DISCUSSION PAPER SERIES

DP14340
(v. 2)

DISCRIMINATORY LENDING: EVIDENCE FROM BANKERS IN THE LAB

Michelle Brock and Ralph De Haas

DEVELOPMENT ECONOMICS

FINANCIAL ECONOMICS

ISSN 0265-8003

DISCRIMINATORY LENDING: EVIDENCE FROM BANKERS IN THE LAB

Michelle Brock and Ralph De Haas

Discussion Paper DP14340
First Published 22 January 2020
This Revision 11 April 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

DISCRIMINATORY LENDING: EVIDENCE FROM BANKERS IN THE LAB

Abstract

We implement a lab-in-the-field experiment with 334 Turkish loan officers to test for the presence, and learn about the mechanisms, of gender discrimination in small business lending. Each officer reviews multiple real-life loan applications in which we randomize the applicant's gender. While provisional approval rates are the same for male and female applicants, we detect a more subtle form of discrimination. Loan officers are 30 percent more likely to make approval conditional on the presence of a guarantor when we present an application as coming from a female instead of a male entrepreneur. This discrimination is concentrated among young, inexperienced, and gender-biased loan officers. Discrimination is also most pronounced for loans that performed well in real life, making it costly to the bank.

JEL Classification: D81, D83, D91, G21

Keywords: Gender Discrimination, guarantors, Implicit bias, lab-in-the-field, banks

Michelle Brock - brockm@ebrd.com
EBRD and CEPR

Ralph De Haas - dehaasr@ebrd.com
EBRD and Tilburg University and CEPR

Acknowledgements
We thank Francesca Dalla Pozza, Gozde Esen, Melanie Koch, Victoria Robinson and Alex Stepanov for excellent research assistance and Martin Acht, Sofia Amaral, Thorsten Beck, Miriam Bruhn, Francois Derrien (discussant), Guido Friebel, Selim Gulesci, Sergei Guriev, Sandra Orozco-Aleman (discussant), Alexander Popov, and seminar participants at the EBRD, Surrey Business School, Ghent University, ZEW Mannheim, Southern Economic Association Annual Meetings, 8th EFI Network Workshop (Brussels) and 4th EBC Network Workshop in Banking and Corporate Finance (Lancaster) for useful comments. The views expressed are the authors' and not necessarily those of the institutions they are affiliated with.

Discriminatory Lending: Evidence from Bankers in the Lab*

J. Michelle Brock†
EBRD and CEPR

Ralph De Haas‡
EBRD, CEPR, and Tilburg University

April 11, 2020

Abstract

We implement a lab-in-the-field experiment with 334 Turkish loan officers to test for the presence, and learn about the mechanisms, of gender discrimination in small business lending. Each officer reviews multiple real-life loan applications in which we randomize the applicant's gender. While provisional approval rates are the same for male and female applicants, we detect a more subtle form of discrimination. Loan officers are 30 percent more likely to make approval conditional on the presence of a guarantor when we present an application as coming from a female instead of a male entrepreneur. This discrimination is concentrated among young, inexperienced, and gender-biased loan officers. Discrimination is also most pronounced for loans that performed well in real life, making it costly to the bank.

JEL codes: D81; D83; D91; G21; G41; L26

Keywords: Gender discrimination, implicit bias, lab-in-the-field, guarantors, banks

*We thank Francesca Dalla Pozza, Gözde Esen, Melanie Koch, Victoria Robinson and Alex Stepanov for excellent research assistance and Martin Acht, Sofia Amaral, Thorsten Beck, Miriam Bruhn, François Derrien (discussant), Guido Friebel, Selim Gulesci, Sergei Guriev, Sandra Orozco-Aleman (discussant), Alexander Popov, and seminar participants at the EBRD, Surrey Business School, Ghent University, ZEW Mannheim, Southern Economic Association Annual Meetings, 8th EFI Network Workshop (Brussels) and 4th EBC Network Workshop in Banking and Corporate Finance (Lancaster) for useful comments. The views expressed are the authors' and not necessarily those of the institutions they are affiliated with. IRB approval was obtained from the Heartland Institutional Review Board (project 160920-23).

†European Bank for Reconstruction and Development. One Exchange Square, EC2A 2JN, London. Email: brockm@ebrd.com. Tel.: +442073387193. Also affiliated with CEPR.

‡Corresponding author. European Bank for Reconstruction and Development. One Exchange Square, EC2A 2JN, London. Email: dehaasr@ebrd.com. Tel.: +442073387213. Also affiliated with CEPR and Tilburg University.

1 Introduction

Throughout the world, many more men than women use financial services. Low-income countries in particular remain characterized by large gender gaps in account ownership and the use of bank credit (Demirgüç-Kunt et al., 2018). Whether such gaps reflect economic inefficiencies depends on whether they stem from differences in the demand for or the supply of finance. On the demand side, women may select into industries that are less capital intensive or operate at a smaller scale and hence require less external finance (Demirgüç-Kunt, Beck and Honohan, 2008). On the supply side, gender discrimination is often cited as a key driver of women's financial exclusion (The Economist, 2013). Discrimination occurs when lenders treat male and female loan applicants differently even when they are equal in terms of all business-related characteristics. Aspiring or existing female entrepreneurs then face overly tight credit constraints and their productive capacity remains underutilized.

In this study, we test for direct and indirect gender discrimination in the supply of small business loans. Loan officers who do not reject female loan applicants directly may still apply gender-specific conditions that make credit de facto unattainable for many women. Such indirect discrimination is particularly difficult to detect empirically. We therefore implement a lab-in-the-field experiment in which loan officers evaluate multiple real-life loan applications where the gender of the applicant has been (randomly) manipulated by us.[1] Bringing experienced loan officers into a controlled laboratory environment allows us to carefully track their lending decisions and to trace the mechanisms through which gender discrimination can materialize. We not only observe whether officers grant credit but also the conditions under which they are willing to do so. Here our focus is on guarantor requirements—where a co-signer has to underwrite the loan—as these requirements are potentially an important source of indirect gender discrimination (Alesina, Lotti and Mistrulli, 2013).[2]

We conduct our experiment with 334 loan officers of a large Turkish bank. Turkey provides a particularly well-suited setting to study gender discrimination in lending. It is a large and growing emerging market with a professional and competitive banking system.

[1] Gneezy and Imas (2017) define a lab-in-the-field study as one "conducted in a naturalistic environment targeting the theoretically relevant population but using a standardized, validated lab paradigm".

[2] Unlike passive collateral, guarantors actively monitor borrowers to ensure repayment (Banerjee, Besley, and Guinnane, 1994) and monitoring is often leveraged by the threat of social sanctions (Bond and Rai, 2008). The use of guarantors is not only widespread in Turkey and other emerging markets but also common in Europe and the United States.

The country scores well in terms of *de jure* gender equality: there exist few legal obstacles that restrict women's ability to become an entrepreneur (Klapper, Munoz and Singh, 2014). Yet, at the same time, the country remains characterized by conservative gender norms. It only ranks 130 out of 149 countries in terms of *de facto* gender equality (WEF, 2018). This discrepancy between gender-related laws on the book and actual attitudes within society is characteristic of many other emerging markets as well.

We find that while provisional approval rates are very similar for male and female loan applicants, loan officers discriminate against women in an indirect way. In particular, they are 30 per cent more likely to require a guarantor when we present an application as coming from a female instead of a male entrepreneur. Young and inexperienced loan officers and those with a high implicit gender bias (measured with an Implicit Association Test, IAT) are most likely to apply this double standard. Crucially, because we use real-life loan applications that our partner bank received in the recent past, we can trace how loans perform in reality. We find that discrimination is concentrated among loans that do well in real life, making it potentially costly to the bank.

These results help us to advance the literature in three important ways. First, we build on earlier empirical work on gender discrimination in credit markets. This literature typically applies multivariate regression analysis to observational data. The dependent variable is then a lending outcome and the main independent variable the gender of the applicant or borrower. Identification relies on controlling for all other applicant characteristics that impact loan outcomes and may correlate with gender. Applying this approach to data from an Italian bank, Bellucci, Borisov and Zazzaro (2010) show that female entrepreneurs face tighter credit availability and collateral requirements but not higher interest rates. Alesina, Lotti and Mistrulli (2013) access the Italian credit registry and find that female-owned firms *do* pay higher rates. Women also need to post a guarantee more often even though their firms are not riskier. In contrast, studies using administrative data from the U.S. generally find no evidence of gender discrimination in small business lending.[3]

Our lab-in-the-field methodology advances this literature by eliminating or reducing a

[3] See Blanchflower, Levine and Zimmerman (2003), Blanchard, Zhao and Yinger (2008) and Asiedu, Freeman and Nti-Addae (2012). A number of papers on the U.S. do detect discrimination against African-Americans and Hispanics (Munnell et al., 1996 and Blanchflower et al., 2003). Ferguson and Peters (1995) provide an insightful discussion of the conclusions one can and cannot draw about discrimination on the basis of loan denial and default rates in administrative data.

number of identification concerns. In particular, we need not worry about omitted variables bias because we can vary applicant gender while keeping all other characteristics of loan applications equal. Our experimental approach also allows us to isolate the supply side of the credit market. This is important because a lower use of credit by female enterprises can simply reflect lower demand. Some women may be more risk averse, more easily discouraged from applying for credit (Ongena and Popov, 2016), or self-select into industries that need less external finance. Another concern is that in administrative data clients may not be randomly matched to loan officers, which can lead to unreliable estimates of discrimination among officers. One way to address this is to exploit rotation policies that generate exogenous matching between officers and borrowers (Fisman, Paravisini and Vig, 2017). We instead randomly assign applications to loan officers so that there is by construction no bias due to endogenous matching.

A second way in which we advance the literature is by bringing loan officers to the lab (or rather: by bringing a lab to the loan officers) so that we can measure key loan officer characteristics that normally remain unobserved. Beck, Behr and Guettler (2013) and Beck, Behr and Madestam (2018) use data from an Albanian microfinance institution to study the role of loan officer gender. The first paper shows that female loan officers make lending decisions that result in fewer arrears. Interestingly, this advantage only emerges with experience when women lend to male borrowers. The second paper finds, in line with an own-gender bias among officers, that new borrowers matched with opposite-sex loan officers pay higher interest rates. The authors of the first paper conclude that *"not only the institutional and governance structure of financial institutions matters, but also the gender of the people operating in a given bank structure"* (p. 5). Yet they acknowledge that a performance gap between male and female loan officers may reflect unobserved differences. By bringing officers to the lab, we can measure personality traits that usually remain unobserved by the econometrician, such as risk aversion and implicit gender bias. This yields new insights into how discrimination varies within the loan officer population and into the nature of discrimination itself.

Third, we contribute to the broader literature on discrimination, which distinguishes between taste-based (Becker, 1957), implicit (Bertrand, Chugh and Mullainathan, 2005), and statistical discrimination (Phelps, 1972; Arrow, 1973).[4] Taste-based discrimination takes

[4]Most papers focus on racial discrimination (for example Neal and Johnson, 1996 and Bertrand and

place when decision makers (say, loan officers) are prejudiced against a group (say, women) and consciously avoid interacting with them. Implicit discrimination is similar but occurs without individuals being aware of it and is more likely when decisions are made under conditions of ambiguity, time pressure, or distraction (Bertrand, Chugh and Mullainathan, 2005). In contrast, statistical discrimination refers to decision makers relying on a group attribute (such as gender) as a signal of unobserved individual characteristics. Such stereotyping takes place in the absence of reliable individual-level information.[5] Statistical discrimination may be stronger when decision makers evaluate "out-group" individuals (say, male loan officers screening female loan applications) because it can be especially difficult to interpret signals from out-groups (Cornell and Welch, 1996). Distinguishing between different types of discrimination is challenging as they can occur simultaneously and exacerbate each other. For instance, if decision makers collect less information about individuals in certain groups due to taste-based or implicit discrimination, this may over time result in statistical discrimination (Bertrand and Duflo, 2017).

We find that loan officer heterogeneity in implicit gender bias and in experience correlates with discriminatory behavior. At the same time, varying the information available to loan officers does not affect their lending in a gender-specific way. While we cannot rule out taste-based discrimination, these findings are indicative of an indirect expression of implicit bias (in the form of gender-biased guarantor requirements) and a limited role for statistical discrimination (which is attenuated by experience).

A novel aspect of our study is also that we integrate several features of earlier experimental work within one setting. We conduct a lab-in-the-field experiment in which we observe actual loan officers (rather than, say, students) working on a task that closely resembles their everyday work. We randomly manipulate the gender of loan applications as is typical of this approach.[6] Yet, unlike most correspondence studies we use real-world loan applications instead of fictitious ones. This ensures that the applications are realistic and that

Mullainathan, 2004) and gender discrimination (for example Altonji and Blank, 1999 and Bayard et al., 2013) in the labor market. See Neumark (2018) for a survey of the experimental literature on labor market discrimination.

[5]While taste-based and implicit discrimination are inefficient, statistical discrimination can be rational and efficient.

[6]In their seminal correspondence study, Bertrand and Mullainathan (2004) use fictitious resumés with randomly assigned African-American or White-sounding names that are sent in response to a job ad. They find that White-sounding names receive 50 percent more call-backs for interviews.

we can observe what happened to them in real life (Cole, Kanz and Klapper (2015) follow a similar approach).[7] Moreover, we incentivize all decision making in a way that closely resembles loan officers' regular remuneration system. This allows us to mimic the costs associated with certain forms of discrimination and contrasts with existing experimental work on discrimination, which typically does not impose costs for making inefficient discriminatory choices (Neumark, 2018). Lastly, combining correspondence studies with IATs, as we do in this paper, is still relatively rare.[8] In sum, we combine a highly realistic setting (actual loan officers, real-life loan applications, regular incentive scheme) with experimental elements (randomization of applicant gender and an IAT) to understand how gender discrimination can occur in small business lending.

The rest of the paper is structured as follows. Section 2 describes our setting and experimental design. Section 3 then summarizes the data generated by the experiment and outlines our estimation strategy. Section 4 presents the results and Section 5 concludes.

2 Experimental context and design

2.1 Context: Participants and loan application process

We conducted our experiment in cooperation with a large, private (and non-Islamic) commercial bank in Turkey. We held 22 experimental sessions with a total of 334 bank employees across eight cities over a two-month period.[9] The bank operates a regional office in each of these cities and participants were randomly selected from all bank employees involved in small business lending (which makes up two-thirds of the bank's loan portfolio). Figure 1

[7] Related to our work, Alibhai et al. (2019) conduct an email survey of 77 Turkish loan officers in which respondents had to distribute a total loan amount among four fictitious and highly stylized loan applications with randomized applicant gender. The authors find that in this non-incentivized setting, female applications receive a smaller portion of the total available funding.

[8] Reuben, Sapienza and Zingales (2014) conduct a lab experiment in which "employers" hire "workers" for a mathematical task. They find that employers are more likely to hire men when they have no information about the ability of hires. This pattern is reduced when information on prior performance is made available. The researchers also administer an IAT based on stereotypical associations between gender and words related to science versus liberal arts. They find more bias, and less updating based on performance data, for those who are more gender biased according to the IAT. Implicit biases may thus interact with information availability and statistical discrimination.

[9] These cities were Adana, Ankara, Antalya, Bursa, Gaziantep, Istanbul, Izmir, and Trabzon. We also conducted a pilot with 32 employees in Istanbul but do not use these data.

shows the location of the regional offices and the number and gender of the participating bank employees.[10]

Our sample includes employees at two levels of seniority: loan officers (192) and supervisors (142). The lending process at our partner bank is highly standardized and both roles are clearly defined. Loan officers establish contact with a new (potential) borrower and conduct the initial screening. They collect documentation on business performance (income statements and balance sheets). They also check the availability of collateral and guarantors and request a credit score from the Turkish credit registry (KKB). Loan officers enter this objective screening information into an electronic application form. At this stage they can also voluntarily add subjective notes to the form, such as about the client's perceived trustworthiness, experience, or social standing. If the loan officer deems a client creditworthy in principle, they pass on the electronic application form to their supervisor (typically the branch manager) with a proposal for a maximum credit limit. Crucially, at this point they also include their view as to whether a guarantor is required. That is, loan officers can recommend that the loan application is approved unconditionally or made conditional on the presence of a guarantor. The supervisor reviews the loan application and can reject or provisionally approve it. In the latter case, they need to send it to the bank's headquarters for formal sign off.[11] Both loan officers and their direct supervisors are thus based in the same regional office and involved in the screening of potential borrowers. Henceforth we refer to the total experimental population as either "participants" or simply "loan officers".

A relevant contextual question is whether there are systematic differences between the creditworthiness of male and female-owned small businesses in Turkey. Such differences could justify a different treatment of male and female loan applications. To look into this, we randomly sample 250 loan applications (stratified by gender) out of all applications the bank received from small firms in recent years. We then compare the credit score of male and female applicants (a higher score implies less risk). The average score is 1,035 for men and 1,023 for women and this small difference is not statistically significant (p=0.80). Appendix

[10] We stratified by gender, so that the participants' gender composition does not exactly reflect that of the local universe of bank employees.

[11] Branches can approve loans below a certain size threshold but in practice only 10 per cent of micro loans and 0.5 per cent of loans to small and medium sized enterprises (SMEs) are signed off in a branch. Micro clients are those with an annual turnover below TRY 2.5 million (≈US$ 700k) and a credit limit below TRY 750k (≈US$ 210k). The application process is fast, with applications typically approved within 1.5 days.

Table A2 presents OLS regressions for the 243 files for which these credit scores were available (the dependent variable). The first column confirms there is no significant difference between female and male applicants. This holds when we include sector fixed effects (column 2) or sector and region fixed effects (column 3) and when we control for firm size (column 4) and the loan amount requested (column 5).

2.2 Experimental design

For the main task in our experiment participants evaluated two rounds of four loan applications (eight in total).[12] Applications were randomly presented as coming from a female or a male entrepreneur. Participants had to decide whether to approve or reject each application and, in case of initial approval, whether to request a guarantor or not. For each loan application, participants also had to provide a subjective risk rating between 0 and 100. We did not constrain the time participants had to evaluate each application. The sessions were framed as a general training exercise and no gender-related issues were mentioned or discussed during the sessions.

The task closely mimicked the choices the participants make in their daily work. To be consistent with real life lending decisions, all loan applications were presented to the participants electronically and in the format of the standard application forms that they normally process on their computer. The forms (henceforth called "loan applications") contained all the information available at the time the application was processed.[13]

We use 100 unique loan applications in the experiment, selected from an initial sample of 250.[14] These 250 applications were a stratified random sample of all applications by existing SMEs (that is, no start-ups) that the bank had received in the three to six years before the experiment.[15] Using this earlier period allows us to track what happened to each of these

[12]Loan officers made decisions on loan applications worth US$ 81.1 million in total.

[13]These forms are at the heart of the decision making about whether the bank is willing to lend, what the maximum credit exposure will be, and whether a guarantor is required. Only after this stage, do the loan officer and client negotiate about specific product types, such as credit lines and fixed-term loans. The maturity and pricing of individual products is also determined at this later stage. This means that during the experiment we could collect data on willingness to lend, maximum amount granted and the need for a guarantor, but not on the interest rate and maturity of specific credit products.

[14]Figure A1 in the Appendix shows that also among these 100 application files, the distribution of credit scores—an indicator of ex ante credit risk—is similar among applications from male and from female entrepreneurs.

[15]When participants evaluated the files, they did not see the real application date but a date in the year of

applications in real life. The strata were region, gender, firm size, and whether the application was accepted in real life. By only using applications from first-time loan applicants, who had never before borrowed from our partner bank, we minimize the potential influence of soft information generated over time.

All applications were gender neutral except for the randomly assigned applicant name. To achieve this, we only used applications from Turkish industries in which both female and male ownership is relatively common. For instance, we excluded applications from construction companies (a male dominated sector) and beauty parlors (a female dominated sector). Some files were eliminated because they contained explicit references to the applicant's gender (other than their name) or because they were incomplete.

All 100 files occur in the experimental data multiple times: each loan application was on average evaluated by 13.4 different participants per round, half of the time as a female and half of the times as a male file. This is important because it allows us to obtain a within-application estimate of gender discrimination. By asking participants to review both male and female applications, we preserve external validity as no one at the bank sees only male or female clients. We indicate the gender of each file by assigning new names, randomizing between male ones (Ahmet, Ali, Mehmet, Mustafa) and female ones (Ayse, Emine, Fatma, Zeynep). These names are common across Turkey and are well represented among working-age adults across regions.[16] No one saw the same file or same name more than one time.

We held constant the ratio of performing, non-performing and rejected files that each participant saw, at 2-1-1. This 2-1-1 ratio does not reflect the bank's actual application flow, but we used this ratio so that participants evaluated at least one file of each type in each round of decision making. Names were randomized such that each participant saw one performing loan and one "bad" loan application (either a non-performing loan or a declined

---

the experiment. We did so to avoid recall bias—loan officers did not have to think back about the economic situation in the past. This of course introduced a slight disconnect between loan performance in real life and the application evaluated during the experiment. To check whether this disconnect matters empirically, we regress our outcomes (loan rejection or guarantor requirement) on the difference between the loan application date and the time of the experiment, interacted with applicant gender. These interaction effects are never significant, indicating that the small timing difference does not have any gender-specific impact.

[16] We checked which names had the highest frequencies in the relevant cohorts and across regions using information from the Turkish General Directorate of Population and Citizenship Affairs (https://www.nvi.gov.tr/isim-istatistikleri) and an additional online data source (https://www.isimarsivi.com/). When we include name fixed effects in our regressions we fail to reject the null that these effects are jointly equal to zero.

application) from each gender.[17]

After a first round of four loan applications, the loan officers received a second round of four applications. We again randomized the gender of each application. Moreover, we now also randomized participants into three groups: a control group evaluated applications with all information available (as in the first round), a first treatment group evaluated files from which we had deleted the credit score from Turkey's credit bureau, and a second treatment group evaluated files where a section with subjectively provided information had been removed. This section contains voluntary comments by loan officers about the applicant (such as about how industrious they are or whether they have a good business network). Bank staff provide this information in the application file to strengthen the rationale for lending. It targets the decision makers higher up in the lending hierarchy. If either the objective credit score or the subjective comments section contribute to staff's ability to make fair and objective lending decisions, omitting it may increase statistical discrimination, especially among loan officers with an implicit gender bias (Reuben, Sapienza and Zingales, 2014). On the other hand, if the information itself is perceived with bias, omitting it may potentially reduce discrimination. In the first (second) case, we should see that bias is higher (lower) in the treatment groups than in the control group.[18]

For the second round, we opted for a within-file (in terms of gender randomization) and between-participant (in terms of the information treatment) experimental design for two reasons. First, we wanted to avoid non-linear or heterogeneous order effects. Non-linear order effects are difficult to control for, while controlling for heterogeneous order effects would require a larger participant pool than we had. Second, subjecting all participants to all treatments would have required each participant to complete 12 reviews, and there was not enough time for that.

We incentivized all loan decisions in line with the structure of common bank incentive schemes. Participants earned ten points (equivalent to ten Turkish lira) for each completed review (quantity) and an additional five points when they correctly approved a loan that performed well in real life (quality).[19] In contrast, five points were deducted when they

[17] That is, analogous to Bertrand and Mullainathan's (2004) correspondent study on racial discrimination, we crossed applicant gender with application quality. Due to time constraints participants could not evaluate more than four files and we wanted to ensure the data could be analyzed by application quality.

[18] Heilman (1984) documents inconsistent findings in past experimental studies as to whether giving additional information about job applicants reduces adverse outcomes for women.

[19] This incentive scheme resembles the remuneration system that the bank uses in reality and is also similar

incorrectly accepted a loan that was defaulted on in real life. When participants approved a file that had been declined in real life, we gave them a 50/50 chance that the file was counted as performing, thus yielding the extra five points. We did not penalize incorrect rejections in order to mimic as closely as possible the actual incentive scheme at the bank, and the bank cannot realistically know when a rejection is incorrect.

We aggregated all points per participant at the end of the experiment. Participants were ranked according to their score and split into four quartiles. In line with our instructions at the start of the session, those in the highest quartile could spend their points on higher valued prizes while those in the lower quartiles had to select gifts with lower values. All participants had chosen their preferred prizes from each category prior to the experiment. This ensured they understood how the incentives worked and what the benefit would be of getting into the top quartiles.[20] The incentive scheme was thus both material and competitive.

2.3 Eliciting personality traits

After both rounds of application decisions, we measured participants' risk preferences and implicit gender bias. We follow Binswanger (1982) and Eckel and Grossman (2008) and elicit risk preferences by presenting participants with six risk scenarios from which they had to choose one. Each scenario was depicted as a circle split in two. Each half contained a point outcome and the even split represented that in each scenario there was a 50 per cent chance of getting the left or right outcome. The outcome pairs were 28-28; 20-44; 24-36; 16-52; 12-60; and 2-70. The first scenario has no risk while the last scenario entails most risk. The task was incentivized: an on-site computer drew random draws to determine whether participants would get the low or high number from the circle they selected. We added the resulting points to the participants' point total at the end of the session.

Participants also took an implicit association test (IAT).[21] Participants sorted, as quickly

to the low-powered (baseline) scheme of Cole, Kanz and Klapper (2015).

[20]Specimens of the prizes that participants could buy were on display during the sessions. The prizes were sent to the participants a few weeks later.

[21]See https://implicit.harvard.edu/implicit/takeatest.html and Carney et al. (2007). IATs are by now common in psychology (Greenwald, McGhee and Schwartz, 1998) and economics (Bertrand, Chugh and Mullainathan, 2005; Beaman et al., 2009; Carlana, 2019). A meta-analysis of 184 studies found an average correlation of 0.24 between the IAT score and outcome measures such as judgments, choices, and physiological responses (Greenwald et al., 2009). While attitude IATs are used to detect implicit negative attitudes towards social groups, stereotype IATs measure implicit associations between social groups and specific

as possible, words that appeared sequentially on their tablet by clicking buttons at the right and left sides of the screen. The IAT started with two practice rounds in which participants sorted "career" words into a "career" bucket (left) and "family" words into a "family" bucket (right). This was repeated for male and female words.[22] After these practice rounds, the IAT mixed gender words and career/family words. Male and career words now shared a sorting button while female and family words shared the button on the other side of the screen (the stereotypical task). This was followed by another task where male and family words shared a sorting button while female and career words shared the other button (the non-stereotypical task). The time it took to sort each word was recorded in milliseconds. The assumption is that respondents with a stronger association between two concepts find sorting easier and complete it faster in one task compared to the other. We defined a participant's implicit bias as the normalized difference in mean response times between the non-stereotypical and the stereotypical task. Higher values indicate stronger bias.

3 Data and estimation strategy

3.1 Data

Table 1 summarizes our experimental data (Appendix Table A1 contains all variable definitions). Panel A describes the main characteristics of the 334 participants. Almost half of them are female and their average age is 37 years, ranging between 26 and 53. Forty-three per cent of the participants are supervisors, the others are loan officers. While the average participant has worked at the bank for five years, this varies between zero and 19 years.[23] There is thus substantial variation in the lending experience that loan officers built up over the course of their career.

Our lab-in-the-field experiment allows us to measure participant characteristics that are otherwise difficult to observe. As described in Section 2.3, we use lottery questions to elicit

traits (Bertrand and Duflo, 2017). Our IAT falls in the latter category.

[22]The IAT and all other documentation was provided in Turkish. The family-related words were the Turkish translations for words such as "kitchen", "marriage", and "laundry". Career words included "office", "manager", and "job". To designate "male" we used words like "man", "boy", and "gentleman" and for "female" words we used words such as "woman", "girl", and "lady".

[23]We also asked participants whether they had previous lending experience in a similar role at another bank. All results that follow are robust to using this broader experience definition.

risk aversion and an IAT to gauge participants' implicit bias against women in business. Table 1 reveals substantial variation in these measures. The categorical variable *Participant risk aversion* ranges between 1 (risk loving) and 6 (most risk averse). The average participant scores 4.1. A large literature has documented that, on average, women tend to be more risk averse than men (for example Eckel and Grossman, 2008). Evidence from the financial services industry indicates that female decision makers take less risk on average (Sunden and Surette, 1998; Agnew, Balduzzi and Sunden, 2003). The correlation matrix in Appendix Table A3 shows that this is the case in our setting as well. The average risk aversion score is 4.32 for females and 3.92 for males.

The IAT score is transformed so that it ranges between -1 and 1 with zero indicating no implicit gender bias. While the scores vary widely, a large majority of lending staff (87 per cent) has a positive IAT score, indicating that they subconsciously associate business more with men than with women. This tendency is stronger among women than among men (Figure 2). A two-sample Kolmogorov-Smirnov test confirms that the distributions are significantly different (see also Appendix Table A3). The average IAT score is 0.39 for women and 0.28 for men and this difference is statistically significant at the 5 per cent level.

Panel B of Table 1 summarizes the real-life characteristics of the 100 loan files that we use in the experiment. By design, half of these files refer to loans that in real life were paid back on time (performing loans), a quarter refers to loans that in real-life were defaulted upon (non-performing loans), and another quarter consists of loan applications that were rejected in real life (declined applications). Panel C summarizes the experimental outcomes at the participant-file decision level. We show separate statistics for round 1 (no information treatment) and round 2 (information treatments). In both rounds, almost forty per cent of the loan applications were rejected outright whereas, conditional on provisional acceptance, a guarantor was requested in 27 per cent of the cases. Panel D shows that we withheld either subjective or objective applicant information in a third of the decisions in the second round.

For each credit application, the participant was asked to estimate, on a 0-100 scale, the probability that the borrower would repay. This helps us to verify that the experimental task was meaningful in the sense that loan officers could infer credit risk based on the information in the loan file. Figure 3 provides a scatterplot of the 100 files used in the experiment, based on data from the first round only. The horizontal axis indicates the average subjective repayment probability (each file was evaluated by 13.4 participants on average in round 1) while the vertical axis shows the share of participants that rejected the application. Figure 3 reveals a tight negative correlation between expected repayment probability and the likelihood of loan rejection. This suggests that our incentive scheme worked and that participants thought the task realistic and paid attention to the information provided.

Equally important is whether the decision making in our lab-in-the-field correlates with what happened to the loan applications in real life. We find that this is the case. Overall, 72 per cent of all applications that had resulted in loans that performed well in real life were approved during the experiment. This percentage is significantly lower for applications that resulted in non-performing loans (53 per cent) and for applications that were rejected in real life (47 per cent). As a result, files that in real life were non-performing (gray dots) or declined (white) are concentrated in the upper left corner of Figure 3 while performing loans (black) are concentrated in the lower right-hand corner. This indicates that across the board participants correctly identified loans that performed well or badly in real life and made decisions in line with these subjective perceptions of loan quality.

Appendix Table A3 provides a correlation matrix of the participant characteristics and the rejection dummy. We have already discussed that female participants are on average more risk averse and more gender biased. There are two other interesting correlations in this table. First, participants higher up in the lending hierarchy (supervisors) are more often female, older and more experienced, as well as more implicitly gender biased. Second, older participants (unsurprisingly) tend to be more experienced but also more gender biased, suggesting that cohort effects are important. Lastly, it is reassuring that whether a file was presented as male or female (*Female applicant*) is uncorrelated with any of the participant characteristics. This reflects that the randomization process was successful.

Appendix Table A4 assesses the correlates of implicit gender bias in a multivariate setting. When we "horse race" the participant characteristics in this way, participants' own gender is the main variable that explains implicit gender bias. Even when controlling for a participant's experience, age, hierarchical position, and risk aversion, we continue to find that female bank employees are on average 0.124 points (on the [-1,1] scale) more biased against female entrepreneurs as compared with male bank employees.

3.2 Estimation strategy

To test for the presence of gender discrimination in the lending behavior of loan officers, we run linear probability regressions at the decision level:

$$y_{il} = \alpha + \beta \cdot G_{il} + \sum_{k=1}^{K} \gamma_k \cdot X_i + \varphi_l + \varphi_c + \epsilon_{il} \tag{1}$$

Where y_{il} is a lending outcome of interest when participant i evaluates loan application l; α is a constant; G_{il} is the randomly assigned applicant gender to loan application l as seen by participant i; X_i is a vector of K participant characteristics (gender, experience, age, a *Supervisor* dummy, risk aversion, and IAT score); φ_l are loan application (file) fixed effects; φ_c is a set of fixed effects for the eight cities where the experiment took place; and ϵ_{il} is a stochastic error term.[24]

Due to the experimental design, applicant gender is the only loan application characteristic that (randomly) varies across decisions about the same application. The loan application (file) fixed effects thus absorb all observed and unobserved file characteristics aside from applicant gender. Unobservables here include all (combinations of) features of the written loan applications that the econometrician might ignore but that loan officers consciously or unconsciously care about. In this sense the experimental design and associated analytical specification provide stronger identification compared with observational studies where the data do not allow for within-file estimates.

Standard errors are heteroskedasticity robust. Because we randomize applicant gender at the file level there is no need to cluster standard errors. The standard robust variance estimator yields correct inferences (Abadie, Athey, Imbens and Wooldridge, 2017).

[24]We do not include participant fixed effects because the matching of files to participants was random and everyone evaluated the same proportion of male and female applications, of both high and low quality. Individual heterogeneity is therefore not systematically correlated with the treatment variable (that is, the gender of the loan applicant).

4 Results

4.1 Applicant gender and the rejection of loan applications

Table 2 presents linear probability regressions based on Equation 1. The dependent variable, *Rejection dummy*, is "1" if an application was outright rejected by a participant and "0" if approved. The independent variable of interest, *Female applicant*, is a dummy whether the application was presented as coming from a female ("1") or male ("0") entrepreneur. All specifications include application (file) fixed effects, city fixed effects, and the following participant covariates: *Participant is female*; *Participant experience*; *Participant age*; and whether they are a supervisor or loan officer (*Participant is supervisor*). In columns 2 and 3, we also control for *Participant risk aversion* and implicit gender bias (*Participant IAT score*), respectively. Column 4 includes both. All data are from the first experimental round.[25]

Table 2 shows that we cannot reject the null hypothesis of no significant treatment effect of *Female applicant* on loan rejection. Across the four specifications, the coefficient for *Female applicant* is close to zero and, if anything, negative. Since we include file fixed effects, our results show that *the same* loan application does not have a higher chance of being rejected when we present it with a woman's name rather than a man's name. In short, we find no evidence of direct gender discrimination.

Turning to the participant covariates, we find that older participants are slightly less likely to reject loan applications, even when controlling for their experience.[26] This holds at the 10 per cent level of statistical significance. Moreover, participants who in real life are supervisors are 10 percentage points less likely to accept a loan application as compared with loan officers. This is a large difference as the unconditional acceptance rate in our experiment is 61 per cent. This finding is significant at the 1 per cent statistical level and likely reflects that the main role of supervisors is to validate (or overrule) the initial lending decisions made by more junior loan officers.[27] Lastly, we do not find that participants' risk aversion or

[25] All results also hold when we combine the observations from the first round with those from the control group of the second round (in which we did not delete any information).

[26] Both variables are positively correlated (p=0.50). See Appendix Table A3.

[27] Appendix Table A5 shows that this higher conservatism among supervisors is independent of the (randomized) gender of the loan applicant: the interaction between *Female applicant* and *Participant is supervisor* is never statistically significant. The same holds when we run separate regressions for split samples of loan officers and supervisors.

their implicit gender bias have an independent effect on the rejection probability. We include columns 2 and 3 to illustrate that this lack of significance is not due to multicollinearity. Thus, while implicit gender bias is widespread (and varied) among our loan officer population, according to the IAT scores, this bias does not translate into explicit discrimination in the decision of whether or not to give credit.

We also assess whether this null result holds in various sub-groups. We cut the data in six ways—by participant gender; above/below median experience; above/below median age; supervisors versus loan officers; above/below median risk aversion; and above/below median implicit gender bias—and run sample-split regressions (unreported). There is no evidence of direct gender discrimination in any of these sample splits.[28]

4.2 Applicant gender and guarantor requirements

Our results so far indicate that, when considering whether or not to offer credit, loan officers treat loan applications in the same way regardless of whether we present them as coming from male or female entrepreneurs. We next assess whether there exist more indirect forms of discrimination against female applicants. In particular, an earlier (non-experimental) literature suggests that the approval of female loan applications may often be made conditional on the presence of a guarantor. Such indirect discrimination can negatively impact female entrepreneurs who do not have access to a guarantor. It can also be disadvantageous to the bank if such entrepreneurs are, in fact, good credit risks. Our experimental set up allows us to investigate both these dimensions.

We start in Table 3 by analyzing whether loan officers are more likely to request a guarantor when the application comes from a female instead of a male entrepreneur (all else equal). The structure mirrors that of Table 2. The sample is smaller as the decision to require a guarantor is conditional on provisional loan approval. We find strong evidence of indirect gender discrimination: loan officers are seven percentage points more likely to make final loan approval conditional on the presence of a guarantor when the application is shown as coming from a female instead of a male entrepreneur. This is a substantial effect as only

[28]When we partition continuous variables, the below-median sample contains values strictly below the median while the above-median sample contains values at the median and above. All results remain unchanged when we instead allocate at-the-median observations to the below-median group or when we partition at other points. All results are also the same when we run interaction regressions rather than sample-split regressions.

27 per cent of all pre-approved applications were required to have a guarantor. Unlike the case for loan approval, none of the participant characteristics impact the probability that a guarantor is required.

Figure A2 in the Appendix depicts coefficient estimates similar to those in column 4 of Table 3 while dropping one city at a time from the sample. In each case the coefficient indicates a 6 to 10 percentage points higher likelihood that a guarantor is requested in the case of female applicants. The coefficients are ordered, from top to bottom, by decreasing average disposable household income in the excluded city. There is no apparent relationship between indirect gender discrimination and the local level of economic development.

4.3 Indirect gender discrimination: Participant heterogeneity

Table 4 presents a heterogeneity analysis for unequal guarantor requirements. We only present the coefficient for *Female applicant* but all regressions include the same covariates and fixed effects as in Tables 2 and 3. We find a consistent and intuitive pattern of statistically significant heterogeneous treatment effects. When we present the application as coming from a woman instead of a man, loan officers are more likely to ask for a guarantor when they are less experienced (columns 3-4); younger (columns 5-6); and/or display more implicit gender bias in our IAT (columns 11-12). For instance, loan officers with a below-median level of lending experience are 11 percentage points more likely to ask women (as compared with men) for a guarantor than their more experienced colleagues. This suggests that experience reduces the extent to which loan officers use gender as a mental shortcut to determine whether a loan application requires a guarantor or not.

Likewise, loan officers with above-median levels of implicit gender bias are 12 percentage points more likely to demand a guarantor when we present a file as coming from a female instead of a male entrepreneur.[29] This indirect gender discrimination is also concentrated among loan officers at the lower hierarchical level, that is, those who in real life make the initial screening decisions (columns 7-8). This indicates that businesswomen with good loan propositions, but without a guarantor, may already be rejected at the initial screening

[29] A few (42) loan officers display a negative gender bias, meaning that they associate women—rather than men—with a career. In line with symmetric interaction effects, we find that these officers are *less* likely to request a guarantor when we present an application as coming from a woman.

stage.[30] The t-test p-values at the bottom of Table 4 confirm that we can reject equality of coefficients in these column pairs.[31]

Importantly, columns 1 and 2 of Table 4 show that there is no difference between male and female participants in how they treat female applicants. This holds independently of whether we control for other participant characteristics or not. There is also no statistically significant difference between participants with higher versus lower levels of risk aversion, a characteristic strongly correlated with participant gender (columns 9 and 10).

4.4 Indirect gender discrimination and loan quality

The applications that loan officers reviewed during the experiment were real applications that had been processed by the bank in the recent past. We therefore know what happened to these applications: whether they were rejected or approved and, if approved, whether the loans were repaid or not. We already know from Figure 2 that applications that were rejected or that turned out to be non-performing loans in real life were less likely to be approved during the experiment. In contrast, applications that were approved and subsequently performed well in real life were more likely to be accepted during the experiment.

We now ask whether the higher probability that female loan applicants are required to have a guarantor is driven by loans that performed well in real life or by those that did less well. Figure 4 gives a non-parametric answer to this question. We divide all loan applications into those that were accepted in real life and performed well (dark gray bars); those that were accepted and became non-performing (medium gray) and those that were declined in real life (light gray). The data pattern is striking. When we present files as coming from

[30] While supervisors are more conservative than loan officers in terms of approving loans, their conservatism is "gender neutral". In contrast, when it comes to guarantor requirements, loan officers are not only more conservative than supervisors but there is also a bias against women.

[31] We summarize results from equivalent fully interacted regression models in Appendix Figure A3. The independent variables in these regressions include the *Female applicant dummy*, an interaction between this dummy and a participant characteristic (such as *Participant experience*), and a full set of additional interaction terms between this participant characteristic and all other controls, including the file and city fixed effects. The bars in Figure A3 show the coefficients for the *Female applicant dummy* and the interaction of this dummy with the respective participant characteristic. The black dots indicate the sum of these two coefficients. For instance, 0.106 in the second column corresponds with the equivalent *Female applicant dummy* coefficient in column 3 of Table 4. This number indicates the impact of presenting a file as female among loan officers with below-median experience. The dark bar depicts this impact for officers with above-median experience and its height (0.032) corresponds exactly with the coefficient in column 4 of Table 4, based on sample-split regressions.

male loan applicants (left-hand side), loan officers clearly and strongly differentiate between high-quality and lower-quality loans. For loans that were repaid in real life, men are asked for a guarantor in only 20.1 percent of the cases during the experiment. This number is substantially higher for non-performing loans and applications that were declined in real life, at 28.6 and 32.9 per cent respectively (these percentages are statistically different from that for performing loans with p=0.10 and p=0.02, respectively).

When we instead present the same files as coming from female loan applicants (right-hand side), the higher-quality loan applications do *not* benefit from lower guarantor requirements at all. It appears that women are held to a higher standard: even in the case of high-quality loan applications, there is still a 30 per cent likelihood that a guarantor is requested. This is about the same percentage as for *low-quality* applications from male applicants. The data therefore show that it is among the best-performing loans that loan officers discriminate between male and female applicants in terms of guarantor requirements. A similar picture emerges when we split the sample into applicants with an above or below median subjective repayment probability (Appendix Figure A4, Panel A) or when we split the sample into applicants with low, median, or high ex ante credit risk as measured by their KKB credit score (Figure A4, Panel B). In both cases, gender discrimination in terms of requested guarantors is concentrated among applications with less ex ante credit risk.

In Table 5 we perform this analysis parametrically. Column 1 confirms that even when we control for loan officer covariates as well as file and city fixed effects, women are 12.4 percentage points more likely to be asked for a guarantor in case of high-quality loans (column 1). This difference is absent for loans that were either rejected or non-performing in real life (column 2). The indirect gender discrimination that loan officers display in terms of requesting guarantors is therefore not driven by the lower-quality segment of the application pool. Instead, double standards are applied in the case of relatively good loans that were paid back in real life.

Similar to Table 4, we can assess which participant types are responsible for this gender discrimination. Columns 3 to 14 show similar heterogeneity as before. High-quality female loan applications are about 15 percentage points more likely to be asked for a guarantor compared to identical male applications if the participant is relatively inexperienced (columns 5-6); relatively young (columns 7-8); a loan officer rather than a supervisor (columns 9-10); relatively risk averse (columns 11-12); and has a relatively high implicit gender bias (columns

13-14).[32] The difference in the sub-sample coefficients is large throughout Table 5 but less precisely estimated due to the smaller sample (performing loans only). This is reflected in the t-test p-values at the bottom of Table 5.

In summary, the discriminatory behavior of less experienced and more biased loan officers in terms of guarantor requirements is concentrated among loans that were repaid in real life, not among low-quality loans. This suggests that loan officers, especially less experienced and more biased ones, resort to the applicant's gender as a heuristic when there are no clear indications that a loan is risky.

In unreported further analysis, we assess whether the (randomized) gender of the applicant has an impact on the ability of loan officers to correctly identify loans that were non-performing in real life (that is, avoiding type II errors) and/or to correctly identify loans that performed well in real life (avoiding type I errors). Overall, we find that applicant gender does not affect either form of prediction accuracy. This holds for all types of loan officers. The one exception is that loan officers are slightly less successful in identifying performing files when we present these as coming from a female entrepreneur. This effect is driven by relatively inexperienced loan officers. This is in line with the results in Table 5 showing that inexperienced loan officers are more likely to require a guarantor from women in the case of loans that performed well in real life.

4.5 Gender and the availability of applicant information

In the second round of the experiment, we randomize the type of information that loan officers have access to. When a correspondence study is rich in applicant characteristics, statistical discrimination is less likely (Neumark, 2018). Vice versa, withholding some information may increase gender differences in loan conditions. We would interpret such a finding as evidence for statistical discrimination in which loan officers rely more on applicants' group membership (gender) whenever less individual information is available. For instance, in the model of Aigner and Cain (1977), statistical discrimination can arise when decision makers put less weight on individual indicators and more weight on group means for the group for which signals about individual productivity are relatively weak. In our context, this would imply that loan officers may rely more on (their subjective expectations of) gender-specific

[32] There is, again, no difference by participant gender: the coefficients in columns 3 and 4 are not significantly different.

average repayment probabilities when we experimentally weaken the signal about individual loan applications.[33]

Table 6 assesses whether restricting the information that is available to loan officers has a disproportionate impact on female loan applications, all else equal. In columns 1 and 2 (3 and 4), we present linear probability regressions where the dependent variable is our *Rejection dummy* (*Guarantor dummy*). In specifications 1 and 3, we include dummy variables that indicate whether in a particular decision we randomly withheld subjective (*No subj.*) or objective (*No. obj.*) loan application information. The former refers to subjective information that had been voluntarily entered by loan officers (in real life) at the earliest stage of client contact. The latter is the credit score from the Turkish credit registry. All specifications include our standard participant covariates and fixed effects.

The main take-away from Table 6 is that varying the available objective and subjective applicant information does not affect gender discrimination.[34] The interaction terms between *Female loan applicant* and the information treatments are not statistically significant in columns 2 and 4. We do find, however, that withholding subjective borrower information increases the likelihood of loan rejection by 6 percentage points (column 1). Yet, this effect does not differ between male and female loan applicants (column 2). The subjective information that loan officers can voluntarily add to an application file thus increases the willingness to lend among those who subsequently review the file.

Where does this leave us in terms of the nature of the discrimination we detect in the form of gender-biased guarantor requirements? First, while we cannot explicitly rule out taste-based discrimination, this is unlikely to drive the gender-biased guarantor requirements we observe. If taste-based discrimination would be important, we would expect it to already manifest itself in the provisional loan rejection decisions. Second, the fact that discrimination in guarantor requirements is concentrated among loan officers with a high IAT score is a clear sign that implicit gender bias plays a role. Third, our findings tell a nuanced story as regards statistical discrimination. On the one hand, the results in Table 6 indicate that experimental variation in the amount of information does not impact loan officer behavior in a gender-

[33] Kaas and Manger (2012) use a correspondence study in Germany to study the effect of Turkish-sounding names on call back rates for job interviews. Discrimination was eliminated when a reference letter, containing indirect information about productivity (such as applicants' conscientiousness and agreeableness) was added. The authors interpret this evidence as consistent with statistical discrimination.

[34] Unreported results show this holds regardless of the strength of the implicit gender bias of loan officers.

specific way. This goes against a role for statistical discrimination. On the other hand, loan officers' experience does significantly reduce gender-biased guarantor requirements. This does suggest that learning from experience over time can make decision making more objective.

5 Conclusions

Across many emerging markets women are much less likely to use formal financial services. An important policy question is whether (part of) this gender gap reflects supply-side constraints in general and lender discrimination in particular. To the extent that the latter is the case, a lack of finance may prevent women from developing their entrepreneurial potential.

To assess the empirical relevance and mechanisms of gender discrimination in small business lending, we implemented a randomized lab-in-the-field experiment with a large sample of Turkish loan officers. The files assessed in the experiment were real-life loan applications. Accordingly, we were able to directly link experimental choices to real-world outcomes and to pay participating loan officers performance incentives based on the realized outcome of the loans they approved. Because gender was randomly (re-)assigned to each application, the same application was sometimes linked to a male and sometimes to a female entrepreneur.

While we do not find any evidence of different unconditional approval rates of loan applications from male versus female entrepreneurs, there exists a more subtle form of discrimination. All else equal, the approval of female applications is up to 30 per cent more likely to be made conditional on the presence of a guarantor. This indirect gender discrimination is almost exclusively concentrated among loans that in real life performed well, making it costly to the bank. A number of loan officer traits—their experience, risk aversion, and implicit gender bias—turn out to be strongly correlated with indirect discrimination. These deeper (and usually unobservable) traits appear to be more important than loan officers' gender. For instance, while we find that female loan officers are significantly more gender biased on average, it is this variation in gender bias that drives indirect discrimination rather than loan officers' gender as such. These results warn against drawing too strong conclusions about the relevance of loan officer gender for lending outcomes (including discriminatory behavior), without assessing deeper characteristics such as risk aversion, experience and implicit biases.

Overall, our results are most in line with models of implicit and, less clear-cut, statistical discrimination. The finding that discrimination in guarantor requirements is concentrated among loan officers with a high IAT score is a clear indicator that implicit gender bias plays a role. We find more mixed evidence for statistical bias. On the one hand, varying the information available during the experiment, thus manipulating the signal strength about individual applicants, did not have a gender-specific impact. In the presence of statistical gender discrimination, we would have expected stronger effects from this treatment. We also show that discriminatory lending decisions do not boost loan quality as one would expect in a model of statistical discrimination. On the other hand, we find that discriminatory lending decisions are concentrated among less experienced loan officers (even when controlling for age). Learning through experience can mitigate or even eliminate statistical discrimination over time (Aigner and Cain, 1977; Altonji and Pierret, 2001).[35] Thus, our results may reflect a form of statistical bias that is mitigated by the cumulative impact of learning from experience, rather than by information provided in real time.

Our findings suggest a number of policy options for banks that want to prevent or mitigate gender discrimination among their loan officers. First, interventions that increase (the reliability of) information about loan applicants may not be of first-order importance. Our results instead suggest that discrimination may be less prevalent among more experienced loan officers who rely less on mental short-cuts. Adding such officers to relatively junior teams may be an effective way to reduce the risk of discriminatory lending practices. Supervisors and branch managers may also be tasked with monitoring and, where necessary, challenging the guarantor decisions of more junior colleagues.

Second, policies to mitigate the real-world impact of loan officers' implicit biases may be called for. This could include simply making sure that loan officers have sufficient time to evaluate loan applications. Banks could also set bank-wide or branch-wide goals for lending to women without a guarantor. Management could then hold those that deviate from this norm accountable. Third, banks can make successful female entrepreneurs more visible to loan officers, for instance by integrating them in banks' internal communication and training programs. Interventions like these may be more effective than explicit diversity training, which makes gender differences more salient and can even generate a backlash (Bohnet, 2016). Fourth, banks could also conduct IATs with loan officers and reveal the results to those who hold implicit stereotypes in order to counteract biased lending behavior. Alternatively,

[35]Earlier experimental evidence indicates that more experienced loan officers also acquire more financial information when lending to fictitious small firms (Andersson, 2004).

they could provide loan officers with factual information about gender discrimination in the loan officer population as a whole.[36] Measuring the relative (cost) effectiveness of such interventions to contain the negative impact of implicit gender bias among loan officers provides a fruitful area for future experimental research.

[36] Alesina et al. (2018) test how the former intervention mitigates bias among teachers who evaluate children while Boring and Philippe (2019) test the latter intervention among students who evaluate their teachers.

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2017). When Should You Adjust Standard Errors For Clustering? NBER Working Paper No. 24003.

Agnew, J., Balduzzi, P., and Sunden, A. (2003). Portfolio Choice and Trading in a Large 401(k) Plan. *American Economic Review*, 93(1):193–215.

Aigner, D. J. and Cain, G. G. (1977). Statistical Theories of Discrimination in Labor Markets. *Industrial and Labor Relations Review*, 30(2):175–187.

Alesina, A., Carlana, M., La Ferrara, E., and Pinotti, P. (2018). Revealing Stereotypes: Evidence from Immigrants in Schools. NBER Working Paper No. 25333.

Alesina, A. F., Lotti, F., and Mistrulli, P. E. (2013). Do Women Pay More for Credit? Evidence from Italy. *Journal of the European Economic Association*, 11(s1):45–66.

Alibhai, S., Donald, A., Goldstein, M., Oguz, A. A., Pankov, A., and Strobbe, F. (2019). Gender Bias in SME Lending. Experimental Evidence from Turkey. World Bank Policy Research Working Paper No. 9100.

Altonji, J. and Blank, R. (1999). *Handbook of Labor Economics*, volume 3, part C, chapter Race and Gender in the Labor Market, pages 3143–3259. Elsevier, New York.

Altonji, J. G. and Pierret, C. R. (2001). Employer Learning and Statistical Discrimination. *Quarterly Journal of Economics*, 116(1):313–350.

Andersson, P. (2004). Does Experience Matter in Lending? A Process-Tracing Study on Experienced Loan Officers' and Novices' Decision Behavior. *Journal of Economic Psychology*, 25(4):471–492.

Arrow, K. J., Ashenfelter, O., and Rees, A. (1973). Discrimination in Labor Markets. *The Theory of Discrimination*, pages 3–33.

Asiedu, E., Freeman, J. A., and Nti-Addae, A. (2012). Access to Credit by Small Businesses: How Relevant are Race, Ethnicity, and Gender? *American Economic Review*, 102(3):532–37.

Banerjee, A. V., Besley, T., and Guinnane, T. W. (1994). Thy Neighbor's Keeper: The Design of a Credit Cooperative with Theory and a Test. *Quarterly Journal of Economics*, 109(2):491–515.

Bayard, K., Hellerstein, J., Neumark, D., and Troske, K. (2003). New Evidence on Sex Segregation and Sex Differences in Wages from Matched Employee-Employer Data. *Journal of Labor Economics*, 21(4):887–922.

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., and Topalova, P. (2009). Powerful Women: Does Exposure Reduce Bias? *Quarterly Journal of Economics*, 124(4):1497–1540.

Beck, T., Behr, P., and Guettler, A. (2012). Gender and Banking: Are Women Better Loan Officers? *Review of Finance*, 17(4):1279–1321.

Beck, T., Behr, P., and Madestam, A. (2018). Sex and Credit: Do Gender Interactions Matter for Credit Market Outcomes? *Journal of Banking and Finance*, 87:380–396.

Becker, G. S. (1957). *The Economics of Discrimination*. University of Chicago Press.

Bellucci, A., Borisov, A., and Zazzaro, A. (2010). Does Gender Matter in Bank–Firm Relationships? Evidence from Small Business Lending. *Journal of Banking and Finance*, 34(12):2968–2984.

Bertrand, M., Chugh, D., and Mullainathan, S. (2005). Implicit Discrimination. *American Economic Review*, 95(2):94–98.

Bertrand, M. and Duflo, E. (2017). *Field Experiments on Discrimination*, volume 1, pages 309–393. Elsevier.

Bertrand, M. and Mullainathan, S. (2004). Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, 94(4):991–1013.

Blanchard, L., Zhao, B., and Yinger, J. (2008). Do Lenders Discriminate Against Minority and Woman Entrepreneurs? *Journal of Urban Economics*, 63(2):467–497.

Blanchflower, D. G., Levine, P. B., and Zimmerman, D. J. (2003). Discrimination in the Small-Business Credit Market. *Review of Economics and Statistics*, 85(4):930–943.

Bohnet, I. (2016). *What Works. Gender Equality by Design*. Harvard University Press.

Bond, P. and Rai, A. S. (2008). Cosigned vs. Group Loans. *Journal of Development Economics*, 85(1-2):58–80.

Boring, A. and Philippe, A. (2019). Reducing Discrimination in the Field: Evidence from an Awareness Raising Intervention Targeting Gender Biases in Student Evaluations of Teaching. Technical report, mimeo.

Carlana, M. (2019). Implicit Stereotypes: Evidence from Teachers' Gender Bias. *Quarterly Journal of Economics*, 134(3):1163–1224.

Carney, D., Nosek, B. A., Greenwald, A. G., Banaji, M. R., Baumeister, R., and Vohs, K. (2007). Implicit Association Test (IAT). *Encyclopedia of Social Psychology*.

Cole, S., Kanz, M., and Klapper, L. (2015). Incentivizing Calculated Risk-Taking: Evidence from an Experiment with Commercial Bank Loan Officers. *Journal of Finance*, 70(2):537–575.

Cornell, B. and Welch, I. (1996). Culture, Information, and Screening Discrimination. *Journal of Political Economy*, 104(3):542–571.

Demirgüç-Kunt, A., Honohan, P., and Beck, T. (2008). *Finance for All?: Policies and Pitfalls in Expanding Access*. World Bank Policy Research Report, World Bank, Washington, D.C.

Demirgüç-Kunt, A., Klapper, L., Singer, D., Ansar, S., and Hess, J. (2018). *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*. World Bank, Washington, D.C.

Eckel, C. C. and Grossman, P. J. (2008). Differences in the Economic Decisions of Men and Women: Experimental Evidence. *Handbook of Experimental Economics Results*, 1:509–519.

Ferguson, M. F. and Peters, S. R. (1995). What Constitutes Evidence of Discrimination in Lending? *Journal of Finance*, 50(2):739–748.

Gneezy, U. and Imas, A. (2017). Lab in the Field: Measuring Preferences in the Wild. In *Handbook of Economic Field Experiments*, volume 1, pages 439–464. Elsevier.

Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. (1998). Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74(6):1464.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., and Banaji, M. R. (2009). Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity. *Journal of Personality and Social Pychology*, 97(1):17.

Heilman, M. E. (1984). Information as a Deterrent Against Sex Discrimination: The Effects of Applicant Sex and Information Type on Preliminary Employment Decisions. *Organizational Behavior and Human Performance*, 33(2):174–186.

Kaas, L. and Manger, C. (2012). Ethnic Discrimination in Germany's Labour Market: A Field Experiment. *German Economic Review*, 13(1):1–20.

Klapper, L. and Singh, S. (2014). The Gender Gap in the Use of Financial Services in Turkey. mimeo.

Munnell, A. H., Tootell, G. M., Browne, L. E., and McEneaney, J. (1996). Mortgage Lending in Boston: Interpreting HMDA Data. *American Economic Review*, 86(1):25–53.

Neal, D. A. and Johnson, W. R. (1996). The Role of Premarket factors in Black-White Wage Differences. *Journal of Political Economy*, 104(5):869–895.

Neumark, D. (2018). Experimental Research on Labor Market Discrimination. *Journal of Economic Literature*, 56(3):799–866.

Ongena, S. and Popov, A. (2016). Gender Bias and Credit Access. *Journal of Money, Credit and Banking*, 48(8):1691–1724.

Phelps, E. S. (1972). The Statistical Theory of Racism and Sexism. *American Economic Review*, 62(4):659–661.

Reuben, E., Sapienza, P., and Zingales, L. (2014). How Stereotypes Impair Women's Careers in Science. *Proceedings of the National Academy of Sciences*, 111(12):4403–4408.

Sunden, A. E. and Surette, B. J. (1998). Gender Differences in the Allocation of Assets in Retirement Savings Plans. *American Economic Review*, 88(2):207–211.

The Economist (2013). Discrimination Abounds. Women are Being Excluded from Finance Across the Developing World. https://www.economist.com/free-exchange/2013/11/19/discrimination-abounds. [Accessed 2019-09-24].

World Economic Forum (2018). The Global Gender Gap Report 2018. https://www.weforum.org/reports/the-global-gender-gap-report-2018. [Accessed 2019-09-24].

Tables and Figures

Table 1: Summary statistics

	N	Mean	Median	Sd.	Min	Max
Panel A: Participant characteristics						
Participant is female	332	0.47	0.00	0.50	0	1
Participant experience (years)	324	4.99	4.00	3.89	0	19
Participant age (years)	321	37.30	36.00	5.84	26	53
Participant is supervisor	334	0.43	0.00	0.50	0	1
Participant risk aversion	333	4.11	4.00	1.37	1	6
Participant gender bias (IAT)	325	0.33	0.34	0.32	-0.93	1.00
Panel B: Loan-file characteristics						
Real life performing	100	0.50	0.5	0.50	0	1
Real life non-performing (NPL)	100	0.25	0	0.44	0	1
Real life declined	100	0.25	0	0.44	0	1
Panel C: Decision characteristics						
First round						
Rejection dummy	1,336	0.39	0.00	0.49	0	1
Guarantor dummy	814	0.27	0.00	0.44	0	1
Subjective repayment probability	1,329	60.11	70.00	30.81	0	100
Second round						
Rejection dummy	1,334	0.36	0.00	0.48	0	1
Guarantor dummy	860	0.27	0.00	0.44	0	1
Subjective repayment probability	1,324	61.48	70.00	30.41	0	100
Panel D: Treatment characteristics (second round)						
No subj.	1,334	0.34	0	0.47	0	1
No obj.	1,334	0.33	0	0.47	0	1

Notes: This table displays summary statistics for the variables used in the empirical analysis. Panel A summarizes the main characteristics of all participants who took part in the experiment. Panel B displays summary statistics for the 100 loan application files used in the experiment. Panel C displays summary statistics at the decision level (participant-file combination). Panel D displays summary statistics for the information treatments in the second round of the experiment.

Table 2: Applicant gender and loan rejection

Dependent variable: Rejection dummy				
	[1]	[2]	[3]	[4]
Female applicant	-0.013	-0.013	-0.010	-0.010
	(0.023)	(0.023)	(0.024)	(0.024)
Participant is female	0.023	0.029	0.021	0.026
	(0.023)	(0.023)	(0.024)	(0.024)
Participant experience (years)	-0.002	-0.003	-0.003	-0.003
	(0.004)	(0.004)	(0.004)	(0.005)
Participant age (years)	-0.005*	-0.005*	-0.005*	-0.005*
	(0.003)	(0.003)	(0.003)	(0.003)
Participant is supervisor	0.100***	0.101***	0.099***	0.100***
	(0.032)	(0.032)	(0.032)	(0.032)
Participant risk aversion		-0.012		-0.012
		(0.010)		(0.010)
Participant IAT score			-0.000	-0.003
			(0.044)	(0.044)
Constant	0.552***	0.604***	0.553***	0.607***
	(0.098)	(0.103)	(0.101)	(0.107)
File FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
R-squared	0.014	0.015	0.014	0.015
N	1,272	1,272	1,240	1,240

Notes: The dependent variable is a *Rejection dummy* that equals '1' if the participant declines the credit application and '0' if the participant approves it. The sample is restricted to the first round of the experiment. Robust standard errors are in parentheses. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively.

Table 3: Applicant gender and guarantor requirements

Dependent variable: Guarantor dummy				
	[1]	[2]	[3]	[4]
Female applicant	0.068**	0.068**	0.069**	0.070**
	(0.029)	(0.029)	(0.030)	(0.030)
Participant is female	-0.026	-0.033	-0.020	-0.027
	(0.032)	(0.031)	(0.033)	(0.032)
Participant experience (years)	0.002	0.003	0.003	0.003
	(0.005)	(0.005)	(0.006)	(0.006)
Participant age (years)	0.002	0.002	0.001	0.002
	(0.004)	(0.004)	(0.004)	(0.004)
Participant is supervisor	0.036	0.035	0.044	0.042
	(0.042)	(0.042)	(0.043)	(0.043)
Participant risk aversion		0.014		0.015
		(0.012)		(0.013)
Participant gender bias (IAT)			-0.038	-0.038
			(0.063)	(0.062)
Constant	0.036	-0.031	0.065	-0.007
	(0.120)	(0.137)	(0.121)	(0.138)
File FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
R-squared	0.064	0.063	0.062	0.061
N	772	772	752	752

Notes: The dependent variable is a *Guarantor dummy* that equals '1' if the participant approves the credit application but requests a guarantor and '0' if the participant approves it without requesting a guarantor. The sample is restricted to the first round of the experiment. Robust standard errors are in parentheses. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively.

Table 4: Applicant gender and guarantor requirements: Participant heterogeneity

Dependent variable: Guarantor dummy

	Participant gender		Participant experience		Participant age	
	Female	Male	Below median	Above median	Below median	Above median
	[1]	[2]	[3]	[4]	[5]	[6]
Female applicant	0.082	0.078	0.106**	0.032	0.121**	0.013
	(0.052)	(0.049)	(0.052)	(0.046)	(0.050)	(0.040)
R-squared	0.107	0.080	0.097	0.077	0.136	0.037
N	338	414	341	411	325	427
t-test p-value	0.473		0.108		0.035	

	Participant position		Participant risk aversion		Participant gender bias	
	Officer	Supervisor	Below median	Above median	Below median	Above median
	[7]	[8]	[9]	[10]	[11]	[12]
Female applicant	0.130***	-0.022	0.067	0.087*	0.022	0.119**
	(0.038)	(0.061)	(0.065)	(0.044)	(0.051)	(0.046)
R-squared	0.117	0.034	0.161	0.041	0.063	0.090
N	471	281	214	538	381	371
t-test p-value	0.008		0.389		0.055	
Participant covariates	Yes	Yes	Yes	Yes	Yes	Yes
File FE	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is a *Guarantor dummy* that equals '1' if the participant approves the credit application but requests a guarantor and '0' if the participant approves it without requesting a guarantor. The sample is restricted to the first round of the experiment. When partitioning continuous variables the "Below median" sample corresponds to strictly below the median while the "Above median" sample corresponds to values at the median and above. For the *Participant risk aversion* variable, higher values indicate greater risk aversion so that participants with above median risk aversion are the most risk averse. *Participant gender bias* measures implicit gender bias based on an implicit association test (IAT). Higher IAT values indicate that participants associate men more with careers and women more with household tasks. The t-test *p-value* corresponds to one-sided tests. Robust standard errors are in parentheses. All regressions include the same participant covariates as in column (4) of Table 2. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively.

Table 5: Applicant gender, guarantor requirements, and real-life loan performance

Dependent variable: Guarantor dummy								
	All		Performing loans					
	Loan in real life		Participant gender		Participant experience		Participant age	
	Performing	NPL & Declined	Female	Male	Below median	Above median	Below median	Above median
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Female applicant	0.124***	-0.022	0.119**	0.113	0.145**	0.076	0.157**	0.092*
	(0.040)	(0.047)	(0.057)	(0.071)	(0.063)	(0.057)	(0.069)	(0.049)
R-squared	0.083	0.064	0.139	0.114	0.132	0.100	0.170	0.063
N	449	303	207	242	208	241	201	248
t-test p-value	0.008		0.466		0.175		0.196	
			Participant position		Participant risk aversion		Participant gender bias	
			Officer	Supervisor	Below median	Above median	Below median	Above median
			[9]	[10]	[11]	[12]	[13]	[14]
Female applicant			0.161***	0.035	0.059	0.137**	0.102	0.161***
			(0.051)	(0.077)	(0.078)	(0.057)	(0.063)	(0.053)
R-squared			0.174	0.062	0.243	0.059	0.105	0.130
N			280	169	122	327	217	232
t-test p-value			0.060		0.171		0.207	
Participant covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
File FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

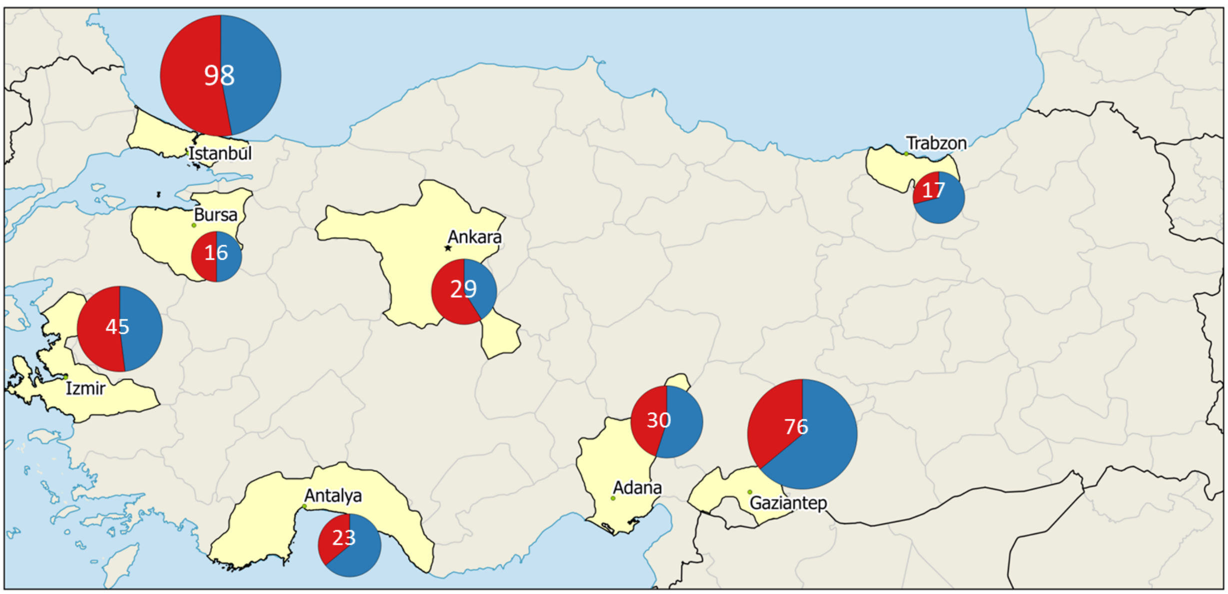
Notes: The dependent variable is a *Guarantor dummy* that equals '1' if the participant approves the credit application but requests a guarantor and '0' if the participant approves it without requesting a guarantor. The sample is restricted to the first round of the experiment. When partitioning continuous variables the "Below median" sample corresponds to strictly below the median while the "Above median" sample corresponds to values at the median and above. For the *Participant risk aversion* variable, higher values indicate greater risk aversion so that participants with above median risk aversion are the most risk averse. *Participant gender bias* measures implicit gender bias based on an implicit association test (IAT). Higher IAT values indicate that participants associate men more with careers and women more with household tasks. The t-test p-value corresponds to one-sided tests. Robust standard errors are in parentheses. All regressions include the same participant covariates as in column (4) of Table 2. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively.

Table 6: Availability of borrower information and gender bias

Dependent variable:	Rejection dummy		Guarantor dummy	
	[1]	[2]	[3]	[4]
Female applicant	-0.017	0.014	0.035	0.023
	(0.022)	(0.044)	(0.031)	(0.051)
No subj.	0.055*	0.087**	-0.031	-0.053
	(0.029)	(0.041)	(0.037)	(0.050)
No obj.	-0.049	-0.035	-0.017	-0.014
	(0.036)	(0.048)	(0.033)	(0.048)
No subj. × Female applicant		-0.064		0.043
		(0.063)		(0.074)
No obj. × Female applicant		-0.028		-0.006
		(0.068)		(0.074)
Participant covariates	Yes	Yes	Yes	Yes
File FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
R-squared	0.036	0.037	0.066	0.066
N	1,238	1,238	802	802

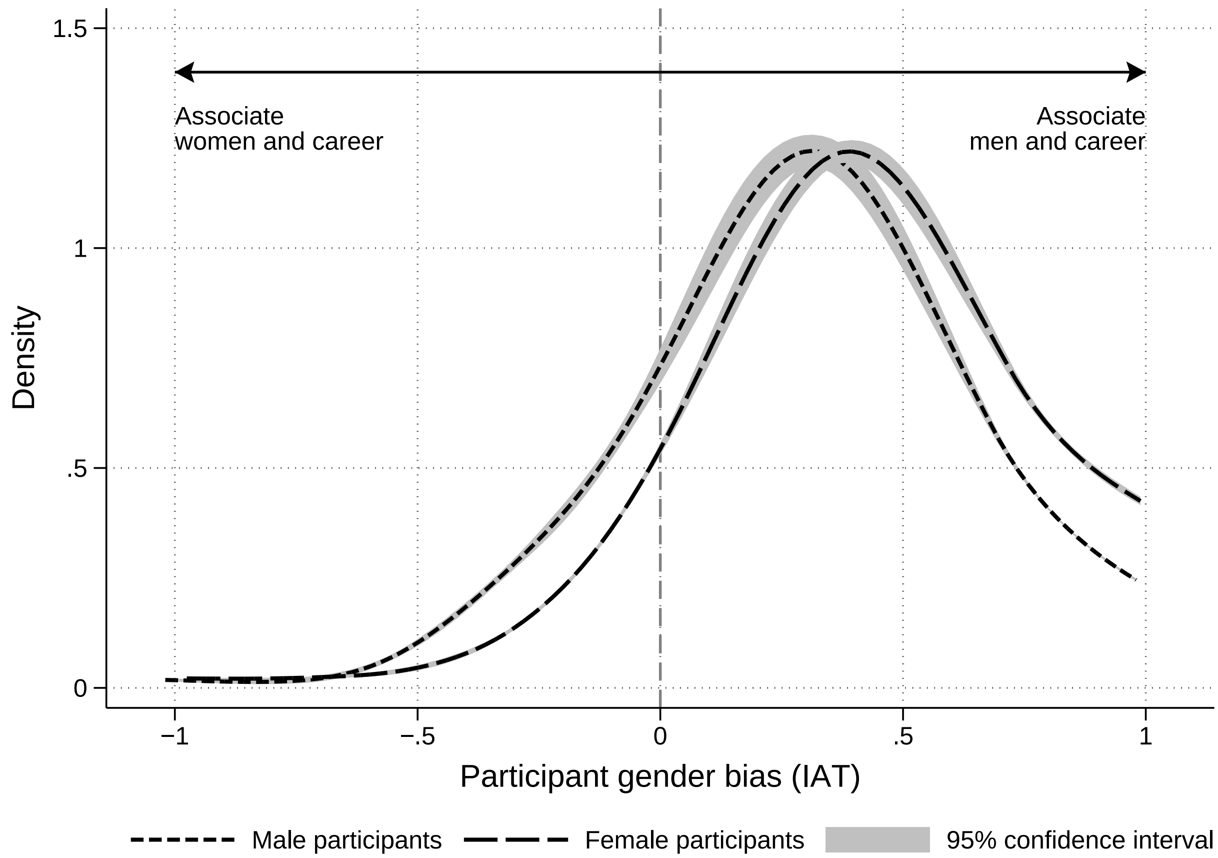
Notes: The dependent variable in columns [1] and [2] is a *Rejection dummy* that equals '1' if the participant declines the credit application and '0' if the participant approves it. The dependent variable in columns [3] and [4] is a *Guarantor dummy* that equals '1' if the participant approves the credit application but requests a guarantor and '0' if the participant approves it without requesting a guarantor. The sample is restricted to the second round of the experiment. All regressions include the same participant covariates as in column (4) of Table 2. Robust standard errors are in parentheses. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively.

Figure 1: Geographical distribution of participants across the Turkish bank branches



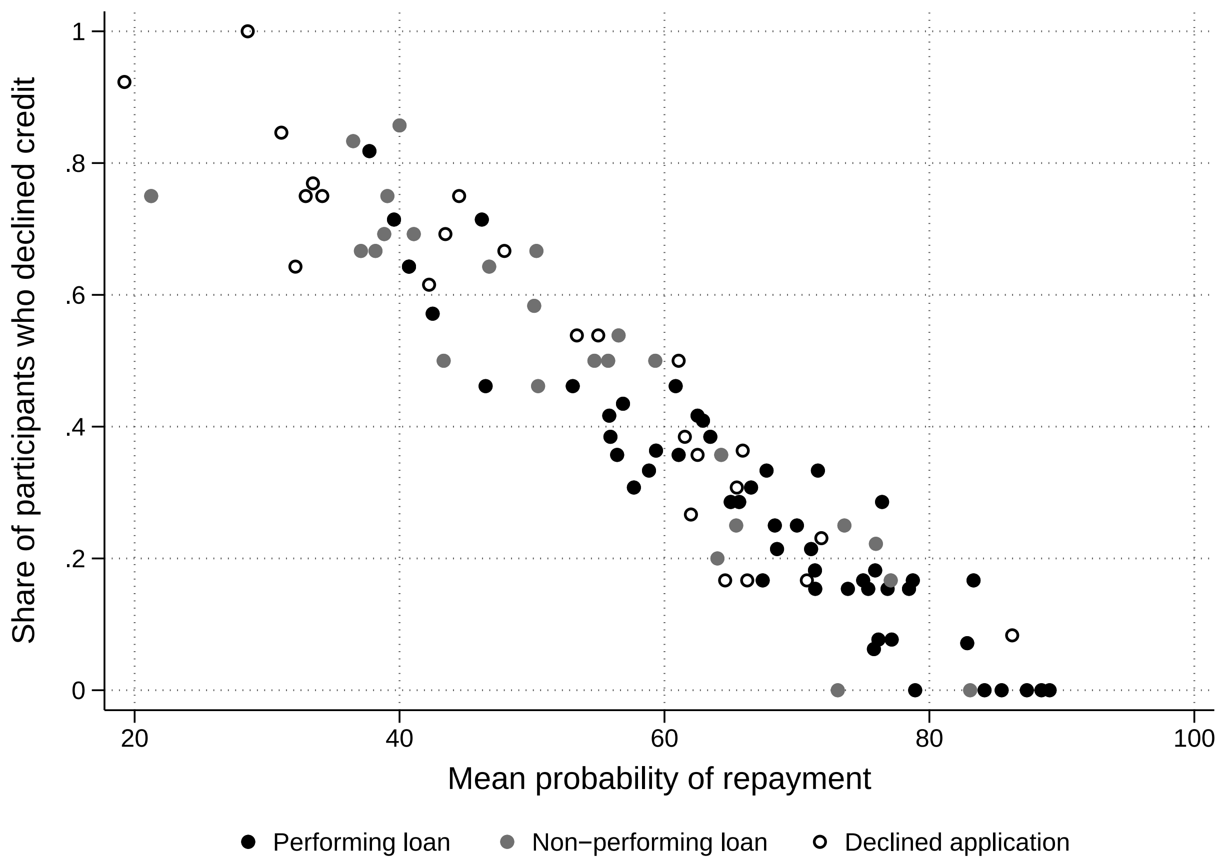
Notes: This map shows the number and gender composition of the participants in the eight Turkish regional bank branches that participated in the experiment. Circle size is proportional to the number of participants. The percentage of female (male) participants is shown in red (blue).

Figure 2: Participant gender bias (IAT), by participant sex



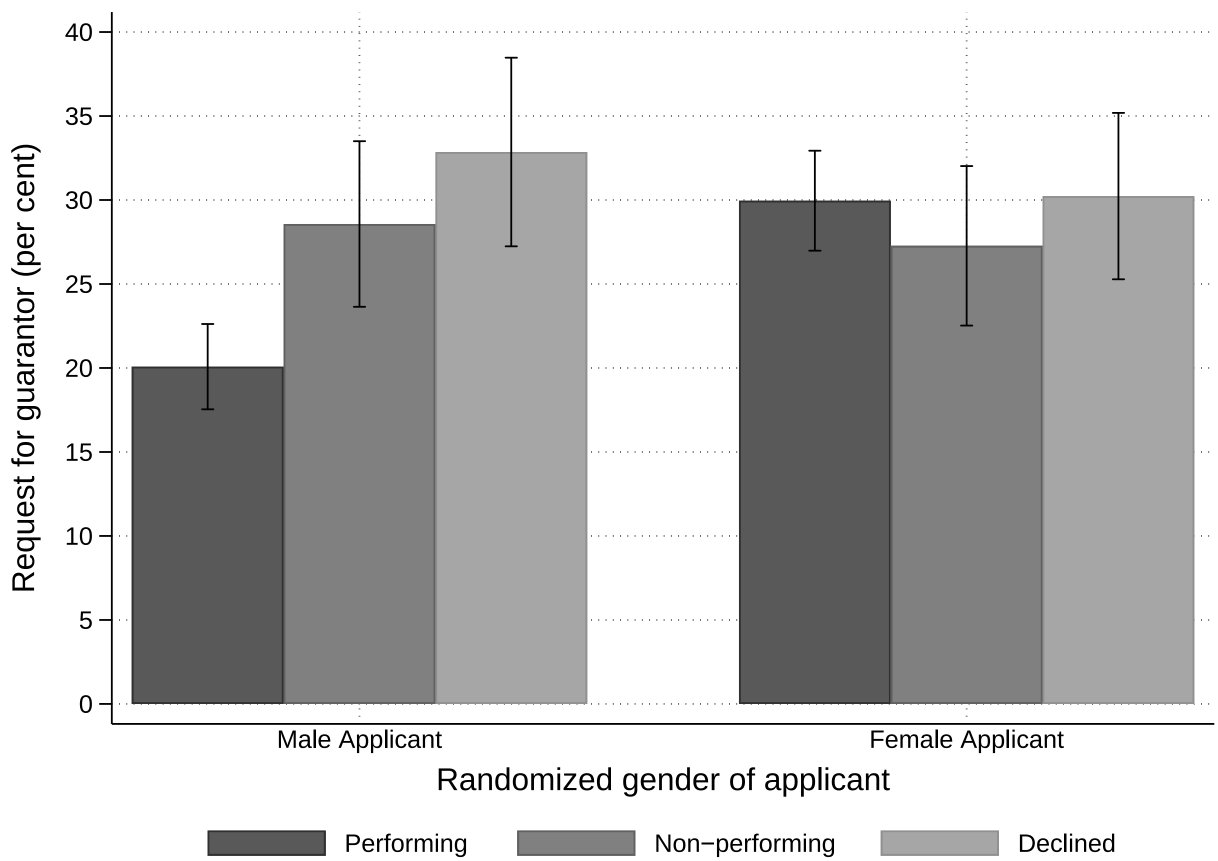
Notes: This figure shows a local polynomial smooth with 95 per cent confidence intervals of the variable *Participant gender bias (IAT)* for male (short dash) and female (long dash) participants, respectively. The combined two-sample Kolmogorov-Smirnov test statistic is 0.181 and has a p-value of 0.01.

Figure 3: Expected repayment and loan rejection rates



Notes: The x-axis is the within-file mean, across participants, of the subjective repayment probability. The y-axis is the share of participants who declined the loan application. The figure is based on the first round of the experiment only.

Figure 4: Guarantor requirements, by loan quality and applicant sex



Notes: This figure shows the percentage of loan applications approved during the experiment and for which participants requested a guarantor. Bars are shown for approved loans repaid in real life (dark gray), approved loans that were defaulted on in real life (medium gray), and loan applications rejected in real life (light gray). Bars indicate applications that were shown to participants as coming from a female (right) or male (left) entrepreneur. Whiskers indicate one binomial standard error. The sample is restricted to the first round of the experiment.

Appendices

Table A1: Variable definitions

Panel A: Participant characteristics	
Participant is female	Dummy variable equal to 1 for female and 0 for male participants.
Participant experience (years)	Number of years the participant has been an employee in the bank's credit division.
Participant age (years)	Age of the participant in years.
Participant is supervisor	Dummy variable equal to 1 for participants who are a supervisor/branch manager, 0 for those who are a loan officer.
Participant risk aversion	Integer variable ranging from 1 to 6, with 1 indicating risk loving and 6 indicating the highest level of risk aversion.
Participant gender bias (IAT)	Takes values from -1 to 1. Positive (negative) values indicate that the participant associates careers and entrepreneurship with being male (female). A score of zero indicates no implicit gender bias.
Panel B: File characteristics	
Female applicant	Dummy variable equal to 1 if the randomized gender of the loan application is female and 0 otherwise.
Female applicant (original)	Dummy variable equal to 1 if the gender of the real-life loan application was originally female and 0 otherwise.
Real life performing	Dummy variable equal to 1 if the loan was performing in real life, 0 otherwise.
Real life NPL	Dummy variable equal to 1 if the loan was non-performing in real life, 0 otherwise.
Real life declined	Dummy variable equal to 1 if the loan application was declined by the lending staff in real life, 0 otherwise.
Micro	Dummy variable equal to 1 if the credit file was from a micro firm and 0 if the credit file was from an SME firm.
Log of credit demanded	Logarithm of the amount of credit requested by the applicant.
Credit score	Credit score as taken from the KKB credit registry. Higher values indicate less ex ante credit risk.

Table A1 continued on next page

Panel C: Decision characteristics	
Rejection dummy	Dummy variable equal to 1 if the participant rejects the loan application, 0 otherwise.
Guarantor dummy	Dummy variable equal to 1 if the participant offers credit conditional on the presence of a guarantor and 0 if the participant offers credit but does not request a guarantor.
Panel D: Treatment characteristics	
No subj.	Dummy variable equal to 1 if information subjectively provided by lending staff is removed from the loan application file, 0 otherwise.
No obj.	Dummy variable equal to 1 if objective information (the credit score) from the credit bureau is removed from the loan application file, 0 otherwise.

Table A2: Applicant gender and credit score

Dependent variable: Credit score					
	[1]	[2]	[3]	[4]	[5]
Female applicant (original)	-12.845	51.042	59.297	66.736	79.874
	(49.441)	(67.354)	(67.639)	(67.332)	(67.102)
Micro				-136.459*	-39.468
				(70.387)	(96.173)
Log of credit demand					68.671*
					(36.547)
Constant	1,035.730***	1,065.000***	964.336***	1,115.907***	299.568
	(29.942)	(0.000)	(138.865)	(158.469)	(486.487)
Sector FE	No	Yes	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	Yes
R-squared	0.000	0.212	0.233	0.250	0.273
N	243	243	243	243	243

Notes: The dependent variable is *Credit score* (taken from the KKB credit registry) where higher values indicate less ex ante credit risk. The sample includes the 250 loan files from which the 100 loan files used in the experiment were drawn. Robust standard errors are in parentheses. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively.

Table A3: Correlation matrix

	Participant is supervisor	Participant is female	Participant age (years)	Participant risk aversion	Participant experience (years)	Participant gender bias (IAT)	Female applicant	Rejection dummy
Participant is supervisor	1.000							
Participant is female	0.091**	1.000						
Participant age (years)	0.570***	0.032	1.000					
Participant risk aversion	0.027	0.153***	-0.007	1.000				
Participant experience (years)	0.360***	0.031	0.503***	0.026	1.000			
Participant gender bias (IAT)	0.093**	0.196***	0.086**	-0.007	0.014	1.000		
Female applicant	0.000	0.000	-0.000	0.000	-0.000	0.000	1.000	
Rejection dummy	0.075**	0.037	0.014	-0.013	-0.015	0.008	-0.017	1.000

Notes: The sample is restricted to the first round. *, **, *** indicate significance at the 10, 5 and 1 per cent level, respectively.

Table A4: Predictors of participant gender bias

Dependent variable: Participant gender bias (IAT)

	[1]
Participant is female	0.124***
	(0.036)
Participant experience (years)	-0.004
	(0.005)
Participant age (years)	0.004
	(0.004)
Participant is supervisor	0.033
	(0.044)
Participant risk aversion	-0.009
	(0.014)
Constant	0.164
	(0.146)
R-squared	0.049
N	310

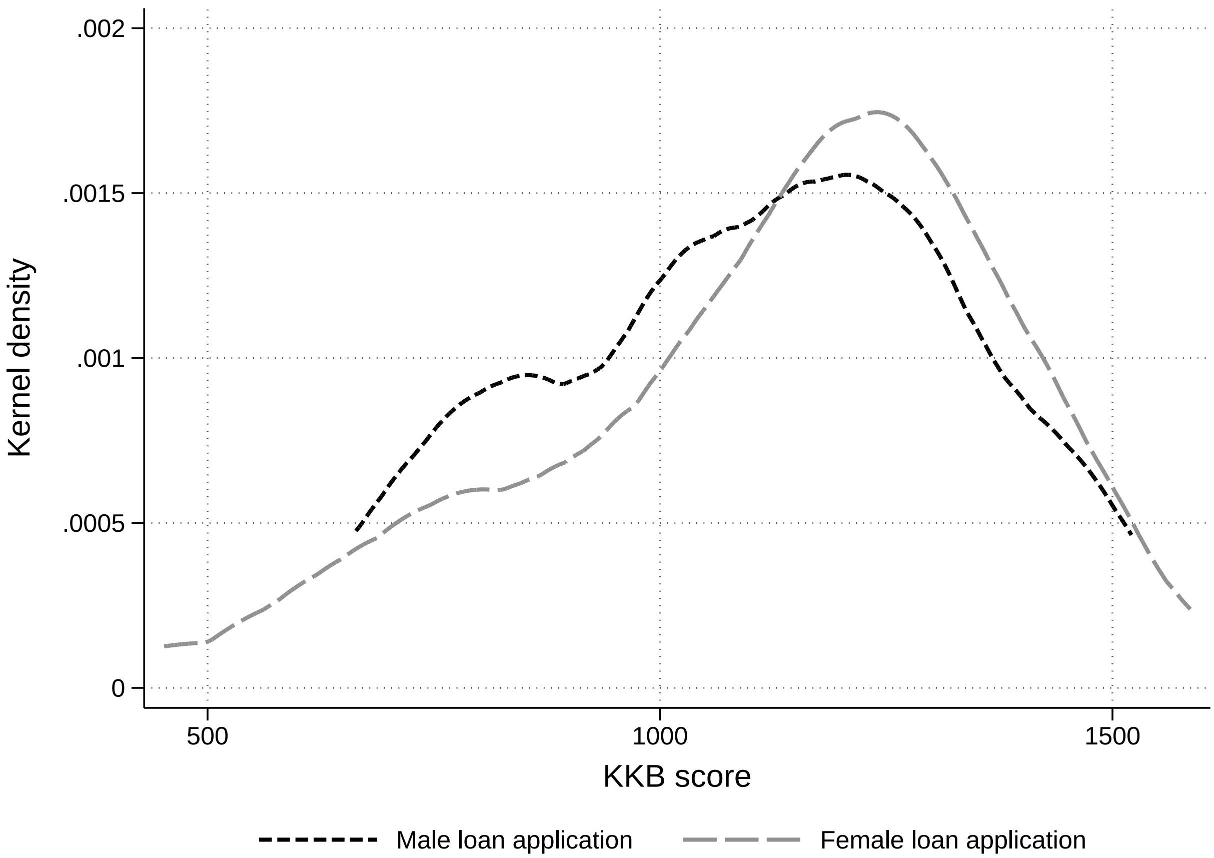
Notes: The dependent variable is *Participant gender bias (IAT)* which takes values from -1 to 1. Positive (negative) values indicate that the participant associates careers and entrepreneurship with being male (female). A score of zero indicates no implicit gender bias. The sample is restricted to the first round round of the experiment. Standard errors are in parentheses. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively.

Table A5: Gender and loan rejection: The role of supervisors

Dependent variable: Rejection dummy				
	[1]	[2]	[3]	[4]
Female applicant	-0.041	-0.041	-0.039	-0.039
	(0.029)	(0.029)	(0.029)	(0.029)
Participant is female	0.024	0.029	0.021	0.027
	(0.023)	(0.023)	(0.024)	(0.024)
Participant experience (years)	-0.003	-0.003	-0.003	-0.003
	(0.004)	(0.004)	(0.004)	(0.005)
Participant age (years)	-0.005*	-0.005*	-0.005*	-0.005*
	(0.003)	(0.003)	(0.003)	(0.003)
Participant is supervisor	0.067*	0.067*	0.064	0.065
	(0.039)	(0.040)	(0.040)	(0.040)
Female applicant × Participant is supervisor	0.067	0.068	0.070	0.071
	(0.044)	(0.045)	(0.046)	(0.046)
Participant risk aversion		-0.012		-0.012
		(0.010)		(0.010)
Participant IAT score			-0.000	-0.003
			(0.044)	(0.044)
Constant	0.564***	0.617***	0.566***	0.620***
	(0.099)	(0.105)	(0.102)	(0.108)
File FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
R-squared	0.016	0.017	0.016	0.017
N	1,272	1,272	1,240	1,240

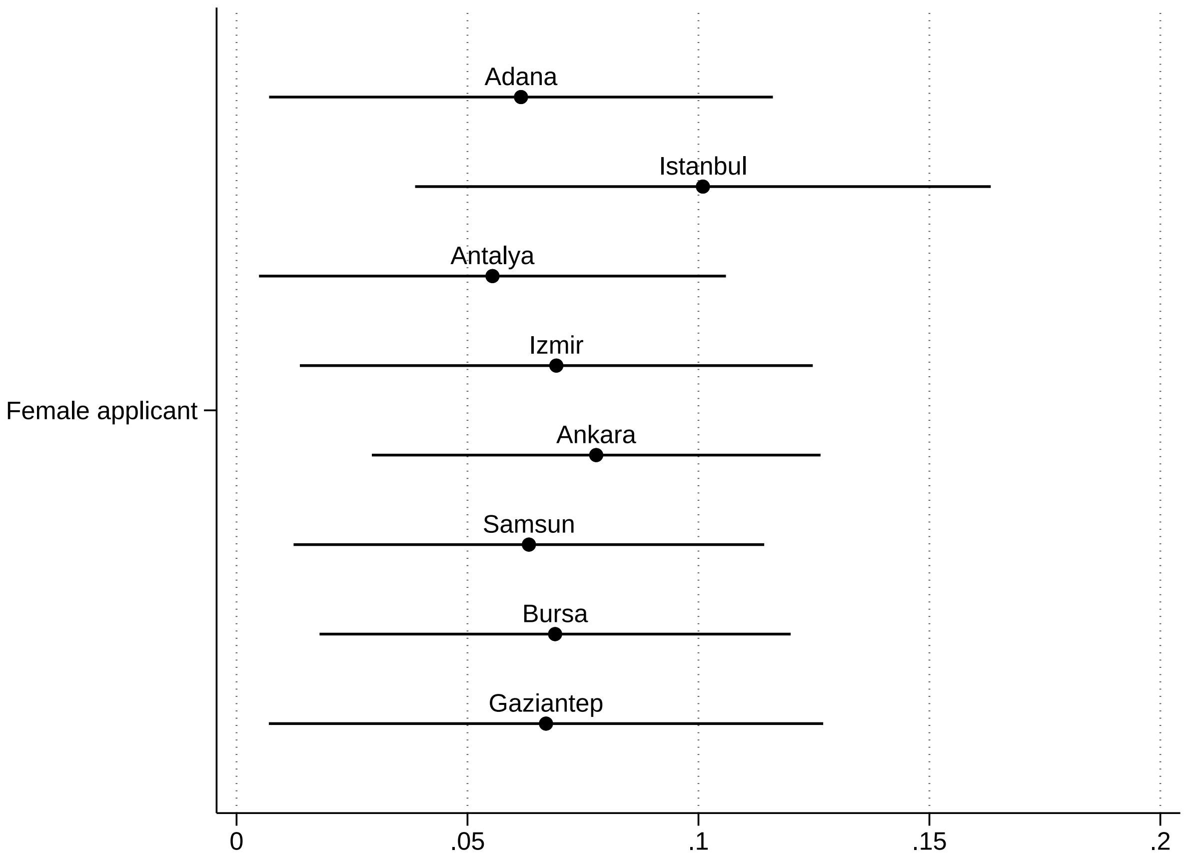
Notes: The dependent variable is a *Rejection dummy* that equals '1' if the participant declines the loan application and '0' if the participant approves it. The sample is restricted to the first round of the experiment. Robust standard errors are in parentheses. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively.

Figure A1: KKB credit score by original gender of loan application



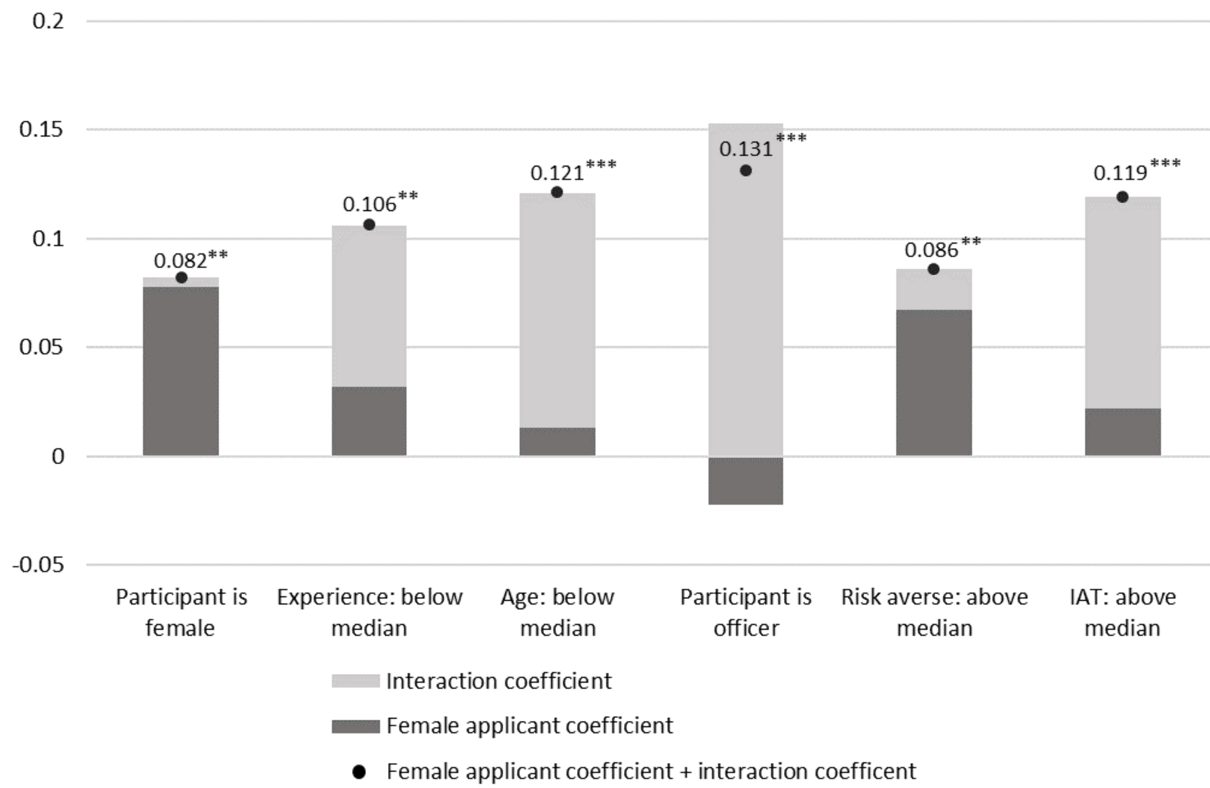
Notes: This figure shows the kernel density curves of the variable *KKB score* for loan application files that were male (short dash) and female (long dash) in real life, respectively. The figure is based on the 100 loan application files used in the experiment. The combined two-sample Kolmogorov-Smirnov test statistic is 0.152 and has a p-value of 0.728.

Figure A2: Indirect gender discrimination: City variation



Notes: This figure shows estimated coefficients for *Female applicant* using the same specification as in column [4] of Table 3. Each dot reflects the coefficient based on the full sample minus the observations from the indicated city. The dependent variable is a *Guarantor dummy* which equals '1' if the participant approved the credit application but requests a guarantor and '0' if the participant approved it without requesting a guarantor. The sample is restricted to the first round of the experiment. The horizontal lines reflect 90% level confidence intervals. Coefficients are ordered, from top to bottom, from highest to lowest regional household disposable income in 2016 (with Adana and Gaziantep having the highest and lowest income levels, respectively). Household disposable income is the total of disposable household income divided by household size and comes from the Turkish Statistical Institute's "Income and Living Conditions Survey Regional Results".

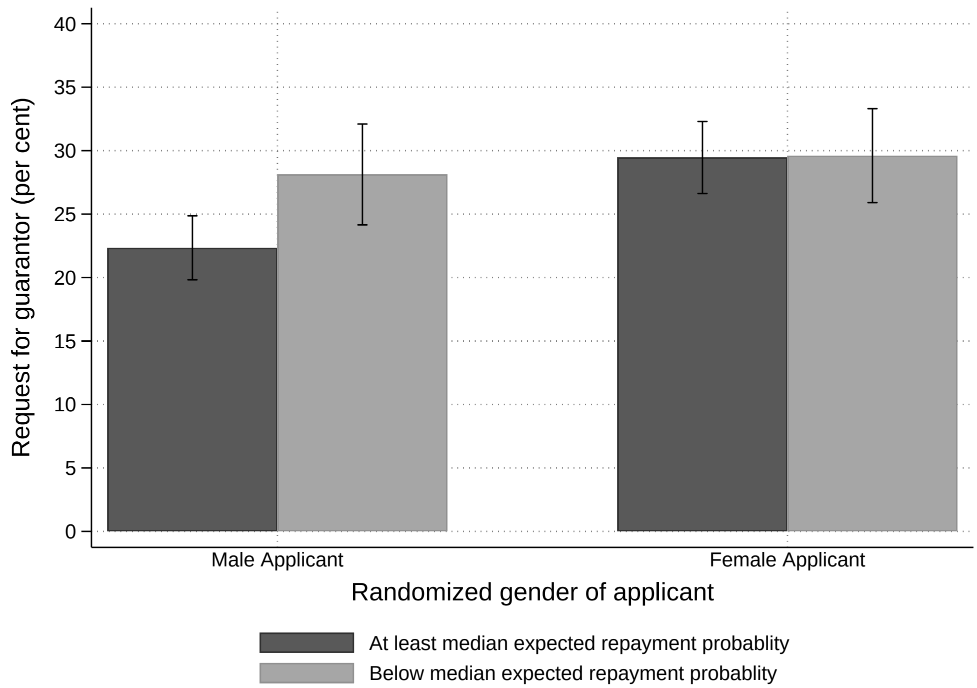
Figure A3: Heterogeneous guarantor requirements: Fully interacted models



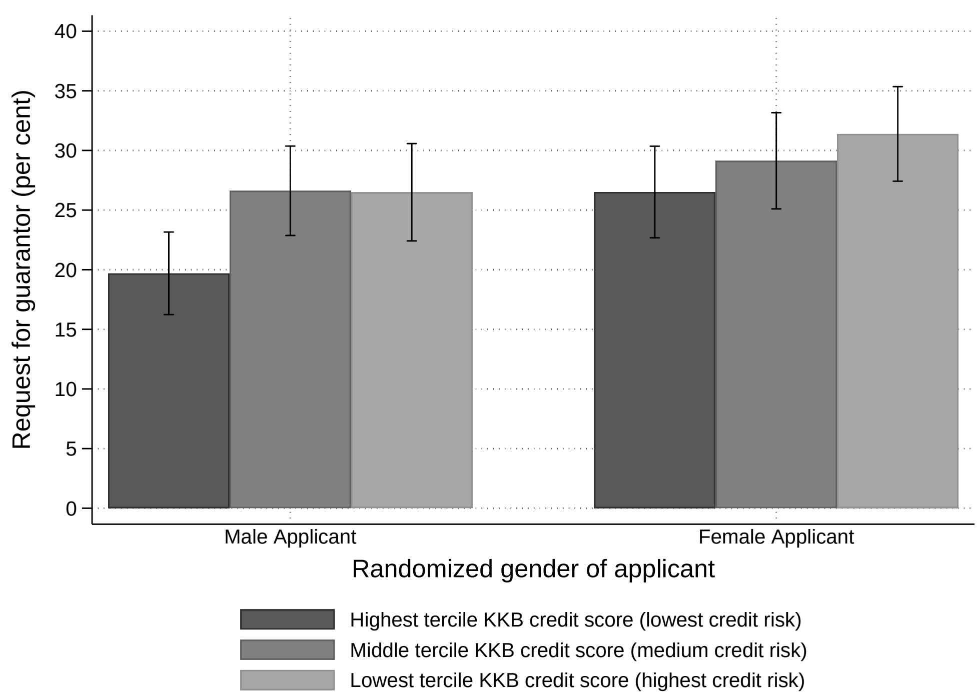
Notes: This figure shows coefficients from linear fully interacted models where the dependent variable is a *Guarantor dummy* that equals '1' if the participant approves the application but requests a guarantor and '0' if the participant approves without a guarantor. The sample is restricted to the first round of the experiment. Each bar corresponds to coefficients from a separate regression where we regress the *Guarantor dummy* on *Female applicant*, a given *Participant characteristic* interacted with *Female applicant* and the given *Participant characteristic* interacted with all other controls in column [4] of Table 3 including the file and city fixed effects. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively, and refer to t-tests of the null that (*Female applicant* + *Female applicant*$\times$*Participant characteristic*)=0.

Figure A4: Guarantor requirements, by loan quality and applicant gender

Panel A



Panel B



Notes: This figure shows the percentage of loan applications that were approved during the experiment and for which participants requested a guarantor. Panel A: bars indicate applications to which participants assigned a repayment probability at/above the median (dark gray) or below the median (light gray). Panel B: bars indicate loan applications with a KKB credit score in the highest tercile (lowest credit risk, dark gray); middle tercile (medium credit risk, medium gray); or lowest tercile (highest credit risk, light gray). Whiskers indicate one binomial standard error. The sample is restricted to the first round of the experiment.