY	Fairfax County, VA
Total Population	1,383,726	1,138,331
Per capita personal income	$76,442	$101,400
Females as percent of total	50.70%	49.97%
White population share	70.79%	63.31%
Black or African American share	13.38%	10.92%
American Indian and Alaska Native share	0.62%	0.56%
Asian population share	12.92%	20.91%
Native Hawaiian and Other Pacific Islander share	0.12%	0.09%
Identifies as two or more races	2.16%	4.21%
Hispanic or Latino identification	17.90%	16.83%
White, not Hispanic or Latino	55.81%	48.52%

Note. Population: Census Bureau (July 1, 2022), Annual County Resident Population Estimates, available at: https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-detail.html. Income: Bureau of Economic Analysis (Nov. 16, 2023), Personal Income by County and Metropolitan Area, 2022, available at https://www.bea.gov/data/income-saving/personal-income-county-metro-and-other-areas. Per capita personal income for Fairfax County, VA combines Fairfax, Fairfax City, and Falls Church.

TABLE 2: EMPLOYER BUSINESS CHARACTERISTICS BY COUNTY

Employer Business Characteristic	Nassau County, NY	Fairfax County, VA
All employer businesses	*43,969*	*24,631*
Male-owned	29,597 (67.3%)	14,230 (57.8%)
Women-owned	8,722 (19.8%)	5,373 (21.8%)
Non-white, non-Hispanic-owned	9,122 (20.8%)	8,207 (33.3%)
White, non-Hispanic-owned	32,415 (73.7%)	13,678 (55.5%)

Note. Census Bureau (2017) Annual Business Survey Estimates. Numbers do not sum to "all employer businesses" totals, and percentages do not sum to 100%, because respondents could identify as members of multiple demographic groups including groups other than those listed here.

3.2.2. Bank and branch selection methodology

Testing occurred at large banks with one or more operating physical branches that offered small business credit services in at least one of the tested regions.[33] While non-bank lenders (e.g., financial technology companies) provide a growing source of small business credit, banks remain the most likely source of credit for small business owners. Data contained in the "2023 Report on Employer Firms: Findings from the 2022 Small Business Credit Survey" from the member banks of the Federal Reserve found that small businesses were most likely to apply for loans at large banks (defined as banks with $10 billion or more in total deposits), followed by small banks (those banks with less than $10 billion in deposits), fintech lenders, commercial finance companies, credit unions, and community development financial institutions (CDFIs).[34] Eligibility for testing and inclusion in the present research was restricted to large banks, specifically, banks with a minimum of $10 billion in assets.[35]

By applying these eligibility criteria, 23 banks qualified for testing. These consisted of 18 banks in Fairfax County, VA, and 14 banks in Nassau County, NY, with nine of the qualified banks present with branch operations in both counties.

After identifying banks for inclusion, 50 specific branches from these banks were randomly selected for testing. To mitigate the risk of detection of testing by bank staff, only one matched-pair test was conducted at a given branch and the number of tested branches per financial institution was restricted to between one and four, based on how many branches a bank operated in the target county.

In case testing at a specific branch was not feasible after it was selected, contingency branches were also randomly selected. A selected branch was replaced with a contingency branch when it was determined that the selected branch served only private banking clients or when the selected branch had recently closed, which occurred in only a few cases.

---

[33] Presence of an operating, physical branch in at least one of the tested counties was verified using FDIC data on federally-insured banks with full-service branch locations at the county level, available at https://banks.data.fdic.gov/bankfind-suite/bankfind. Small business credit offerings were verified in two ways: (1) availability on a bank's website, and (2) using 2020 Community Reinvestment Act (CRA) data to screen out financial institutions that had originated fewer than 5 small business loans over the last 12-month period in the relevant county.

[34] According to the 2023 Federal Reserve System report on employer firms covering the roughly 12-month period between September-November 2021 and September-November 2022, 43% of small businesses applied to large banks for credit, 30% to small banks, 22% to fintech (online lenders), 13% to commercial finance companies and 12% to others. See *p.* 15, "2023 Report on Employer Firms: Findings from the 2022 Small Business Credit Survey." 2023. *Small Business Credit Survey*. Federal Reserve Banks. Available at https://doi.org/10.55350/sbcs-20230308.

[35] This criterion excluded smaller community banks and credit unions operating in the two counties, which would be promising areas for future research. Larger credit unions were excluded because credit unions generally operate under a legal cap on their ability to make commercial loans, called the aggregate member business loan limit. The aggregate limit on a federally insured credit union's net member business loan balances is the lesser of 1.75 times the actual net worth of the credit union, or 1.75 times the minimum net worth required under section 1790d(c)(1)(A) of the Federal Credit Union Act.

3.3 Tester Recruitment, Training, and Protocols

The Department of Justice (DOJ) coordinated the testing in Fairfax County, Virginia, and the CFPB in conjunction with the Fair Housing Justice Center (FHJC) coordinated the testing in Nassau County, New York. Both DOJ and the FHJC are experienced in and regularly conduct civil rights testing, and they had experienced test coordinators train the testers and coordinate the tests.

Testers were generally recruited from pools of individuals with prior training and experience in a field setting, and some testers had experience as small business owners or as employees of small businesses.

Prior to conducting their branch visits, testers received training that included curriculum on small businesses. Trainings were conducted in-person and consisted of oral presentations by trainers on topics including testing procedures, tester profiles, and the survey that testers would complete after conducting each visit. During the training, testers also participated in role-playing a simulated test visit and viewed a video of a pre-test visit conducted at a small business lender to expose them to situations and questions they might encounter during a visit.

Following training, each tester also conducted a practice visit at a banking branch that was selected separately from the list of branches selected for testing. These additional "practice" visits were not included in the data for analysis. Practice tests provided an opportunity for testers to gain experience with protocols and their assigned profile. The practice tests also improved protocol quality control by allowing test coordinators to assess whether testers were complying with protocols and provide feedback to testers on needed corrections if warranted.

3.3.1. Testing Protocols

Protocols were established to ensure that tests were completed consistently. The objectives of these protocols were to ensure that testers asked similar questions during their visits, that testers were prepared to answer questions asked by bank representatives consistently, and that interactions were fully captured for later analysis. Each test was supervised and reviewed by a test coordinator, who confirmed that a completed test met protocol standards, provided data that could be analyzed, and captured a complete interaction.

A paired team consisted of a Black tester and a white tester of the same sex. Each tester was assigned a profile as the owner of a small business that had existed for a similar period of time in a comparable industry.[36] On each test visit, testers were instructed to inform the bank employee that they were

[36] Some tester pairs were assigned profiles indicating that they had received a Paycheck Protection Program ("PPP") loan. If asked, these testers were instructed to answer they had received a PPP loan and it had been forgiven.

looking to expand their business and to inquire specifically about financing through business loans and business lines of credit.

For each matched pair, testers and test coordinators either scheduled in-person appointments via a lender's website or by phone prior to each visit or both testers conducted a drop-in visit at the branch to speak with a small business lender and set up an appointment if one was required. Testers' appointments within a matched pair were separated by several days.

All calls and visits with the lenders were audio recorded. Immediately following each visit, each tester completed an online survey about their experience and provided all materials to the test coordinator, including materials received from the lender, tester assignment forms, and any notes taken during test preparation or the visit. Testers were instructed to report all the information received regarding small business financing on the online survey, which typically took a little over 30 minutes to complete. The test coordinators monitored the phone and email accounts assigned to the testers and documented any follow-up communications received from the lenders.

3.3.2. Tester Profiles

The profiles assigned to each tester were designed to reflect closely comparable small business owners who were well-qualified for the credit they were seeking. Each profile comprised elements including the type of business (e.g., technology, health care), amount of credit requested, purpose for the request, time in business, number of employees, and annual business revenue. Where appropriate, the values for these elements fell within ranges that were typical for small business owners in the New York City metropolitan and Northern Virginia areas. For example, the range of possible values for the number of employees in each profile was based on data from the 2020 Census Annual Business Survey, and selected industries and credit request characteristics were based on the Federal Reserve's 2022 Annual Report on Employer Firms. Across all matched-pairs, testers inquired about a desired credit amount of less than $100,000, or typically around 20 percent of their previous year's business revenue. Table 3 presents the key elements of a profile along with a description of the typical values used and the sources used when choosing these values.

Critically, profiles for each matched pair were designed to be closely comparable but not identical. This is to maintain the integrity of the methodology while mitigating the risk of detection by bank employees. To maintain comparability, differences in profiles within a matched pair were designed to be small enough not to meaningfully affect qualifications for credit. For example, testers were assigned high personal credit scores that differed by only a few points. The test design also adopted an accepted methodological strategy of making differences within a matched pair slightly favor the Black tester

relative to the white tester.[37] This strategy makes statistical testing for differential treatment favoring white over Black testers more conservative, since any differences in treatment attributable to the small differences in non-race profile characteristics should go in the opposite of the hypothesized direction (favoring Black over white testers).

[37] See Lubin, P. *Protecting Main Street: Measuring the Customer Experience in Financial Services for Business and Public Policy*. Routledge, 2010.

TABLE 3: TESTER PROFILE CHARACTERISTICS

Business Characteristic	Values in Tester Profiles
Business Revenue (Annual)	Between $100,000 and $400,000
Number of Employees	Fewer than ten employees
Time in Business	Less than five years
Typical Industry of Business	Including but not limited to healthcare, social services, technical services, construction
Credit Request (amount)	Less than $100,000, typically around 20 percent of business revenues
Product Requested	Business-purpose closed-end loan or line of credit for a small business (No preference between these stated)
Purpose of Borrowing	Operating expenses, expansion, and asset purchase
Personal Characteristic	
Credit score	A well-qualified score, typically over 700
Total household assets	Over $300,000
Household annual income	Over $70,000

Note. Tester profile values were selected to maximize comparability between and within test pairs. The characteristics provided in this table are described more broadly than what was assigned during testing. Sources used to inform business profile characteristics included the 2020 Census Annual Business Survey (Data Year 2019), the Federal Reserve's 2022 Report on Employer Firms (Data Year 2021), and the Federal Reserve's 2022 Report on Firms Owned by People of Color (Data Year 2021).[38]

[38] See National Center for Science and Engineering Statistics (NCSES). 2022. "*Annual Business Survey: 2020 (Data Year 2019).*" NSF 22-344. Alexandria, VA: National Science Foundation. Available at https://ncses.nsf.gov/pubs/nsf22344/; "2022 Report on Employer Firms: Based on the 2021 Small Business Credit Survey." 2022. Small Business Credit Survey. Federal Reserve Banks. Available at https://doi.org/10.55350/sbcs-20220222; "2022 Report on Firms Owned by People of Color: Based on the 2021 Small Business Credit Survey." 2022. Small Business Credit Survey. Federal Reserve Banks. Available at https://doi.org/10.55350/sbcs-20220629

3.4 Outcome Measures and Analytic Approach

Our analytic approach seeks to answer the primary research question: When inquiring about obtaining small business credit, do otherwise-matched Black and white testers receive equivalent treatment? We obtained data via the post-visit online survey completed by each tester, which contained over 70 individual items about the tester's experience.[39] These individual items were designed to capture testers' experiences and treatment in the following domains: (1) level of **encouragement/discouragement** to apply for financing; (2) information provided to the tester about available credit **products and potential steering** toward product types; (3) overall quality of treatment or **customer service**; and (4) **business and credit information** requested of the tester.

Answering this research question across and within four domains of interest requires an analytic approach suited to both narrow applications (e.g., targeted outcomes of interest) and broad ones (e.g., all data collected on relevant measures). Flexible and distinct analyses can strengthen conclusions related to the primary hypothesis, which we pursue by examining the data through different lenses.[40] To this end, our analytic approach (1) applies standard regression models to indices and ratings within each domain, generated using paired testers' survey responses, (2) tests whether the totality of the data supports the hypothesis (agnostic to any single measure or domain), and (3) provides individual supporting examples that clearly illustrate differential treatment within a pair.

While prior government reports on matched-pair testing (e.g., HUD/Urban 2012) have focused on analyzing objective outcomes (i.e., observable events such as verbal statements made by lenders), academic research studies have also analyzed subjective measures as outcomes in matched-pair testing—including in the small business lending context (e.g., Scott et al., 2023). Subjective measures are widely used in social science research to better understand individuals' experiences.[41] Subjective measures can, for example, supplement objective measures by revealing whether a person's personal experience correlates with observable events. Furthermore, subjective measures may provide a window

[39] Items that were not relevant to the hypothesis (e.g., the cleanliness of the branch being tested) were excluded from all analyses.

[40] Strengthening conclusions in this way improves *triangulation*—or the use of different analytic approaches to evaluate a single hypothesis—around research findings, which increases both internal and external validity. This increases the replicability and trustworthiness of evidence generated from individual research studies. See comment Munafò, M. R., and Davey Smith, G. (2018). "Robust Research Needs Many Lines of Evidence," *Nature* 553(7689): 399-401. Available at https://www.nature.com/articles/d41586-018-01023-3.; Yarkoni, T., (2022). "The Generalizability Crisis." *Behavioral and Brain Sciences* 45: e1. Available at https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/generalizability-crisis/AD386115BA539A759ACB3093760F4824.

[41] Social science researchers routinely elicit subjective outcomes from research participants as a proxy for subjective outcomes in the ultimate population of interest. See e.g., discussion of external validity and inferences about causal relationships outside of an experimental sample, p. 83, Shadish, W., Cook, T., and Campbell, D. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Houghton Mifflin, 2002.

into outcomes that cannot be directly observed. In this research we examine several inherently subjective constructs, for example, whether testers felt encouraged or discouraged to submit a credit application.[42] Results from these subjective measures should not be generalized given the study's limited sample size and scope, and because testers' reports of their subjective experiences during testing may differ from those of typical consumers.

Our analysis focuses on ten primary outcome measures spanning these four domains, which are summarized in **Table 4.** Specifically, each domain includes two or three survey outcomes of two main types: The first type—what we refer to as an "index"—combines several objective yes/no questions about whether an event occurred during the visit (e.g., "were you thanked for coming in?"), while the second type captures testers' overall rating of a given subjective construct, (e.g., "how high quality was the customer service?").[43]

The first two rows of **Table 4** summarize the first domain: **encouragement/discouragement** to apply for credit. The index measure of this domain captures testers' responses to 16 objective yes/no questions about what occurred during the visit and included ten potentially encouraging actions by a bank representative (e.g., "Provided an application or details to access an application") and six potentially discouraging items (e.g., "Encouraged or told me to contact other lenders"). The score, or "**encouragement/discouragement index**" outcome is defined as the count of potentially encouraging behaviors experienced during a visit (0 to 10) minus the count of potentially discouraging behaviors experienced (0 to 6) plus six, for a possible range from 0 (most negative level of encouragement/discouragement) to 16 (most positive level). In contrast, the "**encouragement/discouragement rating**" outcome captures the tester's response to the subjective question, "Considering your visit from start to finish, do you feel like the bank representative was *discouraging, encouraging, or neither* toward your potential application for a small business loan or

[42] One limitation to these measures is that, within a test pair, individual differences in how testers interpret and respond to a subjective question cannot be fully controlled. For example, one tester may adopt a stricter criterion for what they consider "encouraging" when reporting how encouraged they felt, compared to their paired counterpart. This limitation is partially addressed by (1) our regression approach that controls for differences between (but not within) test pairs (see section 4.2.1), (2) our analysis of the relationship between objective and subjective measures (see Figure 3), and (3) the triangulation we achieve by including multiple measures and analytic approaches to test a single hypothesis (see footnote 40).

[43] In defining index measures, we assume a relationship between the measures and the functional form of the construct we are studying. For example, we assume a positive relationship between testers' reported levels of encouragement/discouragement to submit an application and their actual likelihood to submit an application if they were truly pursuing credit. We approach this issue by imposing the simplest possible linear form with unit weights on each predictor; that is, we treat each level of a given measure as equally important or predictive as all others. This approach is informed by decision science research on the advantage of unit-weighted linear predictive applications; See Dawes, R., (1979). "The Robust Beauty of Improper Linear Models in Decision Making." *American Psychologist,* 34(7): 571-582.

line of credit at this institution?" on a scale ranging from 1 (very discouraging) to 7 (very encouraging) with a midpoint of 4 (neither encouraging nor discouraging).[44]

The third and fourth rows of the table summarize the second domain: information provided to the tester about available credit **products and potential steering** toward product types. First, the "**Discussed a requested product**" outcome captures whether the lender discussed at least one of the two main product categories that testers were instructed to seek: a business loan (BL) or business line of credit (BLOC). Responses are coded as 1 if the tester reported that the lender discussed at least one of these products and 0 if the tester reported that the lender did not discuss either product. Second, the "**Discussed a non-requested product**" outcome captures whether the lender discussed a product other than BL or BLOC (e.g., a business credit card, a real estate (e.g., personal residence) secured loan or home equity line of credit, a non-real estate personal loan or line of credit, personal credit card, or alternative funding sources such as borrowing from friends or family).[45] The first outcome measure captures whether testers received information about the types of products they sought, while the second outcome measure captures whether testers received information about products that they did not explicitly request. These non-requested products tend to offer less favorable terms and usefulness than the requested products.[46]

The fifth through seventh rows of the table summarize the third domain: overall quality of treatment or **customer service**. First, the "**customer service behaviors index**" outcome captures whether the tester experienced six objective events reflecting polite treatment upon branch arrival and departure (e.g., "A bank employee greeted me"). Second, the "**customer service visit rating**" outcome captures scale ratings by the tester in response to two subjective questions: "how high quality do you feel the customer service was?" and "how valued did you feel as a potential customer?" (ranging from 1: not at all, to 4: very). Third, the "**bank representative rating**" outcome averages the tester's scale ratings of the bank representative on how friendly, helpful, professional, patient, and informative the tester found them.[47]

---

[44] A scale of this type is often referred to as a Likert scale, or a rating scale used to measure attitudes and other constructs that contains scale points associated with numeric or verbal responses to a statement or question.
[45] The survey also captured whether a referral was made to another lender such as a Community Development Financial Institution (CDFI).
[46] Relative to the requested products (business term loans and business lines of credit), business credit cards are generally less favorable for the profiles presented because they often include lower credit limits than the loan amounts sought, typically charge higher interest rates on any revolving balance, and require quicker repayment to avoid higher interest rates. Personal real-estate backed and other personal credit products are less favorable because they introduce risk to a borrower's personal credit and assets through their business. For example, while BLOCs typically do not require collateral, HELOCs are typically collateralized by the small business owner's home and could threaten overall household financial stability. An assumption of this outcome is that non-requested products are often less favorable, but we acknowledge that some of the non-requested products discussed can meet specific business needs.
[47] These scales included response options with two poles, meaning that response options varied starting with the opposite of the trait adjective being rated (e.g., 1: "very unfriendly" to 7: "very friendly").

The eighth and ninth rows summarize the fourth domain: **business and credit information** requested of the tester. First, the "**business and credit information requested index**" outcome counts how many of 16 specific items were requested from the tester (e.g., years in business, personal credit score). Second, the "**business and credit information requested rating**" outcome asked testers to rate "how much information you discussed [with the bank representative] related to your business and business finances" on a scale from 1 (none) to 4 (a great deal).

In addition to the domain summaries in the first nine rows of the table, the bottom row describes one **non-survey outcome:** whether testers received any post-visit email or phone contact from bank representatives at the contact details testers provided.

TABLE 4: PRIMARY OUTCOME MEASURES FOR STATISTICAL ANALYSIS

Domain	Outcome	Type	Description	Scale
1. Encouragement/ discouragement	Encouragement/ discouragement index	Objective	Index summing 16 yes/no items: 6 potentially discouraging (reverse-coded), 10 potentially encouraging[48]	0 to 16
	Encouragement/ discouragement rating	Subjective	Tester's scale rating of application discouragement / encouragement experienced at branch	1 to 7
2. Products and potential steering	Discussed a requested product	Objective	Was at least one business loan (BL) or business line of credit (BLOC) product discussed?	0 (no) 1 (yes)
	Discussed a non-requested product	Objective	Was at least one product from a category other than BL, BLOC discussed? (business c/c, personal credit, alt. finance)	0 (no) 1 (yes)
3. Customer service	Customer service behaviors index	Objective	Index summing 6 politeness behaviors upon branch arrival and departure[49]	0 to 6
	Customer service visit rating	Subjective	Sum of two 1-4 scale ratings: service quality; sense of value as potential customer	2 to 8
	Bank rep. rating	Subjective	Average of 5 scale ratings of bank representative descriptors (e.g., friendly, helpful, patient, informative, professional)	1 to 7
4. Business and credit information requested of tester	Business and credit info requested index	Objective	Index summing 16 yes/no pieces of information requested[50]	0 to 16
	Business and credit info requested rating	Subjective	Rating of amount of information discussed related to tester business and finances	1 to 4
Post-visit follow-up *(non-survey)*	Post-visit follow-up	Objective	Did the bank follow up via phone or email after test?	0 (no) 1 (yes)

[48] The index captures 16 items with the prompt, "During your visit, did the bank representative do any of the following?" for 10 potential actions that may encourage submitting an application for credit at the bank. These included: "Stated that they would recontact or follow up with me after our visit," "Said or implied that I should formally apply [either during visit OR after visit," "Said or implied that I should formally apply during the visit," "Provided an application or details to access an application," "Offered to schedule a follow-up appointment for more information or to start an application," "Clearly described what information I would need to apply," "Informed me that they were interested in my application," "Informed me about how long it would take to approve my application," "Offered to assist with starting or completing a small business loan application," "Made plans for follow-up contact," and 6 potential actions that may discourage applying for credit at the bank: "Said or implied that my outcome could be better if I improve my credit before applying," "Said or implied that I should pursue other financing options outside this institution," "Informed me that they were not accepting or reviewing applications at this time," "Informed me that the relevant bank staff were not in today," "Encouraged or told me to contact other lenders," "Was unable to meet with a bank representative to discuss my business and financing needs." The index is defined as the count of potentially encouraging items [0-10] – (count of potentially discouraging items [0-6] + 6.

[49] The index captures 6 items: "A bank employee greeted me," "A bank employee asked for my name," "A bank employee told me their name," "A bank employee asked me to take a seat," "A bank employee asked how they could assist me," "When concluding your visit, were you thanked for coming in?"

[50] The index includes 16 items: years in business, number of employees, type of business, role in business, financial institution (FI) for business banking, business indebtedness/expenses, business assets/revenue, business website, current banking customer of this FI, tax returns, education, employment history, credit score, other sources of income, marital status, homeownership status.

4. Statistical Analyses and Results

This section presents and discusses statistical analyses that test the research hypothesis. Our statistical analyses, several of which were planned before viewing any data, are separated into three main sections. First, we provide a brief descriptive overview of all test visits. Second, we report our primary statistical analyses for the four domains we studied. Third, we analyze all hypothesis-relevant outcomes using an aggregate analysis with minimal assumptions. Throughout these sections, we provide illustrative examples of individual test pair cases that likely reflect the kinds of differential treatment evidenced—but not descriptively conveyed—by statistical analyses.

4.1 Descriptive Overview of Test Visits

One hundred bank branch visits were conducted over several months in 2023. Visits were conducted by 18 distinct testers (nine male and nine female), who posed as small business owners seeking credit across multiple tests at different banking institutions. Each tester conducted between one and thirteen tests. Specific test pair combinations could vary during the project but tended to remain consistent.[51] No business profile was used more than once at any one financial institution, and an individual tester never visited more than one branch of the same financial institution. This resulted in 50 complete matched-pair tests where one Black tester and one white tester visited a single bank branch (25 tests in Nassau County, NY and 25 in Fairfax County, VA).[52] Table 5 summarizes basic characteristics of the full sample of 100 test visits. On average, test visits lasted just over half an hour (32 minutes). Most testers (78 percent) waited under 5 minutes at the branch before speaking with someone. In about half of the tests (56 percent), paired testers were each able to meet with the same bank representative at a specific branch. This was determined by reviewing the tester surveys and any other records (see footnote 54).

[51] Eleven testers were always paired with the same counterpart, six testers were paired with two unique counterparts throughout the project, and one tester conducted tests with three unique counterparts.
[52] In four visits, the tester was unable to meet with a bank representative. This revises our total sample size down to 96 visits for some, but not all, outcome measures.

TABLE 5. OVERALL SUMMARY OF TEST VISITS

Variable	Observations	Average/Percent
Test Descriptives		
Test visit length (min) [7, 135]	100 visits	32.40
Average number of tests by each tester [1, 13]	18 testers	7.00
Sex of tester (1=female,0=male)	18 testers	50%
Able to meet with a bank representative to discuss credit needs? (1=yes, 0=no)	100 visits	96%
Description of bank representative[53]		
Race and National Origin		
White	72 responses	18%
Black	72 responses	17%
Other	72 responses	56%
Unsure	72 responses	9%
Title		
Loan Officer	72 responses	2%
Branch Manager	72 responses	22%
Banker	72 responses	30%
Other	72 responses	35%
Met with same employee as paired tester?[54]	100 visits	56%
Products Discussed During Visits		
Product discussed? (1=yes, 0 = no/don't know)		
Small business loan	96 visits	68%
Small business line of credit	96 visits	93%
Business credit card	94 visits	39%
Personal loan	93 visits	3%
HELOC	94 visits	6%
Other LOC	94 visits	4%
Personal credit card	94 visits	4%
Alternative funding	93 visits	4%
Other	92 visits	16%

Note. Brackets indicate minimum and maximum values for the relevant variable. "Observations" provide the denominator used to compute percentages in the third column. "Observation" numbers change across variables because in some cases testers did not provide this information (e.g., reporting details about bank representative), and in other cases the test ended before reaching this topic (e.g., discussing specific products). Percentages do not always sum to 100% within categories because some testers did not answer these questions while other provided multiple responses.

[53] Bank representatives' race, national origin, and title are constructed based on open-ended survey responses completed by testers. Race and national origin are based on subjective interpretations (e.g., "Caucasian", "perhaps Hispanic? unsure"), which limit the accuracy of this measure. Employer title is organized by core function (i.e., "Vice President Branch Manager" is classified as "Branch Manager").

[54] A matched test pair was identified as having met with the same employee based on an assessment of the tester's open-ended survey description of the bank representative. A representative was identified to be the same for both tests if they had the same first name (last names were not captured) and gender and had similar descriptions for race, national origin, and title.

4.2 Primary Statistical Analyses

4.2.1 Regression Framework

To test our research hypothesis, our analyses compare each outcome (see Table 4) across Black and white testers in a fixed-effects Ordinary Least Squares (OLS) regression framework. Equation (1) shows our model specification for estimating the difference in average outcomes for Black relative to white testers, where variation between test pairs is held constant (i.e., each pair is treated as a fixed effect):

$$(1) \qquad Y_{ij} = \beta_0 + \beta_1 Black_{ij} + \delta_i + \varepsilon_{ij}$$

Y_{ij} denotes the outcome of interest for the test visit indexed by test i (1 through 50) and tester j (1,2), *"Black,"* is an indicator for tester race in that visit and is equal to 1 if the tester is Black and 0 if the tester is white, and δ_i are fixed effects for each of the 50 test pairs in the study. The coefficient of interest, β_1, isolates the average effect of being Black relative to white on the outcome Y from other attributes of tester and business profile characteristics that are matched by design within pairs—in other words, estimating the predictive effect of race while controlling for all other relevant factors. β_1 is statistically significant if its probability of occurrence under a null hypothesis of no difference is less than .05 (reported as $p < 0.05$).[55] β_0 is a standard intercept term and ε_{ij} is the error term for the model.

4.2.2 Summary of Findings

We find statistically significant evidence that Black testers were treated less favorably relative to paired white testers on some—but not all—primary outcomes. Table 6 presents a descriptive summary of results in each domain, and we summarize the results on both objective (index) and subjective (rating) outcomes below.[56]

- We find that lenders provided more favorable levels of objective encouragement/discouragement to apply for credit to white testers as compared to paired Black testers, although both Black and white testers were encouraged to apply, on average.
- White testers also subjectively reported feeling more encouraged to apply than Black testers, although both Black and white testers reported feeling encouraged, on average.
- Bank representatives were similarly likely to discuss the requested products with both white and Black testers. However, these bank representatives were more likely to discuss non-requested products with Black testers.

[55] As specified in our analysis plan and recommended in past literature (Lubin 2010; Lakens, 2022), we compute *p*-values and infer statistical significance using one-tailed statistical tests with the directional hypothesis that white testers will receive more favorable outcomes than Black testers. Lakens, D., (2022). Improving Your Statistical Inferences. Available at https://zenodo.org/records/6409077

[56] Refer to section 3.4 for our characterization of objective vs. subjective measures.

- White testers subjectively reported higher levels of customer service than Black testers, but there were no statistically significant differences in the objective index measure of observed customer service behaviors (e.g., whether testers were greeted or thanked).
- We did not observe statistically significant differences between Black and white testers on the amount of information requested during the visit or on rates of follow-up contact after the test visit occurred.

TABLE 6. OVERVIEW OF RESULTS BY DOMAIN

Domain	Statistical Significance Supporting Hypothesis	Effect Characterization
1. Encouragement/ discouragement	Yes: both on objective index and subjective rating of encouragement/discouragement	Overall, most testers felt encouraged, and received favorable levels of encouragement/discouragement, but white testers felt more encouraged and received more favorable levels of encouragement/discouragement than Black testers.
2. Products and potential steering	Yes: in objective likelihood of discussing non-requested loan products	Bank representatives tended to discuss requested loan products with all testers, but bank representatives discussed non-requested products in 40 percent of visits with a white tester and 58 percent of visits with a Black tester.
3. Customer service	Mixed: significant on testers' subjective ratings of customer service, but not on objective index of politeness behaviors	No significant difference observed in objective index of politeness behaviors. Subjective customer service ratings tended to be high overall, but white tester ratings were significantly higher than Black tester ratings
4. Business and credit information requested	No	No effect observed
Post-visit follow-up	No	No effect observed

Table 7 quantifies these results. Each row details results for a given outcome, including measurement scale and sample size (*N*); outcome measure averages for white and Black testers; and the regression coefficient (see Equation 1). Each row also displays whether the coefficient was statistically significant at the 5 percent level, indicated by the associated *p*-value.[57] We proceed by describing and discussing these results for each outcome.

[57] *p*-values are computed on each regression coefficient of interest and represent the probability of observing a given coefficient or one larger under a null hypothesis that there is no difference between groups. Smaller *p*-values convey greater levels of statistical significance but do not necessarily correspond to larger effect sizes (i.e., average differences between groups).

TABLE 7. PRIMARY OUTCOME MEASURES BY RACE AND RACE DISPARITY ESTIMATES

					Outcome by race		Race disparity estimate (1=Black)	
Domain	**Outcome**	**Measure type**	**Scale**	**Sample size (*N*)**	**White**	**Black**	**Disparity estimate (scale units)**	***p*-value of estimate**
Encouragement/ discouragement	Encouragement/ discouragement index	Objective	0 to 16	100	9.56	8.80	-0.76 behaviors	0.03*
Encouragement/ discouragement	Encouragement/ discouragement rating	Subjective	1 to 7	94	5.79	5.19	-0.55 rating points	0.04*
Products & Potential Steering	Discussed a requested product (BL or BLOC)	Objective	Percent likelihood	94	98%	94%	-4 percentage points	0.16
Products & Potential Steering	Discussed a non-requested product (Other than BL, BLOC)	Objective	Percent likelihood	94	39%	59%	19-20 percentage points^	0.02*
Customer Service	Customer service behaviors index	Objective	0 to 6	100	4.30	4.32	0.02 behaviors	0.47
Customer Service	Customer service rating	Subjective	2 to 8	94	6.79	6.31	-0.4 rating points	0.08
Customer Service	Bank representative rating	Subjective	1 to 7	94	6.17	5.66	-0.49 rating points	0.03*
Business and credit info requested	Business and credit info requested index	Objective	0 to 16	94	4.81	4.75	-0.09 pieces of info	0.43
Business and credit info requested	Business and credit info requested rating	Subjective	1 to 4	94	2.73	2.65	-0.11 rating points	0.23
Post-visit follow-up *(non-survey)*	Post-visit follow-up	Objective	Percent likelihood	94	27%	33%	6 percentage points	0.19

Note. * denotes statistical significance at the five percent level. ^ denotes a disparity estimate presented as a range due to rounding error in the individual estimates.

4.2.3 Encouragement/Discouragement

Did white testers receive more favorable levels of encouragement/discouragement to apply for a loan than otherwise-matched Black testers? We found statistically significant race-based differences in this domain on our 16-point index of potentially encouraging or discouraging behaviors ($\mu_{white} = 9.56$,

$\mu_{black} = 8.80, \beta_1 = -0.76, p = 0.03$)[58], indicating that Black testers received worse objective treatment than white testers. This pattern also emerged when analyzing the 7-point, subjective rating measure of encouragement/discouragement ($\mu_{white} = 5.79,\ \mu_{black} = 5.19, \beta_1 = -0.55, p = 0.04$) where—consistent with the objective measure—Black testers reported *feeling* less encouragement than white testers. It is important to qualify this pattern by noting that most testers experienced and reported favorable levels of encouragement/discouragement overall. On average, both the objective index and subjective rating measures were above the midpoint on these measures, possibly reflecting a general pattern of encouragement to apply. However, the finding that white testers tended to receive more favorable treatment than Black testers on this measure supports our hypothesis.

> **For example:** In one paired test, the Black tester was told his business was too small to qualify for any of the bank's small business loan products. The paired white tester, with a business of the same size, met with the same bank representative at the same banking office and was encouraged to apply for a small business line of credit at the bank for the amount of 10 percent of his business revenues.

Statistical significance alone does not convey the practical magnitude of differences between groups or variability in individual outcomes. How consistent and how large are differences in encouragement within individual test pairs? We provide insight into this question for each domain by plotting individual outcomes within each test pair (see Figure 1 for encouragement/discouragement).

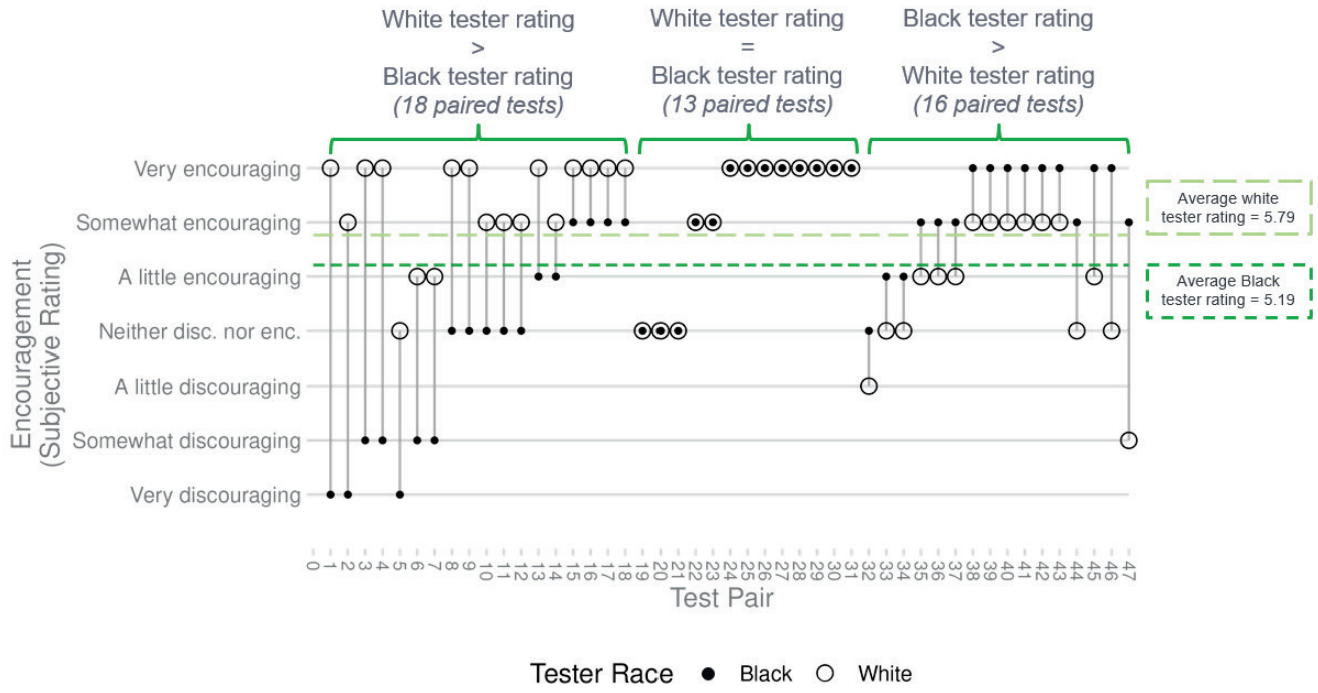
Beginning with the subjective encouragement/discouragement ratings, Figure 1 plots each test pair along the horizontal axis and shows how encouraged/discouraged each tester reported feeling along the vertical axis. Test pairs are sorted from left to right in order of difference in this rating: the lefthand side of the figure shows the 18 tests where the white tester reported a higher level of encouragement than the paired Black tester, the middle shows the 13 tests where both testers reported the same level of encouragement, and the righthand side shows the 16 tests where the Black tester reported a higher level of encouragement than the paired white tester. The orange and green dashed lines show the average rating across all white (M = 5.79) and Black (M = 5.19) testers. The typical rating for both racial groups is between "a little encouraging" and "somewhat encouraging," supporting the earlier characterization that explicit feelings of *dis*couragement were rare—despite a hypothesis-consistent difference on average.

Plotting outcomes in this way reveals insights beyond statistical significance in the differences between group averages. This heterogeneity in test pairs' experiences suggests that white testers did not consistently feel greater encouragement than paired Black testers, but that differences in this outcome

[58] μ_{white} and μ_{black} denote the population mean estimated by the regression model for white and Black testers, respectively.

tended to be larger when favoring white testers over Black (i.e., the lefthand side of Figure 1) than when Black testers were favored over white (i.e., the righthand side of Figure 1).

FIGURE 1. SUBJECTIVE ENCOURAGEMENT/DISCOURAGEMENT RATING

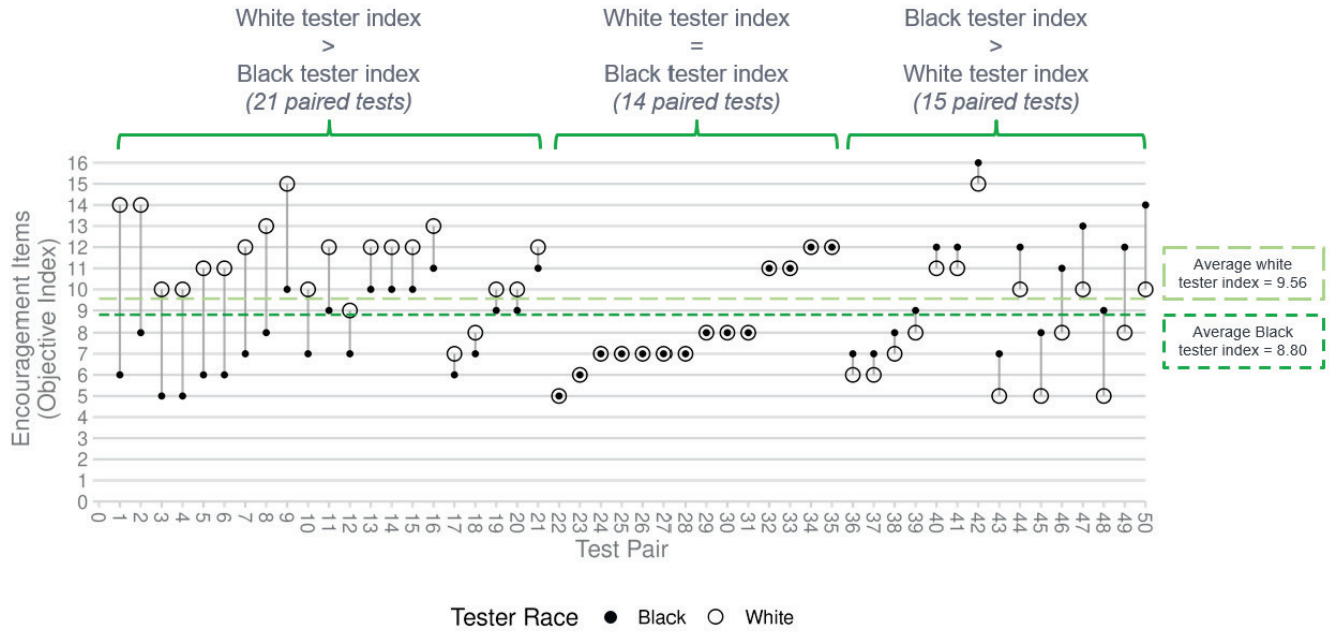


Note. This figure plots subjective encouragement/discouragement ratings responding to the prompt: "Considering your visit from start to finish, do you feel like the bank representative was **discouraging, encouraging**, or **neither**, toward your potential application for a small business loan or line of credit at this institution?" Test pair numbers are ordered by difference sizes (largest differences favoring the white tester to smallest) and change with each plot. Testers provided incomplete ratings or the test ended before these ratings could be provided in three tests, resulting in a total sample of 47 tests.

> **For example:** In one test, the white tester was told he qualified for a loan with the bank by a representative who provided encouragement to apply. The Black tester, with a similar business and credit profile, met with the same representative, who told him he did not qualify and suggested he go to a Small Business Administration loan office instead.

Figure 2 displays the objective encouragement/discouragement index using a similar plot. Like the subjective rating, most scores on this measure were above the scale midpoint, reflecting generally favorable levels of encouragement/discouragement. Once again, however, white testers tend to receive higher scores than paired Black testers (21 tests showing this vs. 15 tests reflecting the opposite pattern), and differences favoring white over Black testers tended to be larger than differences favoring Black over white testers.

FIGURE 2. OBJECTIVE ENCOURAGEMENT/DISCOURAGEMENT BEHAVIORS INDEX



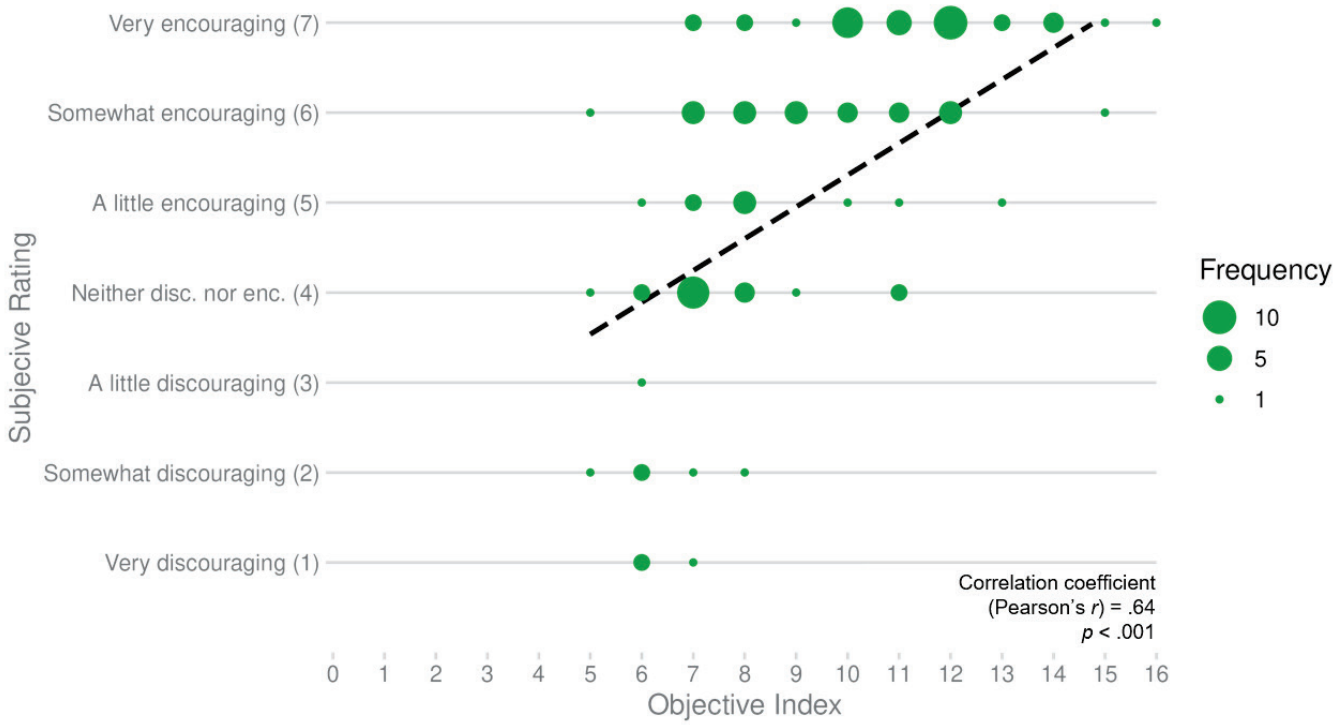
Note. This figure plots objective encouragement/discouragement index scores for all testers, with higher scores indicating more favorable treatment in this domain. The index combines 16 items with the prompt: "During your visit, did the bank representative do any of the following?" *[16 possible actions: yes = 1, no or don't know = 0]*. Test pair numbers are ordered by difference sizes (largest differences favoring the white tester to smallest) and change with each plot.

The subjective and objective outcome measures produce similar results, reflecting a qualitative degree of concordance between them. To quantify this relationship, we calculated the correlation between each tester's score on these two outcomes. The resulting coefficient is positive, strong, and statistically significant ($r = 0.64$, $p < 0.01$), confirming a high degree of correspondence between testers' objective experiences and subjective feelings or beliefs. Figure 3 displays this relationship. Each point on the figure displays test visits that resulted in a specific combination of the objective encouragement/discouragement index and the subjective encouragement/discouragement rating. For example, the point in the top right of the figure shows that one tester received the highest level on the objective index (16/16) and provided the highest subjective rating 7/7, or "Very encouraging." Larger points indicate more observations of a given objective index score and subjective rating.

This relationship indicates that for each additional potentially encouraging behavior (and absence of each potentially discouraging behavior) out of 16, the average score on the 7-point encouragement/discouragement rating scale increases by nearly half a point (0.41). This finding strengthens our conclusions in the encouragement/discouragement domain and increases their validity. First, it suggests that the subjective rating measure is consistent with the objective index measure, which was constructed using observable behaviors like those studied in prior research (e.g. Bone et al.,

2019). Second, it rules out the possible conclusions that testers felt discouraged but in fact received equivalent treatment, or that testers were objectively discouraged from applying but failed to notice.[59]

FIGURE 3. ENCOURAGEMENT/DISCOURAGEMENT: RELATIONSHIP BETWEEN OBJECTIVE INDEX AND SUBJECTIVE RATING MEASURES



Note. This figure displays a scatterplot of encouragement/discouragement index scores and subjective ratings provided by all testers. The size of each point is weighted by the frequency of test visits; larger points represent more test visits. The line of best fit is generated from a linear model predicting subjective ratings from objective index scores (with no other variables entered into the prediction). The correlation coefficient (Pearson's *r*) for the relationship between these measures is .64, which is statistically significant at the .1 percent level.

4.2.4 Products and Potential Steering

The second domain where we find reliable statistical differences relates to products and potential steering. Nearly all bank representatives discussed at least one requested business credit product with the tester they met with, with no statistically significant difference between white and Black testers ($\mu_{white} = 98\%,\ \mu_{black} = 94\%, \beta_1 = -4$ pct. pts., $p = 0.16$). This finding confirms that nearly all testers

[59] Although there was a high degree of correspondence between reported (subjective) and observed (objective) encouragement in this study, a similar relationship should not necessarily be expected in future studies or tests. This is because credit shopping environments—and testers' experiences—are likely to vary with contextual factors.

discussed the products they were trained to request. Importantly, however, bank representatives were significantly more likely to discuss at least one non-requested product with Black testers than with white testers ($\mu_{white} = 40\%,\ \mu_{black} = 58\%, \beta_1 = 19$ pct. pts., $p = 0.02$).[60] This difference is large: our analysis indicates that the bank representatives were 19 percentage points more likely to discuss a non-requested product with a Black tester (58.5 percent) than an otherwise equivalent white tester (39.4 percent). Figure 4 visualizes this finding by plotting the two product discussion outcomes (light green bars for Black testers, dark green bars for white testers; requested products on the left of the plot, non-requested products on the right).[61]

> **For example:** In one test, the bank representative directed the Black tester toward a home equity line of credit product but recommended that the white tester apply for a business line of credit. Both the white and Black tester requested business credit. A home equity credit product is not business purpose credit and can place the owner's personal residence at risk if the loan should default.

To summarize, we found that while bank representatives were similarly likely to discuss a requested product with Black and white testers, they were substantially more likely to discuss a non-requested product with Black testers, potentially directing or steering them toward different or alternative credit products.

This finding is related to results reported by Scott et al. (2023) on the likelihood that Black and white testers received recommendations for a product judged to be more favorable (BLOC) or less favorable (HELOC). They found that bank representatives were more likely to recommend a BLOC to white testers but were no more or less likely to recommend a HELOC to either tester group.[62] Both results suggest practices that may be associated with less favorable outcomes for Black testers.[63] We note, however, that we measured only whether products were discussed, not the quality (e.g., depth, clarity, or length) of these discussions. Additionally, because testers were instructed to explicitly request information about specific credit products during their visits, it remains unclear if differences might emerge in product discussions in the absence of these instructions. Future research

[60] As described earlier, we define "requested" products as those that testers were trained to explicitly inquire about during their visit: a business loan (BL) or business line of credit (BLOC). "Non-requested" products here are defined as all other financial products that could be raised during the visit. These include, for example, HELOCs and personal credit cards, which often carry less favorable terms.

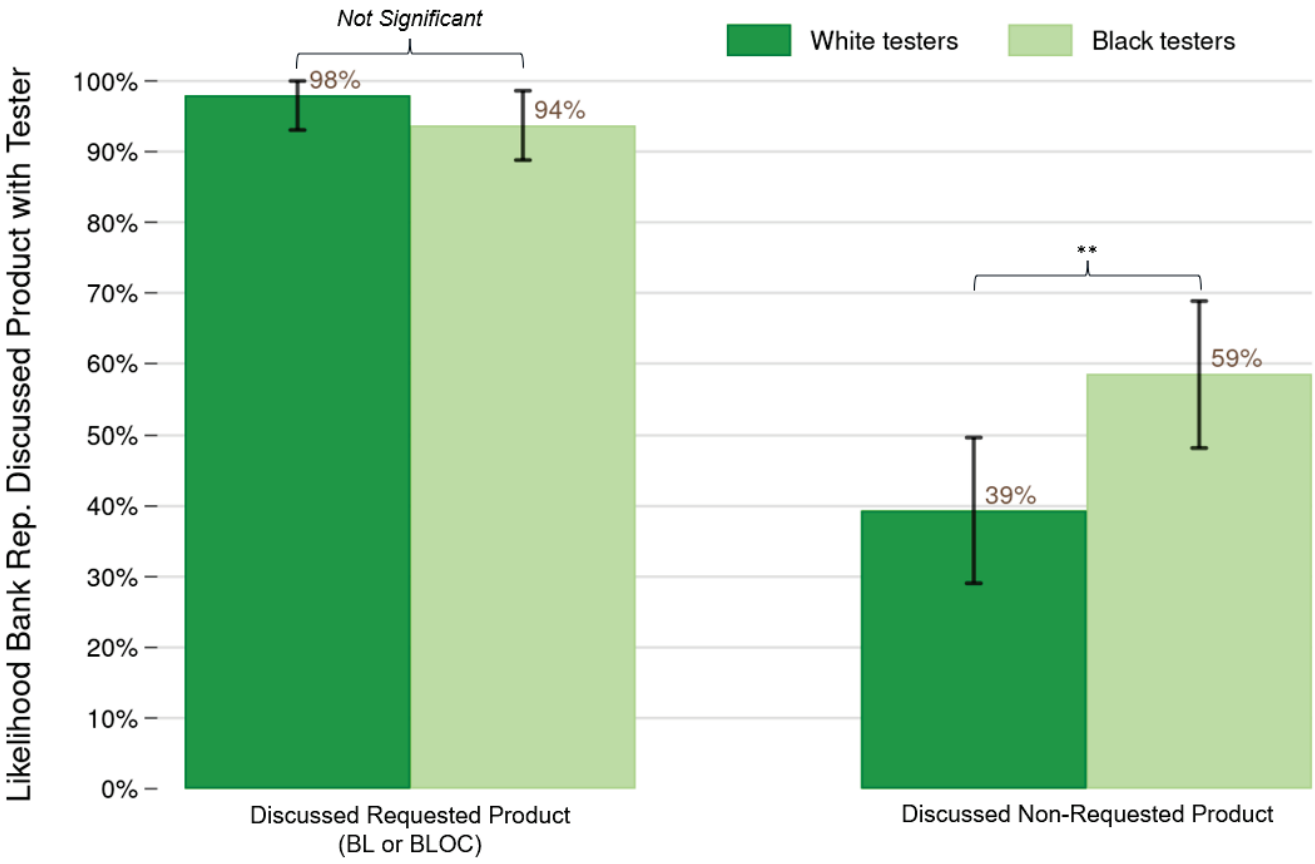
[61] These outcomes cannot be plotted using the same style as Figures 1 and 2 because encouragement/discouragement was measured on a continuous scale with multiple levels between zero and the scale maximum, whereas "products discussed" was measured as a binary outcome with only two levels ("discussed" or "not discussed.") A third option, "don't remember," was excluded from analysis. This option was selected for all requested or non-requested products in four test visits and its exclusion does not affect the overall finding or conclusions.

[62] The tests in Scott et al. (2023), referenced in footnote 1, were conducted in a lower-rate environment than the present study and a different region (Atlanta metropolitan area), which may have produced different tester experiences.

[63] Our analysis focuses on all categories of products discussed rather than on specific products recommended. While the survey also captured products that bank representatives recommended, testers found it difficult to answer whether a specific product was explicitly recommended, and more than one third of testers reported that no recommendation was made. Among testers who received explicit product recommendations, the patterns are directionally consistent with those for products discussed.

may consider evaluating and testing hypotheses related to differences in products discussed and the quality of these discussions.

FIGURE 4. LIKELIHOOD OF DISCUSSING REQUESTED AND NON-REQUESTED SMALL BUSINESS CREDIT PRODUCTS, BY RACE



Note. This figure plots results from the regression analysis that indicate how likely bank representatives were to discuss requested products (left two bars) and non-requested products (right two bars) with Black (light green bars) and white (dark green bars) testers. These likelihoods were obtained from the regression specification in Equation 1 and control for between-pair variation (i.e., test pair fixed effects). Raw (i.e., non-regression-adjusted) percentages produce similar results (each regression-adjusted likelihood is within 1 percentage point of the percentage obtained in the raw data). Error bars display 90% Confidence Intervals. Stars denote statistical significance from one-sided tests: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, ns (not significant) $p > 0.10$.

4.2.5 Customer Service

Turning to the customer service domain, we find a mix of some statistically significant differences and some instances of no (or only a weak) difference. The strongest findings come from the bank representative rating, which measured testers' subjective ratings of bank representatives along several dimensions (friendliness, helpfulness, informativeness, patience, and professionalism, each on a 7-point scale with two poles; see footnote 47) and averaged these ratings together. Bank representatives

were rated significantly higher by white testers than by Black testers on this measure ($\mu_{white} = 6.17$, $\mu_{black} = 5.66, \beta_1 = -0.49, p = 0.03$), although these ratings tended to be positive (i.e., above the scale midpoint) regardless of tester race.

The subjective customer service visit rating (the sum of two 1-4 point scales of how valued the tester felt and the level of customer service the tester reported receiving, resulting in a 2-8 point scale) were also generally high across testers of both races, but were modestly higher for white than for Black testers ($\mu_{white} = 6.79$, $\mu_{black} = 6.31, \beta_1 = -0.40, p = 0.08$).[64]

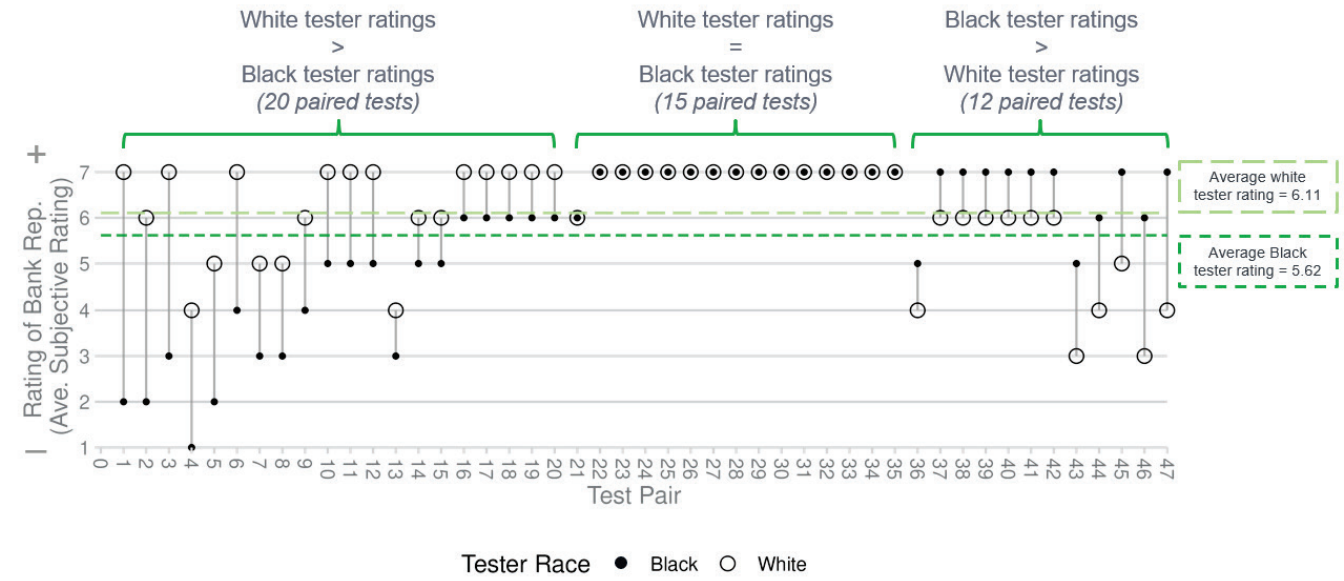
Finally, there was no significant difference on the 6-item objective customer service behaviors index ($\mu_{white} = 4.30,\ \mu_{black} = 4.32, \beta_1 = 0.02, p = 0.47$). One speculative reason is that this outcome measure captured polite behaviors when arriving at and leaving the branch (e.g., "a bank employee greeted me," "thanked for coming in") but did not capture more substantive aspects of the interaction that are likely better captured by the two subjective ratings (e.g., bank representative friendliness or conversational tone).

For example: In one test, the Black tester was not greeted or acknowledged upon entering the branch office and had difficulty finding an employee who could discuss his needs. The bank representative he met with also told him that he couldn't answer many of his questions. The white tester was greeted upon entry and waited only briefly to meet with the same bank representative. The representative brought in their manager to talk with the white tester, who offered to assist the tester in starting a loan application, sent the tester's information to a small business loan specialist, and told the tester to call the manager the next day if he did not hear from the specialist. On the follow-up survey the white tester gave a customer service rating of 7 out of 8, while the Black tester gave a rating of 2 out of 8.

Figures 5-7 present paired plots for the subjective bank representative ratings, the subjective customer service visit ratings, and the objective customer service behaviors index. These plots convey the positive experiences encountered by most testers (i.e., most points are above the scale midpoint for all three measures). For the two outcomes where statistically significant or suggestive race differences emerged (subjective bank representative ratings, Figure 5; subjective customer service visit rating, Figure 6), these effects are driven by a few tests with large differences and several additional tests with generally small differences that tend to favor white testers.

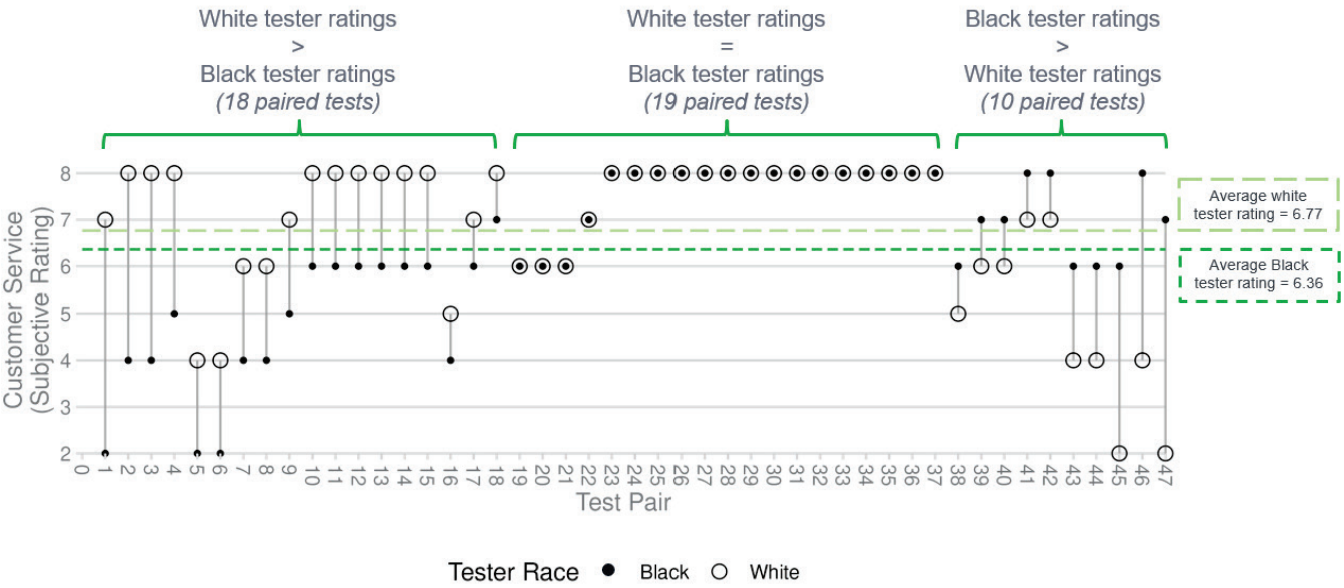
[64] Although this difference is in the hypothesized direction, it does not meet our criterion of $p < .05$ for statistical significance. We consider this result "suggestive," in that it may be indicative of support for the hypothesis, but the evidence for this conclusion is weak.

FIGURE 5. SUBJECTIVE BANK REPRESENTATIVE RATINGS



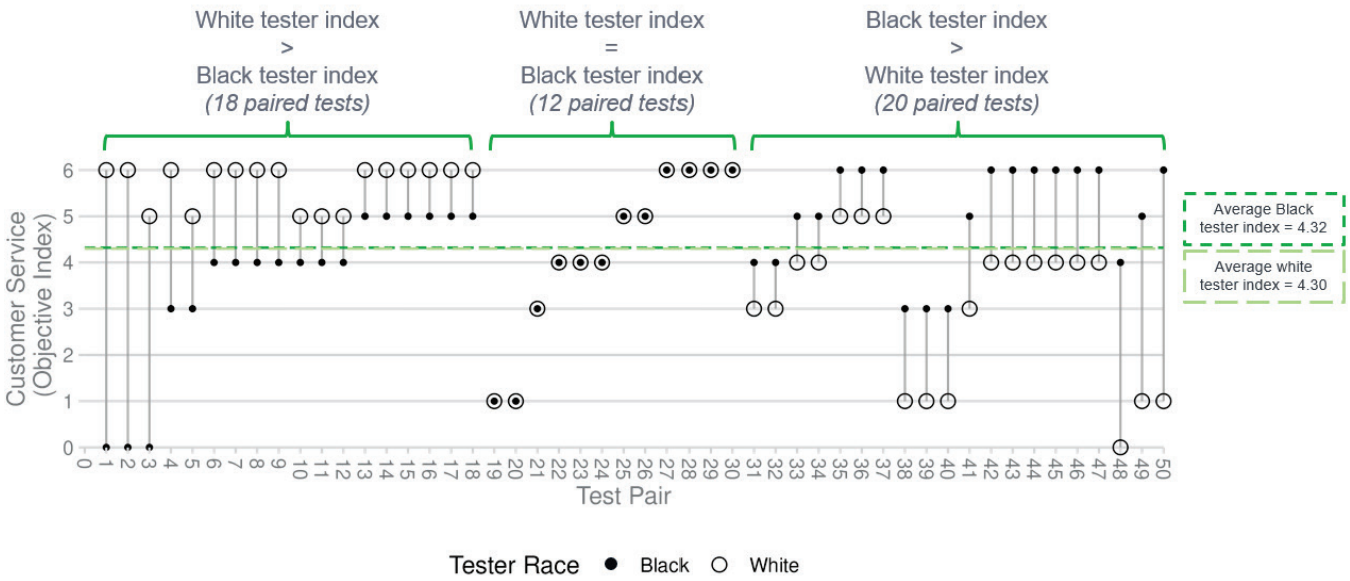
Note. This figure plots each tester's average subjective rating of the bank representative. Testers were asked: "Please rate the bank representative you met with on the following characteristics" Unfriendly/Friendly, Unhelpful/Helpful, Unprofessional/Professional, Impatient/Patient, Uninformative/Informative. *[7 possible responses to each: Negative (1), Neutral (4), to Positive (7)]*. Testers provided incomplete ratings or the test ended before these ratings could be provided in three tests, resulting in a total sample of 47 tests for this measure.

FIGURE 6. SUBJECTIVE CUSTOMER SERVICE VISIT RATINGS



Note. This figure plots subjective ratings of customer service. A single measure ranging from 2-8 was derived by adding testers' responses to two questions: "How high quality do you feel the customer service was?" and "How valued did you feel as a potential customer?" *[4 possible responses to each: not at all (1) to very (4)]*. Testers provided incomplete ratings or the test ended before these ratings could be provided in three tests, resulting in a total sample of 47 tests for this measure.

FIGURE 7. OBJECTIVE CUSTOMER SERVICE BEHAVIORS INDEX



Note. This figure plots objective customer service index scores capturing 6 items. Testers were asked: "What interactions took place in the first two minutes after you entered the branch?" (And: "were you thanked for coming in?") *[6 possible behaviors: yes = 1, no or did not notice = 0]*

4.2.6 Information Requested of Tester

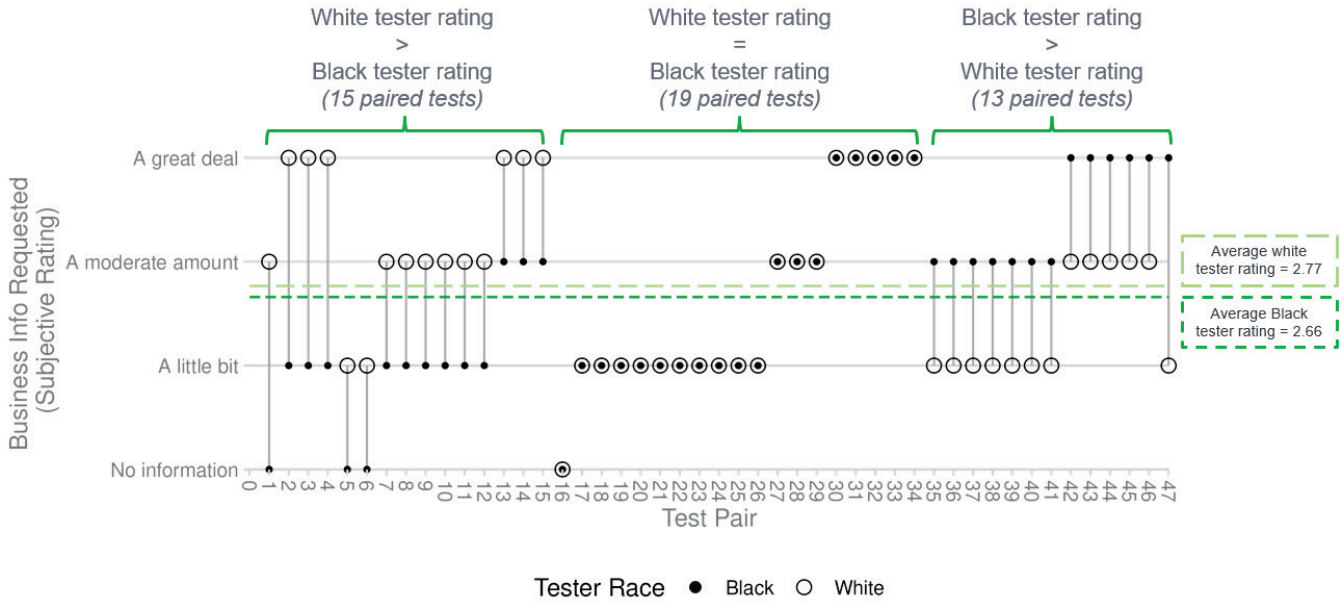
We found no statistically significant differences by race in the domain related to the amount of business and credit information requested of the tester. There was no reliable difference for the 16-item objective business and credit information requested index ($\mu_{white} = 4.81,\ \mu_{black} = 4.75, \beta_1 = -0.09, p = 0.43$) or for the subjective business and credit information requested rating, which captured testers' subjective ratings of the amount of business information discussed on a 4-point scale ($\mu_{white} = 2.73,\ \mu_{black} = 2.65, \beta_1 = -0.11, p = 0.23$).[65] Figure 8 presents the paired plot of the subjective rating and Figure 9 presents the analogous plot for the objective index (see footnote 50 for the specific behaviors). These findings suggest, counter to our hypothesis, that there did not appear to be statistically significant differences favoring white testers over Black testers (or vice versa) across our sample for this domain.[66]

The absence of a statistically significant finding in this domain does not mean that testers received similar treatment on all tests. For example, an exploratory analysis did reveal a statistically significant difference where Black testers were more frequently asked to provide W2 forms than white testers (see Appendix A, Appendix Table 1). Although this was a large difference in our sample, it was not enough to

[65] As previously discussed, we would not always expect correspondence between objective and subjective outcomes in future research.
[66] However, this does not suggest that testers always received similar treatment on all tests.

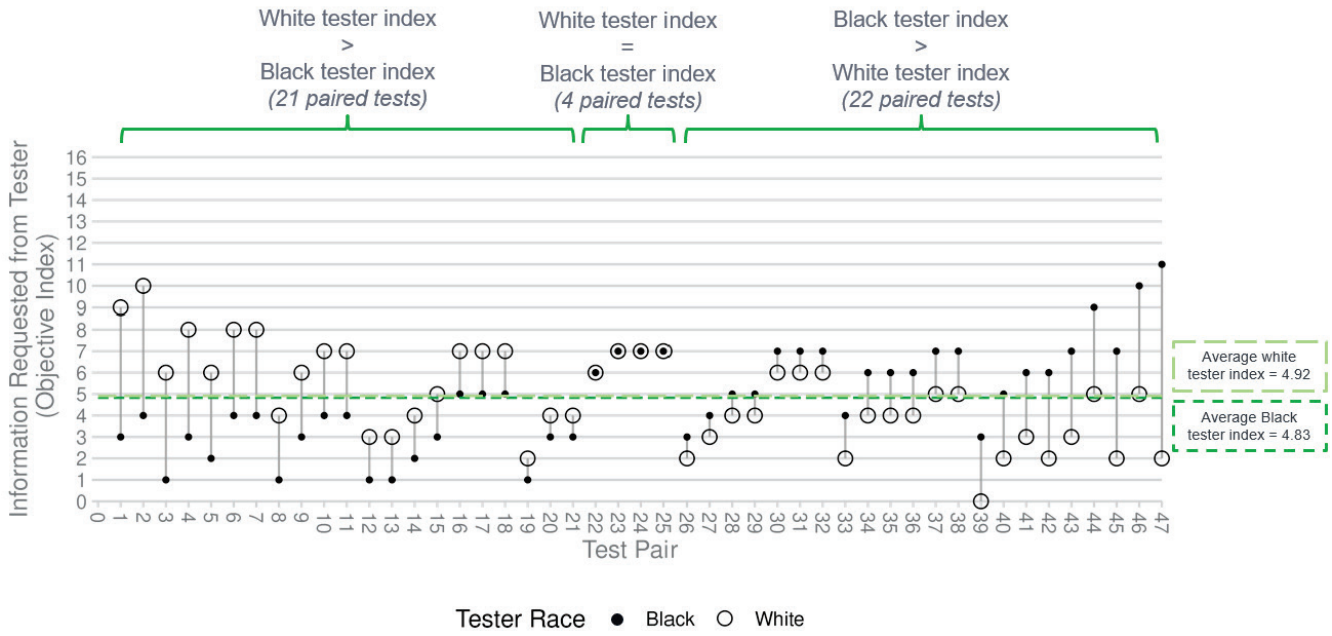
drive any overall effect of information requested on the objective index measure. Additionally, specific within-pair observations can reveal examples of meaningful, differential treatment on an individual (i.e., qualitative) level. In one test, a single bank representative told the Black tester, but not the similarly qualified white tester, that the representative was required to ask about whether the tester had tax liens, backpay, or late child support payments. Tests at two other branches had opposite outcomes: in one test the Black tester was provided more detailed information than the paired white tester, and in the other test the white tester was provided more information than the paired Black tester. In both examples the testers met with the same bank representative, but in the test where the representative provided the Black tester with more information, the representative also told the Black tester that he needed a preexisting relationship to pursue a business loan. This barrier was not communicated to the white tester.

FIGURE 8. SUBJECTIVE BUSINESS AND CREDIT INFORMATION REQUESTED RATING



Note. This figure plots subjective ratings of the amount of business information requested from testers. Testers were asked: "Think back to your full conversation with the bank representative. Which of the following best describes how much information you discussed related to your business and business finances?" *[1 (None) to 4 (A great deal)].* Testers provided incomplete ratings or the test ended before these ratings could be provided in three tests, resulting in a total sample of 47 tests for this measure.

FIGURE 9. OBJECTIVE BUSINESS AND CREDIT INFORMATION REQUESTED INDEX



Note. This figure plots the objective index of information requested from tester. Testers were asked: "Please indicate what information the bank representative requested about you and your household" *[16 possible pieces of info: info requested = 1, info not exchanged or don't know = 0].* Testers provided incomplete ratings or the test ended before these ratings could be provided in three tests, resulting in a total sample of 47 tests for this measure.

How do we reconcile the absence of a statistically significant (i.e., average) difference in this domain with the observation that testers were treated differently in several specific matched-pair tests? Informed in part by a 2019 case study of matched-pair testing methodology in the small business lending context,[67] we hypothesized that greater information requested would reflect greater burden on potential applicants. However, the type and amount of information requested by lenders early in the credit process likely depends on many unobserved factors including local contexts and broader economic trends. It is not clear whether this absence of an average difference would replicate in a future study. More research is necessary to improve the interpretation of results in this domain—for example, to distinguish when information requests are facilitative and expedient or when they impose unnecessary burdens that potentially discourage the borrower. Future research should also consider developing protocols to measure requests for unnecessary information, or information that is not pertinent to business financing.

4.2.7 Non-Survey Outcome: Post-Visit Follow-Up

As shown in the last row of Table 7, there were no statistically significant differences by race in the likelihood that testers received a post-visit follow-up email or phone call from a bank representative. Overall, a little under one third of test visits resulted in a follow-up email or call ($\mu_{white} = 0.27,\ \mu_{black} = 0.33, \beta_1 = 0.06, p = 0.19$).

4.2.8 Exploratory Regression Analyses on All Individual Survey Outcomes

The analyses presented above were designed to prioritize informativeness while minimizing the number of statistical tests by combining similar measures. Our approach accomplished this by combining related, hypothesis-relevant survey measures into ten total outcomes across four domains. In addition to analyzing these index and rating measures, we also conducted exploratory analyses to look for differences between white and Black testers on each individual survey measure. These analyses offer descriptive evidence on the underlying items used to construct our main measures, but they include a substantial limiting caveat: repeating the same regression analysis on more than 70 individual measures will inflate the likelihood of falsely concluding that a difference exists on at least one of these measures.[68]

[67] Bone et al., (2019). See footnote 30.
[68] When conducting multiple statistical tests (i.e., running the same regression analysis on many different variables), some results will achieve statistical significance by random chance alone. It is generally considered a statistical best practice to run only as many tests as necessary to answer a research question, which can minimize "false-positive" findings. We prioritized this in our approach to analyzing combined index and rating measures.

We present these exploratory analyses in Appendix A. Table A1 presents an exploratory analysis for each hypothesis-relevant individual measure captured by the survey form. Each row shows results for the individual survey measure, including the sample size for each measure, the average of each measure for Black and white testers, and each measure's regression coefficient or "disparity estimate," and this estimate's associated *p*-value, which conveys whether each difference was statistically significant.

4.3 Aggregate Analysis of All Relevant Measures

The analyses presented above are designed to capture reliable (i.e., statistically significant) race-based differences in tester outcomes in specific domains of treatment. However, these regression analyses have limitations. We describe an important limitation below and then present an aggregate analysis of all relevant outcome measures as a robustness check on our overall conclusions.

An important limitation of the regression framework is that our final sample size provides statistical power to detect only sizable differences in treatment across the test sample (i.e., 15-20 percentage point differences between paired Black and white testers). We designed our research approach around these effect sizes to balance expectations based on prior research (Bone et al. 2019) with available resources for testing. The limitation of this approach is that standard regression analyses may not achieve statistical significance when applied to outcome measures with small differences between testers; these analyses may "miss" small differences that are nevertheless real, important, and reflect a persistent pattern of unfavorable treatment that would otherwise not be detected.[69] The only way to reliably detect these smaller differences on a single outcome measure, if they exist, would be to increase sample size by conducting more tests (see, e.g., Urban Institute 2012 report prepared for HUD with 8,000 paired tests on housing discrimination).[70] In the absence of a larger sample size given resource limitations, we can reframe the research question to ask if a more general pattern emerges that considers *all* hypothesis-relevant measures we collected.

To address this limitation, we conducted an analysis of conditional differences between Black and white testers following an approach that imposes fewer assumptions than the regression framework.[71] This

[69] An outcome where a statistical test does not achieve significance even though an effect exists in the broader population is called a Type II error. This inferential error can be inflated when samples are small.
[70] Note, however, that housing markets provide a much larger possible sample of tests than other markets including, e.g., small business lending. The ideal number of tests to conduct in any market depends on contextual factors. See Turner et al., (2013) "Housing Discrimination Against Racial and Ethnic Minorities 2012" referenced in footnote 27.
[71] Such a "non-parametric" analysis does not attempt to create statistical estimates (or parameters) for the relationship between one variable and another; instead, it characterizes a pattern and returns the probability of that pattern occurring by chance. See Heckman and Siegelman (1993) referenced in footnote 29.

analysis asks: given that an individual measure reveals a difference favoring one tester race over the other, is that difference more likely to favor white or Black testers?

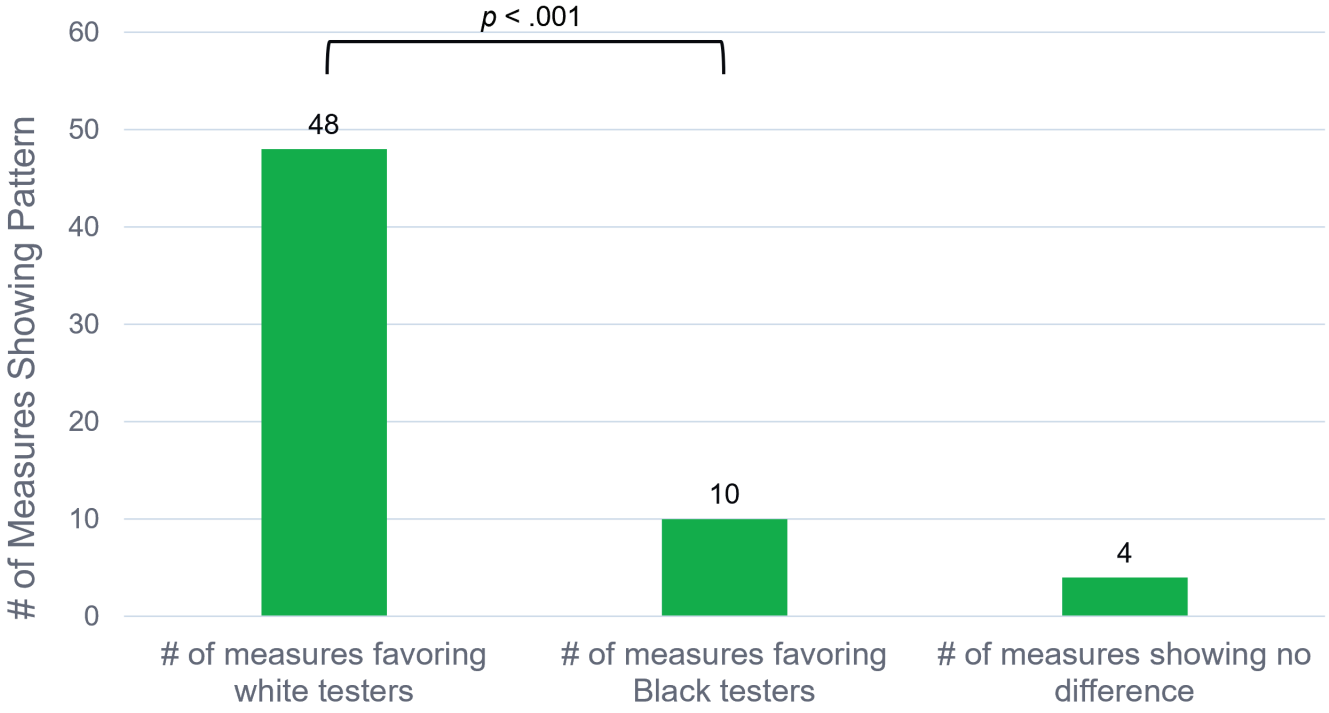
To answer this question, we retained all individual survey outcome measures that were relevant to the primary hypothesis.[72] Measures were deemed relevant if (1) they were plausibly related to differential treatment based on race and (2) a prediction could be made for which direction would reflect favorable treatment.[73] This yielded 62 individual survey measures (out of approximately 72), including both objective and subjective measures.

For each of these 62 individual survey measures, we assessed across all tests whether the measure revealed a difference favoring white testers (e.g., 40 percent of white testers vs 23 percent of Black testers were informed that the bank was interested in their loan application), a difference favoring Black testers (e.g., 88 percent of Black testers reported being greeted when they began their visit vs. 82 percent of white testers), or no difference (e.g., 94 percent of white testers and 94 percent of Black testers were thanked for coming in when concluding their visit). We then counted the number of individual measures that fell into each of these three categories. Across this collection of measures, we would conclude that a pattern of unequal treatment favoring white over Black testers exists if (1) measures reflect differences between matched-pair testers and (2) differences favoring white testers are more common than differences favoring Black testers.

Only four out of 62 measures (6.5 percent) revealed exactly no difference between Black and white testers (see Figure 10). Among the 58 measures that revealed some degree of difference, ten of these (17 percent) directionally favored Black testers over white testers. However, 48 of these measures (83 percent) reflected treatment favoring white testers over Black testers. This analysis reveals that, conditional on a difference in treatment between testers, differences in treatment were 4.8 times as likely to favor white over Black testers than the opposite on measures that are determined to be directly relevant to the research question—a highly statistically significant result ($z = 4.86$, $p < .001$).[74] Although this analysis includes both objective and subjective measures, the pattern of results is consistent across both measure types.

[72] Some measures were converted from an ordinal or interval scale to a binary indication so that differences could be observed on a consistent scale (e.g., percentage point difference between Black and white testers). For example, we converted the ordinal subjective encouragement measure (1-7 scale) to a binary scale (0 or 1) based on whether a tester reported a value greater than the median.
[73] For example, a measure of "personal information exchanged" during the visit was not retained because it was not clear whether greater personal information exchanged would reflect treatment that is favorable or unfavorable.
[74] Heckman and Siegelman (1993), referenced in footnote 29, review the approaches to conducting statistical analyses of differences in matched-pair testing environments and argue that the conditional sign (z) test is an appropriate one.

FIGURE 10. DIRECTIONAL PATTERN FOR AGGREGATE ANALYSIS OF ALL INDIVIDUAL SURVEY MEASURES



Note. $p < .001$ denotes statistical significance at the 0.1 percent level and is computed using the ratio of measures favoring white testers over Black testers to measures favoring Black testers over white testers. This analysis includes objective and subjective measures and neither the general pattern nor specific conclusions are fully attributable to one of these measure types.

We repeated this approach by classifying each survey measure as belonging to one of the same four treatment domains used for our regression analyses: encouragement/discouragement, products and potential steering, customer service, and business and credit information requested of tester.[75] Figure 11 displays the results, which are generally consistent with the primary regression analyses. Differences favoring white testers on individual measures were more likely than differences favoring Black testers for encouragement/discouragement (by a ratio of five to one, $p < .007$), products and potential steering (by a ratio of three to one, $p < .05$), and customer service (on all 10 measures, $p < .002$), but not for business and credit information requested of testers ($p < .11$).

[75] To do this, we specified which of the four domains each of the 62 individual measures *best* reflected. These specifications were mutually exclusive. Although some measures could reflect more than one domain, this approach required defining a best-fitting domain for every survey measure.

FIGURE 11. DIRECTIONAL PATTERN FOR ALL INDIVIDUAL SURVEY MEASURES WITHIN EACH DOMAIN

Note. The figure in each panel displays the result of an aggregate analysis conducted within each of the four domains studied in this research.

The analyses presented above provide supporting evidence that Black testers received less favorable treatment than white testers, agnostic to the individual measures considered. This strengthens our overall conclusions by producing similar results using an alternative analytic approach. However, these analyses are limited because they adopt a liberal criterion for assigning a "difference" between testers; measures could receive a label of "no difference" only if the proportions of Black and white testers are exactly equal. A rigorous statistical evaluation should ask whether differences in treatment were observed across a series of increasingly demanding analytic thresholds, so we conducted a series of robustness checks on this analysis.

We repeated the aggregate analysis using two increasingly conservative criteria for categorizing a measure as revealing a difference. Specifically, we asked whether differences continued to favor white over Black testers when "differences" were defined using a 5-percentage point threshold and again using a 10-percentage point threshold.

Table 8 displays the results of these analyses. As the threshold for what counts as a "difference" between testers becomes more demanding, the number of measures showing "no difference" increases. However, the ratio of the number of measures showing a difference favoring white testers over Black

testers remains high—and even increases with more demanding thresholds. For example, with the lowest (most liberal) difference threshold (zero percentage points), 58 total measures show a difference, and 48 (83 percent) of these reflect differences favoring the white tester. In contrast, when considering the highest (most conservative) threshold (ten percentage points) most measures (42) are categorized as "no difference," but 18 (90 percent) of the 20 remaining measures reflect treatment favoring white over Black testers.

TABLE 8. ROBUSTNESS CHECK FOR AGGREGATE ANALYSIS OF ALL RELEVANT SURVEY MEASURES

Difference Threshold	Domain	# of measures favoring white testers	# of measures favoring Black testers	# of measures showing no difference	Conditional Difference Ratio
0 percentage point difference threshold (liberal)	Total	48	10	4	4.8
	Encouragement/Discouragement	15	3	0	5
	Products and Potential Steering	15	5	0	3
	Customer Service	10	0	2	10:0
	Information Requested of Tester	8	2	2	4
5 percentage point difference threshold (moderate)	Total	32	6	24	5.3
	Encouragement/Discouragement	11	1	6	11
	Products and Potential Steering	8	4	8	2
	Customer Service	8	0	4	8:0
	Information Requested of Tester	5	1	6	5
10 percentage point difference threshold (conservative)	Total	18	2	42	9
	Encouragement/Discouragement	5	0	13	5:0
	Products and Potential Steering	5	1	14	5
	Customer Service	6	0	6	6:0
	Information Requested of Tester	2	1	9	2

Note. "Conditional difference ratio" is computed as the number of measures favoring white testers divided by the number of measures favoring Black testers.

In summary, even when considering only large differences between Black and white testers (i.e., greater than ten percentage points), measures that show differences between testers are much more likely to reflect treatment favoring white testers than treatment favoring Black testers.

5. Conclusion

5.1 Summary of Results

This report summarizes analyses of matched-pair testing conducted to test for differential treatment of well-qualified Black and white small business owners seeking credit at large bank lenders in two select counties. Fifty tests were conducted, comprising 100 in-person visits to 25 bank branches in Nassau County, NY and 25 bank branches in Fairfax County, VA.[76] Each matched-pair test recorded the treatment and experiences of one Black tester and one white tester who visited the same banking office to inquire about business financing for expanding their small business.

Using the data from these tests, we examined whether Black testers received less favorable treatment than their white counterparts in four domains: (1) **level of encouragement/discouragement** to apply for financing; (2) information provided to the tester about **available credit products and potential steering** toward product types; (3) overall quality of treatment or **customer service**; and (4) **business and credit information requested** of the tester.

We found statistically significant evidence that Black testers received worse treatment than similarly qualified white testers in two domains. First, while testers tended to receive favorable levels of encouragement/discouragement regardless of their race, white testers received more favorable treatment in this domain. This pattern is consistent across both an objective index of potentially encouraging/discouraging behaviors and testers' subjective rating of the encouragement/discouragement they experienced. Second, while bank representatives were equally likely to discuss the credit products that testers requested as part of their inquiry (business loans or business lines of credit), these bank representatives were statistically significantly more likely to discuss non-requested products with Black testers. This was a large difference: bank representatives discussed non-requested products with nearly 3 in 5 Black testers (59 percent) but fewer than 2 in 5 white testers (39 percent). The third domain—customer service—revealed mixed results, and we did not observe a statistically significant difference in the fourth domain (business and credit information requested from the tester).

Given the limited scope of the testing, these findings should not be generalized to the total U.S. small business lending market nor to any specific financial institutions. Within this scope, however, our

[76] These counties are located within the New York-Newark-Jersey City-Yonkers, NY-NJ-PA and Washington-Arlington-Alexandria, DC-VA-MD-WV MSAs, respectively.

analyses reveal evidence of—and provide a framework for detecting—less favorable treatment of well-qualified Black small business owners relative to otherwise matched white small business owners at large banks in two counties.

5.2 Limitations

The most important limitations to our study relate to the scope of testing. First, because of the scale of tests and the focus on two select counties, our statistical results should not be generalized to the total U.S. market for small business credit. As our sample typically included only one to two branches within a financial institution, the statistical results should not be generalized to any specific financial institutions without additional investigatory testing within an institution. Since testing focused on banking institutions with $10 billion or more in assets, it does not address treatment by small lenders such as smaller community banks and credit unions or by non-bank depository lenders (e.g., financial technology companies and commercial finance companies). However, these limitations also mean that our results could *understate* differential treatment relative to the overall market as prior evidence suggests that Black-owned businesses face greater barriers at smaller institutions than larger ones (Howell et al. 2024).[77]

Second, testing focused on small business owners who were well-qualified to receive financing. This characteristic does not reflect the experience of all consumers in the market including those with little or unfavorable credit history. Prior research also shows that marginally qualified testers tend to experience greater discrimination than well-qualified testers,[78] indicating that results with well-qualified testers likely underestimate the average level of discrimination in the market (HUD/Urban 2012).[79] With this limitation in mind, future research could consider studying borrowers who are only marginally qualified for small business credit or present weaker business profiles.

[77] Howell, S., Kuchler, T., Snitkof, D., Stroebel, J. and Wong, J. (2024). "Lender Automation and Racial Disparities in Credit Access." *The Journal of Finance* 79(2): 1457–1512.
[78] Hunter, W., and Walker M. (1996). "The Cultural Affinity Hypothesis and Mortgage Lending Decisions." *Journal of Real Estate Finance and Economics* 13(1): 57-70.
[79] See Turner et al. (2013) "Housing Discrimination Against Racial and Ethnic Minorities 2012" referenced in footnote 27.

5.3 Implications for Research and Industry

5.3.1. Research Implications

This study contributes to the growing body of academic and applied research on field testing methods (including matched-pair testing) for detecting differential treatment of protected classes. Future studies could build on the present work by targeting larger sample sizes, other regions, and additional protected class categories. We encourage replication and extension of these results, potentially broadening the generalizability of findings, for example, to other regional small business lending markets, other loan providers, and other borrower demographics of interest. Acknowledging that in-person matched-paired testing is costly, complementing in-person testing with other contact methods such as phone, video call, or email may improve the scalability of testing methods. However, in-person testing provides uniquely rich datasets and insights into tester experiences that likely would not be possible using phone or email testing alone; we support the continued use of this methodology to answer research questions about differential treatment and discrimination.

Additional testing could extend our findings by varying tester sex and gender identity to test for differential treatment on these characteristics. Our study design held tester sex constant within test pairs (i.e., both testers were female or both testers were male), which strengthened inferences based on tester race but precluded within-pair comparisons based on sex. Sex and gender identity may play an important role in the interpersonal dynamics of small business financing, and they may also interact with race in ways that could not be observed in our study design.[80] Exploring this possibility will require additional research.

Finally, including both objective and subjective outcome measures enhanced the research conclusions of this study by helping to understand how testers' perceptions and experiences intersected. In some cases, testers' objective and subjective experiences converged (e.g., in encouragement/discouragement); in others, only the subjective measures revealed differences between testers (e.g., in customer service). Because interpersonal treatment and discrimination are complex and nuanced behaviors, it is unlikely that objective or subjective measures alone can fully capture the experience of a small business owner. For example, vocal tone and non-verbal behaviors are important aspects of encouragement, assistance, or customer service that a bank representative or other agent involved in the credit process may provide (or withhold). We therefore suggest that future research consider the unique strengths and limitations

[80] Lederer, A., and Asante-Muhammad, D. "Racial and Gender Mystery Shopping for Entrepreneurial Loans: Preliminary Overview." (2020). National Community Reinvestment Coalition. Available at https://ncrc.org/wp-content/uploads/2020/02/NCRC-Mytery-Shopping-Race-and-Gender-v8.pdf

of objective and subjective approaches to measuring tester experiences. Doing so may unlock insights that remain hidden when considering only one type of measure in isolation, and emerging technologies (e.g., Artificial Intelligence-assisted analysis of test recordings) may further elucidate the relationship between lender behaviors and small business owners' experiences.

5.3.2. Industry Implications

The Equal Credit Opportunity Act (ECOA) prohibition against discrimination is not limited to consumer transactions; it also applies to business-purpose credit transactions, including credit extended to small businesses. The results presented in this report highlight the existence of differential treatment between Black and white testers, which can also highlight the potential importance of effective compliance management. The CFPB has previously identified several compliance management practices that may serve to mitigate the risk of an ECOA violation in a financial institution's small business lending program.[81] These include:

- Active oversight by the board of directors and management of the institution's compliance management system (CMS) framework.
- Comprehensive risk-focused policies and procedures for small business lending.
- Periodic reviews of policies and procedures in place to address the risk of an ECOA violation, with revisions as needed and supported by documentation.
- Small business lending monitoring programs and risk assessments.

A financial institution may also conduct its own testing to identify the risk of an ECOA violation. For instance, institutions may adopt a testing methodology such as the matched-pair testing framework used in this pilot study. Customer interactions with financial institutions may occur in banking offices, phone contact centers, or at small business locations and each of these interactions may raise the possibility of differential treatment, as demonstrated by this report. Hence, testing is one tool that can help financial institutions identify the risk that customers experience unlawful discrimination.

[81] See Summer, 2018 issue 17 of Supervisory Highlights for observations from examinations of supervised small business lender practices as to how risks of ECOA violation can be effectively managed.

5.3.3. Final Statement

Our findings highlight the potential for differential treatment when well-qualified Black and white small business owners inquire about small business credit, potentially resulting in disproportionally greater barriers to credit for Black small business owners. We encourage continued development and use of matched-pair testing as a useful tool to complement other data sources in the small business credit market. Doing so can enable a richer and more rigorous evidentiary base for understanding differential treatment and discrimination in small business lending.

APPENDIX A: SUPPLEMENTARY ANALYSES AND RESULTS

APPENDIX TABLE A1. EXPLORATORY ANALYSES: INDIVIDUAL SURVEY MEASURES, SUMMARY STATISTICS, AND RACE DISPARITY ESTIMATES WITH ASSOCIATED STATISTICAL SIGNIFICANCE LEVELS

Outcome Measure	N, White	Mean, White	N, Black	Mean, Black	Disparity estimate	*p*-value of estimate
Stated that they would recontact or follow up with me after our visit	47	0.47	47	0.43	-0.04	0.47
Said or implied that I should formally apply during visit	47	0.19	47	0.13	-0.07	0.19
Said or implied that I should follow up by applying after visit	46	0.61	45	0.42	-0.19	0.61
Provided an application or details to access an application	47	0.30	48	0.23	-0.04	0.29
Offered to schedule a follow-up appointment for more information or to start an application	48	0.21	46	0.24	0.07	0.19
Asked me to return to this location with more information	47	0.17	48	0.08	-0.07	0.16
Clearly described what information I would need to apply	45	0.71	48	0.54	-0.16	0.71
Informed me that they were interested in my application	48	0.40	47	0.23	-0.15	0.39
Informed me about how long it would take to approve my application (if yes, enter length)	48	0.42	48	0.31	-0.11	0.42
Informed me about how long it would take to approve my application (if yes, enter length)	48	0.06	48	0.00	-0.06	0.06
Said or implied that my outcome could be better if I improve my credit before applying	47	0.04	48	0.13	0.09	0.04

Outcome Measure	N, White	Mean, White	N, Black	Mean, Black	Disparity estimate	*p*-value of estimate
Said or implied that I should pursue other financing options outside of this institution	47	0.00	47	0.02	0.02	0.00
Informed me that they were not accepting or reviewing applications at this time	47	0.11	48	0.13	0.00	0.12
Discussed submitting a business plan	48	0.42	48	0.35	-0.06	0.42
Discussed business checking account options	48	0.17	48	0.13	-0.04	0.17
Asked to start a BL application during visit or during follow up to complete	48	0.33	48	0.23	-0.09	0.32
Encouraged to contact other lenders	48	0.04	48	0.08	0.04	0.04
Discussed plans for follow-up contact	48	0.38	48	0.44	0.06	0.27
Number of pieces of info requested from tester	48	0.52	48	0.52	0.00	0.50
Number of pieces of info volunteered by bank rep	48	0.81	48	0.73	-0.06	0.80
Provided a business card or other contact information	47	0.87	48	0.83	0.00	0.70
Asked for my contact information	48	0.58	48	0.48	-0.09	0.57
Informed me that the relevant bank staff were not in today	48	0.02	48	0.10	0.09	0.02
Lender requested other information	48	0.00	48	0.06	0.06	0.00
Asked if a current customer with the lender	48	0.73	48	0.75	0.00	0.41
Bank rep requested social security number	48	0.02	48	0.02	0.00	0.02
Bank rep requested driver's license	48	0.10	48	0.15	0.02	0.11

Outcome Measure	N, White	Mean, White	N, Black	Mean, Black	Disparity estimate	*p*-value of estimate
Bank rep requested any Personal/Business Tax Returns	48	0.04	48	0.17	0.13	0.02
Bank rep requested none of the above	48	0.88	48	0.73	-0.13	0.87
Number of "politeness behavior" measures tester answered yes to	48	0.56	48	0.54	-0.02	0.56
Thanked by the bank rep or someone else at the bank	48	0.94	48	0.94	0.00	0.50
Given the impression would qualify for any type of loan or line of credit (continuous)	44	3.75	32	3.75	0.00	0.50
Felt the bank rep was discouraging, encouraging, or neither	48	0.67	48	0.56	-0.09	0.66
Bank rep was friendly	48	0.75	48	0.58	-0.17	0.75
Bank rep was professional	48	0.73	48	0.54	-0.19	0.73
Bank rep was patient	48	0.79	48	0.56	-0.26	0.80
Bank rep was helpful	48	0.71	48	0.67	-0.02	0.67
Bank rep was informative	48	0.67	48	0.60	-0.04	0.66
The quality of the service	48	0.60	48	0.40	-0.19	0.60
How valued they felt as a potential client	48	0.63	48	0.46	-0.15	0.62
Likelihood of returning	48	0.79	48	0.73	-0.06	0.76
How much info discussed (subjective)	48	0.54	48	0.50	-0.04	0.54
Discussed a small business loan	48	0.60	48	0.75	0.15	0.06
Discussed a small business line of credit other than a credit card	48	0.96	48	0.90	-0.06	0.88
Discussed a business credit card	48	0.35	46	0.43	0.09	0.21
Told about options would not qualify for	48	0.04	48	0.06	0.02	0.04
Discussed a personal loan	47	0.02	46	0.04	0.05	0.01
Discussed a HELOC or other real-estate backed loan	47	0.02	47	0.11	0.09	0.02

Outcome Measure	N, White	Mean, White	N, Black	Mean, Black	Disparity estimate	*p*-value of estimate
Discussed a personal line of credit (non-real estate) other than a credit card	48	0.02	46	0.07	0.04	0.02
Discussed a personal credit card	47	0.00	47	0.09	0.09	0.00
Discussed an alternative funding sources outside of this institution (e.g., Kickstarter or GoFundMe)	47	0.02	46	0.07	0.05	0.02
Discussed "other"	47	0.11	45	0.22	0.14	0.07
Discussed needing more information	5	0.00	11	0.18	0.00	0.13
Told would qualify to obtain any type of small business financing option	31	1.00	31	0.87	-0.14	0.98
Recommended small business loan, including in "other" text	33	0.09	28	0.18	0.05	0.11
Recommended a small business line of credit other than a credit card, including in "other" text	33	0.82	28	0.61	-0.14	0.79
Recommended a business credit card, including in "other" text	33	0.15	28	0.25	0.14	0.13
Recommended a HELOC, including in "other" text	33	0.03	28	0.07	0.05	0.03
Recommended a vehicle loan, including in "other" text	33	0.03	28	0.07	0.05	0.03
Recommended "other" product, including in "other" text	33	0.24	28	0.29	0.05	0.24
Discussed business checking account options	48	0.38	48	0.40	0.02	0.37
Bank rep assisted in determining amount qualified for	48	0.31	48	0.19	-0.11	0.30

Note. a *p*-value of .05 or less indicates statistical significance at the five percent level. However, given the large number of exploratory tests run on the measures presented in this table, any single "significant" result in this table has an inflated chance of achieving statistical significance based on random chance alone (see footnote 68). The survey instrument testers completed is available on the report landing page.