Discriminatory Lending: Evidence from Bankers in the Lab*

J. Michelle Brock†
EBRD and CEPR

Ralph De Haas‡
EBRD, CEPR, and Tilburg University

March 31, 2021

Abstract

We implement a lab-in-the-field experiment with 334 Turkish loan officers to document gender discrimination in small business lending and to unpack the mechanisms at play. Each officer reviews multiple real-life loan applications in which we randomize the applicant's gender. While unconditional approval rates are the same for male and female applicants, loan officers are 26 percent more likely to require a guarantor when we present the same application as coming from a female instead of a male entrepreneur. A causal forest algorithm to estimate heterogeneous treatment effects reveals that this discrimination is strongly concentrated among young, inexperienced, and gender-biased loan officers. Discrimination mainly affects female loan applicants in male-dominated industries, indicating how financial frictions can perpetuate entrepreneurial gender segregation across sectors.

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

Keywords: Gender bias, bank credit, implicit association test, lab-in-the-field, causal forest

*We thank Francesca Dalla Pozza, Gözde Esen, Melanie Koch, and Victoria Robinson for excellent research assistance and Martin Acht, Sofia Amaral, Thorsten Beck, Tobias Berg (discussant), Miriam Bruhn, Zoe B. Cullen (discussant), François Derrien (discussant), Guido Friebel, Selim Gulesci, Sergei Guriev, Leora Klapper (discussant), Elif Kubilay (discussant), Sheisna Kulkarni (discussant), Sandra Orozco-Aleman (discussant), William Parienté, José-Luis Peydró, Alexander Popov, Siddharth Vij (discussant), Luana Zaccaria, and seminar participants at the London School of Economics, King's College London, University of Oxford, Goethe University, Ghent University, University of Southampton, University of Edinburgh, ZEW Mannheim, 2020 NBER Summer Institute, 2020 European Finance Association (EFA) Annual Meeting, Swiss Winter Conference on Financial Intermediation (Lenzerheide), Southern Economic Association Annual Meetings, 8th EFI Network Workshop, 4th EBC Network Workshop in Banking and Corporate Finance, MoFiR Virtual Seminar on Banking, 2nd London School of Economics Workshop on the Political Economy of Turkey, 6th IWH-Fin-Fire Workshop, FDIC Center for Financial Research Seminar, and the 2020 Fixed Income and Financial Institutions Conference (Darla Moore School of Business) 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

Across the world, female entrepreneurs borrow much less from banks than male entrepreneurs do (Demirgüç-Kunt et al., 2018). Whether this financial gender gap is inefficient depends on whether it reflects differences in the demand for or the supply of loans. On the demand side, women may select into smaller and less-capital intensive firms that require little credit (Demirgüç-Kunt, Beck and Honohan, 2008). On the supply side, discrimination by lenders is often cited as contributing to women's financial exclusion (The Economist, 2013; OECD, 2016). In the latter case, female entrepreneurs face too tight credit constraints and their productive capacity may remain underutilized. Such a misallocation of entrepreneurial talent can in turn hamper economic growth (Hsieh et al., 2019).

Discrimination in small business lending occurs when loan officers treat male and female applicants differently even if they are equal in all business-related aspects. Loan officers may hold female applicants to a higher standard by either directly rejecting women who do not meet this standard or by applying onerous conditions that make credit unattainable. Such indirect discrimination is particularly difficult to detect empirically. To test for the presence of both direct and indirect gender discrimination in small business lending, we 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 loan officers into a controlled environment allows us to carefully track their decisions and to trace the mechanisms through which gender discrimination materializes.

We conduct our experiment with 334 loan officers of a large Turkish bank. Turkey provides a particularly suitable setting to study gender discrimination in lending. It is a large and growing emerging market with a competitive banking system. The country scores well in terms of *de jure* gender equality: Few legal obstacles restrict women's ability to become an entrepreneur (Klapper and Singh, 2014). 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 tension between gender-related laws on the book and actual attitudes within society characterizes many other emerging markets too.[2]

[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] Turkey's Civil Code protects women's rights related to inheritance and marriage but contains no law that explicitly prohibits gender discrimination in lending (like the Equal Credit Opportunity Act in the U.S.).

We start by testing whether loan officers discriminate directly against female applicants. We find no evidence for such outright discrimination. Unconditional loan approval rates are very similar when we present the same application as coming from a male or a female entrepreneur. We next investigate whether loan officers discriminate in a less direct way. We find strong evidence that they do. In particular, loan officers are 26 percent more likely to make final loan approval conditional on the presence of a guarantor when we present the same application as coming from a female instead of a male entrepreneur. Because we use real-life loan applications that our partner bank received in the recent past, we can trace how loans performed in reality. We find that discrimination is concentrated among loans that were fully repaid in real life, making lending biases potentially costly to the bank.

To shed light on the mechanisms at play, we start by investigating whether biased lending is widespread across the loan officer population or concentrated among certain types. We first estimate conditional average treatment effects using sample-split and fully interacted regressions. We then apply machine learning—Wager and Athey's (2018) causal forest estimator—to more flexibly and efficiently explore heterogeneous impacts. The algorithm identifies who discriminates most by predicting individual treatment effects based on loan officer covariates. We find that younger and less experienced officers, and especially those with stronger implicit stereotypes against entrepreneurial women (measured through an Implicit Association Test), are more likely to impose discriminatory guarantor requirements.

We proceed by exploring two mechanisms that may underpin our results: gender differences in (actual or perceived) credit risk and loan officers acting on biased beliefs. We find no evidence supporting the idea that loan officers are concerned about higher credit risk among female entrepreneurs. The distribution of credit scores across male and female applicants is very similar and loan officers themselves do not perceive female entrepreneurs to be riskier than equivalent male ones. We also find no evidence of statistical discrimination. Experimentally varying the borrower information that is available to loan officers does not influence the gender bias in guarantor requirements.

The second mechanism concerns biased beliefs as reflections of social stereotypes (Bohren, Imas and Rosenberg, 2019; Bordalo et al., 2019). Our finding that loan officers with stronger implicit biases against entrepreneurial women are more likely to discriminate in terms of guarantor requirements already suggests that stereotypes play a role. To dig deeper, we divide our loan applications into those in relatively male-dominated versus female-dominated industries. We find that in stereotypically male industries, but not in female industries, loan approval is 15 percentage points more likely to be made conditional on a guarantor when we present the application as coming from a female instead of a male entrepreneur.

We again build a causal forest to learn about treatment effect heterogeneity. The algorithm helps to disentangle the role of loan officers' implicit gender bias, age and work experience. We find that these moderators play distinct roles depending on whether women apply for a loan in a traditionally male or female industry. In female-dominated industries, individual treatment effects range between zero and 10 percentage points. The algorithm reveals a tight negative relationship between loan officers' age and work experience and the predicted treatment effect. Once loan officers reach an age of about 43 (or, equivalently, two decades of work experience) they no longer discriminate against female applicants—that is, as long as applicants stick to traditionally female industries. In sharp contrast, in male-dominated industries, age and experience do little to attenuate discrimination. Here, the treatment effect is consistently above 10 percentage points and we find a tight positive relationship between the strength of individual officers' stereotypes and their predicted bias in guarantor requirements. In sum, social stereotypes underpin biased guarantor requirements but do so in a context-specific way.

Our results advance the literature on several fronts. First, we help to fill an important gap in the literature on gender discrimination in entrepreneurial finance. Recent work using administrative data (Ewens and Townsend, 2020) and experiments (Brooks et al., 2020; Hu and Ma, 2020) documents an investor bias against female entrepreneurs in need of venture capital. Herbert (2020), using French administrative data, shows that this equity funding gap reverses in female-dominated industries. An experiment by Gornall and Strebulaev (2020) finds that female entrepreneurs receive *more* replies to pitches to venture capitalists, suggesting an absence of discrimination at this initial stage.

There is less work on gender discrimination in entrepreneurial lending and most of it relies on observational data.[3] Analyzing the loan portfolio of an Italian bank, Bellucci, Borisov and Zazzaro (2010) show that women face tighter credit availability and collateral

[3]Two recent papers focus on discrimination in consumer lending. Dobbie et al. (2020) use administrative data from a UK lender and find evidence for discrimination against immigrants and older applicants (but not women) due to an incentive scheme that biases loan officers against illiquid applicants. Montoya et al. (2020) randomly match stylized consumer loan requests to male and female individuals who then apply by email for a small consumer loan. Requests submitted by women are less likely to be approved.

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. Similar studies using administrative data from the U.S. find no gender discrimination.[4] Lastly, Alibhai et al. (2019) conduct an email survey of 77 Turkish loan officers in which respondents have to distribute a loan amount among four stylized applications with randomized gender. Female applications received less funding.

We build on this work by bringing loan officers to the lab and measuring traits that are typically unobservable—including implicit gender bias, risk preferences, and work experience. Employing recent advancements in causal machine learning, we show that some of these characteristics are first-order determinants of biased lending. This sheds new light on work by Beck, Behr and Guettler (2013) and Beck, Behr and Madestam (2018) who use data from an Albanian lender. The first paper shows that lending decisions by female loan officers result in fewer arrears, while the second finds that 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 performance differences between male and female loan officers may in fact reflect unobserved characteristics. We provide evidence to this effect by measuring such characteristics and quantifying their relative importance. Our causal forest shows that loan officers' implicit gender bias and their work experience is six and three times, respectively, more important than their own gender as drivers of discriminatory guarantor requirements.

Our experimental approach also reduces some identification concerns inherent to observational studies. In particular, we need not worry about omitted variables bias since we vary applicant gender while keeping all other characteristics of applications equal. We can also cleanly isolate the supply side of the credit market. This is important because a lower use of credit by female enterprises may simply reflect lower demand (Ongena and Popov, 2016). Lastly, in administrative data, clients are typically not randomly matched to loan officers, which can bias estimates of discrimination. One way to address this is to exploit rotation policies that generate exogenous matching between officers and borrowers (Fisman,

[4]See Blanchflower, Levine and Zimmerman (2003), Blanchard, Zhao and Yinger (2008) and Asiedu, Freeman and Nti-Addae (2012). Ferguson and Peters (1995) discuss the conclusions one can and cannot draw about discrimination on the basis of loan denial and default rates in administrative data.

Paravisini and Vig, 2017). We instead randomly assign applications to loan officers so that there is no endogenous matching by construction.

A second line of work we contribute to deals with inaccurate beliefs as drivers of discrimination. Economists have traditionally distinguished between taste-based and statistical discrimination. Taste-based discrimination (Becker, 1957) occurs when decision makers (say, loan officers) are prejudiced against a group (say, women) and consciously avoid interacting with them.[5] In contrast, statistical discrimination takes place when decision makers rely on a group attribute as a signal of unobserved individual quality. For example, loan officers may believe that the creditworthiness of men and women differs on average (Phelps, 1972; Arrow, 1973) or has a different variance (Aigner and Cain, 1977).[6] More recently, Bohren, Imas and Rosenberg (2019) and Bohren et al. (2020) distinguish between statistical discrimination based on correct versus incorrect beliefs. Statistical discrimination is inefficient if it reflects inaccurate beliefs about outcome distributions (such as credit risk) across groups. Such miscalibrated beliefs can take the form of gender stereotypes which, even if they contain a "kernel of truth" (Bordalo et al., 2016), exaggerate average differences and contribute to gender gaps in decision making (Bordalo et al., 2019).

Recent empirical work provides evidence for inaccurate statistical discrimination in the market for entrepreneurial equity (Hebert, 2018; Ewens and Townsend, 2020; Hu and Ma, 2020). We instead investigate biased beliefs in the market for entrepreneurial debt. Bohren et al. (2020) suggest two methods to distinguish between different forms of discrimination: collecting data on the subjective beliefs of evaluators and varying the information supplied to them. We do both. We first administer an Implicit Association Test to generate a direct measure, at the level of individual loan officers, of implicit stereotypes against entrepreneurial women.[7] Such stereotypes may be most salient in male-centric domains (Reuben, Sapienza and Zingales, 2014). In line with this, we find that stereotypes have the strongest impact when women apply for a loan in a male-dominated sector. We also experimentally vary the applicant information available to loan officers. We find that information availability

[5]Implicit discrimination is similar but takes place without individuals being aware of it (Bertrand, Chugh and Mullainathan, 2005).

[6]Statistical discrimination can be especially salient when decision makers evaluate out-group individuals (say, male loan officers screening female entrepreneurs) because it is more difficult to interpret signals about out-groups (Cornell and Welch, 1996).

[7]Attitude IATs measure implicit negative attitudes towards social groups. Stereotype IATs—like the one we use—measure implicit associations between social groups and specific traits (Bertrand and Duflo, 2017).

does not affect lending decisions in a gender-specific way. Together, these results point to inaccurate beliefs (in the form of implicit stereotypes against entrepreneurial women) as the main mechanism explaining the biased guarantor requirements we document in general and in male-dominated industries in particular.

Third, we contribute to research on the underrepresentation of women among entrepreneurs and on gender segregation across industries. For the U.S., Gompers and Wang (2017) document that women constitute less than 10 percent of the entrepreneurial and venture capital labor pool. Moreover, women entrepreneurs cluster in specific sectors and this accounts for a large part of the gender wage gap (Blau and Kahn, 2017). A separate strand of work explains the labor-supply decisions of women and men as a function of deep-rooted social norms about the approprivate behavior of women (Alesina, Giuliano and Nunn, 2013; Giuliano, 2018; Grosjean and Khattar, 2019) and men (Baranov, De Haas and Grosjean, 2020). These norms can transmit across generations (Bisin and Verdier, 2001) and lead men and women to self-select into occupations that best match their self-perceived gender identity (Akerlof and Kranton, 2010); to forego entrepreneurial opportunities at odds with prevailing norms (Field, Jayachandran and Pande, 2010); and to be restricted in their choices because social norms have been codified into discriminatory laws (Naaraayanan, 2020). Our contribution is to connect both lines of literature by showing how social stereotypes about gender and entrepreneurship can generate financial frictions in the form of biased guarantor requirements, especially in traditionally male industries. Such frictions may then perpetuate an inefficient allocation of entrepreneurial talent across industries.

Lastly, our results add to a small literature on social collateral and third-party guarantees in lending. A guarantor takes legal responsibility for repayment in case the borrower fails to do so. The use of guarantors is not only widespread in emerging markets but also in Europe and the United States where borrowers without a credit history are often asked for a guarantor. 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). This makes guarantees particularly effective in mitigating moral hazard (Pozzolo, 2004).[8] The other side of the coin is that when a loan applicant is requested to ask a family member or friend to guarantee a loan, they put

[8]A related literature analyzes joint-liability contracts in microfinance, where groups of borrowers monitor each other's project choice and effort, thus reducing moral hazard (Stiglitz, 1990 and Karlan, 2007).

their social capital and reputation at risk. Guarantees thus tend to come at a social or psychological cost to the borrower. Recent experimental evidence from Vietnam shows that borrowers are willing to pay up to nine percent of their monthly income to prevent repayment difficulties from being disclosed to their guarantor (Diep-Nguyen and Dang, 2020). We show how implicit biases among loan officers expose female loan applicants considerably more to such onerous guarantor requirements than otherwise identical male applicants.

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, outlines our estimation strategy, and introduces the causal forest algorithm. Section 4 presents the results after which Section 5 discusses mechanisms. Section 6 concludes.

2 Experimental context and design

2.1 Context

We conducted our experiment in cooperation with a large commercial bank in Turkey. Over a two-month period, 22 experimental sessions were held with a total of 334 bank employees across eight cities.[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 shows the location of the regional offices and the number and gender of the participating bank employees.[10]

Bank employees at two seniority levels participated in the experiment: loan officers (192) and supervisors (142). Both are located in branches and involved in the screening of borrowers. Loan officers establish contact with potential borrowers, conduct the initial screening, and 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 then enter this information into an electronic application form. They can also voluntarily add subjective notes to this form, such as about the client's perceived trustworthiness, experience, or social standing. If the loan

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

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

officer deems a client creditworthy in principle, they pass on the electronic application form to their supervisor (typically the branch manager) with a proposed maximum credit limit.

Crucially, at this point they also give 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. According to discussions with Turkish loan officers, the main function of requiring a guarantor is to leverage borrowers' social capital and, in doing so, to attenuate ex ante moral hazard and reduce the probability of default. If a borrower defaults nonetheless, banks can in principle start legal proceedings to recover (part of) the loan from the borrower and the guarantor at the same time. In practice, however, loss given default of guaranteed versus non-guaranteed loans tends to be similar, also because the legal process to recover a loan is lengthy. The main mechanism through which guarantees work is therefore to prevent borrowers from defaulting in the first place.

The supervisor reviews the loan application and can reject or approve it. In the latter case, the application is sent to the bank's headquarters for formal sign off.[11] Henceforth we refer to the total experimental population as either "participants" or "loan officers".

2.2 Experimental design

Participants evaluated two rounds of four loan applications.[12] We randomly presented these applications as coming from a female or a male entrepreneur. Participants then 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 repayment probability between 0 and 100. We did not constrain the time participants had to evaluate the applications. The sessions were framed as a general training exercise and no gender-related issues were mentioned before or 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") con-

[11] Branches can approve loans below a certain size threshold but in practice only 10 percent of micro loans and 0.5 percent of loans to small and medium sized enterprises (SMEs) are formally 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 loans typically approved within 1.5 days.

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

tained all the information required for determining creditworthiness of an applicant and that was available at the time the application was processed.[13]

We use 100 loan applications in the experiment, selected from an initial sample of 250. These 250 applications were a stratified random sample of all applications by existing SMEs (that is, no start-ups) that the bank received in the three to six years before the experiment.[14] Using this earlier period allows us to track what happened to each application in real life. The strata were region, gender, firm size, and whether the application was accepted in real life. By 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. Some of the initial 250 files were eliminated because they contained 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 application was on average evaluated by 13.4 participants per round, half of the time as a female and half of the time as a male file. This is important because it allows us to obtain a within-application estimate of gender discrimination. Moreover, 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.[15] No one saw the same file or the same name more than once.

[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 or maturity of specific credit products.

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

[15]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

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 application) from each gender.[16]

After the first round, officers received another four applications. We again randomized the gender of each. Inspired by Bernstein, Korteweg and Laws (2017), who measure the impact of different types of information on investors' decision to fund start-ups, we now also experimentally vary the information available to loan officers. Officers were randomized into one of 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. This score aggregates hard financial data that may help to predict default. A second treatment group evaluated files where we had removed a section with subjective information. 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 to strengthen the rationale for lending. The availability of nonstandard, subjective and less quantifiable information may be most important when evaluating lower-quality borrowers (Iyer et al., 2016).

If either the objective credit score or the subjective comments section contribute to officers' ability to make fair and objective lending decisions, omitting it may increase statistical discrimination. 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.

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

(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.

[16] 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.

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 lending decisions in line with 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).[17] In contrast, five points were deducted when they 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 each experimental session and participants then exchanged points for prizes. 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.[18] 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 possible outcome, in points, and the even split represented that the two outcomes were

[17]This incentive scheme resembles the remuneration system that the bank uses in reality and is also similar to the baseline scheme of Cole, Kanz and Klapper (2015). The incentives allow us to mimic the costs associated with certain forms of discrimination and contrasts with existing experimental work, which typically does not impose costs for making inefficient discriminatory choices (Neumark, 2018).

[18]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.

equally likely. The outcome pairs were 28-28; 20-44; 24-36; 16-52; 12-60; and 2-70. 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.

Participants also took a stereotype Implicit Association Test (IAT).[19] They had to sort, as quickly as possible, words that appeared sequentially on their tablet by clicking buttons at the right and left 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.[20] 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). We recorded the time it took to sort each word 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 define a participant's implicit stereotype against entrepreneurial women as the normalized difference in mean response times between the non-stereotypical and the stereotypical task. Higher values indicate stronger stereotypes.

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

[19] IATs are now common in psychology (Greenwald, McGhee and Schwartz, 1998) and economics (Bertrand, Chugh and Mullainathan, 2005; Beaman et al., 2009; Glover, Pallais and Parienté, 2017; Carlana, 2019). A meta-analysis 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).

[20] The IAT and all other documentation was provided in Turkish. The family-related words were 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".

percent of the participants are supervisors, the others are loan officers. While the average participant has worked as a loan officer for almost nine years, this varies between zero and 32 years. There is thus substantial variation in the lending experience that loan officers have built up over the course of their career.[21]

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 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 A2 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 A2). The average IAT score is 0.39 for women and 0.28 for men and this difference is statistically significant at the 5 percent 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 percent of

[21]We also asked participants for how long they had worked at this particular bank. The average employment duration was five years and ranged between zero and 19 years. All results that follow are robust to using this narrower experience definition.

the loan applications is rejected outright whereas, conditional on provisional acceptance, a guarantor is requested in 27 percent of the cases.

For each credit application, the participant was asked to estimate, on a 0-100 scale, the probability that the borrower would repay. The median estimated repayment probability is 70 percent (Table 1). These data also help 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 loan applications in real life. We find that this is the case. Overall, 72 percent of all applications that 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 percent) and for applications that were rejected in real life (47 percent). 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 A2 provides a correlation matrix of the participant characteristics and the rejection dummy. We already discussed that female participants are on average more risk averse and more implicitly gender biased. Table A2 also shows that participants higher up in the lending hierarchy (supervisors) are more often female, older and more experienced, as well as more gender biased. Lastly, it is reassuring that whether a file was presented as male or female (*Female applicant*) is completely uncorrelated with any of the participant characteristics. This reflects that the randomization was successful.

Appendix Table A3 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 0.11 points (on the [-1,1] scale) more biased against female entrepreneurs as compared with male bank employees.

3.2 Estimation strategy

To test for biased lending behavior of loan officers, we regress the loan application outcomes of interest on G_{il}, the randomly assigned applicant gender of loan application l as seen by participant i. We start with a parsimonious linear probability model with loan application (file) fixed effects, φ_l, to obtain within-file estimates of gender discrimination:

$$y_{il} = \alpha + \beta \cdot G_{il} + \varphi_l + \epsilon_{il} \tag{1}$$

where y_{il} is the lending outcome of interest when participant i evaluates loan application l. Standard errors, ϵ_{il}, are heteroskedasticity robust and clustered at the participant level.

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.

We then run progressively more saturated specifications to further improve the efficiency of the estimates and add structure to the error term.[22] We first add stratification controls: a dummy variable for whether a participant is a supervisor or loan officer (*Participant is supervisor*) and fixed effects for the cities where the bank's regional offices are based (and where the experimental sessions took place). Next, we include additional participant characteristics and, in the most saturated specification, individual-specific random effects. Individual effects control for participant-level unobserved heterogeneity. The use of random effects is appropriate here because the matching between files and participants was random

[22] While randomization ensures our estimate of β is unbiased, adding covariates can improve precision.

and all participants evaluated the same proportion of male and female applications, of both high and low quality. Also, because participants were randomly selected from the total loan officer population, any individual specific effects are the result of a random draw and orthogonal to the treatment variable, G_{il}. Individual heterogeneity is therefore not systematically correlated with G_{il} and random effects will be unbiased and efficient.[23]

The most restrictive model that we estimate thus includes a vector of K participant characteristics (gender, experience, age, a supervisor dummy, risk aversion, and IAT score), X_i; file fixed effects, φ_l; a set of fixed effects for the eight cities where the experiment took place, φ_c; and participant random effects, η_i.

$$y_{il} = \alpha + \beta \cdot G_{il} + \sum_{k=1}^{K} \gamma_k \cdot X_i + \varphi_l + \varphi_c + \eta_i + \epsilon_{il} \tag{2}$$

α is a constant and ϵ_{il} is a cluster robust error term.

3.3 Heterogeneous treatment effects

Equation 2 provides estimates of the Averate Treatment Effect (ATE). We are also interested in conditional average treatment effects (CATE) for subgroups of the loan officer population. In particular, we want to assess heterogeneity by loan officers' gender; work experience; age; position (loan officer versus supervisor); risk aversion; and their implicit bias against entrepreneurial women. We follow two approaches. First, we present traditional sample-split regressions where we estimate Equation (2) on subsamples (the Appendix also summarizes equivalent fully interacted regression models).

Second, we use supervised machine learning in the form of an honest causal forest algorithm to assess how impacts vary across loan officers (Athey and Imbens, 2016; Wager and Athey, 2018; Athey et al., 2019). This has two important advantages (Davis and Heller, 2017). First, causal forests can combine multiple explanatory variables in a data-driven, nonlinear but disciplined way. This gives us a more efficient, and hence statistically more powerful, tool to estimate heterogeneous treatment effects. In a setting like ours, with multiple dimensions of potential heterogeneity (which may also interact or have non-linear effects) and subpopulations that differ in size, this is especially useful. Second, the algorithm tells us

[23] Using participant fixed effects is an alternative that would entail an unnecessary loss of efficiency.

how useful each loan officer trait is in growing the forest from which heterogeneous impacts are predicted. This allows us to gauge the relative importance of these traits as moderators of the causal effect between applicant gender and outcomes. Moreover, we can plot the value of these traits against the predicted treatment effect at the level of individual officers because the algorithm provides us with a complete distribution of these effects.

Conceptually, the honest causal forest algorithm creates a random forest of causal trees. Each tree grows from a random (bootstrapped) subsample of training data, the root node. The tree then recursively splits into increasingly smaller nodes that share similar covariates until it arrives at a set of terminal nodes (called leaves). The algorithm makes splits that produce the biggest difference in treatment effects across leaves while still yielding an accurate estimate of the full treatment effect.[24] If splitting a node would not result in an improved fit, that node is not split further and forms a final leaf. This approach is honest in the sense that for each training subsample (that is, for each tree) the observations are separated into a splitting sample (to determine where to place the splits) and an estimating sample (to estimate the within-leaf treatment effects).

We use the generalized random forest *grf* package for R by Tibshirani et al. (2020) to estimate a forest with 20,000 trees based on a random training sample of 70 percent of the full dataset. To grow each tree, we split the training sample into a splitting and estimating sample of equal size. This step is repeated 20,000 times to grow the complete forest. In a final step, the 30 percent of the full dataset that was left aside is fed through all trees. For each observation, we determine to which leaf it belongs based on the loan officer's various traits. We then assign it the predicted treatment effect of that particular leaf. The average prediction across all trees is then the predicted treatment effect at the loan officer level.

4 Results

4.1 Applicant gender and the rejection of loan applications

Table 2 presents linear probability regressions based on Equations 1 and 2. The dependent variable is a *Rejection dummy*, which is "1" if an application was outright rejected by a participant and "0" if approved. The independent variable of interest, *Female applicant*,

[24] A regular random forest determines the splitting by minimizing the mean squared error of the outcome.

is a dummy whether the application was presented as coming from a female ("1") or male ("0") entrepreneur. Column 1 shows the most parsimonious specification with only file fixed effects. Column 2 adds stratification controls while column 3 also includes the participant characteristics *Participant is female*; *Participant experience*; and *Participant age*. Column 4 then adds two potentially important, but typically unobserved, variables that we measured in the laboratory: risk aversion and implicit bias against entrepreneurial women. Finally, we add individual random effects in column 5. 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 five specifications, the coefficient for *Female applicant* is close to zero and, if anything, negative.[26] Since we include file fixed effects, our results show that the *same* loan application does not have a higher chance of being outright 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.

Before turning to the participant covariates, we stress that applicant gender is the only variable that we manipulated experimentally and that hence can be interpreted causally. With this caveat in mind, we find that supervisors are nine percentage points less likely to accept a loan application as compared with regular loan officers. This difference, significant at the 1 percent statistical level, is large: the unconditional acceptance rate in our experiment is 61 percent. It likely reflects that the main role of supervisors is to validate (or overrule) the initial lending decisions by more junior loan officers.[27] We do not find that participants' risk aversion or their implicit gender bias correlate with the rejection probability.

We also assess whether the null result applies to 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

[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] Our experiment was not powered to detect such a small effect and the 95 percent confidence interval is therefore quite wide at [-0.055, 0.040]. To achieve 80 percent power to detect whether $\beta = -0.008$ is statistically non-zero would have required over 10,000 decisions—ten times our current sample.

[27] Unreported results show that this conservatism among supervisors is independent of the (randomized) gender of the applicant. The interaction between *Female applicant* and *Participant is supervisor* is never statistically significant. The same holds when we run separate split-sample regressions for loan officers and supervisors.

direct gender discrimination in any of these sample splits.[28]

4.2 Applicant gender and guarantor requirements

We next test for a more indirect form of discrimination against female applicants. In particular, in Table 3 we assess whether the approval of a loan application is more likely to be made conditional on the presence of a guarantor when the application comes from a woman instead of a man, all else equal. The structure mirrors Table 2 but the sample is smaller as the decision to require a guarantor is conditional on provisional loan approval.[29]

We find strong evidence of indirect gender discrimination: loan officers are more than six 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. The statistical as well as economic significance of this effect is stable across the increasingly saturated specifications. The effect is also substantial as only 27 percent of all pre-approved applications are required to have a guarantor. This more indirect form of discrimination implies that female entrepreneurs without access to a guarantor remain deprived of credit, even if the loan officer in principle expresses a favorable view of the application. To the extent that such female entrepreneurs are, in fact, good credit risks such a bias will also be disadvantageous to the bank. And even for female borrowers who are able to provide a guarantor, putting their social capital at risk comes at a cost.

We next assess the stability of these estimates across geographies and sectors. Panel A of Appendix Figure A1 depicts coefficient estimates similar to those in column 5 of Table 3. Each estimate reflects a sample in which we drop all observations from one city where an experimental session took place (and where the participating loan officers are based). This visualizes whether the results are stable across the experimental locations. We find that in all cases the coefficient indicates a 5 to 10 percentage point higher likelihood that a guarantor is requested from female applicants. The coefficients are ordered, from top to bottom, by decreasing average disposable household income in the excluded city. There is no apparent

[28] Results are the same when we run interaction regressions rather than sample-split regressions. Note that while a causal interpretation of the covariate coefficients themselves is problematic, the results of sample-split regressions based on those covariates, or of regressions where *Female applicant* is interacted with these covariates, can be interpreted as conditional average treatment effects.

[29] 98 percent of the participants occur in both rejection and guarantor estimations so the second regression does not suffer from a notable self-selection problem.

relationship between indirect gender discrimination and local economic development.

Panel B of Figure A1 repeats this exercise but now takes the perspective of the region of the loan application. For each of the 100 files in the experiment we determine in which region the real-life loan applicant was based.[30] We then drop one region at a time and plot the estimated coefficients, ordering them from the highest (top) to the lowest (bottom) regional income level per capita in 2016. We again find little geographic heterogeneity: in each case the probability that a guarantor is required is between 5 and 10 percentage points higher when we present the same application as coming from a female rather than a male entrepreneur. Lastly, in Panel C of Figure A1, we exclude one of the following macro sectors at a time: Retail, services, manufacturing, wholesale, and other industries. The results again show a coefficient that lies between 5 and 10 percentage points, with a slightly lower coefficient when we exclude the retail sector. We now assess whether biased guarantor requirements occur in the loan officer population as a whole or are instead concentrated among particular types of loan officers.

4.3 Indirect gender discrimination: Participant heterogeneity

4.3.1 Heterogeneous treatment effects: Sample splits

Table 4 investigates heterogeneity in biased guarantor requirements through the lens of sample-split regressions. We focus on the coefficient for *Female applicant* but all regressions include the same covariates, fixed effects, and random effects as column 5 of Table 3.[31] We find a consistent and intuitive pattern of significant conditional average treatment effects. When we present the application as coming from a woman instead of a man, officers are more likely to ask for a guarantor when they are younger (columns 5-6); in a more junior position (columns 7-8); and/or display more implicit gender bias in our IAT (columns 11-12). T-tests confirm that we can reject equality of coefficients in these column pairs.[32] There is also some

[30] Turkey is divided into the following regions: Marmara, Aegean, Central Anatolia, Mediterranean, Black Sea, Eastern Anatolia, and Southeastern Anatolia.

[31] When we partition non-binary 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.

[32] We summarize results from equivalent fully interacted regression models in Appendix Figure A2. The independent variables 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

evidence that participants with a below-median level of lending experience are more likely to ask women (as compared with men) for a guarantor than their more experienced colleagues (columns 3-4). Together these results suggest that age and seniority, possibly summarized by experience, reduce the extent to which loan officers use gender as a mental shortcut to determine whether a loan application requires a guarantor or not.

Importantly, columns 1 and 2 of Table 4 show 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 significant difference between participants with higher versus lower levels of risk aversion, a characteristic strongly correlated with participant gender (columns 9 and 10). We instead find an important role for the implicit stereotypes about entrepreneurial women that participants (both men *and* women) harbor. Columns 11 and 12 show that loan officers with above-median levels of implicit gender bias are nine percentage points more likely to request a guarantor when we present a file as coming from a female instead of a male entrepreneur.[33]

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 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 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. This number is substantially higher for non-performing loans and applications that were declined in real life,

this participant characteristic and all other controls, including the file and city fixed effects. The bars 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.

[33] 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.

at 28.6 and 32.9 percent 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 percent 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 better-quality loans that 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 A3, 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, the gender bias 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 and random effects, as well as file and city fixed effects, women are 10.7 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 bias that officers display when requesting guarantors is therefore not driven by the low-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.

We again assess which participant types are responsible for this gender discrimination. Columns 3 to 14 reveal similar heterogeneity as before. High-quality female loan applications are 10 to 16 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) and revealed strong gender stereotypes in our implicit association test (columns 13-14).[34] In summary, especially more junior and more biased loan officers tend to resort to the applicant's gender as a heuristic when there are no clear indications that a loan is risky.

[34] There is again no difference by participant gender: the coefficients in columns 3 and 4 are very similar. We note that where the sub-sample coefficients differ substantially in size, this difference is in some cases 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.

4.3.2 Heterogeneous treatment effects: Honest causal forests

The previous sub-section provided a first analysis of conditional average treatment effects. While the results form a consistent pattern, applying linear regressions to split samples limits statistical power, especially when covariates such as age and experience are strongly (though far from perfectly) correlated. We now introduce a causal forest algorithm to more flexibly and efficiently disentangle how loan officer traits play distinct moderating roles in the causal relationship between applicant gender and guarantor requirements.

Figure 5 (Panel A) depicts the distribution of the predicted treatment effects. In the absence of treatment heterogeneity, this distribution would cluster tightly around the average treatment effect (ATE) of 6 percentage points (vertical dashed line). Instead the causal forest reveals a broad distribution of treatment effects underlying the ATE. They vary from slightly negative to a 13 percentage points higher probability of requesting a guarantor when we present a loan application as coming from a female instead of a male entrepreneur.

Panel B of Figure 5 ranks loan officer traits by their relative importance as moderators (drivers of treatment heterogeneity). We define a trait's relative importance as the weighted sum of the number of times it is used to split at each depth in the forest. The more a trait is used to split subsamples, the more predictive power it has. We find that loan officers' implicit stereotypes against entrepreneurial women, measured as their IAT score, are by far the most important driver of treatment heterogeneity. In fact, in exactly a third of all trees the algorithm picks an officer's implicit bias to make the first split. The second and third most important drivers are loan officer age and experience, which our algorithm—unlike linear regressions—can neatly disentangle. The three other traits—risk aversion, gender, and hierarchical position—are much less important drivers of treatment heterogeneity. Most of these results are consistent with those based on split-sample regressions. Both show that implicit stereotypes, age, and experience are important and they both tell us that loan officers' own gender is *not* an important driver of discriminatory guarantor requirements. An interesting exception is *Participant is supervisor*. Linear sample-split regressions suggest this variable correlates strongly with bias in guarantor requirements. Yet, the causal forest tells us this is not the case once we account for non-linearities and the fact that being a supervisor correlates with age and work experience.

Figure 6 plots the predicted treatment effects against the three most important loan officer traits. We fit smooth local polynomial functions in each scatterplot. The patterns are striking. Panel A shows how the predicted treatment effect increases when officers' implicit stereotypes are stronger. The causal forest reveals a discrete jump of 2.5 percentage points in the predicted treatment effect (that is, a higher probability of requiring a guarantor when we present a loan application as coming from a female instead of a male entrepreneur) at an IAT score of around 0.25. From a policy perspective, this indicates that there is a distinct group of biased loan officers that may be targeted by, for example, debiasing interventions. Panels B and C of Figure 6 show a tight negative correlation between age and work experience, respectively, and the predicted treatment effect. This relationship is much more linear: The probability that a loan officer engages in discriminatory guarantor requirements declines steadily with age and, independently, with work experience.[35]

5 Interpretation and mechanisms

When we present one and the same file as coming from a female instead of a male entrepreneur, loan officers are on average six percentage points (or 26 percent) more likely to require a guarantor. This biased behavior is concentrated among younger and less experienced loan officers and especially among those who harbor a stronger bias against female entrepreneurs. We now consider two mechanisms that may underpin this result: gender differences in credit risk and loan officers using mental shortcuts that reflect gender stereotypes.

5.1 Gender differences in credit risk

We first investigate whether gender differences in credit risk could justify a different treatment of male and female loan applications. We offer several pieces of evidence that consistently show that the distribution of credit risk across male and female borrowers is very similar and, importantly, that loan officers themselves do in fact not judge female borrowers to be riskier than equivalent male ones.

[35]Unreported results indicate very similar patterns when we grow and use the causal forest on the basis of only those loans that performed well in real life.

5.1.1 Gender differences in credit scores

We first compare the credit scores (from the Turkish credit registry) of the male and female applicants in our random sample of 250 loan applications. Recall that these were sampled from all applications the bank received from entrepreneurs in recent years. The score captures an entrepreneur's borrowing and repayment history and is therefore a good indicator of credit risk. The data reflect the real-life applications and the actual gender of the applicant, so they are non-experimental. Since the sample is stratified by gender, firm size, region and quality of the original application, the distributions can be compared. The average score is 1,035 for men and 1,023 for women (a higher score implies less risk) and this small difference is not statistically significant (p=0.80). Appendix Table A4 presents OLS regressions for the 243 files for which 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 amount requested (column 5). Figure 7 shows that the full distribution of these credit scores is also very similar for male and female loan applicants (as confirmed by a Kolmogorov-Smirnov test).

5.1.2 Gender differences in subjective repayment probabilities

Even if the distribution of ex ante credit risk is objectively very similar, loan officers may still *perceive* women to be riskier (and hence be more demanding in terms of guarantor requirements). To see whether this is the case, Figure 8 shows a binned scatter plot of credit scores (horizontal axis) and loan officers' view of an applicant's repayment probability (vertical axis). Dark gray dots (light gray diamonds) show bin averages for loan applications presented as coming from male (female) entrepreneurs. Confidence intervals (95 percent) are based on a cubic regression spline of subjective repayment probability on the credit score.

Two main messages emerge. First, we observe a tight correlation between credit score and subjective repayment probability along the risk distribution. That is, when loan officers assess lower risk applications (higher credit scores), they systematically perceive these applications to have a higher repayment probability. Second, this tight correlation between objective and subjective credit risk holds independently of whether we present a file as coming from a male or a female entrepreneur. This holds true along the risk distribution: At no point is there a statistically significant disconnect between how loan officers translate male versus female credit risk into subjective repayment probabilities. This is further corroborated by Figure 9 and Appendix Table A5. Figure 9 provides a Kernel density plot of the subjective repayment probability that loan officers assign to male and to female versions of the same applications. Both distributions are very similar, as confirmed by a formal Kolmogorov-Smirnov test. Appendix Table A5 displays regression specifications similar to those in Tables 2 and 3 but with *Subjective repayment probability* as the dependent variable. As expected, there is no significant impact of the (randomized) gender of the loan applicant on the credit risk as perceived by loan officers themselves.

5.1.3 Gender and risk: Evidence from a separate risk module

Next, we present evidence from a separate risk module that we implemented during the experimental sessions. As part of this module, loan officers were randomly matched with a male or a female (real-life) entrepreneur. We informed the officers about the gender, age, and industrial sector of the entrepreneur they had been matched with. Prior to the experimental sessions, we had asked these entrepreneurs to pick one out of six projects that were increasing in riskiness, in the spirit of Eckel and Grossman (2008). They had to do so for a project financed with a loan and for a project financed without debt. During the experiment, loan officers were then asked to guess which risky projects their matched entrepreneur had chosen. We paid loan officers if they chose correctly.

The ordered probit specifications in Appendix Table A6 regress the participants' perceptions of their matched entrepreneur's risk taking (on a 1-6 scale) on the gender of the entrepreneur. We control for the entrepreneur's age and industrial sector. For projects not funded with bank credit (column 1), loan officers believe that the entrepreneur they were matched with picked a slightly *less* risky project if that entrepreneur was female. The statistical significance of this gender difference disappears, however, when we ask loan officers about the risk they think entrepreneurs took for projects financed with bank credit (column 2). In either case, the evidence from this risk module is clearly at odds with loan officers perceiving female entrepreneurs to be *more* risky.

5.1.4 Gender, applicant information, and statistical discrimination

We next consider the decisions made during the second round of the experiment, when we also randomized the types of applicant information that loan officers had access to. Even when officers do not perceive female entrepreneurs to be more risky on average, they may still find it more difficult to judge applications from individual women. They may, for example, encounter relatively few such applications and hence be less sure of the complete risk distribution among entrepreneurial women. This makes it more difficult to interpret signals about the quality of individuals. Rational loan officers may then put less weight on traits of individual female applicants (which to them are weaker signals of creditworthiness) and more weight on group means (Aigner and Cain, 1977).

To investigate whether such statistical discrimination can explain the results from the first round of our experiment, we randomly varied the types of applicant information that loan officers had access to in the second round. Reducing the richness of applicant characteristics, can make statistical discrimination more pronounced (Kaas and Manger, 2012; Neumark, 2018). Table 6 assesses whether restricting the information available to loan officers has a disproportionate impact on female loan applications. 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, file and city fixed effects, and participant random effects.

We find no evidence of statistical gender discrimination. The interaction terms between *Female loan applicant* and the information treatments are far from statistically significant in columns 2 and 4. Column 2 does provide some weak evidence that the subjective information that loan officers can voluntarily add to an application file increases the willingness to lend among those who review the file. Yet, this effect does not differ between male and female loan applicants.

In sum, we analyze objective credit-registry scores; subjective repayment probabilities that loan officers assigned during our experiment; a risk module in which loan officers estimate the amount of risk taking by a real-life entrepreneur; and the information treatments in the second experimental round. None of these exercises returns compelling evidence supporting the hypothesis that gender differences in real or perceived credit risk are a key mechanism to explain the strong gender bias in guarantor requirements that we document.

5.2 Gender stereotypes

5.2.1 Gender stereotypes and biased guarantor requirements across sectors

We now investigate an alternative mechanism: biased beliefs due to social stereotypes (Bohren, Imas and Rosenberg, 2019 and Bordalo et al., 2019). Recent work has shown how decision making can be biased when women are judged in stereotypically male domains.[36] Such stereotypical beliefs can result in a type of gender bias that cannot be explained by standard models of statistical discrimination.

We first identify the 2-digit ISIC industry of each of the 100 loan applications used in the experiment. This gives us fourteen unique industrial sectors. Each of these we then classify as being either a male-dominated or a female-dominated industry. To do so, we use data from the 5th and 6th rounds of the World Bank-EBRD Business Environment and Enterprise Performance Survey (BEEPS). This data set contains information on the gender of the owner of 44,540 firms across 48 middle-income countries in Emerging Europe, Central Asia and North Africa.[37] For each industry, we measure the proportion of SMEs owned by women and then rank all industries. We define male-dominated (female-dominated) industries as those with a share of female-owned SMEs below (above) the median.[38] Examples of female-dominated sectors include the manufacturing of textiles and the manufacturing of

[36]For example, Guiso et al. (2008); Carrell, Page and West (2010), Reuben, Sapienza and Zingales (2014); Bohren, Imas and Rosenberg (2019) and Carlana (2019). Alan, Ertac and Mumcu (2018) show how traditional gender views among Turkish elementary school teachers negatively affect girls' math and verbal test performance.

[37]The survey design uses a comprehensive sample frame (typically the business registry) of all formal private-sector firms with at least five employees. Three stratification criteria are used: sector of activity, firm size, and geographical location. Size stratification divides the population into small (5-19 employees), medium (20-99) and large firms (100 and more employees). Importantly, the design ensures the sample adequately represents the sectoral and geographical distribution of each country's SME population.

[38]Appendix Table A7 provides our sector breakdown and the classification into male- versus female-dominated industries. Our results are robust to only using BEEPS data for Turkey plus other middle-income countries along the Southeastern Mediterranean. For seven loan applications we cannot determine the typical gender composition of the industry. Our results are robust to classifying these applications as coming from either male or female-dominated sectors.

food products and beverages, whereas male-dominated industries include the manufacturing of rubber and plastic products as well as the construction sector.

In the first two columns of Table 7, we test whether the substantially (6 percentage points) higher guarantor requirements that we observe for female loan applications on average, are equally present in male- and female dominated sectors. In case social stereotypes play an important role, we would expect biased guarantor requirements to be mainly or exclusively concentrated in male-dominated sectors. This is indeed what we find. In stereotypically male industries, the approval of a female loan application is almost 15 percentage points more likely to be made conditional on the presence of a guarantor (column 1). In stereotypically female industries, on the other hand, women entrepreneurs face no such bias (the coefficient is three times smaller and not statistically significant).[39]

Theory predicts that the impact of social stereotypes is most salient for loan officers who harbor stronger implicit gender biases. In columns 3 through 6, we therefore split the decisions for stereotypically male sectors (columns 3-4) and for stereotypically female sectors (columns 5-6) into those taken by loan officers with a below median score on our Implicit Association Test (columns 3 and 5) and those with an above-median score (columns 4 and 6). For female-dominated sectors, we do not find a statistically significant difference between more and less implicitly gender-biased loan officers. In sharp contrast, in the case of male-dominated industries, we find that the higher guarantor requirements for women are driven by loan officers with a strong implicit gender bias. Among these officers, there is a 27 percentage points gender difference in the probability of a guarantor request in stereotypically male industries. Unreported regressions show no relationship between applicant gender, on the one hand, and subjective repayment probability in either male- or female-dominated industries on the other hand. This again indicates that the stricter guarantor requirements

[39] When we randomize applicant gender, we create loan applications where the match between gender and industry is by construction artificial. Yet, the resulting applications reflect gender-industry combinations that are all observed in real life. More specifically, among the 250 files from which we draw our 100 loan applications, the percentage male (female) applicants in male-dominated industries is 64 (36) percent. These numbers are 41 and 59 percent in female industries. This shows that while men (women) are clearly overrepresented in male-dominated (female-dominated) industries, there is sufficient overlap between these industries to create realistic experimental gender variation within both industry types. We also note that female applicants in male industries are not more risky—in terms of credit score—than male applicants in such industries. In our sample of 250 files, female entrepreneurs in male industries are in fact slightly *less* risky, although this difference is not statistically significant (p=0.30). In female industries, women have slightly better credit scores as well (p=0.06).

do not reflect loan officers' concerns about higher credit risk for female applicants, even if these women apply in stereotypically male industries. Instead, our results offer strong support in favor of social stereotypes underpinning our average treatment effects.

5.2.2 Stereotypes, industries and guarantors: Heterogeneous treatment effects

We return to the causal forest to investigate heterogeneous treatment effects across industries. Figure 10 shows the distribution of the predicted treatment effects in female-dominated industries (dark grey bars, with the ATE indicated by the dashed vertical line) and male-dominated industries (light grey bars and a solid line). We again observe a substantial spread in the conditional treatment effects around the ATEs. Interestingly, both distributions hardly overlap. Only the largest predicted treatment effects in female industries overlap with the smallest ones in male industries. This indicates that loan officers systematically judge female entrepreneurs differently—they apply a different standard—in male- versus female-dominated industries.

Figure 11 depicts the relative importance of loan officer traits as drivers of biased guarantor requirements in female-dominated industries (Panel A) and male-dominated ones (Panel B). The same traits as before play a key role: gender stereotypes (IAT score), age, and work experience. Yet, when female entrepreneurs apply in a male-dominated industry (Panel B), loan officer age and experience are less important relative to their gender stereotypes, which stands out as the main driver of biased guarantor requirements.

Figure 12 visualizes even more starkly the difference between male- and female-dominated industries in terms of the relationship between, on the one hand, gender stereotypes (top panels), age (middle) and experience (bottom) and, on the other hand, the predicted treatment effects across loan officers. A first clear difference concerns implicit stereotypes. In female sectors (left), individual treatment effects vary between zero and 10 percentage points, but without an apparent relationship with officers' stereotypes. In contrast, in male-dominated sectors, the treatment effect is not only systematically above 10 percentage points but there is also a strong positive relationship between individual officers' implicit stereotypes and their predicted bias in guarantor requirements. This illustrates how stereotypes about female entrepreneurs can be context-dependent (Coffman, 2014) and only manifest themselves when women apply in male sectors.

Strikingly, we observe the opposite pattern for loan officer age (middle) and work experience (bottom). The causal forest algorithm can disentangle the two and shows how both lead to a monotonic decline in biased lending behavior in female-dominated sectors. When loan officers reach an age of around 43, or have almost two decades of work experience, they typically no longer display a bias against female applicants—as long as these entrepreneurs stick to traditionally female industries. In sharp contrast, the beneficial effect of age and experience is absent in male-dominated sectors (right). In these industries, independent of a loan officer's age or experience, the predicted gender bias in guarantor requirements consistently fluctuates between 10 and 15 percentage points.

6 Conclusions

We implement a randomized lab-in-the-field experiment to gain insights into the nature of discrimination in small business lending. While we find no evidence of direct discrimination in terms of unconditional approval rates, we uncover a strong gender bias in loan requirements. All else equal, the approval of female applications is 26 percent more likely to be made conditional on the presence of a guarantor. Exploring heterogeneity in treatment effects helps to understand the mechanisms underpinning this indirect form of discrimination. A causal forest algorithm reveals that specific loan officer traits—their implicit stereotypes about entrepreneurial women, their work experience, and their age—independently and strongly correlate with the intensity of discrimination, considerably more so than their gender.

What do these results tell us about the nature of the discrimination we observe? While we cannot rule out taste-based discrimination, we believe this is unlikely to drive the bias in guarantor requirements. If taste-based discrimination would be of first-order importance, we would expect it to already rear its head in the unconditional loan approval decisions. In contrast, our finding that discrimination is concentrated among loan officers who adhere to strong implicit stereotypes about entrepreneurial women is a clear sign that implicit gender bias plays an important role. Our causal forest reveals clearly how variation in stereotypical thinking across the loan officer population is the main determinant of biased guarantor requirements. These implicit stereotypes have the strongest impact on decision making when women apply for a loan in relatively male-dominated sectors.

We note, however, that discriminatory guarantor decisions are also concentrated among less experienced loan officers (even when controlling for age). Learning through experience can mitigate statistical discrimination over time (Aigner and Cain, 1977; Altonji and Pierret, 2001; Botelho, Madeira and Rangel, 2015) and our causal forest suggests this is indeed the case—but only when female entrepreneurs apply for loans in gender-congruent sectors. This indicates that statistical and implicit discrimination may interact. To the extent that loan officers with biased beliefs do not make an effort to screen and collect information about female loan applicants in male sectors (but continue to rely on mental shortcuts in the form of stereotypes), the accumulation of work experience will not attenuate their implicit biases about women in such sectors. In contrast, in female-dominated sectors, the intensity of discrimination declines linearly as work experience gets accumulated.

Because biased guarantor requirements are concentrated among loans that perform well in real life, discrimination may be costly to the bank. If creditworthy female applicants cannot provide a guarantor, profitable projects go unfunded. In equilibrium, women may avoid applying for credit altogether—as they anticipate being asked for a personal guarantor. Moreover, in those cases where women do manage to come up with a guarantor so as to provide additional comfort to (biased) loan officers, there will be a cost for these entrepreneurs themselves as they are asked to put scarce social capital on the line.

We sketch three courses of action for banks that want to mitigate gender discrimination among loan officers. First, our results show clearly that discrimination is less prevalent among older and more experienced loan officers (at least in female-dominated sectors). Adding more senior officers to relatively junior teams can then be a straightforward way to reduce the risk of discriminatory lending.

Second, policies to mitigate the real-world impact of implicit biases may be called for. For example, banks can set branch-level goals for lending to women without a guarantor and hold those branches that do not meet this goal accountable. Successful female entrepreneurs can also be made more visible to loan officers, for instance by integrating them in banks' internal communication and training programs. This holds in particular for female entrepreneurs in stereotypically male industries. Banks can also conduct IATs with loan officers and reveal the results to those who hold implicit stereotypes. Alternatively, they can provide loan officers with factual information about gender discrimination in the loan officer population as a whole.[40]

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

Third, banks might consider replacing human with algorithmic decision-making altogether. While algorithmic credit scoring can reduce face-to-face discrimination in markets prone to implicit and explicit biases, it may fail to reduce (or even increase) disparities between and within social groups in lending terms (Bartlett et al. 2019; Fuster et al., 2020).

We end this paper with an observation about the generalizability of our findings. As mentioned before, Turkey is a middle-income country with a competitive banking sector but relatively conservative social norms. To what extent are our results portable across borders? One way to answer this question is to identify countries that are similar to Turkey in terms of economic and financial development as well as gender norms.[41] This yields a broad and varied group of countries across the world, including Egypt, Jordan, Lebanon, Morocco, Qatar, and the United Arab Emirates in the Middle East and North Africa; the Dominican Republic, Guatemala, and Paraguay in Latin America; Greece and Hungary in Europe; and Cambodia and Sri Lanka in Asia. In all these countries, discrimination by (parts of the) loan officer population may contribute to women's financial exclusion and, therefore, to a misallocation of entrepreneurial talent.

[41] More specifically, we identify the intersection of all countries within one standard deviation from Turkey in terms of GDP per capita, domestic credit to the private sector as a percentage of GDP, and the World Economic Forum Global Gender Gap Index.

References

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.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and Identity. *Quarterly Journal of Economics*, 115(3):715–753.

Alan, S., Ertac, S., and Mumcu, I. (2018). Gender Stereotypes in the Classroom and Effects on Educational Outcomes. *Review of Economics and Statistics*, 100(5):876–890.

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., Giuliano, P., and Nunn, N. (2013a). On the Origins of Gender Roles: Women and the Plough. *Quarterly Journal of Economics*, 128:469–530.

Alesina, A. F., Lotti, F., and Mistrulli, P. E. (2013b). 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. G. and Pierret, C. R. (2001). Employer Learning and Statistical Discrimination. *Quarterly Journal of Economics*, 116(1):313–350.

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.

Athey, S. and Imbens, G. (2016). Recursive Partitioning for Heterogeneous Causal Effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized Random Forests. *The Annals of Statistics*, 47(2):1148–1178.

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.

Baranov, V., De Haas, R., and Grosjean, P. (2020). Men. Causes and Consequences of Masculinity Norms. *mimeo*.

Bartlett, R., Morse, A., Stanton, R., and Wallace, N. (2019). Consumer-Lending Discrimination in the FinTech Era. *mimeo*.

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.

Bernstein, S., Korteweg, A., and Laws, K. (2017). Attracting Early-Stage Investors: Evidence from a Randomized Field Experiment. *Journal of Finance*, 72(2):509–538.

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.

Bisin, A. and Verdier, T. (2001). The Economics of Cultural Transmission and the Dynamics of Preferences. *Journal of Economic Theory*, 97(2):298–319.

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.

Blau, F. D. and Kahn, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, 55:789–865.

Bohren, A., Imas, A., and Rosenberg, M. (2019). The Dynamics of Discrimination: Theory and Evidence. *American Economic Review*, 109(10):3395–3436.

Bohren, J. A., Haggag, K., Imas, A., and Pope, D. G. (2020). Inaccurate Statistical Discrimination: An Identification Problem. *mimeo*.

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

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2016). Stereotypes. *Quarterly Journal of Economics*, 131(4):1753–1794.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2019). Beliefs about Gender. *American Economic Review*, 109(3):739–773.

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.

Botelho, F., Madeira, R. A., and Rangel, M. A. (2015). Racial Discrimination in Grading: Evidence from Brazil. *American Economic Journal: Applied Economics*, 7(4):37–52.

Brooks, A. W., Huang, L., Kearney, S. W., and Murray, F. E. (2014). Investors Prefer Entrepreneurial Ventures Pitched by Attractive Men. *PNAS*, 111(12):4427–4431.

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

Carrell, S. E., Page, M. E., and West, J. E. (2010). Sex and Science: How Professor Gender Perpetuates the Gender Gap. *Quarterly Journal of Economics*, 125:1101–1144.

Coffman, K. B. (2014). Evidence on Self-Stereotyping and the Contribution of Ideas. *Quarterly Journal of Economics*, 129(4):1625–1660.

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.

Davis, J. M. and Heller, S. B. (2017). Using Causal Forests to Predict Treatment Heterogeneity: An Application to Summer Jobs. *American Economic Review: Papers and Proceedings*, 107(5):546–550.

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.

Diep-Nguyen, H. and Dang, H. (2020). Social Collateral. *mimeo*.

Dobbie, W., Liberman, A., Paravisini, D., and Pathania, V. (2020). Measuring Bias in Consumer Lending. *mimeo*.

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.

Ewens, M. and Townsend, R. R. (2020). Are Early Stage Investors Biased Against Women. *Journal of Financial Economics*, 135(3):653–677.

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

Field, E., Jayachandran, S., and Pande, R. (2010). Do Traditional Institutions Constrain Female Entrepreneurship? A Field Experiment on Business Training in India. *American Economic Review: Papers & Proceedings*, pages 125–129.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., and Walther, A. (2020). Predictably Unequal? The Effects of Machine Learning on Credit Markets. *mimeo*.

Giuliano, P. (2018). *Gender: A Historical Perspective*, chapter 26, pages 645–672. Oxford University Press, New York.

Glover, D., Pallais, A., and Parienté, W. (2017). Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores. *Quarterly Journal of Economics*, 132(3):1219–1260.

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.

Gompers, P. A. and Wang, S. Q. (2017). Diversity in Innovation. *NBER Working Paper No. 23082*.

Gornall, W. and Strebulaev, I. A. (2020). Gender, Race, and Entrepreneurship: A Randomized Field Experiment on Venture Capitalists and Angels. *mimeo*.

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.

Grosjean, P. and Khattar, R. (2019). It's Raining Men! Hallelujah? The Long-Run Consequences of Male-Biased Sex Ratios. *The Review of Economic Studies*, 86(2):723–754.

Guiso, L., Monte, F., Sapienza, P., and Zingales, L. (2008). Culture, Gender, and Math. *Science*, 320:1164–1165.

Herbert, C. (2020). Gender Stereotypes and Entrepreneur Financing. *mimeo*.

Hsieh, C.-T., Hurst, E., Jones, C. I., and Klenow, P. J. (2019). The Allocation of Talent and US Economic Growth. *Econometrica*, 87(5):1439–1474.

Hu, A. and Ma, S. (2020). Human Interactions and Financial Investment: A Video-Based Approach. *mimeo*.

Iyer, R., Khwaja, A. I., Luttmer, E. F., and Shue, K. (2016). Screening Peers Softly: Inferring the Quality of Small Borrowers. *Management Science*, 62(6):1554–1577.

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

Karlan, D. (2007). Social Connections and Group Banking. *Economic Journal*, 117:F52–F84.

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

Turkey. mimeo.

Montoya, A. M., Parrado, E., Solis, A., and Undurraga, R. (2020). Bad Taste: Gender Discrimination in Consumer Credit Markets. *mimeo*.

Naaraayanan, S. L. (2020). Women's Inheritance Rights and Entrepreneurship Gap. *mimeo*.

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

OECD (2016). Entrepreneurship at a Glance 2016.

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.

Pozzolo, A. (2004). The Role of Guarantees in Bank Lending. *EFMA 2004 Basel Meetings Paper, European Financial Management Association, Basel, Switzerland*.

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.

Stiglitz, J. (1990). Peer Monitoring and Credit Markets. *World Bank Economic Review*, 4(3):351–366.

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].

Tibshirani, J., Athey, S., Friedberg, R., Hadad, V., Hirshberg, D., Miner, L., Sverdrup, E., Wager, S., and Wright, M. (2020). grf package. *https://grf-labs.github.io/grf/*.

Wager, S. and Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

World Economic Forum (2018). The Global Gender Gap Report.

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)	326	8.67	8.00	5.77	0	32
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

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). Appendix Table A1 contains all variable definitions.

Table 2: Applicant gender and loan rejection

Dependent variable: Rejection dummy	[1]	[2]	[3]	[4]	[5]
Female applicant	-0.008	-0.008	-0.015	-0.012	-0.012
	(0.024)	(0.024)	(0.025)	(0.025)	(0.025)
Participant is supervisor		0.062**	0.095***	0.094***	0.094***
		(0.025)	(0.033)	(0.033)	(0.033)
Participant is female			0.023	0.024	0.024
			(0.026)	(0.027)	(0.027)
Participant experience (years)			-0.002	-0.002	-0.002
			(0.003)	(0.003)	(0.003)
Participant age (years)			-0.005	-0.005	-0.005
			(0.003)	(0.003)	(0.003)
Participant risk aversion				-0.012	-0.012
				(0.010)	(0.010)
Participant IAT score				0.006	0.007
				(0.043)	(0.043)
Constant	0.395***	0.368***	0.552***	0.602***	0.471***
	(0.017)	(0.022)	(0.106)	(0.114)	(0.172)
R-squared	0.259	0.267	0.280	0.282	0.282
N	1,336	1,336	1,280	1,248	1,248
File FE	✓	✓	✓	✓	✓
City FE		✓	✓	✓	✓
Participant RE					✓

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. Standard errors are shown in parentheses and clustered at the participant level. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively. Appendix Table A1 contains all variable definitions.

Table 3: Applicant gender and guarantor requirements

Dependent variable: Guarantor dummy

	[1]	[2]	[3]	[4]	[5]
Female applicant	0.063**	0.060**	0.067**	0.068**	0.062**
	(0.030)	(0.030)	(0.030)	(0.031)	(0.029)
Participant is supervisor		0.051	0.051	0.061	0.043
		(0.041)	(0.056)	(0.057)	(0.057)
Participant is female			-0.027	-0.027	-0.018
			(0.039)	(0.040)	(0.041)
Participant experience (years)			0.002	0.003	0.003
			(0.005)	(0.005)	(0.005)
Participant age (years)			0.001	0.000	-0.000
			(0.006)	(0.006)	(0.006)
Participant risk aversion				0.016	0.014
				(0.014)	(0.015)
Participant IAT score				-0.045	-0.052
				(0.066)	(0.069)
Constant	0.236***	0.217***	0.170	0.130	0.536**
	(0.022)	(0.026)	(0.176)	(0.197)	(0.250)
R-squared	0.152	0.191	0.200	0.201	0.192
N	814	814	778	758	758
File FE	✓	✓	✓	✓	✓
City FE		✓	✓	✓	✓
Participant RE					✓

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. Standard errors are shown in parentheses and clustered at the participant level. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively. Appendix Table A1 contains all variable definitions.

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	0.073*	0.071*	0.115**	0.060	0.120**	0.013
	(0.042)	(0.043)	(0.056)	(0.040)	(0.047)	(0.042)
R-squared	0.409	0.293	0.356	0.295	0.406	0.270
N	344	414	354	404	325	433
t-test p-value	0.490		0.215		0.044	
	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	0.118***	-0.036	0.083	0.061*	0.016	0.104**
	(0.036)	(0.056)	(0.077)	(0.034)	(0.048)	(0.043)
R-squared	0.301	0.357	0.416	0.204	0.343	0.323
N	474	284	217	541	387	371
t-test p-value	0.010		0.399		0.087	
File FE	✓	✓	✓	✓	✓	✓
City FE	✓	✓	✓	✓	✓	✓
Participant covariates	✓	✓	✓	✓	✓	✓
Participant RE	✓	✓	✓	✓	✓	✓

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 non-binary 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. Standard errors are shown in parentheses and clustered at the participant level. All regressions include the same participant covariates as in column [5] of Table 2. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively. Appendix Table A1 contains all variable definitions.

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	0.107***	-0.020	0.090	0.100*	0.150**	0.115**	0.161***	0.067
	(0.038)	(0.056)	(0.057)	(0.060)	(0.071)	(0.051)	(0.061)	(0.056)
R-squared	0.195	0.238	0.378	0.314	0.319	0.307	0.363	0.270
N	453	305	211	242	214	239	201	252
t-test p-value	0.031		0.451		0.344		0.127	
			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			0.143***	-0.017	0.074	0.097**	0.099	0.142**
			(0.051)	(0.066)	(0.092)	(0.046)	(0.062)	(0.055)
R-squared			0.333	0.317	0.509	0.203	0.380	0.311
N			282	171	124	329	221	232
t-test p-value			0.028		0.410		0.304	
File FE	✓	✓	✓	✓	✓	✓	✓	✓
City FE	✓	✓	✓	✓	✓	✓	✓	✓
Participant covariates	✓	✓	✓	✓	✓	✓	✓	✓
Participant RE	✓	✓	✓	✓	✓	✓	✓	✓

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 non-binary 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. Standard errors are shown in parentheses and clustered at the participant level. All regressions include the same participant covariates as in column [5] of Table 2. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively. Appendix Table A1 contains all variable definitions.

Table 6: Availability of borrower information and gender bias

Dependent variable:	Rejection dummy		Guarantor dummy	
	[1]	[2]	[3]	[4]
Female applicant	-0.015	0.014	0.036	0.028
	(0.025)	(0.043)	(0.029)	(0.050)
No subj.	0.056	0.087**	-0.039	-0.056
	(0.035)	(0.043)	(0.048)	(0.059)
No obj.	-0.049	-0.037	-0.016	-0.013
	(0.036)	(0.046)	(0.048)	(0.057)
No subj. × Female applicant		-0.063		0.033
		(0.059)		(0.074)
No obj. × Female applicant		-0.024		-0.005
		(0.062)		(0.067)
R-squared	0.230	0.230	0.235	0.235
N	1,246	1,246	808	808
File FE	✓	✓	✓	✓
City FE	✓	✓	✓	✓
Participant covariates	✓	✓	✓	✓
Participant RE	✓	✓	✓	✓

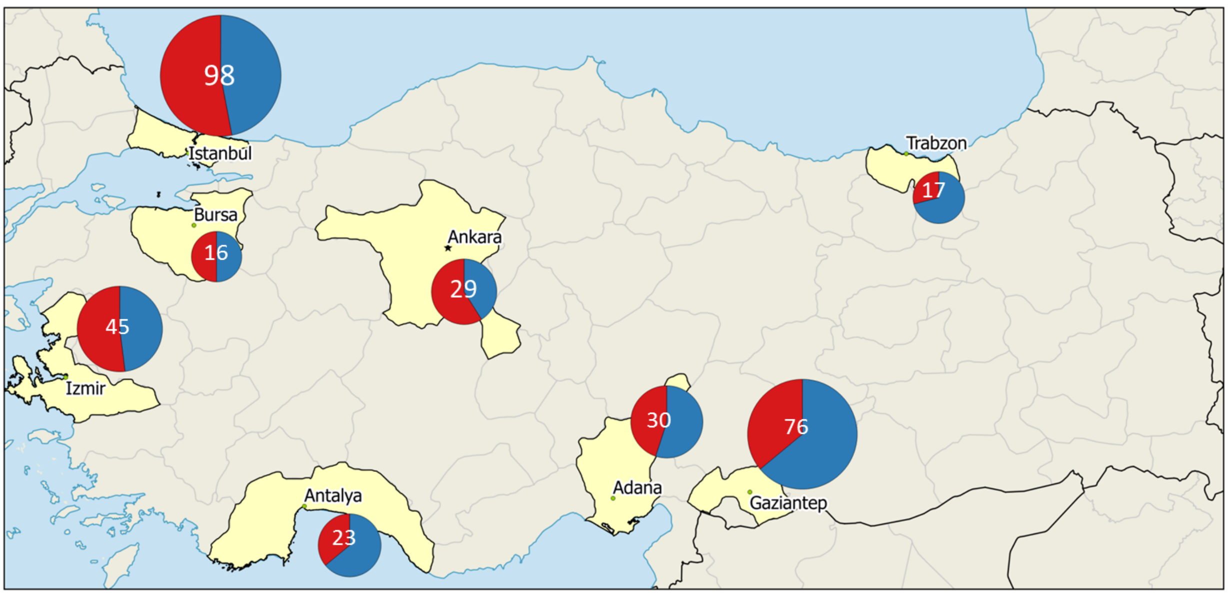
Notes: The dependent variable in columns [1] and [2] is a *Rejection dummy* that equals '1' if the participant declines the credit application and '0' if the participant approves it. The dependent variable in columns [3] and [4] is a *Guarantor dummy* that equals '1' if the participant approves the credit 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 [5] of Table 2. Standard errors are shown in parentheses and clustered at the participant level. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively. Appendix Table A1 contains all variable definitions.

Table 7: Applicant gender, sectoral gender composition, and guarantor requirements

Dependent variable: Guarantor dummy						
	Male-dominated sectors	Female-dominated sectors	Male-dominated sectors		Female-dominated sectors	
			Below median IAT	Above median IAT	Below median IAT	Above median IAT
	[1]	[2]	[3]	[4]	[5]	[6]
Female applicant	0.148**	0.049	0.024	0.270***	0.015	0.090*
	(0.059)	(0.037)	(0.090)	(0.088)	(0.063)	(0.055)
R-squared	0.208	0.222	0.325	0.502	0.344	0.350
N	206	504	106	100	259	245
t-test p-value	0.079		0.025		0.183	
File FE	✓	✓	✓	✓	✓	✓
City FE	✓	✓	✓	✓	✓	✓
Participant covariates	✓	✓	✓	✓	✓	✓
Participant RE	✓	✓	✓	✓	✓	✓

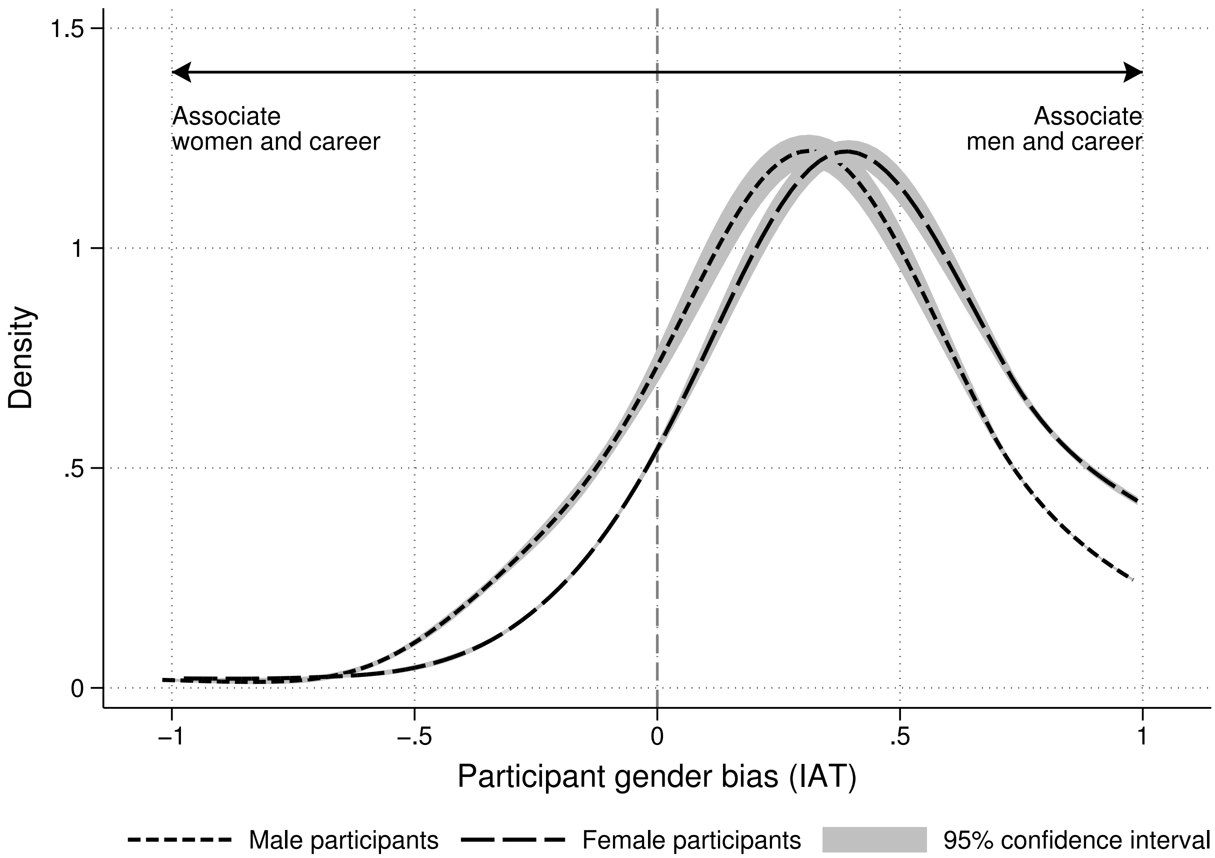
Notes: The dependent variable is a *Guarantor dummy* that equals '1' if the participant approves the credit application but requests a guarantor and '0' if the participant approves it without requesting a guarantor. Female- and male-dominated sectors are defined by the share of firms with majority female ownership at the 2-digit ISIC industry level using data from the EBRD–World Bank Banking Environment and Performance Survey (BEEPS) V and VI. The sample is restricted to the first round of the experiment. Standard errors are shown in parentheses and clustered at the participant level. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively. Appendix Table A1 contains all variable definitions.

Figure 1: Geographical distribution of participants across the bank's regional offices



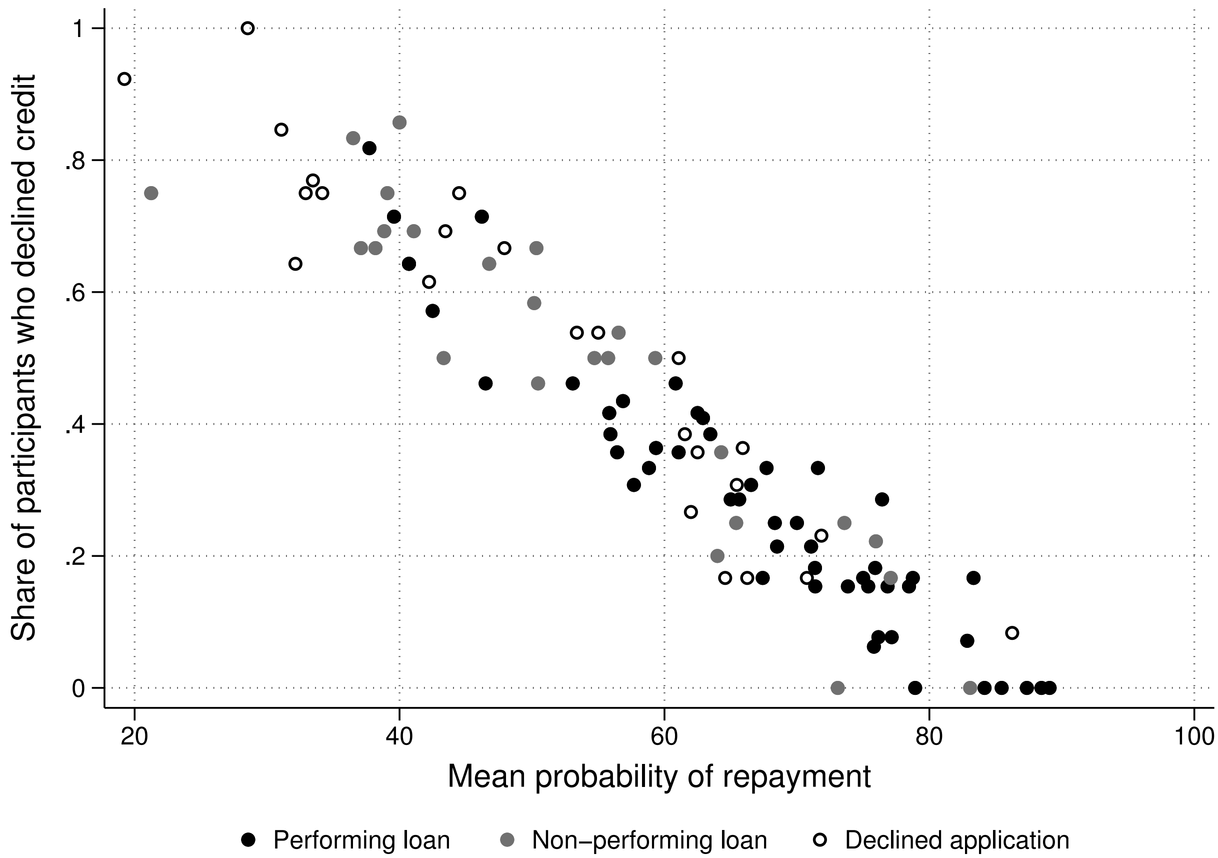
Notes: This map shows the number and gender of the participants in the eight Turkish regional bank offices that participated in the experiment. Circle size is proportional to the number of participants. The percentage of female (male) participants is shown in red (blue).

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



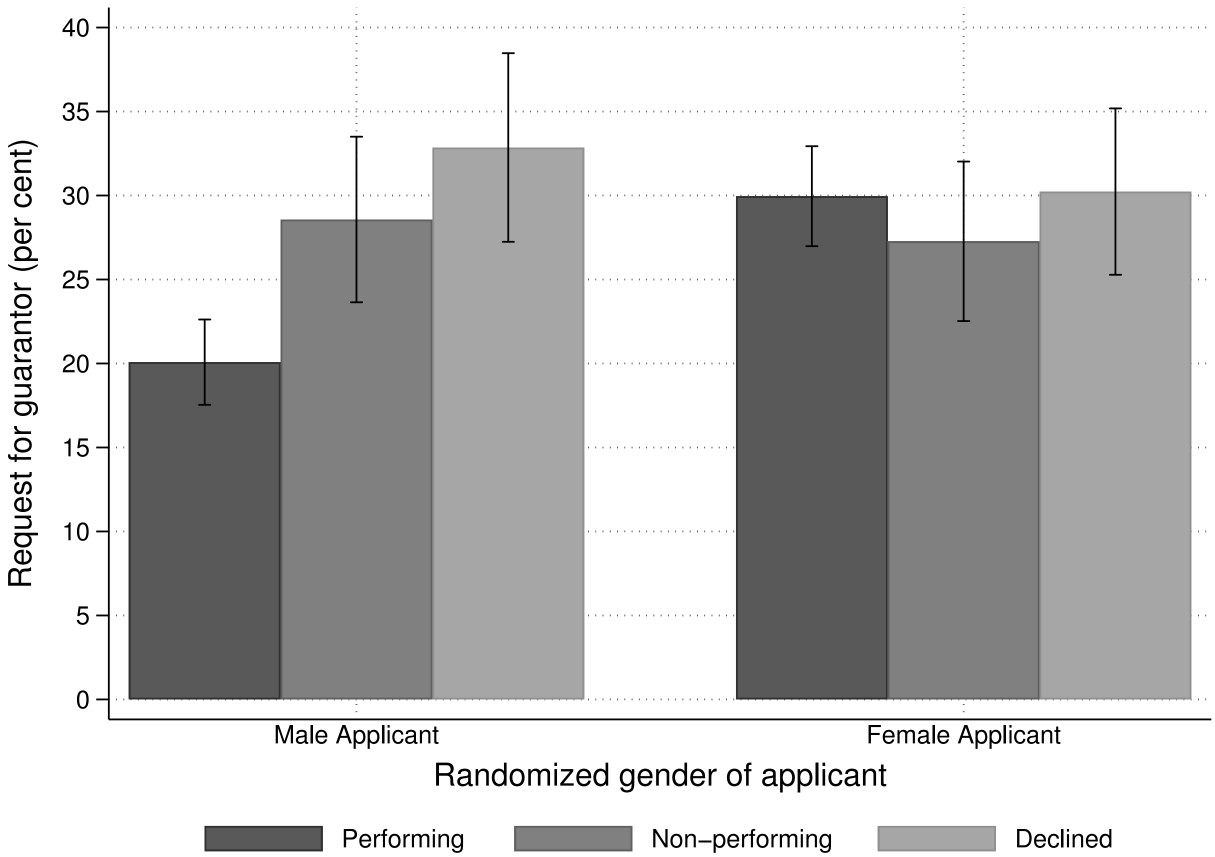
Notes: This figure shows a local polynomial smooth with 95 per cent confidence intervals of the variable *Participant gender bias (IAT)* for male (short dash) and female (long dash) participants, respectively. The combined two-sample Kolmogorov-Smirnov test statistic is 0.181 and has a p-value of 0.01. Appendix Table A1 contains all variable definitions.

Figure 3: Expected repayment and loan rejection rates



Notes: The x-axis shows the within-file mean, across participants, of the subjective repayment probability. The y-axis shows the share of participants who declined the loan application. The figure is based on the first round of the experiment. Appendix Table A1 contains all variable definitions.

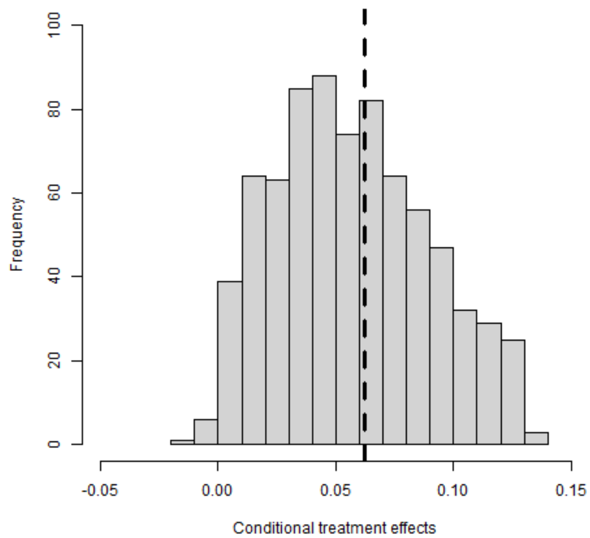
Figure 4: Guarantor requirements, by loan quality and applicant gender



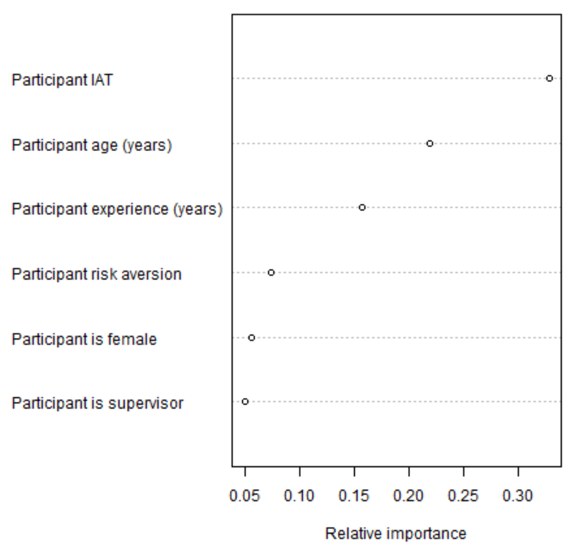
Notes: This figure shows the percentage of loan applications approved during the experiment and for which participants requested a guarantor. Bars are shown for approved loans repaid in real life (dark gray), approved loans that were defaulted on in real life (medium gray), and loan applications rejected 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. Appendix Table A1 contains all variable definitions.

Figure 5: Applicant gender and guarantor requirements – Heterogeneous treatment effects

Panel A: Distribution of conditional treatment effects



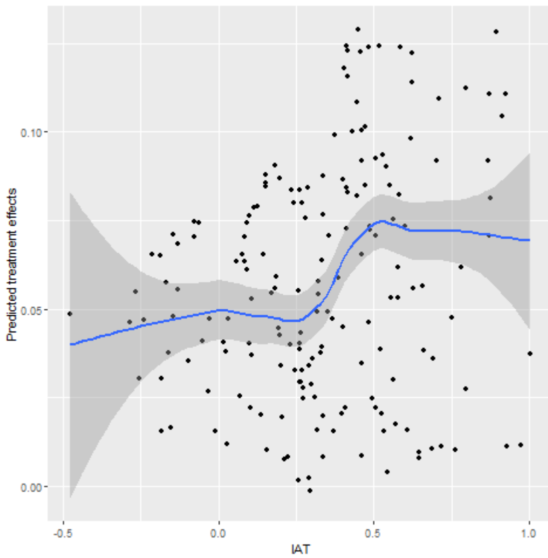
Panel B: Relative importance of covariates



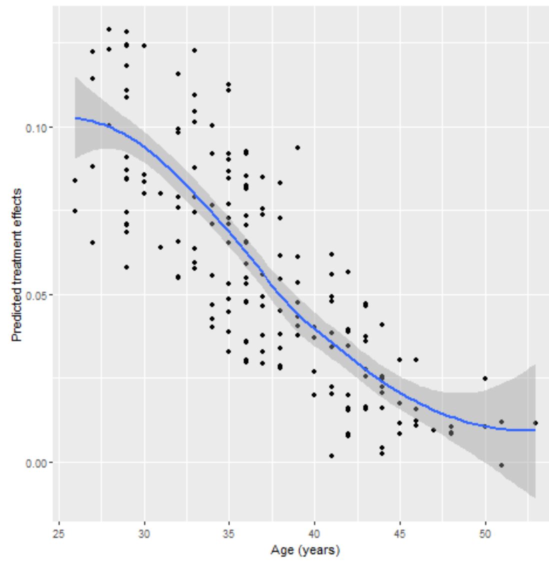
Notes: This figure shows results from a generalized causal forest model with 20,000 trees and honest splitting (Athey, Tibshirani and Wager, 2019). The outcome is the *Guarantor dummy* and the covariates are the participant characteristics in column [5] of Table 3. *Female applicant* is the treatment variable. Panel A shows the distribution of the conditional treatment effects. The black dashed line indicates the average treatment effect from the baseline model (Table 3, column [5]). Panel B shows the variable *Relative importance*. This is a weighted sum of how many times a loan officer trait was used to split at each depth in the forest when estimating treatment heterogeneity.

Figure 6: Predicted treatment effects by implicit gender bias, age, and experience

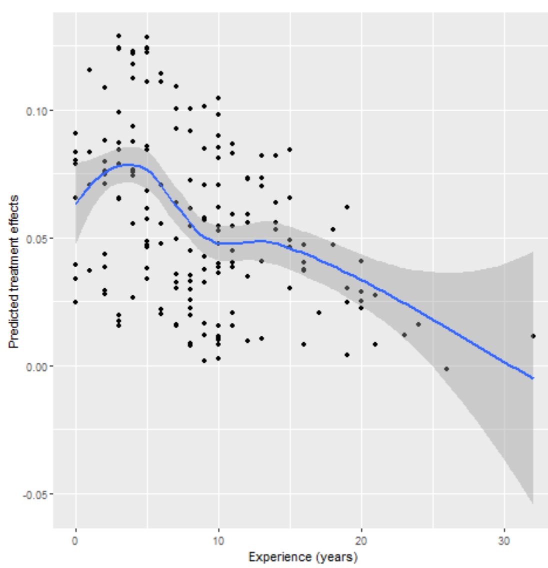
Panel A



Panel B

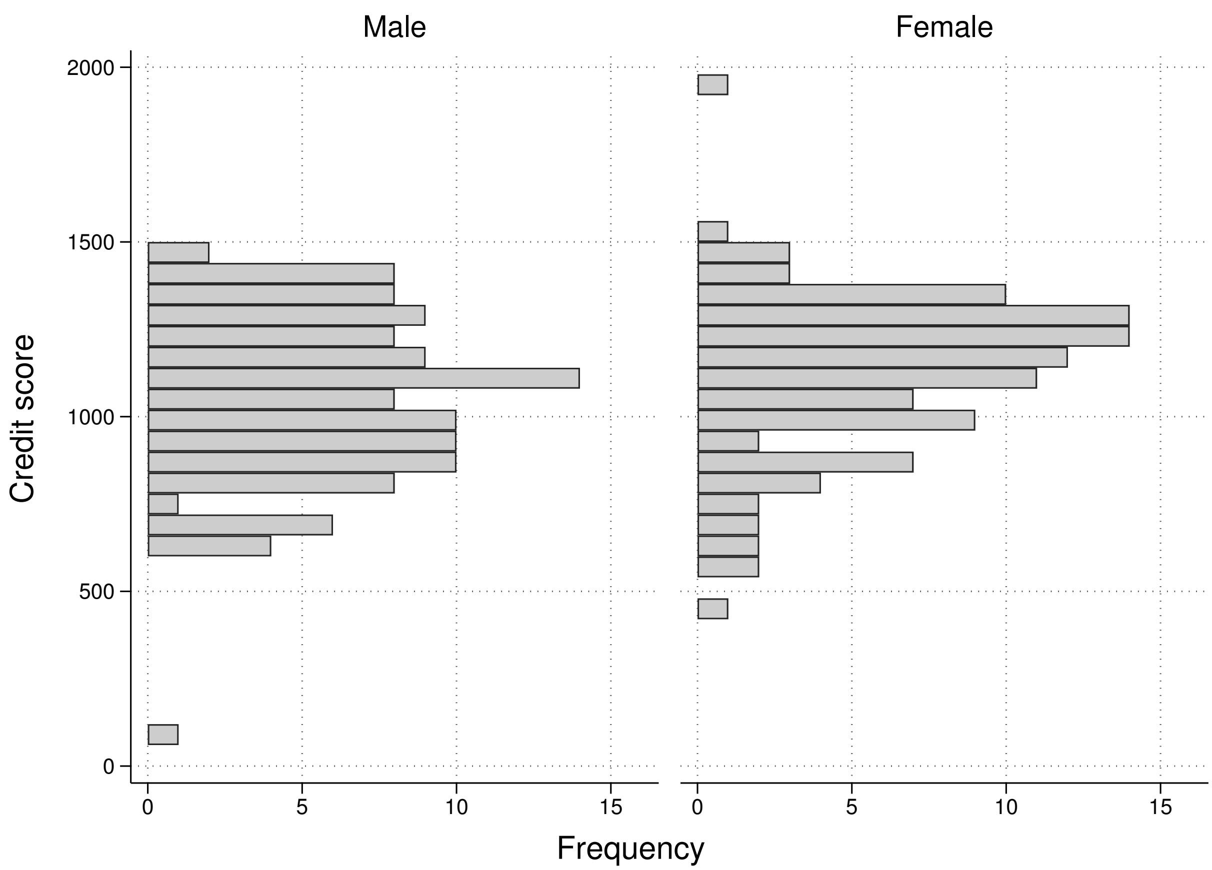


Panel C



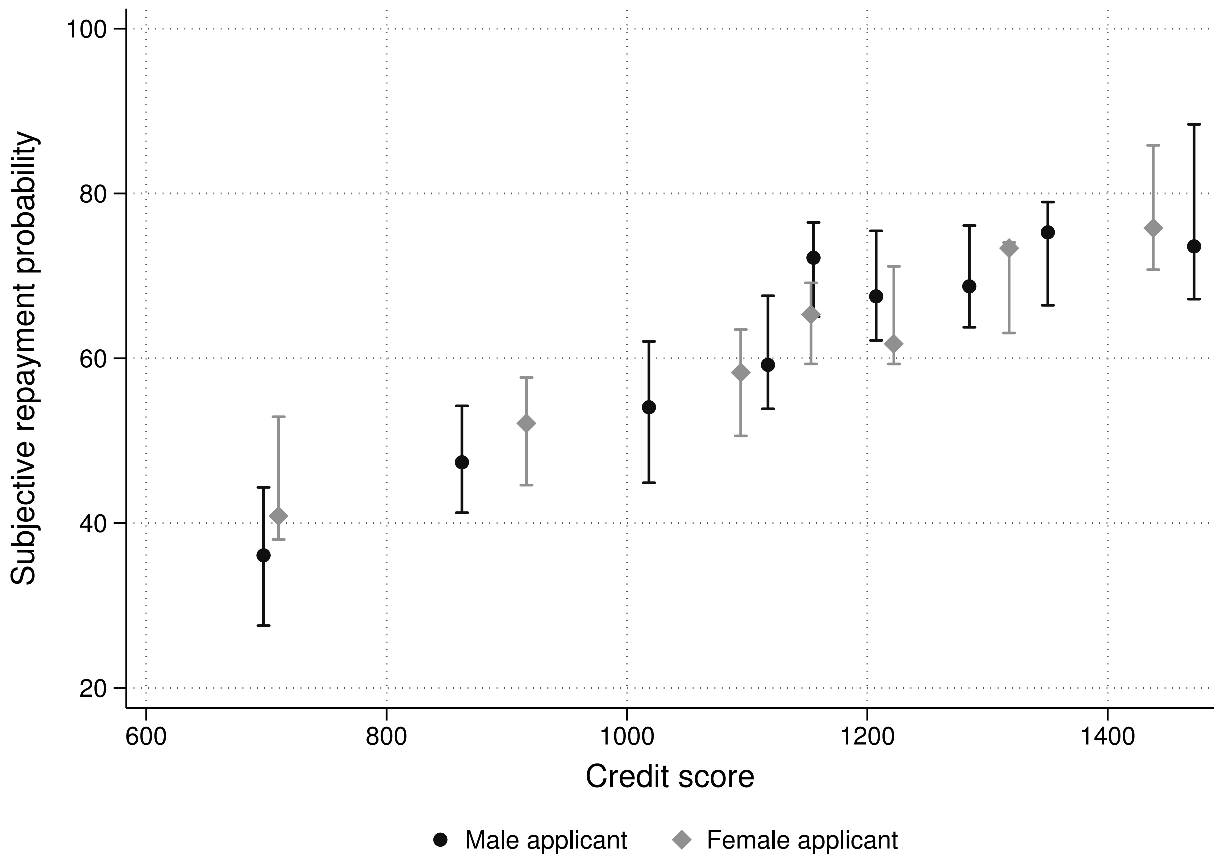
Notes: Plotted points represent individual loan officers. The horizontal axis indicates implicit gender bias (IAT score, Panel A), age (Panel B), and experience (Panel C). These are the three most important treatment moderators according to the causal forest algorithm (cf. Figure 5, Panel B). The vertical axis in each panel indicates the conditional average treatment effect (CATE) predicted by our causal forest. The lines display the local smoothed polynomial relationship between the loan officer trait and the CATE. The treatment effects are predicted by feeding our test sample (30% of the full sample) through the trees grown by the causal forest algorithm on the basis of the splitting sample (70% of the full sample).

Figure 7: Credit score by real-life gender of applicant



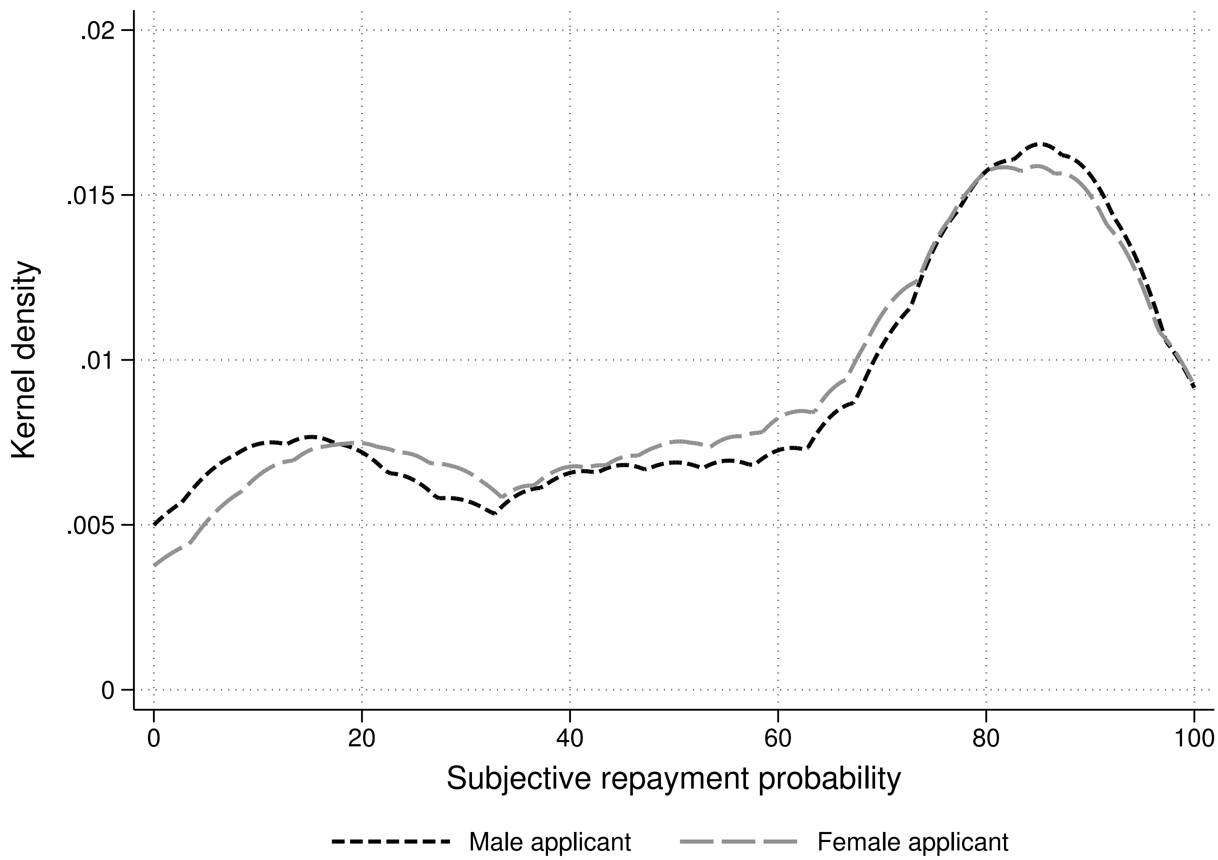
Notes: This figure shows the distribution of the variable *Credit score* for loan application files that were male (left) and female (right) in real life. Credit scores are from the KKB credit registry and higher scores indicate lower credit risk. The figure is based on the 250 loan application files from which the 100 files used in the experiment were drawn. The combined two-sample Kolmogorov-Smirnov test statistic is 0.168 and has a p-value of 0.087. Appendix Table A1 contains all variable definitions.

Figure 8: Credit score and subjective repayment probability, by randomized applicant gender



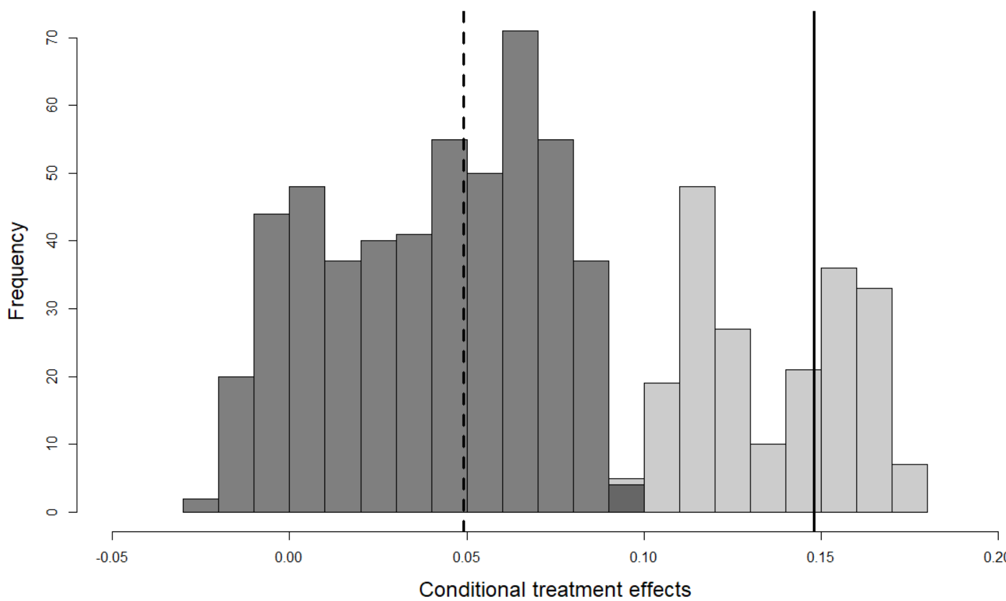
Notes: This figure shows binned scatter plots for male applicants (dark grey dots) and female applicants (light grey diamonds) using robust pointwise confidence intervals. The data reflect all decisions in the first round of the experiment. The number of bins is not pre-determined but data driven and the integrated mean squared errors are minimized. The confidence intervals are at the 95% level and based on a cubic B-spline regression estimate of subjective repayment probability on the credit score. Credit scores are provided by the KKB credit registry and higher scores indicate lower credit risk. Appendix Table A1 contains all variable definitions.

Figure 9: Subjective repayment probability by randomized gender of loan application



Notes: This figure shows the kernel density curves of the variable *Subjective repayment probability* for loan applications that were presented as male (black short dash) and female (gray long dash), respectively. The figure is based on the 1,329 decisions made in the first round of the experiment. The combined two-sample Kolmogorov-Smirnov test statistic is 0.404 and has a p-value of 0.649. Appendix Table A1 contains all variable definitions.

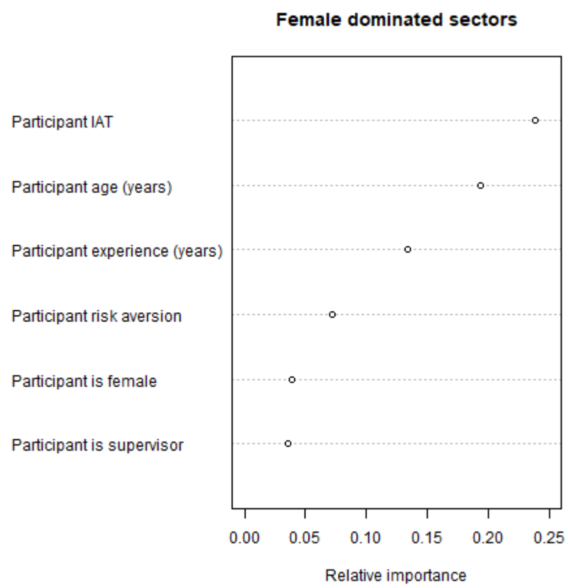
Figure 10: Conditional treatment effects in male- versus female-dominated sectors



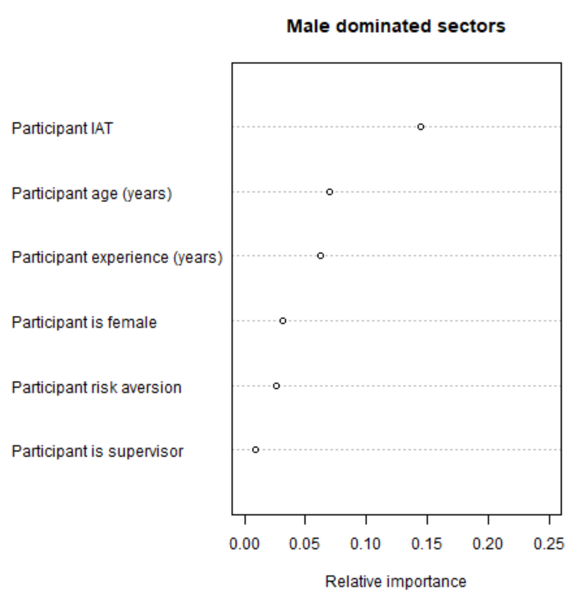
Notes: This figure shows results from two separate generalized causal forest models each with 20,000 trees and honest splitting (Athey, Tibshirani and Wager, 2019). The outcome is the *Guarantor dummy* and the covariates are the participant characteristics in column [5] of Table 3. *Female applicant* is the treatment variable. The dark (light) grey bars show the distribution of the conditional treatment effects for female (male) dominated sectors. The dashed (solid) line indicates the average treatment effect from the baseline model for female (male) dominated sectors as in Table 7, column [2] (Table 7, column [1]).

Figure 11: Heterogeneous treatment effects - Relative importance of covariates

Panel A

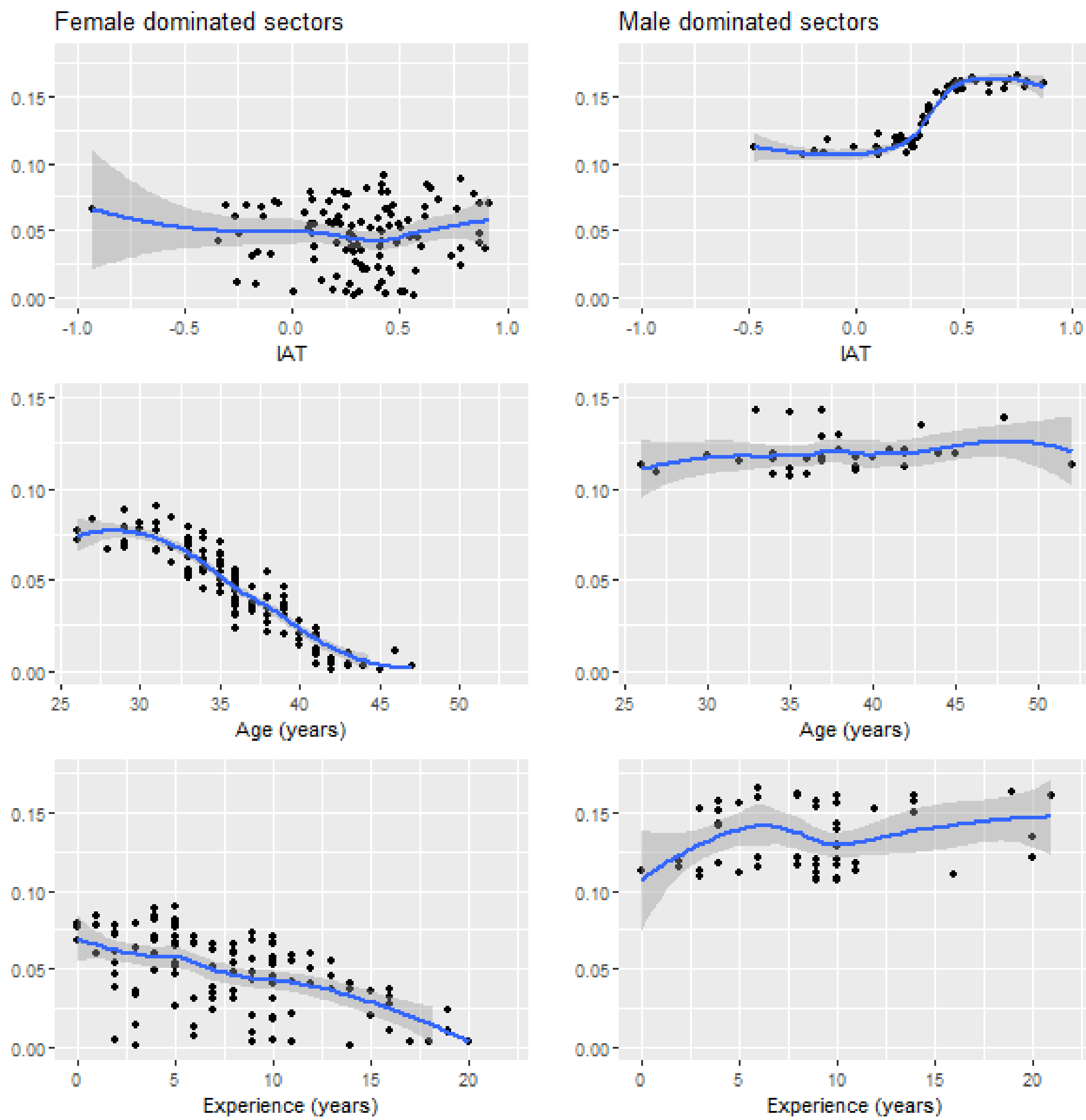


Panel B



Notes: This figure shows results from two separate generalized causal forest models each with 20,000 trees and honest splitting (Athey, Tibshirani and Wager, 2019). The outcome is the *Guarantor dummy* and the covariates are the participant characteristics in column [5] of Table 3. The horizontal axes of Panels A and B show the variable *Relative importance*. This is a weighted sum of how many times a loan officer trait was used to split at each depth in the forest when estimating treatment heterogeneity in female-dominated sectors (Panel A) or male-dominated sectors (Panel B).

Figure 12: Predicted treatment effects across sectors, by implicit bias, age, and experience



Notes: Plotted points represent individual loan officers. The horizontal axis indicates implicit gender bias (IAT score, top), age (middle), and experience (bottom). These are the three most important treatment moderators according to the causal forest algorithm (cf. Figure 11). The vertical axis in each panel indicates the conditional average treatment effects (CATE) predicted by our causal forest. The lines display the local smoothed polynomial relationship between the loan officer trait and the CATE. The treatment effects are predicted for female (male) dominated sectors by feeding our test sample (30% of the sample corresponding to female (male) dominated sectors) through the trees grown by the causal forest algorithm on the basis of the splitting sample (70% of the sample corresponding to female (male) dominated sectors).

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 of any 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

Male-dominated sector	Dummy variable equal to 1 if the share of firms with majority female ownership, in a given industry, is less than or equal to the median industry share; 0 otherwise. The share of female-owned firms is calculated at the 2-digit ISIC level using pooled observations from the EBRD–World Bank BEEPS V and VI surveys.
Female-dominated sector	Dummy variable equal to 1 if the share of firms with majority female ownership, in a given industry, is greater than the median industry share; 0 otherwise. The share of female-owned firms is calculated at the 2-digit ISIC level using pooled observations from the EBRD–World Bank BEEPS V and VI surveys.
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.
Subjective repayment probability	Continuous variable which takes values from 0 to 100. For each decision, the participant estimates the likelihood that the loan would be repaid. Higher values indicate a greater chance of repayment.
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: 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	0.092**	1.000						
Participant age (years)	0.567***	0.037	1.000					
Participant risk aversion	0.033	0.149***	-0.011	1.000				
Participant experience (years)	0.205***	0.066*	0.558***	-0.034	1.000			
Participant IAT score	0.093***	0.188***	0.081**	-0.003	0.118***	1.000		
Female applicant	0.000	-0.000	0.000	0.000	0.000	-0.000	1.000	
Rejection dummy	0.074**	0.035	0.012	-0.012	-0.035	0.010	-0.020	1.000

Notes: The sample is restricted to the first round. *, **, *** indicate significance at the 10, 5 and 1 per cent level, respectively. Appendix Table A1 contains all variable definitions.

Table A3: Predictors of participant gender bias

Dependent variable: Participant gender bias (IAT)	
	[1]
Participant is female	0.114***
	(0.036)
Participant experience (years)	0.006
	(0.004)
Participant age (years)	-0.001
	(0.004)
Participant is supervisor	0.045
	(0.044)
Participant risk aversion	-0.007
	(0.013)
Constant	0.283*
	(0.151)
R-squared	0.051
N	312

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. Appendix Table A1 contains all variable definitions.

Table A4: 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		✓	✓	✓	✓
Region FE			✓	✓	✓
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* as provided by the KKB credit registry. 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. Appendix Table A1 contains all variable definitions.

Table A5: Applicant gender and subjective repayment probability

Dependent variable: Subjective repayment probability (%)					
	[1]	[2]	[3]	[4]	[5]
Female applicant	0.553	0.535	1.187	0.873	0.921
	(1.399)	(1.403)	(1.423)	(1.457)	(1.450)
Participant is supervisor		-1.813	-3.300	-3.059	-3.178
		(1.882)	(2.409)	(2.442)	(2.456)
Participant is female			-2.138	-2.322	-2.349
			(1.819)	(1.910)	(1.921)
Participant experience (years)			0.012	0.023	0.016
			(0.198)	(0.203)	(0.205)
Participant age (years)			0.218	0.132	0.130
			(0.247)	(0.246)	(0.248)
Participant risk aversion				-0.506	-0.480
				(0.661)	(0.663)
Participant IAT score				1.711	1.605
				(2.772)	(2.804)
Constant	59.835***	60.612***	53.525***	58.041***	61.390***
	(1.149)	(1.441)	(8.169)	(8.295)	(12.432)
R-squared	0.268	0.277	0.288	0.290	0.289
N	1,329	1,329	1,273	1,243	1,243
File FE	✓	✓	✓	✓	✓
City FE		✓	✓	✓	✓
Participant RE					✓

Notes: The dependent variable is *Subjective repayment probability* which ranges between 0 and 100. The sample is restricted to the first round of the experiment. Standard errors are shown in parentheses and clustered at the participant level. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively. Appendix Table A1 contains all variable definitions.

Table A6: Gender of the entrepreneur and loan officers' risk perceptions

Dependent variable: Project risk the loan officer expects the entrepreneur to choose

	Loan officer's perception of:	
	Entrepreneur's risk choice	Entrepreneur's risk choice with credit
	[1]	[2]
Female entrepreneur	-0.229**	-0.157
	(0.115)	(0.115)
Pseudo R-squared	0.008	0.006
N	333	333

Notes: This table uses data from a separate experimental module in which participants were randomly matched with a (real-life) entrepreneur. Participants were informed about the gender, age, and sector of the entrepreneur they had been matched with. Prior to the experimental sessions, the entrepreneurs had been asked to pick one out of six entrepreneurial bets that were increasing in riskiness, in the spirit of Eckel and Grossman (2008). They were asked to do so once for a project they would finance with a loan and once for a project financed without debt. During the experiment, loan officers were then asked to guess which risky bet they thought their matched entrepreneur had chosen. They were paid if they guessed correctly. The ordered probit specifications in columns [1] and [2], regress the participant's perceptions of their matched entrepreneur's risk taking (on a 1-6 scale) on the gender of the entrepreneur for a project funded without and with credit, respectively. Both specifications control for the two other known traits of the matched entrepreneur (age and sector).

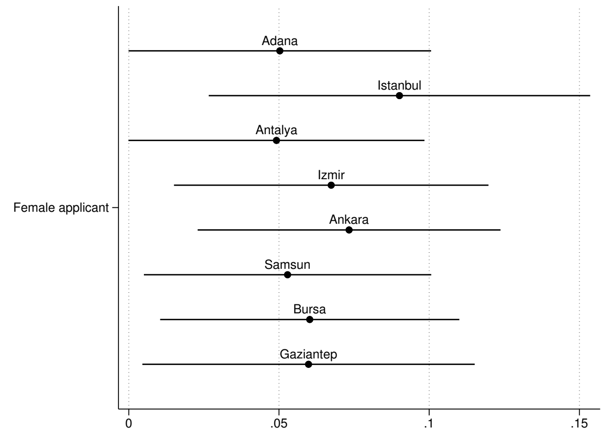
Table A7: Classification of 2-digit ISIC sectors as female- or male-dominated

ISIC code	Sector description	Female-dominated sector	Number of files	Number of decisions: First round	Number of decisions: Second round
15	Manufacture of food products and beverages	1	2	25	27
17	Manufacture of textiles	1	5	64	63
18	Manufacture of wearing apparel; dressing and dyeing of fur	1	7	89	91
25	Manufacture of rubber and plastics products	0	1	14	12
26	Manufacture of other non-metallic mineral products	0	1	16	14
29	Manufacture of machinery and equipment not elsewhere classified	0	1	14	12
36	Manufacture of furniture; manufacturing not elsewhere classified	1	3	37	36
45	Construction	0	1	13	13
50	Sale, maintenance and repair of motor vehicles and motorcycles; retail sale of automotive fuel	0	5	62	63
51	Wholesale trade and commission trade, except of motor vehicles and motorcycles	0	14	189	189
52	Retail trade, except of motor vehicles and motorcycles; repair of personal and household goods	1	36	484	476
55	Hotels and restaurants	1	8	105	116
60	Land transport; transport via pipelines	1	6	78	79
93	Other service activities	0	3	41	40
	Unable to classify		7	105	103

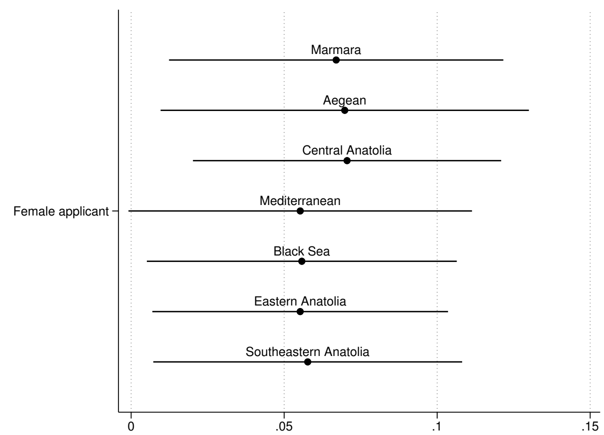
Notes: This table shows, for the 2-digit ISIC codes of the 100 files used in the experiment, whether the sector is classified as being a *Female-dominated sector*, the number of files in each 2-digit sector, and the number of decisions made during the experiment based on the files of each 2-digit sector. Female-dominated sectors are defined by the share of firms with majority female ownership at the 2-digit ISIC industry level using data from the EBRD–World Bank Business Environment and Enterprise Performance Survey (BEEPS) V and VI.

Figure A1: Indirect gender discrimination: Heterogeneity

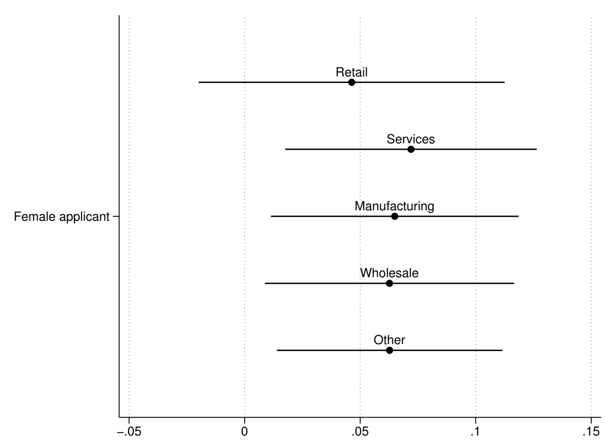
Panel A: Heterogeneity by experiment location



Panel B: Heterogeneity by province of original loan application

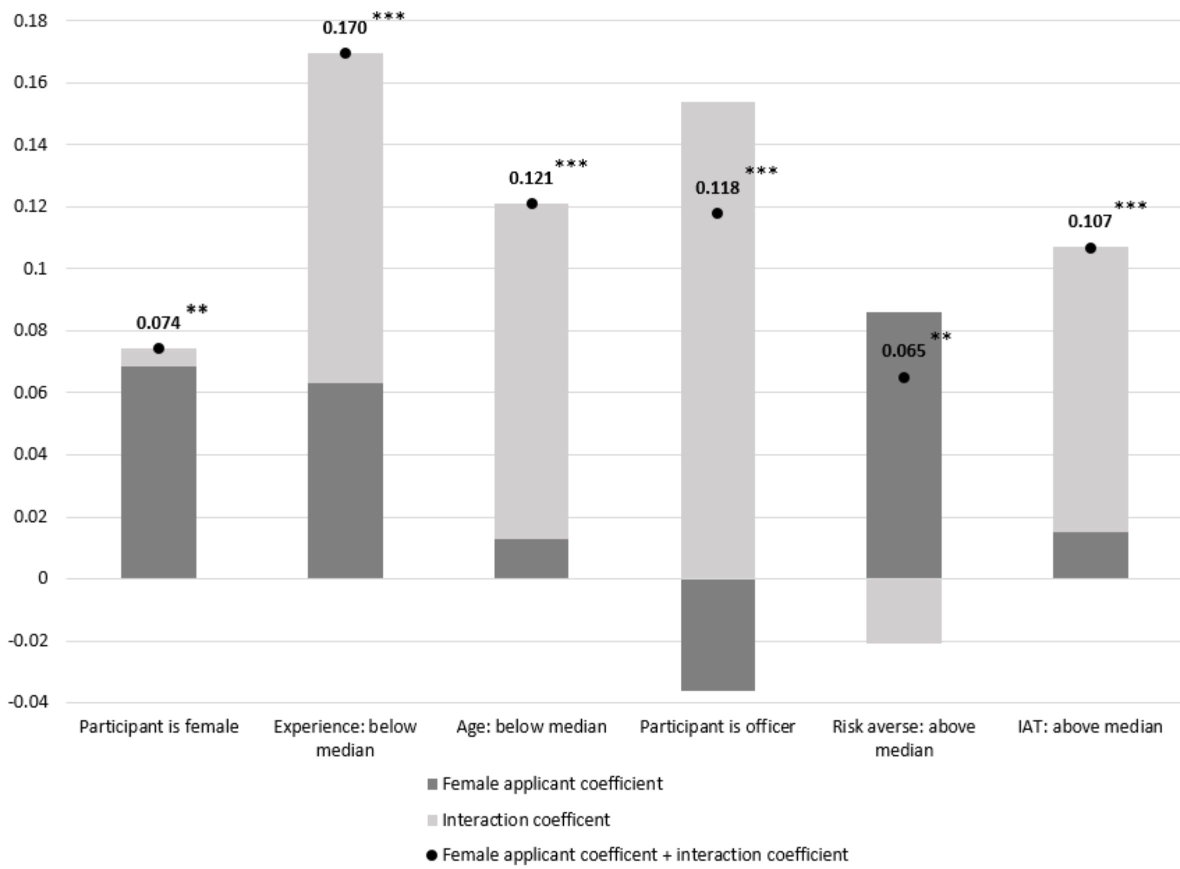


Panel C: Heterogeneity by macro-sectors



Notes: This figure shows estimated coefficients for *Female applicant* using the same specification as in column [5] of Table 3. Each dot reflects the coefficient based on the full sample minus the observations from the indicated city, province, or industry in Panel A, B and C, respectively. 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. In Panel A, the coefficients are ordered from highest (top) to lowest (bottom) regional household disposable income in 2016. 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". In Panel B, the coefficients are ordered from highest (top) to lowest (bottom) regional income level per capita in 2016. Appendix Table A1 contains all variable definitions.

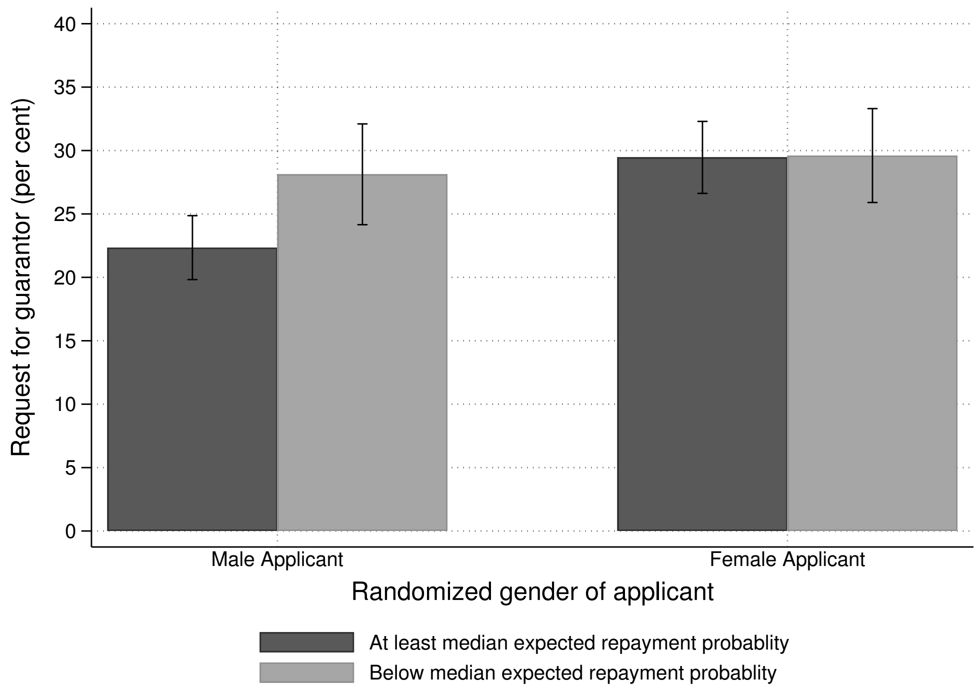
Figure A2: Heterogeneous guarantor requirements: Fully interacted models



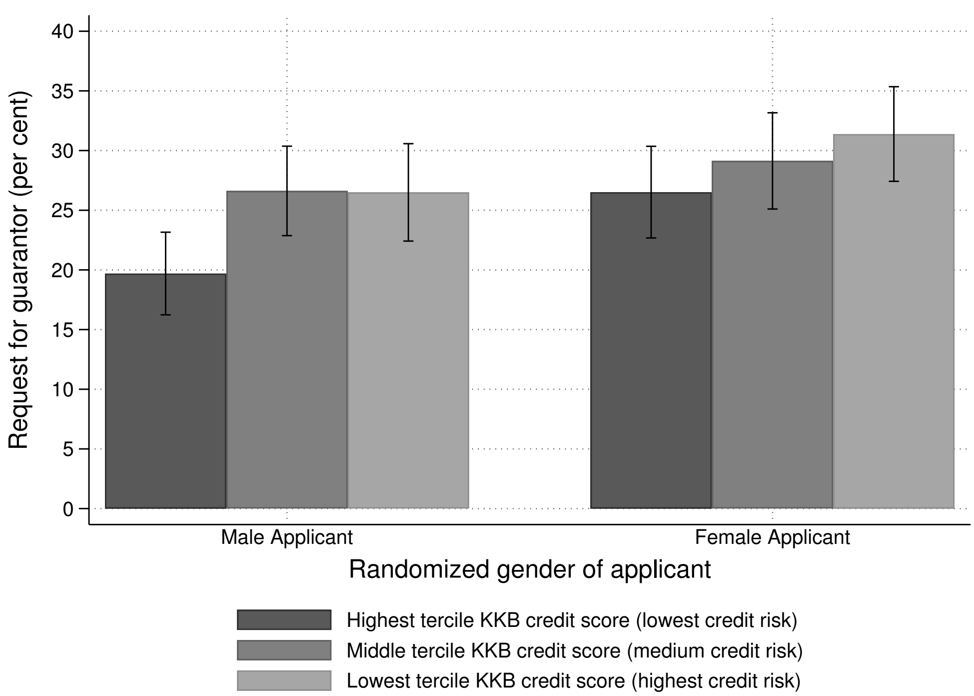
Notes: This figure shows coefficients from linear fully interacted models where the dependent variable is a *Guarantor dummy* that equals '1' if the participant approves the application but requests a guarantor and '0' if the participant approves without a guarantor. The sample is restricted to the first 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 [5] 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. Appendix Table A1 contains all variable definitions.

Figure A3: 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. Appendix Table A1 contains all variable definitions.