How Costly Are Cultural Biases? Evidence From FinTech[*]

Francesco D'Acunto[†]
Boston College

Pulak Ghosh[§]
IIMB

Rajiv Jain[‡]
Faircent

Alberto G. Rossi[¶]
Georgetown University

This version: October 2021

Abstract

We propose a unique field setting to detect and estimate the effects of cultural biases on high-stake economic decision-making—a leading Indian peer-to-peer lending platform paired with an automated robo-advising tool. Comparing the choices lending consumers ("lenders") make with those proposed by the tool, we find that both in-group vs. out-group bias and implicit bias based on religion and caste are pervasive and sizable. Culturally-biased choices and subsequent debiasing are stronger in locations with higher historical inter-ethnic and inter-caste conflict. Cultural biases affect performance negatively: lenders face 14% higher default rates before debiasing and increase the returns of their loan portfolios between 4.5 and 7.3 percentage points after debiasing. The substantially higher risk of the marginal borrowers from favorite demographic groups largely explains the worse performance of culturally-biased choices.

Keywords: Discrimination, Implicit Bias, Cultural Economics, Robo-Advising, Biased Beliefs, Lending, Partisanship, Inter-ethnic Conflict, Social Conditioning, Social Norms, Religion, Caste.

---

[*]For very helpful comments, we thank Tetyana Balyuk, Vicki Bogan, Paul Goldsmith-Pinkham, Luigi Guiso, Apoorv Gupta, Filippo Mezzanotti, Anirban Mitra, Adair Morse, Chris Parsons, Vesa Pursiainen, Antoinette Schoar, Kelly Shue, Lea Stern, Huan Tang, David Thesmar, Ansgar Walther, as well as seminar participants at the University of Rochester, EPFL, Université de Lausanne, HEC Paris, Purdue University, Aarhus University, Arizona State University, City University of Hong Kong, University of Alberta, University of Missouri, the WEFIDEV online seminar series, the JILAEE webinar series, the 2021 Western Finance Association Annual Meeting, the 2021 SFS Finance Cavalcade, the 2021 Midwest Finance Association Meeting, the CEBI Workshop on Subjective Beliefs and Household Finance, and the 2020 Bartunek Research Forum. All errors are our own.

[†]Carroll School of Management, Boston College, Chestnut Hill, MA, USA. e-Mail: dacuntof@bc.edu

[‡]Fairassets Technologies India Private Ltd, Gurugram, Haryana, India.

[§]Indian Institute of Management, Bannerghatta Road- 560076, Bangalore, India. e-Mail: pulak.ghosh@iimb.ac.in

[¶]McDonough School of Business, Georgetown University, Washington, DC, USA. Fellow of the Luohan Academy. e-Mail: agr60@georgetown.edu.

1 Introduction

The cultural norms to which agents are exposed, especially when deep-rooted and based on centuries-long societal customs, can have a long-lived influence on beliefs and economic decision-making (Guiso, Sapienza, and Zingales (2006), Alesina, Giuliano, and Nunn (2013), D'Acunto (2019a); D'Acunto et al. (2019, 2020). Norms can also produce cultural biases—they can shape agents' beliefs about economic signals in a way that deviates from the beliefs of a neoclassical agent thus reducing consumption utility (Guiso, Sapienza, and Zingales (2009), Pursiainen (2020), D'Acunto (2019b)).[1] Examples of cultural biases include taste-based discrimination against members of different social groups (*in-group vs. out-group discrimination*; see, e.g., Tajfel et al. (1979)), discrimination against members of one's own group due to cultural stereotypes (*implicit bias*, see e.g. (Becker (1957),Akerlof and Kranton (2000), Parsons et al. (2011)), and *inaccurate statistical discrimination*—biased beliefs about the quality of counterparts that are revealed incorrect ex post (Bohren et al. (2019)).

Detecting and quantifying the effects of cultural biases on discriminators' economic choices is challenging because, in the absence of full information about agents' quality, the mere fact that somebody belongs to a discriminated group might provide a reliable signal of their quality (*accurate statistical discrimination*) (Phelps (1972), Borjas and Goldberg (1978)), thus improving decision-makers performance ex post. Moreover, belonging to the same social group might improve principals' ability to monitor agents and through this channel make discrimination economically valuable to discriminators (e.g., see (Fisman et al. (2017) and Fisman et al. (2020)). Isolating and quantifying the negative effects of cultural biases on discriminators' choices, if any, requires a setting in which these channels are muted. The ideal setting should also focus on high-stake economic choices that have a direct and quantifiable effect on discriminators' consumption utility, so that we can assess if economic incentives (un)successfully reduce cultural biases.

This paper studies a FinTech peer-to-peer (P2P) lending platform paired with an automated robo-advising tool (D'Acunto et al. (2019), Rossi and Utkus (2020), D'Acunto and Rossi (2020)) to ask whether cultural biases about religion and caste shape high-stakes decision-making under risk and to estimate if and by how much cultural biases affect discriminators' performance. This robo-advising algorithmic tool, which might be prone to statistical discrimination, does not have or reconstruct information about borrowers' religion or caste. Comparing lenders' choices with those the tool proposes and lenders can decide to implement allows us to test for the presence and effects of cultural biases in a context where statistical discrimination plays a limited role.

We find that lenders discriminate based on cultural biases: they are systematically more likely to provide credit to borrowers that belong to their same social group rather than to other available borrowers—an imbalance that disappears once lenders observe borrower matches suggested by the robo-advising tool. Discrimination

[1] These deviations can be optimal if the agent's utility decreases when he/she makes choices that conflict with the cultural norms to which they adhere. Deviations are only suboptimal if individuals would have preferred behaving like a neoclassical agent had they been aware of their cultural bias.

worsens lenders' investment performance: on average, lenders face 14% higher default rates when making unassisted choices and increase the returns of their loan portfolios between 4.5 and 7.3 percentage points after cultural debiasing. These results refer to actual choices in the field based not only on hypothetical algorithmic decisions (Tantri (2021)), but also accounting for the possibility that lenders might not adopt and implement algorithmic advice even when they are provided with it (Bhattacharya et al. (2012)). As we discuss in more detail below, the fact that lenders decide to implement the robo-advisor's suggestions reduces the possibility that they obtained sizable non-consumption utility from biased choices, for instance due to kin altruism.

Our laboratory is a leading P2P platform in India, *Faircent*, whose lenders and borrowers reside in all Indian states. Different from forms of marketplace lending in the US (Paravisini, Rappoport, and Ravina (2017); Vallee and Zeng (2019); Chiu et al. (2018); Dobbie et al. (2020)), this platform includes only individuals who invest their own capital and loan officers are not admitted. The incentives of capital owners and lenders thus coincide in the same individual in our setting. Moreover, contrary to bank-branch lending, the platform provides little scope for local or relationship lending—90% of the lenders provide funds to borrowers who reside in 5 or more Indian states and the platform engages in monitoring and loan servicing after loans are issued, without any interactions between lenders and borrowers.

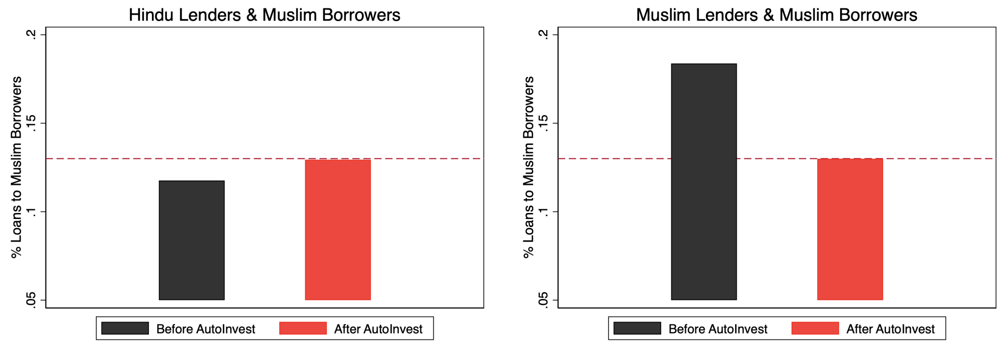
The Indian setting allows us to study two forms of cultural biases. We start with in-group vs. out-group discrimination, whereby the members of two conflicting social groups tend to favor members of their own group (in-group) and disfavor members of the conflicting group (out-group). In the Indian context, as we discuss in detail below, this form of bilateral discrimination has been detected between Hindus and Muslims (Brass, 2011; Tajfel et al., 1979). We then study implicit bias (aka stereotypical discrimination), whereby everybody discriminates against one social group—so much so that even the members of that group have excessively negative beliefs about the quality of their similar relative to others and hence discriminate among each other. We discuss how, in the Indian context, this form of discrimination can arise due to the centuries-old caste-based system.

Figure 1 previews our baseline results in the raw data for the case of in-group vs. out-group discrimination. We compare the average share of Hindu and Muslim borrowers who were financed by Hindu and Muslim lenders before lenders adopted the automated robo-advising tool (black bars) and after lenders accessed the tool (red bars). The raw data reveal three noticeable facts. First, both Hindu and Muslim lenders tend to lend more to borrowers of their same religion, which is consistent with the presence of an in-group vs. out-group bias for at least one of the two groups. Second, once the robo-advising tool makes choices, the shares of Muslim and Hindu borrowers change for both groups of lenders in opposite directions, which suggests that all lenders discriminated against out-group borrowers when making unassisted choices. Third, the shares of borrowers by religion are virtually identical for Hindu and Muslim lenders after robo-advice adoption and correspond to the shares of each religion in the population of borrowers on the platform.[2] This result corroborates the notion

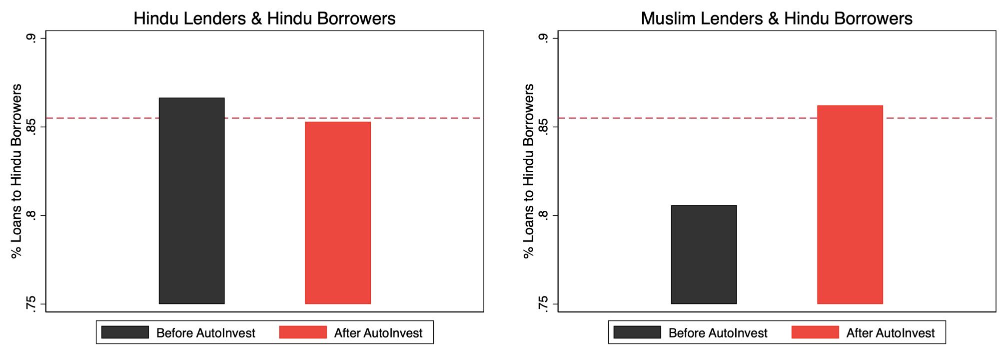
[2]These shares need not be the same as in the broader Indian population if the platform engages in statistical discrimination.

Figure 1: Lending to In-Group vs. Out-Group Borrowers: Before and After Debiasing

Panel A. Probability of Choosing Muslim Borrowers



Panel B. Probability of Choosing Hindu Borrowers



that the robo-advising tool does not use information about borrowers' religions when allocating borrowers to lenders.[3]

A set of additional features make our setting a viable laboratory to study the question at stake. First, the platform screens prospective borrowers based on their risk profile *before* borrowers are visible to lenders, who are not required to screen borrowers. If lenders think that they can screen borrower more successfully than the platform we can test directly whether this belief leads to accurate or inaccurate statistical discrimination based on lenders' performance before and after using the robo-advising tool.[4] Second, the automated robo-advising tool does not have information about borrowers' religion or caste. For this reason, we can compare the portfolios of borrowers that lenders picked before the adoption of robo-advising and the portfolios after adoption, in which

[3]As we discuss below, the robo-advising tool as well as the selection of the borrower pool that accesses the platform might be subject to (accurate) statistical discrimination. We are not arguing that the platform and the tool do not engage in any form of discrimination, but that cultural biases do not appear to be present in the allocation choices of the tool.

[4]If lenders engaged in accurate statistical discrimination, on average the performance of the borrowers they pick on their own should be better than that of the borrowers the tool assigns to them based on the platform's screening procedure.

taste-based discrimination or inaccurate statistical discrimination cannot play a role by construction. Finally, because the platform provides loan servicing and the monitoring of borrowers (Iyer et al. (2016); Tang (2019); Balyuk (2019)), we can assess the sign and size of the economic effects of cultural biases by comparing the performance of loans issued by the same lenders based on their own choices or based on the robo-advising tool's choices.

We find the raw-data patterns in Figure 1 are quite robust: they do not change in multivariate analyses that control for the loan-level characteristics we observe, when we restrict the variation within lenders, and hence absorb unobserved systematic time-invariant differences across lenders such as education levels or cultural background, and when we include time fixed-effects to account for different time-varying shocks before and after the robo-advising tool was available to lenders.[5] We also find that the results hold at the intensive margin: once lenders adopt the tool, they can decide the share of funds for which they want to get an allocation suggestion by the tool. We find that the extent of debiasing (drop in fraction of in-group borrowers after accessing robo-advising) is a monotonic function of the share of funds lender have on the platform for which they ask allocation suggestions by the robo-advising tool.

To further assess whether cultural biases drive these results, we perform several heterogeneity tests in which we vary the extent to which cultural biases should be salient to lenders. For instance, we find that the extent of bias and subsequent debiasing are economically and statistically larger in settings in which the Hindu-Muslim conflict is salient—in cities with a higher presence of Hindu-Muslim riots, in states in which nationalistic parties that foment this conflict obtain higher vote shares, and for lenders who have been more exposed to Hindu-Muslim riots and inter-ethnic political conflict during their formative years.

After detecting the presence of in-group vs. out-group discrimination, we move on to study whether discrimination relates to loans' performance and, if so, through which channels. Under cultural biases—taste or inaccurate beliefs of being able to engage in accurate statistical discrimination—when making unassisted decisions, lenders should dig deeper into the pool of in-group borrowers, who should perform worse than out-group borrowers (Agarwal et al. (2017)). The opposite should be true if lenders engage in accurate statistical discrimination.

We find that the loans lenders grant to in-group borrowers before adopting the robo-advising tool perform systematically worse than the loans they grant to out-group borrowers, in terms of both default rates (extensive margin of performance) and overall returns earned (intensive margin of performance). In-group-borrower loans are about 4.2 percentage points (pp) more likely to default (14.4% of the average default rate). Also, among borrowers who pay at least a portion of the overall amount due, in-group borrowers are less likely to repay in full. And, in back-of-the-envelope calculations we propose in the last part of the paper, we estimate that the cost of in-group vs. out-group discrimination at the lender level amounts to about 6% of the average capital

[5] Note that, because debiasing implies that the shares of Hindus and Muslims move in opposite directions for Hindu and Muslim lenders, systematic cross-sectional or time-series differences between the two groups of borrowers cannot explain our results.

lenders invested in the platform before the robo-advising tool was available.

We document two other results in the loan-return analysis of performance that are consistent with debiasing from in-group vs. out-group discrimination. First, the improvement in loan returns after debiasing relative to before is mostly driven by a higher performance of the left tail of the distribution of in-group borrowers rather than higher percentiles of the distribution. Second, we find that the lower delinquencies and higher returns lenders enjoy after accessing robo-advising are driven almost exclusively by the different risk profiles of the pools of in-group borrowers before and after access, because robo-advising does not provide any incremental performance improvement once we condition on borrowers' risk characteristics.

Note that in our setting several channels that predict a positive association between discrimination and loan performance cannot drive the results by construction. For instance, homophily in monitoring borrowers and relationship lending have no scope in our setting (Iyer et al. (2016); Schoar (2012); Drexler and Schoar (2014); Fisman et al. (2017); Fisman et al. (2020)), because lenders do not service the loans and do not monitor borrowers after the loans are issued. There is also no scope for local or relationship lending in our setting—90% of the lenders issue loans to borrowers across more than 5 Indian states. For these reasons, the incentive effects of social collateral (Karlan, Mobius, Rosenblat, and Szeidl (2009); Diep-Nguyen and Dang (2019)), moral incentives and social image (Bursztyn et al. (2018); Bursztyn et al. (2019)), peer effects (Breza (2019)), familiarity through in-person interactions between lenders and borrowers (Rao (2019)), preferences of physical appearance (Duarte, Siegel, and Young (2012); Ravina (2019)), or systematic ethnic differences in housing collateral value (Avenancio-León and Howard (2019); Naaraayanan (2019))[6] have no scope either.

The second form of cultural bias we study is stereotypical discrimination (a.k.a implicit bias), which, in India arises against members of the lower traditional Hindu caste, namely, *Shudra*.[7] This form of discrimination has received less attention from the recent quantitative literature in Cultural Economics, which typically focuses on settings characterized by strong inter-ethnic conflict and hence in which in-group vs. out-group discrimination is the prevalent form of bias (e.g., see Jha (2014), Hjort (2014), D'Acunto et al. (2019), Hjort et al. (2021)).

We find evidence consistent with substantial stereotypical discrimination: before lenders (including *Shudra* lenders) move to the robo-advising tool, *Shudra* borrowers are less likely to appear in their loan portfolios relative to the share of *Shudra* in the population of borrowers. Once lenders move to the robo-advising tool, the share of *Shudra* borrowers they finance increases substantially and the difference in defaults between *Shudra* borrowers and other borrowers drops. We also find that the discrimination against *Shudra* borrowers is higher for lenders who reside in states in which the share of crimes against lower-caste inhabitants is higher, which we interpret as settings in which the negative stereotypes against lower castes might be more salient.

Studying discrimination against castes is also interesting because the recognizability of one's caste, when

[6]Loans are not backed by collateral on the platform.

[7]Less than 0.5% of borrowers and virtually no lenders belong to the out-caste group of *Dalits*, and hence, unfortunately, we cannot test for discrimination against this group.

not disclosed explicitly as in our setting, varies substantially across individuals. On our platform, lenders can only infer borrowers' caste based on a set of individual characteristics—e.g., borrowers' surnames, locations, and occupations—but the extent to which these characteristics are a precise signal of one's caste varies. We exploit this feature to design an intensive-margin test of implicit bias inspired by the experimental literature (Mobius, Rosenblat, and Wang (2016)). We build on Bhagavatula et al. (2017) and Bhagavatula et al. (2018), who design and test an algorithm that replicates an average Indian's inference problem of which caste individuals of known demographic characteristics belong to. In this way, we obtain variation in the likelihood that lenders might recognize *Shudra* borrowers as effectively belonging to this caste. We find that discrimination against *Shudra* borrowers increases with the likelihood that the borrower is recognizable as a *Shudra*, whereas it virtually disappears for *Shudra* borrowers whose caste is difficult to assess.

The fact that *Shudra* lenders, too, discriminate against *Shudra* borrowers allows us to disentangle our cultural-bias interpretation from kin altruism—the tendency of individuals to take costly actions to favor those who share similar ethnic or other demographic characteristics (e.g., see Simon (1993)).[8] Indeed, if kin altruism was behind our results, we should have observed that *Shudra* lenders favor *Shudra* borrowers, whereas we find the opposite.[9]

We conclude our analysis by assessing the role of debiasing on loan performance also for the case of Shudra borrowers and we detect the same patterns as for the performance after debiasing in-group vs. out-group discrimination. First, lenders face fewer defaults and higher average returns after accessing the robo-advising tool's borrower suggestions. Second, the improvement in returns comes largely from the elimination of a left tail of low returns. Third, the return improvement the tool produces is driven by a drop in the riskiness of the loans that appear in lenders' portfolios after high-risk non-Shudra borrowers are replaced by lower-risk Shudra borrowers.

The fact that virtually none of the lenders in our setting overrides the robo-advisor's proposed allocations, which they could do if they wanted, suggests that lenders are unlikely to obtain substantial non-consumption utility from biased choices, which might overcome the consumption utility loss due to lower financial returns. The behavior also suggests that inaccurate statistical discrimination may be an important driver of the results we document.[10]

Our results also emphasize an unintended role of robo-advising tools (D'Acunto and Rossi (2020)), which are diffusing around the world to facilitate consumers' spending (D'Acunto et al. (2019)), saving (Gargano and Rossi (2020)), borrowing (Agarwal et al. (2019)), and lending decisions. We show that such tools can help discriminating agents avoid the financial losses they face when making (perhaps implicit) culturally biased choices.

[8] We thank David Thesmar for raising this point.

[9] Kin altruism is also unlikely to explain Hindu-Muslim discrimination in our setting, because if lenders obtained substantial non-consumption utility from kin altruism they would not follow through with the robo-advising suggestions of picking borrowers from the out-group, but they would rather replace them with in-group borrowers.

[10] The fact that agents barely modify the allocations proposed by robo-advisors, even when given the chance, is common to other forms of robo-advising, such as those for equity portfolio allocations (D'Acunto et al. (2019)) and debt management (Burke et al. (2021)).

Robo-advising tools might be a viable substitute for financial disclosure and financial literacy in improving the outcomes of consumers and individual investors, because individuals are not required to understand all the aspects of the investment problems they face when making a robo-advised-assisted decision (Adams et al. (2019); D'Acunto et al. (2019b); D'Acunto et al. (2019a)).

2 Institutional Setting

The setting for our analysis is Faircent, a large FinTech platform that specializes in P2P lending in India and whose borrowers and lenders reside across all Indian states.[11] This setting is reminiscent of the recent literature on FinTech adoption in developing countries (e.g., see Agarwal et al. (2019); Crouzet, Gupta, and Mezzanotti (2019); Higgins (2019); D'Andrea and Limodio (2019)).

2.1. Borrowers' Screening by the Platform

Several features of the platform are crucial to the design and interpretation of our tests. First and foremost, the platform screens applicant borrowers before admitting them to the pool lenders can access, and lenders observe the results of this screening. Note that, even though the platform does use racial variables, its screening procedure might incorporate statistical discrimination, which has been detected in other settings in which automation and machine learning are involved in the screening of borrowers (e.g., see Bartlett et al. (2019); Bhutta et al. (2021); Cowgill and Tucker (2020); Fuster et al. (2017); Rambachan et al. (2020)).

Screening by the platform follows two steps. Once a prospective borrower signs up, he/she submits a loan application that includes the proposed amount of the loan, the motivation for the loan, the borrower's credit score, occupation, geographic location, and whether the borrower has dependents (children or elderly). Note that the borrower needs to provide evidence of a bank account or a microcredit account. Unbanked borrowers cannot access the platform. The first (platform-based) screening step starts with an automated algorithmic-based assessment of the borrower's viability and ability to repay, which is largely based on the borrower's credit score, proposed loan amount and maturity, and occupation.[12] Based on this step, prospective borrowers that fall below a set threshold in terms of credit viability are dismissed from the platform.

In the Online Appendix, we provide raw-data evidence about the outcomes of the automated screening procedure in terms of the distributions of credit scores of borrowers who are accepted and those who are rejected (see Figure A.1). We find that rejected borrowers have systematically lower credit scores not only in the overall sample of applications, but also when we focus only on applicants who request loans of the same size. We also provide evidence on the monotonic relationships between borrowers' credit scores and the annual interest rates, maturity, and loan amounts that are assigned by the platform as a direct function of credit scores (see Figure

[11] This platform only hosts individual lenders and no institutional lenders.

[12] Due to confidentiality, we cannot report the details about how the proprietary algorithm screens borrowers.

A.2). The prospective borrowers who are approved and accept the parameters the platform assigns to them proceed to the second screening step. Borrowers who pass this step are assigned a risk category as well as a proposal for the interest rate and maturity of the loan based on the amount the borrower wants to raise.

Faircent's (human) employees perform the second screening step. This step consists of the in-person verification of several borrowers' characteristics. This step aims to eliminate lenders' potential concerns regarding the viability and accuracy of the automated risk profiling process on the platform. The characteristics Faircent employees verify personally include borrowers' identity (through scanned identity cards), an in-person connection via video call, a personal picture to compare with the identity-card picture, the proof of two income paystubs or incoming transactions in a bank account under the borrower's name, the proof of utility payments and addresses, the picture of the borrower's housing location, and the picture of the borrower's work location. Borrowers who fail these verification steps are dismissed from the platform. Borrowers who pass these verification steps are admitted to the borrower pool that lenders can browse.

The two-step screening procedure ensures that in our setting lenders make choices *after* a substantive risk assessment has already been performed, whose outcomes lenders observe. Decoupling borrowers' risk screening from lending decisions, which departs from earlier research that studied the choices of loan officers, is important to tackle the question we propose: if certain demographic groups of borrowers were more represented among high-risk prospective borrowers, they would be rejected more often than other borrowers by the platform, and hence, the scope for accurate statistical discrimination on the part of lenders—the possibility that lenders are more likely to finance certain demographic groups than others because such groups are less likely to belong to high-risk categories in the population—is substantially lower in our setting than in loan-officer-based settings.

Note that lenders might still believe that they can screen borrowers better than the platform, which could lead to incremental accurate or inaccurate statistical discrimination (Balyuk and Davydenko (2019); Vallee and Zeng (2019)). Our tests below will allow disentangling these possibilities: under incremental accurate statistical discrimination by lenders the performance of unassisted choices should be better than the performance of robo-advised choices, and the opposite under biased discrimination. As we describe below, our evidence is inconsistent with accurate statistical discrimination on lenders' part.

Once viable borrowers are vetted they access the borrower pool, from which individual lenders access borrowers' demographic characteristics and the in-depth qualitative and quantitative risk assessment from the screening process (we discuss the information lenders see in section *2.2.* below). In this step, lenders decide whom they want to fund.

Execution, servicing, and monitoring of the borrowers after loan approval also happens within the platform. Once the full amount of the loan is funded by at least 5 lenders who choose the borrower,[13] the platform approves and executes the loan. The loan agreement is a private contract between the borrower and each lender, but

[13]The platform imposes that each loan is financed by at least 5 lenders to enhance the diversification of lenders' loan portfolios and to ensure that the platform cannot be used for money laundering purposes.

the platform produces the electronic forms that lenders and borrowers have to sign. No lenders enjoy any form of seniority. Upon execution, Faircent provides borrowers with an equated monthly installment (EMI)—the monthly payment—and services the loan in house. Faircent monitors the status of loans each month. Loans' status is declared "closed" after full repayment or after repeated delinquency. Borrowers whose loans are closed while delinquent are dismissed from the platform. Faircent's loans are subject to the Reserve Bank of India (RBI) regulatory policies and oversight like traditional financial institutions.

The fact that the platform rather than the individual lenders monitors borrowers reduces the role of relationship lending and/or homophily in monitoring of borrowers' repayment behavior due to belonging to the same religion or caste or living in the same location. In fact, Figure A.3 of the Online Appendix shows the distribution of the number of states in which the borrowers of each lender in our data set reside. The figure reveals that more than 90% of our lenders provide loans to borrowers who live across at least 5 different Indian states. Also, the median lender disburses funds to borrowers who reside in 13 different Indian states. In this respect, thus, our setting is quite different from local-branch relationship lending by professional loan officers, in which soft information about borrowers and social norms/social pressure might influence borrowers' repayment behavior.

2.2. Robo-Advising Tool (Auto Invest)

The second important feature of the platform is that lenders can make their choices unassisted or under the assistance of a robo-advising tool, called Auto Invest.

For unassisted choices, lenders can browse the borrower pool at any point in time. The information about each borrower lenders observe include the coarse risk category assigned by the platform (low, medium, or high risk), the detailed risk assessment of the loan (interest rate, maturity, overall loan amount), and a set of borrowers' demographic characteristics that include names and surnames, location of residence, occupation, education levels, and number of dependents. Upon clicking on the borrowers' profile, the lenders can access the results of the verification step of the borrowers' identity, including borrowers' pictures and evidence of their residence based on utilities and/or bank accounts. Lenders can decide who, if anybody, they want to fund and by how much. Because the platform imposes that each loan is financed by at least 5 different lenders to reduce the scope for money laundering and other criminal uses of the platform, lenders can at most finance 20% of the chosen borrower's loan amounts. Lenders need to have the funds they want to commit deposited on the platform before their loan proposals are posted to the borrowers. Loans proceed to the execution phase only once the financing of the full loan amount is committed.

The second mode of investment is with the assistance of an automated robo-advising tool, *Auto Invest*, which was introduced on the platform in the second half of 2018. After its release, lenders can adopt Auto Invest at any time. Upon adoption, lenders decide the share of their overall funds deposited on the platform for which they want to obtain suggestions of borrowers by Auto Invest and the amount of funds, if any, they want to keep investing unassisted. Lenders are then asked to allocate the funds for which they seek assistance by Auto Invest across the same three borrower-risk categories they see when making unassisted decisions—low, medium, and high risk.

Once funds are allocated across risk categories and lenders run Auto Invest, the tool matches lenders to the borrowers in each risk category who are closest to get their loan amount fully funded, up to exhaustion of the lenders' funds to be allocated. Auto Invest thus does not choose borrowers explicitly based on their interest rate or level of risk within each risk category or based on borrowers' demographic characteristics. We will see below that Auto Invest does not end up incorporating borrowers' demographics implicitly either—for instance, because certain demographic groups are systematically less likely to be close to fully funding their loans than others—because the shares of borrowers from each religion and castes in every lenders' portfolios are on average equalized to the share of such categories in the broader population of borrowers. That is, information about religion or caste does not allow us to predict the allocation choices Auto Invest proposes to lenders.

By contrast, when Hindu and Muslim lenders are unassisted, we will show that they are more likely to choose highly risky borrowers within the high-risk category for the demographic groups they favor, which is important to explain the difference in the performance of unassisted and assisted loan choices and why unassisted choices end up performing worse than assisted choices across demographic groups.

Once Auto Invest proposes a set of borrowers, the lender decides if she wants to proceed with the suggested allocation or change it in part or in full. If the lender decides to make changes, she browses the pool of borrowers and can replace suggested borrowers with other available borrowers. Unfortunately, we do not observe whether each individual loan contribution was made by the lender unassisted or through Auto Invest. For this reason, we cannot compare the choices each individual lender made manually with those based on Auto Invest's suggestions at the same point in time by absorbing lender-by-time variation in our multivariate analyses.

3 Data and Summary Statistics

To perform our analyses, we use seven data sets, each of which covers a different feature of the lending process at Faircent. Our data span the period between January 1, 2018, and March 30, 2020, although, given the large monthly growth of the platform, 60% of the loans in our sample were issued in 2019 and another 19% in the first three months of 2020. Variation in the timing of loan issuance is therefore limited. We limit the sample to the end of March 2020 to avoid covering the COVID-19 pandemic period.

Two data sets—the *Lenders' characteristics data set* and the *Borrowers' characteristics data set*—include cross-sectional data with one observation per individual. Each lender and each borrower is assigned a unique identifier, which allows us to link lenders' and borrowers' characteristics across data sets. For each lender, we observe the individual identifier Faircent assigns at signup, name and surname, the city and state location of residence, and the date of birth. On top of these characteristics, the borrowers' sample also includes information on the borrower's residence type (whether owned or rented), number of dependents, employment type (whether self-employed or not), and credit score.

Faircent does not ask either lenders or borrowers for their religion and/or caste. Our source for information about these dimensions, which are crucial in our analysis, is the *Marriage registry data set* (see Bhagavatula et al. (2017) for an example of an earlier use of these data). These data include demographic information about religion and caste elicited at the time of marriage for a random sample of 2,481,158 Indians. Specifically, it includes names and surnames, date of marriage, state of birth, current city of residence, height (in centimeters), and religion and caste.

We find religion barely varies across individuals who share the same surname. Eighty-nine percent of the unique pairs of surnames and dates of birth in the registry (everybody who was born the same date and shares the same surname) belong to individuals who are assigned the same religion. When we only consider the two religions on which our analysis focuses—Hindu and Muslim—96% of the surname-date-of-birth pairs are matched to only one religion. For these reasons, we assign religions to lenders and borrowers based on surname and date of birth, which we observe in the Faircent data.

Assignment of castes to borrowers and lenders is less straightforward. First, the caste information in the marriage registry is quite dispersed and includes about 540 narrowly-defined partitions, which sometimes merely correspond to the individual's surname. As we discuss in more detail below, to make our analysis of castes meaningful, we need a reliable way to assess to which of the four main *varnas* borrowers and lenders belong so as to split borrowers into those belonging to the Shudra varna—which we argue is subject to stereotypical discrimination—or other varnas.[14] Moreover, we do not find that any of the combinations of characteristics we observe in the Faircent data or the marriage registry, that is, names, surnames, and dates of birth,[15] restrict the possible set of castes enough to proceed in the same way as with religions. To infer borrower and lenders' varnas, we rely on earlier research in computer science that has tackled the same issue of assigning castes to individuals in cross-sectional data sets based on marriage-registry information and individuals' surnames. Specifically, we use the methodology developed by Bhagavatula et al. (2017) and Bhagavatula et al. (2018), which we discuss in more detail below.

The fourth data set we use is the *Lender-Borrower Mapping*, which is a cross-sectional data set at the level of unique lender-borrower-loan triads. This information is critical to our ability to merge individual characteristics to borrowers and lenders who match through a loan that is in part funded by the lender. The data are also critical to merge information about loan characteristics and loan performance to each unique lender-borrower-loan triad. For example, if a borrower requested $10,000 and the money was lent by five different lenders, we

[14]We provide a primer on Indian castes and on how they relate to our analysis below.

[15]Note Faircent observes the borrower's/lender's location at the time they are active on the platform, whereas the location in the marriage registry is the individual's location of residence at the time of marriage.

have information on the amount lent by each individual lender to this borrower and the data set includes five lender-borrower-loan triads for this loan.

Our information about loans comes from a cross-sectional and a panel data set. The *Loan characteristics data set* is a cross-sectional data set at the loan level. For each loan, the data report the borrower to whom the loan was extended as well as the total amount lent, the interest rate, the maturity of the loan, and the proposed monthly payment. Moreover, we observe the loan's status as of March 31, 2020 (active or closed), as well as the whether the last payment happened within the previous 31 days (i.e., whether the loan is in good standing).

The *Loan performance data set* is a panel data set at the loan-month level. In this unbalanced panel, we observe the monthly payments that borrowers provided during the life of the loan each month in which the loan's status was active (including zero if the borrower missed a payment in a month), for both active and close loans as of March 31, 2020.

Finally, the *Auto Invest data set* is a cross-sectional data set at the lender level, which provides us with information on whether lenders have ever activated the robo-advising tool (called Auto Invest) to make automatic lending decisions, and if yes, the activation date and the share of the funds available on the platform that the lender allocated to Auto Invest instead of keeping for manual lending choices. This data set allows us to compare the lending decisions and performance of financed loans at the lender level before versus after their activation of Auto Invest, as well as across lenders who allocated a higher or lower amount of funds to the robo-advising tool.

3.1. Sample Selection

Armed with these seven data sets, we construct the two working samples we use in our analysis.

To study in-group vs. out-group discrimination, we create a sample at the borrower-lender-loan level that includes all the Hindu and Muslim borrowers and lenders in the data for whom we observe no missing information on loan characteristics, the usage of Auto Invest, or loan performance.

A common concern in settings that study the adoption of a new technology is that unobserved lender- or borrower-level characteristics and shocks might at the same time cause lenders' adoption but also the change of behavior after adoption, so that this change of behavior cannot be attributed causally to lenders' use of the new technology. Note that, in our setting, this concern is less compelling than in others, because we test whether different lenders change their behavior in opposite ways after adopting the robo-advising tool. For instance, if Hindu and Muslim borrowers faced different economic shocks over time and for this reason they became more or less viable borrowers, all lenders should react in the same direction in terms of choosing more or less of one of the two categories of borrowers, because borrower-level shocks affect the borrowers available in the pool in the same way, irrespective of whether they end up being chosen by a Hindu or a Muslim lender. Similarly, lender-level shocks might affect the extent to which lenders engage with the platform or the amount of money they invest on the platform over time, but once these dimensions are fixed at the lender level they should not impact the allocation of funds across borrowers of different religions. One possibility is that the lenders that think they have been performing worse on the platform are more likely to adopt the Auto Invest early, but then at a minimum these lenders were not realizing that cultural biases were the source of their low performance, or otherwise they would have changed their allocation across religions and castes before the introduction of Auto Invest. Even in this case, then, comparing the allocation of unassisted choices with the allocation of choices implemented through Auto Invest would provide us with a measure of lender-level cultural biases before they accessed the tool.

Nonetheless, to reduce concerns about the endogenous adoption of Auto Invest, as is customary in the literature,[16] we further select the sample by only including lenders who have activated Auto Invest at some point in time between its introduction on the platform in 2018 and the end of the sample period (March 31, 2020). This selection implies all the lenders in the sample are adopters, and hence we never compare the choices of lenders who never adopted Auto Invest with those of lenders who adopted it at some point in time.[17] At each point in time, the sample includes lenders who have already activated the tool and others who have not yet activated it.

A second way in which the sample allows us to reduce concerns about the endogeneity of adoption is by exploiting the intensive-margin of the use of Auto Invest: we compare lenders who allocate a higher or lower share of their funds to the robo-advising tool, and hence lenders that might have activated the tool at the same point in time but who make a higher or lower fraction of choices unassisted after adoption.

The second working sample allows us to study the role of stereotypical bias in lending decisions. We first select the sample using the same steps discussed above. Then, we further restrict the sample to only include Hindu borrowers, given that other religious groups do not partake the stark caste categorization. Moreover, we only include borrowers for whom we can retrace caste information in the form of one of the four varnas based on the marriage registry data.

3.2. Summary Statistics

The first working sample includes 113,283 unique lender-borrower-loan triads, which are based on 2,818 unique lenders. Panel A of Table 1 reports summary statistics for this sample. Borrowers' religion, consistent with the split between Hindu and Muslim individuals in the general Indian population, is tilted toward Hindus—only 13% of the borrowers are Muslim. The religious imbalance is instead very high on the lenders' side—99% of lenders are Hindu. Despite the small share of Muslim lenders (1% of the sample), the number of observations is large enough to allow a statistically meaningful analysis of Muslim lenders' behavior. One reason for such

[16] For instance, see D'Acunto et al. (2019), DAcunto et al. (2020), and Gargano and Rossi (2021), among others.

[17] All our results and patterns are substantially more pronounced if we also include lenders who have not adopted Auto Invest during the sample period.

a stark imbalance of religion on the lenders' side might be that the precepts of *Sharia* can be interpreted as being against the earning of financial interest on regular loans, which might reduce the willingness of Muslim consumers to sign up as lenders on the platform.[18]

About 45% of the loans in the data set were issued at a time when lenders had activated the robo-advising tool. The average share of funds allocated to the tool is about 60%, but substantial cross-sectional variation exists across lenders. As far as loan characteristics are concerned, the average maturity (tenure) is 22 months and the median maturity is 24 months. The average loan amount is slightly above 130,000 rupees, which corresponds to about $1,770,[19] with a large standard deviation. The average annual interest rate is 24%—similar to the yearly APRs for credit cards in the US over the same period.[20]

The summary statistics so far refer to the sample we use in the main analysis, which includes all unique lender-borrower-loan triads. Considering a set of observables at the lender and loan levels, though, is also important to assess how the sample's characteristics change over time and especially before and after lenders access Auto Invest. We find that the imbalanced share of the 2,818 unique lenders who are Hindu or Muslim reported above is stable throughout the sample period, and in particular we cannot reject the null that the two shares are equal for the period before and after Auto Invest is available either economically or statistically. We find that the shares of Hindu and Muslim borrowers in the borrower population are also stable over time and in particular do not differ systematically before and after the Auto Invest tool is available. Loan's engagement with the platform does not vary substantially, either. For instance, the average lender issue 40 loan offers to any borrowers both before and after accessing Auto Invest (p-value of t-test for equality=0.683). In terms of loan characteristics, when computed at the lender level, we find that the average interest rate of loans issued before using Auto invest is higher than that of loans issued after using Auto Invest (25% vs. 23%, p-value of t-test for equality=0.003) and the default rates lenders face in the raw data are substantially higher before they use Auto Invest (33% vs. 16%, p-value of t-test for equality$< 0.1\%$). Also, lenders offer lower average amounts per loan after accessing Auto Invest (₹2,980 vs. ₹2,123, p-value of t-test for equality$< 0.1\%$). As we will see in the last part of the paper, the fact that the loans produced through Auto Invest are less risky, on average, than those produced through unassisted choices is important to explain why culturally-biased choices deliver lower returns to lenders than robo-advised choices.

Moving on to the second working sample, it includes 62,831 unique lender-borrower-loan triads. Panel B of Table 1 reports the summary statistics for this sample, in which 39% of borrowers belong to the discriminated *Shudra* varna. Despite the smaller size, the summary statistics for the main variables of interest in this sample

[18]We thank Alan Moreira for suggesting this potential explanation. This explanation would imply that the Muslim lenders who sign up to the platform might not a high attachment to the religious precepts of Islam. Note, though, that our tests are based on the ethnic and social identity of being part of the Muslim community and the ethnic conflicts with the Hindu majority, which go above and beyond the mere religious differences between the two groups.

[19]This conversion does not consider the varying exchange rate between the US dollar and Indian rupee in each month our sample covers.

[20]Over our sample period, the nominal interest rate of reference set by the Reserve Bank of India ranged between 4.5% and 6%.

are similar to those described in Panel A. The patterns for the statistics at the lender level, including the lack of changes in the composition of borrower pools before and after Auto Invest is available as well as the lower risk of loan offers lenders make when assisted by Auto Invest, are also the same as for the Hindu-Muslim sample.

4 In-group vs. Out-group Discrimination: Hindu vs. Muslims

The first form of cultural bias we consider is in-group vs. out-group discrimination: agents tend to favor members of their own social group (in-group) over members of conflicting social groups (out-group), where social groups' boundaries are defined based on cultural-identified cleavages (Tajfel et al. (1979); Hewstone et al. (2002); Jenkins (2014)).[21]

A fundamental implication of this theory in terms of economic decision-making, which we can test directly in our setting, is that if favoring in-group vs. out-group members were driven by a cultural bias, lenders would be more willing to choose in-group borrowers even if they had to face a monetary cost from favoring them (D'Acunto et al. (2020)). In our setting, the cultural-bias hypothesis predicts not only that lenders should be more likely to pick in-group borrowers, but also that the in-group borrowers they choose should perform *worse* than the out-group borrowers they choose. Moreover, once lending decisions are automated through the robo-advising tool, the likelihood that in-group or out-group members appear in lenders' portfolios as well as the average performance of these two groups of lenders should converge. These predictions are in stark contrast to what we should find if lenders' favoritism toward in-group borrowers were due to lenders' ability to screen and monitor in-group members better than out-group members, because in this case, in-group borrowers should perform better, on average, when lenders make decisions autonomously.

The Indian setting provides an ideal laboratory to study in-group vs. out-group discrimination in the context of religious conflict, and especially the conflict between Hindus—the religious majority—and Muslims, one of the religious minorities in post-independence India. Acts of in-group vs. out-group discrimination between these two religious groups are deeply rooted in history and pre-date the independence of modern India in 1947 as well as the British rule on the Indian subcontinent (Lorenzen (1999)). Not only have Hindu and Muslim identities developed in contrast over time, but the identity clash has also manifested in acts of conflict, including violent conflict and riots, for decades (e.g., see Engineer (1997)). This conflict has been vivid throughout India's history and has intensified since independence, that is, when the present-day territory of India was separated from the present-day territories of Pakistan and Bangladesh,[22] both of which hosted a stronger Muslim presence.

The Hindu-Muslim conflict has been exacerbated over the last two decades (Graff et al. (2012)) and erupted in violent riots such as the anti-Muslim pogrom in the state of Gujarat in 2002 (Ghassem-Fachandi (2012)). Several

[21] Unfortunately, we cannot provide a comprehensive description of all the facets and decades-long academic debate about this family of theories in this paper, but due to space constraints, we need to focus on the most relevant implications in terms of what we can test directly in our setting. For more comprehensive reviews, see, for example, Hewstone et al. (2002) and Jenkins (2014).

[22] Present-day Bangladesh was part of Pakistan until its independence in 1971.

political scientists and sociologists argue this conflict was exacerbated because right-wing political parties, such as the BJP, that propose Hindu nationalism as a key part of their political platforms, raised to power at the national level (see Kaul (2017) among others). For instance, the approval of the *Citizenship Amendment Act* in 2019 has produced a recent vivid wave of riots and violence between Hindus and Muslims covered by the local and international media (Bhat (2020)). Changes in relative incomes between Hindus and Muslims over the last two decades have also been proposed as a factor contributing to the recent increase in Hindu-Muslim violence (Mitra and Ray, 2014). As we discuss in more detail below, over the last two decades, violent riots and localized conflicts between Hindus and Muslims have been relatively more common in certain Indian states than others.

We exploit the religious conflict between Hindu and Muslim communities to first assess the extent to which Hindu lenders might have been more inclined to finance Hindu borrowers, and Muslim lenders to finance Muslim borrowers, when lenders were making all their decisions autonomously. Then, we compute the change in the propensities to lend to Hindus and Muslims once lenders moved to using the robo-advising tool (Auto Invest), which made decisions automatically and based on characteristics that were unrelated to borrowers' religion. Third, we assess whether the extent of lenders' in-group vs. out-group bias and debiasing were stronger in the areas of India in which the salience of the Hindu-Muslim conflict was higher, where we use the salience of the conflict as a proxy for the extent of cultural bias in this context.

4.1. In-group vs. Out-group Lending Before and After Robo-advising

We start by considering the extent of in-group vs. out-group bias in lending by Hindu and Muslim lenders in the raw data. Specifically, in Figure 1, we report the average share of Hindu and Muslim borrowers within the pool of borrowers for Hindu and Muslim lenders separately. We report these averages for the same lenders at two points in time; that is, the average shares for the period in which the lender was making loan decisions on his/her own and the average shares after the lender started to use the robo-advising tool.

The top-left graph of Figure 1 reports the share of lending to Hindu borrowers by Hindu lenders before (black bar) and after (red bar) using Auto Invest. The top-right subfigure, instead, reports the percentage of lending to Hindu borrowers by Muslim lenders before and after using Auto Invest.

Three broad patterns are worth noticing here. First, consistent with the presence of an in-group vs. out-group bias for at least one of the two religious groups, we find that for each religion, lenders tend to choose a higher share of borrowers of their own religion. The share of Hindu lenders' borrowers who are Hindu is 86%, whereas the share of Muslim lenders' borrowers who are Hindu is only 80%. Conversely, the share of Hindu lenders' borrowers who are Muslim is 12%, whereas the share of Muslim lenders' borrowers who is Muslim is 18%.

A second pattern is that, after they start to use the robo-advising tool, which does not consider borrowers' religion in the lending choice, the shares of Muslim and Hindu borrowers change for all lenders, and the changes are in opposite directions: Hindu lenders' borrowers who are Hindu decrease from 86% to 84%, whereas the share of Muslim borrowers increases from 12% to 13%. At the same time, the share of Muslim lenders' borrowers who are Hindu increases from 80% to 84%, whereas the share of Muslim borrowers decreases from 18% to 13%.

The fact that after using the robo-advising tool, lenders of different religions change their borrowing choices in *opposite* directions is important to reduce the endogeneity concerns inherent to the decision of adopting Auto Invest. This decision might be endogenous to individual- and/or economy-wide time-varying shocks, which would be hard to absorb without a proper identification strategy. But, because we observe that adopting robo-advising affects behavior in opposite directions for lenders of different religions, the most plausible unobserved time-varying economic shocks about which we would worry, that is, shocks that might simultaneously explain the adoption of the tool and a behavioral change in the same direction for all lenders, are by construction unable to explain our results.

The third fact that Figure 1 emphasizes is that the share of borrowers of different religions is equalized for Hindu and Muslim lenders after they use Auto Invest. This fact is emphasized by the red dashed horizontal lines crossing the four graphs of Figure 1. This fact reassures us that unobserved channels that might make in-group borrowers more profitable to lenders than out-group borrowers are unlikely to exist in our setting. Ultimately, the robo-advising tool debiases lenders in that it equalizes the share of borrowers of each religion in each lenders' portfolios to the share of borrowers of each religion in the broader population of borrowers, thus avoiding any favoritism toward in-group borrowers.

The averages in Figure 1 compare lending behavior before and after lenders adopt the robo-advising tool, but Auto Invest allows lenders to choose the share of the funds they have with Faircent that they want to allocate to Auto Invest. If lenders choose a share lower than 1, they can still make choices autonomously for the share of resources they do not allocate to Auto Invest. This institutional detail allows us to consider the extent of debiasing not only at the extensive margin (the choice of borrowers before and after lenders adopt Auto Invest) but also at the intensive margin (the choice of borrowers lenders who allocate a higher or lower share of their funds to Auto Invest make).

We report the analysis of this intensive margin of debiasing in Figure 2. In this figure, we only consider Hindu lenders who use the robo-advising tool.[23] We sort lenders based on the share of their funds with the platform that they decide to allocate to Auto Invest, which is strictly larger than zero and lower than or equal to 1. The solid blue line reports smoothed non-parametric estimates of the relationship between the share of Hindu borrowers in lenders' borrower portfolios (measured on the right y-axis) and the percentage of funds that lenders allocate to Auto Invest. Grey bandwidths refer to 95% confidence intervals around the point estimates of the slope of the curve for each percentage of fund allocation. Consistent with the results at the extensive margin discussed above, even in terms of the intensive margin, a greater use of Auto Invest—which, by construction,

[23] We assess the extensive margin only for Hindu lenders, because we do not have enough Muslim lenders in the sample to obtain a meaningful mass of them at each value of the percentage of funds allocated to Auto Invest.

coincides with a lower share of choices that lenders make autonomously—is associated with a lower share of Hindu borrowers, and the negative relationship is monotonic. The larger the extent of the use of Auto Invest, that is, the lower the share of choices made autonomously, the lower the share of in-group borrowers that appear in lenders' portfolios. We detect a mirroring pattern for Muslim borrowers (green dashed line), which is mechanical given that the sample includes only Hindu and Muslim borrowers. Overall, even in terms of the intensive margin, the data reveal that a higher use of automated lending choices reduces the in-group vs. out-group bias we detected for lenders when they chose borrowers autonomously.

The univariate results discussed thus far suggest automating lending choices reduces lenders' favoritism toward choosing in-group borrowers over out-group borrowers. However, systematically different time-varying trends between the Hindu and Muslim borrower populations might explain the differential shares of lending to borrowers of different religions over time. Note such differential trends should not be deemed relevant to the lending decision by the robo-advising tool to constitute a concern in our setting. Moreover, these trends should change the profitability and risk profiles of borrowers of different religions in the opposite direction for Hindu and Muslim lenders, which is barely plausible and reduces this concern substantially.

To assuage the relevance of this concern in our setting, in Table 2 we report the results for estimating the following multivariate specification:

$$\begin{aligned} Muslim\ Borrower_{i,j} =& \alpha + \beta\ Auto\ Invest_j + \gamma\ Hindu\ Lender_j + \\ & \delta\ Hindu\ Lender_j \times Auto\ Invest_j + \zeta\ \boldsymbol{x}_i + \epsilon_{i,j}, \end{aligned} \tag{1}$$

where $Muslim\ Borrower_{i,j}$ is equal to 1 if borrower i who receives funding from lender j is Muslim, and 0 otherwise; $Auto\ Invest_j$ is equal to 1 if the lender made the loans after activating Auto Invest, and 0 otherwise; $Hindu\ Lender_j$ is equal to 1 if lender j is Hindu; and $\boldsymbol{x}_i$ is a vector of loan-level characteristics that are direct proxies for the risk profiles of the loans that lenders extend to borrowers—loan maturity (measured in months), loan amount, and the interest rate associated with the loan. Importantly, these characteristics are *not* chosen by the lender; rather, the company's algorithm assigns them to borrowers when the loan requests are vetted, borrowers' risk profile is estimated, and requests are approved to be added to the platform.

In terms of statistical inference, in Table 2 we cluster standard errors at the lender level to allow for correlation across the lender-borrower matches that include the same lender. In Table A.1 of the Online Appendix, we show that the results are quite similar if we make different assumptions, such as allowing for double clustering by both lender and borrower, i.e. allowing for correlation also across the residuals of matches within the same loans, allowing for triple clustering at the lender, borrower, and month-of-issuance level, as well as for triple clustering at the level of lender family communities (captured by sharing the same surnames), borrower family communities, and month of issuance.

The coefficients of interest in Table 2 are γ, which measures the likelihood that Hindu lenders had Muslim borrowers in their portfolios relative to Muslim lenders, and δ, which measures the change in the probability of Hindu lenders lending to Muslim borrowers after activating the robo-advising tool.

Column (1) of Table 2 reports the baseline estimates without adding any control variables. We report this specification, which corresponds to the raw-data results of Figure 1, to assess the statistical significance of those results. Indeed, the estimated coefficient $\hat{\gamma}$ is negative and statistically as well as economically significant, indicating Hindu lenders were about 5.8 pp less likely to lend to Muslim borrowers than Muslim lenders before using the robo-advising tool. The constant term—18%—captures the share of Muslim borrowers in Muslim lenders' portfolios.

The estimate of the coefficient β on Auto Invest shows a non-significant effect of Auto Invest on the overall likelihood that Muslim are lent to, on average, in the periods in which all lenders use the robo-advising tool, relative to before they started to use it. The lack of significance is due to the fact that the effect is positive for Hindu lenders and negative for Muslim Borrowers, as we show in the bottom subfigures of Figure 1, so that, overall, the share of Muslim borrowers in the sample does not change but is reallocated among Hindu and Muslim lenders.

The main coefficient of interest, $\hat{\delta}$, is positive and significant, indicating Auto Invest increases the likelihood that Hindu lenders lend to Muslim borrowers by about 4.5 pp. This coefficient denotes a drop in the in-group vs. out-group bias, but not a full debiasing, which would have implied an estimated interaction coefficient of around 5.8 pp. This lack of full debiasing is explained by the fact that, even after moving to Auto invest, lenders can still choose some loans on their own above and beyond what the Auto Invest chooses for them—as we have shown in Figure 2.

In the second column of Table 2, we move on to the multivariate specifications. We find the tenure and the loan amount are significantly related to the probability of the borrower being Muslim. The first is positive and significant, whereas the second is negative and significant. However, the coefficient on the interaction between $Auto\ Invest_j$ and $Hindu\ Lender_j$ has the same statistical significance and point estimate as the baseline result, which excludes that the bias before the use of robo-advising (and hence, the extent of debiasing after using the tool) was driven by heterogeneity in objective proxies for the riskiness of borrowers, such as loans' interest rates and maturity.

We then move on to restricting the variation within lenders by adding a lender fixed effect to the baseline specification (column (3)). In this way, we assess the scope for discrimination by Hindu lenders against Muslim borrowers after absorbing systematic time-invariant differences across lenders, including unobserved time-invariant dimensions such as financial literacy, cognitive skills, and education levels. This specification absorbs the lender's religion, but we can see that the estimated coefficient on the likelihood of choosing Muslim borrowers after using Auto Invest stays almost identical to the results in columns (1) and (2). In column (4), we add year fixed effects to the specification to account for potential time-varying shocks that affect lenders and borrowers over time. Once, again, we find that the coefficient of interest $\hat{\delta}$ barely changes.

In the last two columns of Table 2, we assess the intensive margin of usage of the robo-advising tool. Column (5) reports the multivariate estimates for lenders who allocate less that 40% of their funds to Auto Invest, whereas column (6) reports the results for those who allocate more than 40% of their funds to Auto Invest.[24] Based on the univariate results in Figure 2, we expect that the strength of the effect of debiasing increases with the percentage of funds that lenders allocate to Auto Invest. Consistently, the $\hat{\delta}$ coefficient is small and insignificant for lenders who do not use Auto Invest intensively and positive and significant for other lenders. Comparing the estimates with the corresponding ones in the full sample with lender fixed effects and loan-level characteristics, the positive estimated increase in lending to Muslim borrowers by Hindu lenders after using Auto Invest is 14% larger for the set of lenders who use Auto Invest more intensively.

Overall, both our univariate and multivariate estimates suggest that, once lenders move to using the robo-advising tool, they debias in the sense that they do not display a favoritism for choosing in-group borrowers and avoiding out-group borrowers, and we detect this form of debiasing for both Hindu and Muslim lenders, whose lending choices move in opposite directions once they move to Auto Invest in terms of the religion of the borrowers who enter their portfolios.

4.2. Heterogeneous Salience of the Hindu-Muslim Conflict

The baseline results so far report evidence of in-group vs. out-group bias in lending as well as evidence of debiasing for the average investor in our P2P platform after they start using a robo-advising tool that creates loan portfolios without considering borrowers' religion as a choice variable. Because the bias goes in opposite directions for Hindu and Muslim borrowers before they access robo-advising and hence the reaction of their portfolios also goes in opposite direction after they all access robo-advising, most concerns about the endogeneity of adoption of Auto Invest have little relevance in our context. In particular, standard concerns about unobserved time-varying shocks that might simultaneously explain the decision to adopt robo-advising and the reaction of lenders' portfolios are barely relevant here, because the reactions are very different and the lender's religion determines the direction of the reaction.

At the same time, the Indian setting allows us to dig deeper into our understanding of the role of in-group vs. out-group bias by providing scope for cross-sectional heterogeneity tests. Specifically, based on earlier research in psychology, sociology, and economics, we can isolate sources of heterogeneity in the extent to which the Hindu-Muslim conflict should be salient to lenders, based on the geographic location of lenders and/or their demographic characteristics. For each of these sources of variation, we can thus test the hypothesis that the extent of bias in lending is larger for lenders for whom the Hindu-Muslim conflict is more salient, because for this group the scope for in-group vs. out-group bias is higher. Therefore, under each of the dimensions we introduce

[24]We chose this threshold based on the shape of the relationship we described in Figure 2, which varies systematically above 40%, but the results are similar irrespective of the choice of threshold.

below, based on our interpretation of the baseline results, we would expect the following: (i) Before accessing robo-advising, lenders for whom the conflict is more salient tend to bias their lending against the out-group borrowers and in favor of the in-group borrowers more than other lenders, and (ii) after accessing robo-advising, everybody's shares of lending to each specific religious group are equalized, and hence, the portfolios of lenders for whom the conflict is more salient react more after adopting robo-advising than the portfolios of other lenders.

4.2.1 Hindu-Muslim Riots Across Cities

The first dimension that makes the Hindu-Muslim conflict more salient is localized Hindu-Muslim riots, which, although recorded in India for decades, have substantially increased in frequency and the extent of violence over the last two decades (Oza (2007)). The violence of riots, the aggressive positioning and rhetoric of local politicians on both fronts, and the local media coverage are likely to make the Hindu-Muslim conflict more salient to lenders who are more exposed to such dimensions (D'Acunto et al. (2019)), that is, lenders who reside in locations with a high incidence of riots. Based on this rationale, we expect that the extent of bias against out-group borrowers by in-group lenders is higher for lenders residing in locations that have faced more Hindu-Muslim riots, whereas it is lower for lenders residing in locations in which riots were more limited in number and scope.

For this test, we build on Ticku (2015), who categorizes and collects the occurrence of Hindu-Muslim riots at the local level from 1980 to 2000. The geographic aggregate we consider is the state level. This choice is driven by a few considerations. First, all lenders in our sample are mapped to an Indian state, which they have to choose through a pre-specified menu of available states of residence at the time of signup and is verified personally by Faircent employees based on the lenders' documentation. At the same time, not all lenders report the city in which they reside: If we were considering the city level, we would need to drop all lenders who do not report their city of residence, which would restrict the sample and select it in a way that might potentially correlate with lending choices. Second, regulation and policies that influence the relationship between religious groups are often implemented at the state level; for example, see the case of "anti-conversion laws" (Jenkins (2008); Dhattiwala and Biggs (2012)). Moreover, deep-rooted cultural norms shaped at the local level persistently relate to present-day violence across religious groups in India (Jha (2014)).

Panel A of Figure 3 depicts the cross-state geographic variation we employ based on the incidence of riots as documented by Ticku (2015). In the map, dark-green states (Gujarat, Marahashtra, Karnataka, and Uttar Pradesh) are those in which Hindu-Muslim riots have been most prevalent, whereas riots in other states were less prevalent.

To make the (many) coefficients of interest and their comparison across subsamples easier to track, we report the results of our multivariate heterogeneity tests in graphical form rather than in the form of tables. In Figure 4, we report the results for estimating equation (1) separately for the group of lenders in states with a low incidence of Hindu-Muslim riots over the period 1980-2000 and those in states with a high incidence of riots. In the left plot of panel A, we report estimates of the $\hat{\gamma}$ coefficient, which captures the degree of bias by Hindu lenders towards Muslim borrowers before using Auto Invest. Consistent with our conjecture, the extent of bias was about twice as large for Hindu lenders who reside in states with a higher incidence of Hindu-Muslim riots (6.4 pp), and statistically different from zero, whereas the size of the bias for lenders in other states is low enough that we cannot reject it equals zero statistically. In the left plot of Panel B, we report estimates of the $\hat{\delta}$ coefficient, which captures the size of de-biasing after adopting Auto Invest for the same sets of lenders. As expected, the religious composition rebalancing of the portfolios of lenders who reside in states with a high incidence of Hindu-Muslim riots (4.8 pp, $p<0.01$) was higher than the reaction of other Hindu lenders' portfolios (3 pp, $p>0.1$). The results are similar if we change the threshold of the number of local riots based on which we split lenders into one group or the other.

It is important to note that for this first sample split, in fact we cannot reject the null that the estimated coefficients are equal across subsamples at plausible levels of significance. We report the χ^2 statistic for a Wald-test of equality of the coefficients across the two subsamples on top of the left plot of Panel B. In this case, the statistic is 0.20. As we will see below, instead, the other sample splits, which capture different sources of variation across space and cohorts, allow us to reject this null statistically.

4.2.2 Electoral Support for the BJP Across States

We move on to consider a second dimension that varies in the cross-section of lenders and that earlier research has associated with the salience of the Hindu-Muslim conflict: the local vote share for the *Bharatiya Janata Party* (BJP). A defining feature of the BJP is that the cultural and ideological roots of its platform have always been based on traditional Hindu values (e.g., see Berglund (2004) and Chhibber and Verma (2018)). In particular, the BJP's value and ideological platform includes fostering the notion of *hindutva*, which implies a coincidence between the spheres of Indian culture and traditional Hindu values (e.g., see Prakash (2007) and Chidambaram et al. (2020)). The BJP is the result of a set of mergers of post-independence parties in India and has shared a leading role in Indian national and state-level politics with the *Indian National Congress* since independence (e.g., see Ziegfeld (2020)). In contrast to the BJP, the Indian National Congress has been proposing instances of secularization and, even when supporting traditional Hindu values as defining Indian public life, has been less supportive of conflict between the Hindu majority and Muslim minority (e.g., see Ganguly (2003) and Verma (2016)).

Based on these considerations, we exploit the state-level variation in vote shares of the BJP to capture variation in the extent to which local lenders might see traditional Hindu values as foundational, and hence might be more likely to display an in-group vs. out-group bias against the Muslim population. Note we are not arguing that the BJP vote share is a precise measure of the extent to which each lender supports the Hindu-Muslim conflict, but that, on average, it captures variation in the extent to which the conflict is salient to local lenders.

We obtain data on the official number of voters, residents, and votes cast for various parties for all elections to the national congress and state-level elections from 1977 to 2015 from Bhavnani (2014), whose data set is based on information from the Indian electoral commission. For our test, we compute the vote share of BJP candidates in each election cycle and state and then compute the average BJP vote shares across elections within each state. Panel B of Figure 3 reports the state-level distribution of the average vote share for the BJP between 1977 and 2015.

We compute the average of shares within states, because different states vote for state-level elections in different years and across different cycles. All our results are virtually identical—that is, the allocation of states to those in the high-BJP-vote-share group and low-BJP-vote-share group does not change—if instead we only consider votes cast for national elections as well as if we only consider votes cast over the last two decades instead of since 1977. As is clear from Panel B of Figure 3, using vote shares for the BJP provides us with more variation and a different variation across states than the occurrence of Hindu-Muslim riots depicted in Panel A.

We report the results for these subsamples in the middle plots of Figure 4. When we split lenders into two group based on the vote share for the BJP, we find a stark difference of the estimated effects: the extent of bias against Muslim borrowers by Hindu lenders—reported in Panel A—was 9.4 pp before accessing robo-advising in high-BJP-vote share states, whereas it was only 3.5 pp for lenders in low-BJP-vote-share states. Conversely, after lenders started to use the robo-advising tool, those in high-BJP-vote-share states debiased fully both economically and statistically, whereas other lenders' portfolios barely changed in terms of religious composition of the borrower pool. A χ^2 test further confirms that the degree of de-biasing is statistically different across areas with low and high support for the BJP (χ^2=10.57, p<0.01).

4.2.3 Exposure to Hindu-Muslim Conflict Across Cohorts

The last heterogeneity dimension we consider is not based on variation across space, but on variation across lender cohorts. Here, we exploit the fact that the electoral support for the BJP has increased substantially since the early 2000s (e.g., see Menon and Nigam (2007)) and reached its peak with the increased visibility and popularity of Narendra Modi during his term as Chief Minister of the state of Gujarat from 2001 to 2014, and especially since he became Prime Minister of India in 2014 (Chhibber and Verma (2014)).

The stronger support for the BJP and especially its rise to national power has pushed the issue of Hindu-Muslim relations toward the top of the list of topics in Indian political discourse. To design a cross-sectional test, we build on the recent literature in economics and finance that documents the long-run effect of beliefs and convictions that agents develop during their formative years (late childhood and early adolescence) on their beliefs and convictions later in life (Malmendier and Nagel (2011)). We exploit variation across cohorts of lenders who were exposed to the rise of BJP during their formative years or were only exposed to this phenomenon afterwards, during their adulthood, at a time when their political beliefs and convictions were likely already cemented. Specifically, in our context, we compare lenders who were born after 1990 (i.e., they were 24 or younger when Narendra Modi became Prime Minister) and lenders born before 1990.[25]

[25]The results are similar if we split the sample based on different years around 1990.

We report the results for this third heterogeneity test in the right plots of Figure 4. Panel A shows that the extent of pre-robo-advising bias by Hindu lenders against Muslim borrowers is substantially larger for lenders born after 1990 (7.1 pp) and statistically significant, whereas the bias is smaller (2.6 pp) for older lenders whose convictions were likely formed by the time the BJP's rose to national prominence. The latter effect is so small that we fail to reject the null of no bias. Similar to the other tests, in Panel B we find that after accessing robo-advising, the change in the religious composition of the borrower pool was large for lenders born after 1990 (7.2 pp), who in fact are almost completely debiased, whereas it is economically and statistically zero for older lenders. These two estimated effects are not only economically but also statistically different (χ^2=4.46, p<0.01).

Overall, our heterogeneity tests provide evidence that collectively points in the same direction: the lenders for whom the Hindu-Muslim conflict is likely to be more salient, and for whom the in-group vs. out-group bias based on social conditioning is therefore likely to be stronger, are indeed the lenders who display the stronger bias against out-group borrowers before accessing the robo-advising tool. Consistently, they are also the lenders for whom the bias decreases substantially after using the robo-advising tool, relative to before using it.

On top of providing evidence consistent with lending choices being driven by in-group vs. out-group bias, these heterogeneity results further reduce concerns related to the endogeneity of the adoption of Auto Invest in our context: if the endogenous adoption were explained by unobserved factors that simultaneously explain the change in outcomes, not only should these factor predict changes in outcomes of opposite signs for Hindu and Muslim lenders, ceteris paribus, but also within each group, the unobserved factors should exist for lenders for whom the Hindu-Muslim conflict is more salient but not for other lenders, whether we proxy for varying salience through geographic location or through lenders' cohorts.

5 Stereotypical Discrimination: Shudra (Low-Caste) Borrowers

The second type of cultural bias we study is stereotypical discrimination—the fact that decision-makers systematically discriminate certain social groups, because society attaches negative stereotypes to the members of such groups (Becker (1957), Akerlof and Kranton (2000)). Stereotypical discrimination differs from the in-group vs. out-group bias we discussed above in at least two relevant dimensions: First, everybody shares the negative stereotypes associated with members of the discriminated group, *including* the members of the discriminated group, who might thus engage in actions that discriminate against members of their own group (Jost and Banaji (1994); Nosek et al. (2002); Pritlove et al. (2019)). For example, research finds that not only men but also women tend to rate women's quality and performance in leadership roles lower than men's, even when objective measures of performance across genders are similar, also known as *implicit bias* (Bertrand et al. (2005); Brownstein (2015)).

Moreover, stereotypical discrimination is often based on deep-rooted cultural norms and beliefs that have developed in society and tend to be highly stable over time (e.g., see D'Acunto et al. (2019); Payne et al. (2019)). Whereas, as we saw in the case of Hindu–Muslim conflict, events and shocks might modify the extent of in-group vs. out-group bias, the extent of stereotypical discrimination tends to persist in the long run.

The Indian setting is well suited to the study of stereotypical discrimination, because of the centuries-long and persistent stereotypes attached to members of lower castes. Castes, as we discuss in more detail below, represent social divisions that are deeply ingrained in India's Hindu heritage and that create a hierarchical ranking between groups (Dumont (1980)), which produces a set of positive stereotypes associated with members of the highest-ranking castes and negative stereotypes associated with members of the lowest-ranking castes, which have been studied in sociology for decades (see, e.g., Sinha and Sinha (1967)).

A second unique feature of the Indian setting is that variation exists in the extent to which somebody's caste can be recognized based on observational characteristics such as names and surnames, physical appearance, and occupation, especially without being part of the same caste (Muthukumar (2020)). Certain last names are more recognizable than others as belonging to a specific caste, and certain last names may belong to one caste in a certain community but another caste in other communities. This variation in caste recognizability is why India provides a natural setting in which to test for the extent to which the salience of one's caste based on observational characteristics, rather than the (unobserved to lenders in our setting) actual belonging to a discriminated caste, affects lenders' choices of borrowers. This feature would not exist in other settings such as, for instance, the US, in which instead researchers have exploited the fact that certain names, surnames, and physical traits make one's social group immediately recognizable to any other member of society.

Before motivating the empirical tests in more detail and discussing the results, we provide a concise summary of the institutional feature of the caste system in India and especially of the aspects that are most relevant to our tests.

5.1. A Concise Primer on the Indian Caste System

Based on a set of traditional and foundational Hindu writings, Indian society has been divided into five broad social groups for centuries: four *varnas*, or castes, and a fifth group of "outcasts" or untouchables (Fox (1969)).[26] In the traditional interpretation, these social groups have a strict hierarchical relation to one other. *Brahmins* were the highest caste, and traditionally included the Hindu clerics cast, as well as teachers and researchers, that is, all those who dedicated their lives to contemplative activities rather than engaging directly in administration or manual work. The second caste (*Kshatriyas*) traditionally encompassed warriors and rulers. Members of this caste historically covered governmental and military positions. The third caste, the *Vaishyas*, included farmers, traders, and merchants. Historians have emphasized similarities between the notion of *Vaishyas* varna and the "Third state" in pre-revolutionary France, for instance.

[26] Here, we refer to the traditional scriptures-based notion of *varnas*. It does not coincide with the notion of *jati*, which is as richer and more complex sociological system based on which Hindus are further divided into other castes, tribes, and local social groups.

Against the three top varnas stands the (*Shudra*) caste, which has historically included laborers such as peasants and servants in various roles (Ambedkar (1947)). This caste was explicitly lower ranked than other castes, and its members were employed in roles of service to the benefit of higher castes. The centuries-long strict implementation of societal roles based on the caste system has created a set of cultural stereotypes attached to members of each caste and consequent outright discrimination towards the members of the *Shudra* caste.

Note that members of the outcast group, the *Dalits*, have faced even stronger discrimination and segregation over the centuries, both physically and socially (Maikkēl (1999)). Perhaps as a consequence of this strong discrimination and segregation, unfortunately, less than 2% of the borrowers in our platform are *Dalits*, which makes meaningfully studying the effects of lending between them and members of the four varnas impossible for us. For this reason, our empirical tests and analysis focus on comparing borrowers' lending decisions toward *Shudra* borrowers relative to borrowers who belong to the three higher castes. Importantly, as discussed above, a tenet of stereotypical discrimination is that the members of the discriminated group themselves might have an implicit bias that is so ingrained that they discriminate against other members of their own caste. In our context, this feature suggests *Shudra* borrowers are likely to be discriminated against by both *Shudra* lenders and other lenders.

5.2. Variation in the Recognizability of Borrowers' Caste

As we discussed when introducing the raw-data evidence on stereotypical discrimination, the Indian setting is unique in that substantial variation exists in the extent to which somebody's caste can be inferred from observable information. In our P2P application, lenders observe borrowers' names and surnames, see their face picture, and some other limited set of demographics. The most salient information on the platform relates to the borrowers' risk characteristics, for which we control directly in our multivariate analyses, such as the loans' interest rates, maturity, and amounts, which are ultimately set by the platform's algorithm.

In other cultural environments, such as the US, the recognizability of somebody's ethnicity based on name, surname, and face picture is very high, and in fact, earlier research in psychology, sociology, and economics has exploited the salience of ethnic features observed from these three pieces of evidence to assess how choices depends on agents' ethnicity. In the case of India, instead, these three features do not always provide immediate recognition of one's caste.

Indeed, features such as names, surnames, occupations, and complexion sometimes help agents identify each others' caste, but that's not always the case. Ultimately, whether a potential borrower belongs to the *Shudra* varna is easier or harder to assess on a case-by-case basis.[27]

We exploit these two features—variation in caste recognizability based on demographic information and based on whether lenders and borrowers belong to the same community—to design heterogeneity tests for the

[27] Recall that, due to the lack of Dalit borrowers and lenders on the platform, we do not have this fifth group in our sample.

effects of stereotypical discrimination in lending.

We first focus on proxying for the extent to which a borrowers' caste is easily or minimally recognizable based on demographic characteristics. Ideally, we would have measured directly the extent to which lenders recognized the borrowers' caste, for instance, by asking them to guess the caste when facing a potential borrower and then comparing this lender-reported caste with the actual caste of the borrower. Unfortunately, had we asked explicitly, we would have invalidated our test: making borrowers' caste salient by asking would have interfered with the lenders' decisions, potentially in both directions. For instance, lenders could have willingly changed their discriminating behavior relative to what they would have done without our intervention, to pretend they were not discriminating against lower castes in daily life. For this reason, we could not directly elicit the extent of lenders' ability to recognize castes.

To overcome this issue, we instead implemented an off-the-shelf algorithm that assigns last names and other characteristics to castes and is designed to mimic the decision that a human would make based on the information at hand. Specifically, we rely on the methodology developed and detailed in Bhagavatula et al. (2017), of which we provide a brief summary here.[28] The procedure relies on two unique aspects of the caste system in India. First, castes are endogamous—marriages occur mainly between individuals that belong to the same caste. Second, in India, last names are in part indicative of castes.

In the first step, the procedure collects data from 2.5 million individuals registered on online matrimonial agencies. This data contains information on individuals' last names and varna. Varnas are self-reported, and the possibility of misreporting is virtually non-existent, because varnas are a fundamental variable that matrimonial agencies use to match potential couples, due to the fact that castes tend to be endogamous. In this step, the procedure also groups together all the different variations of the same last name.

In the second step, the procedure assigns one or more castes to each of the last names. The relation between caste and last names is not unique, because the same last name can be associated with different castes especially in different cities or regions. The algorithm assigns the probability of a last name belonging to a given caste to equal the proportion of times the matrimonial website users with that last name self-identify as belonging to that caste in every state and city.

In the third step, we assign a caste (and its probability) to each borrower and lender on the platform on the basis of the user's last name, state, and city.

Figure A.5 in the Online Appendix plots the distribution of the probability of being *Shudra* for the borrowers in our sample for whom such a probability is strictly larger than zero, which includes about 80% of the full sample of borrowers on the platform. Except for a fat right tail of 17% of borrowers whose probability of being *Shudra* is close to 1, the probability is distributed rather homogeneously throughout the support.

[28]We thank Manaswini Bhalla for graciously running the algorithm developed in Bhagavatula et al. (2017) and Bhagavatula et al. (2018) on our data.

5.3. Stereotypical Lending Before and After Robo-advising

We start by showing in Figure 5 the motivating plots on stereotypical discrimination in lending based on borrowers' castes. The top three graphs report the average of *Shudra* borrowers within lenders' portfolios for lenders belonging to any castes. The leftmost graph considers the full set of borrowers in a lender's portfolio of loans; that is, it includes borrowers whose caste is easy to recognize based on names, surnames, and face pictures (which our lenders observe on the platform), as well as borrowers whose caste is not recognizable. In this case, we detect no difference in the share of *Shudra* borrowers before and after lenders use the robo-advising tool: the share of loans going to *Shudra* borrowers is virtually the same before and after access to the tool and is equal to 31%, which is close to the share of *Shudra* in the overall borrower population on the platform. This result is consistent with the possibility that, when castes are barely identifiable, no discrimination exists against *Shudra* borrowers.

Because we instead restrict the sample to subgroups in which the borrowers' caste is more and more recognizable, whatever the case, as we do while moving towards the graphs on the right, we observe a very different pattern. In subfigure (b), the pre-Auto Invest lending to *Shudra* borrowers was only 27%, which was further reduced to 25% in subfigure (c), where we restrict the sample to borrowers whose caste is highly recognizable. Crucially for our interpretation, in terms of stereotypical discrimination, we find the lending after the use of Auto Invest is virtually identical across all three subpopulations and equal to 31%.

Overall, the three graphs on top of Figure 5, which are based on raw data, seem consistent with the presence of stereotypical discrimination in our sample: if castes are not recognizable, *Shudra* borrowers are treated similarly to other borrowers. But as castes become more and more recognizable, whenever lenders make decisions on their own, they discriminate more and more against *Shudra* borrowers.

As discussed above, a defining feature of stereotypical discrimination relative to in-group vs. out-group discrimination is that even the members of the same discriminated category are often subject to the implicit stereotypical bias that is ingrained through social conditioning. For this reason, even *Shudra* lenders should discriminate in favor of higher-caste borrowers and against *Shudra* borrowers, ceteris paribus. The Indian setting, in addition, allows for an additional unique proof of concept: as the extent of caste recognizability varies, one might expect that *Shudra* lenders more easily recognize members of the same caste in cases in which members of other castes might have a harder time, because certain features that a *Shudra* might recognize as familiar might not resonate with other castes' lenders. For this reason, if anything, the extent of discrimination of *Shudra* lenders against *Shudra* borrowers might be *higher* for the same levels of caste recognizability by third parties.

And, indeed, the raw-data plots in the bottom three subfigures of Figure 5, in which we restrict the sample to include only *Shudra* lenders, provide evidence consistent with this conjecture. Not only do *Shudra* lenders behave similarly to *Brahmin*, *Kshatriyas*, and *Vaishyas* lenders to *Shudra* borrowers, but if anything *Shudra* lenders discriminate against *Shudra* borrowers more than other lenders. For all levels of recognizability, in the raw data, *Shudra* lenders tended to provide about 1 pp lower loans to *Shudra* borrowers than to lenders of other castes.

These plots represent interesting motivating evidence but are based on the raw data and do not keep constant borrowers' characteristics, which might vary systematically across lenders' portfolio. Moreover, we need to assess the statistical significance of the raw-data results. For these reasons, we estimate the following multivariate specification:

$$Shudra\ Borrower_{i,j} = \alpha + \beta\ Auto\ Invest_j + \zeta\ \boldsymbol{x}_i + \epsilon_{i,j}, \tag{2}$$

where the regressors are defined as in equation 1.

Because we want to focus on several subsamples of the data based on caste recognizability for both the baseline and heterogeneity results, to facilitate the comparison of many relevant coefficients across specifications, we report these results in graphical form. Panel A iof Figure 6 contains four bars representing different estimates of β in equation (2) across different subsamples of the data.

To estimate the first bar to the left, we do not impose any sample restrictions regarding borrowers' caste recognizability. The β coefficient is positive (0.011) and statistically significant at the 5% level (t=2.54), indicating that using Auto Invest increases the probability of lending to Shudra borrowers. On the flip side, the result suggests that before using Auto Invest, lenders were discriminating against Shudra borrowers.

The second bar imposes that the probability of recognizing the caste by way of the algorithm is at least 40% and shows the results are pretty similar to the first, both economically and statistically. The last two columns progressively restrict the estimating sample to lenders whose belonging to the Shudra caste is more easily recognizable—their probability of being recognized as Shudra is larger than 70% and 80%, respectively. When we impose this restriction, we find the relation between lending to Shudra borrowers and adopting Auto Invest more than doubles, suggesting that before using Auto Invest lenders were discriminating more against *Shudra* borrowers who were easy to recognize as belonging to that caste. The size of the estimated effect doubles within the subset of easily recognizable Shudra borrowers relative to the full sample. These estimated effects are statistically significant at the 1% level, with t-statistics in excess of 3.5.

5.4. Heterogeneous Salience of Negative Stereotypes Attached to Lower Castes

The results discussed so far refer to all lenders and borrowers. Similar to our analysis of in-group vs. out-group discrimination, in this section, we exploit pre-determined variation in the extent of the scope for discriminating members of the Shudra caste on the part of lenders. Specifically, we consider two heterogeneity dimensions: the extent of inter-caste hatred across geography (Indian states), which, based on earlier research, we capture with the local incidence of crimes against members of lower castes, as well as the extent to which borrowers' castes are systematically more or less recognizable based on whether borrowers belong to the same community as the lenders.

5.4.1 Hatred Crimes against Lower Castes Across States

Recent research documents a substantial increase in hatred crimes against members of lower castes over the last decade, which has been heterogeneous across space. Researchers argue that several social factors might have driven this phenomenon, in particular the fact that, even though at a very slow pace, lower castes have increased their economic well-being and have been the target of policies aimed at guaranteeing quotas of public employees in both national and local institutions from lower castes (Sharma (2015); Bapuji and Chrispal (2020)). Ultimately, the slow decrease in disparities between the economic and social conditions of higher castes and lower castes might have triggered the increase in acts of hatred against lower castes.

We conjecture that the stereotypical discrimination of *Shudra* borrowers might be higher in areas in which the conflict between higher and lower castes might be more salient due to the higher incidence and reporting of acts of violence against lower castes. To operationalize this conjecture, we collect the number of crimes against lower-caste victims per 100,000 inhabitants of Indian states from the annual report of the *National Crime Records Bureau (NCRB)* (NCRB (2019)). We report the results when using crime rates based on the 2018 annual report, but all our results are similar when using other recent years, because the ranking of states based on the rate of crime against lower-caste victims is quite stable over the recent years.

Figure 7 plots the cross-sectional variation in crimes against lower-caste victims, and we estimate equation (2) separately for lenders in Indian states above and below the median rate of crimes against lower castes per 100,000 inhabitants in the state-level sample (18.8). We report the results in Panel B of Figure 1. For low levels of caste recognizability, we find very little effects in neither areas with low nor high crimes against Shudra: the β coefficients are not statistically different from zero. Starting from 40% recognizability, we start observing a wedge between the two areas, which becomes more and more marked at higher levels of recognizability. Above 70% recognizability, we find that after the adoption of robo-advising, the share of Shudra borrowers in lenders' portfolio in areas with high crimes against Shudra increases by 2.7 pp, which is about 12% more of the average share of Shudra borrowers in lenders' portfolios before accessing the tool as captured by the constant term in the estimated equation (unreported in the graph). When focusing on the subsample of lenders who live in states below the median by hatred crimes against the Shudra caste, we find that the share of *Shudra* borrowers in portfolios only increases by 1.3 pp, and we fail to reject the null that this estimated increase is equal to zero statistically. As shown in the figure, the difference becomes even starker when we condition on higher levels of caste recognizability.

6 Cultural Debiasing and Lenders' Performance

Sections 4 and 5 have studied the effects of discrimination in two different domains. First, we considered in-group vs. out-group discrimination: Muslim lenders tend to prefer lending to Muslim borrowers, Hindu lenders tend to prefer lending to Hindu borrowers. As a result, when individuals adopt Auto Invest, Muslim lenders lend more to Hindu borrowers (and less to Muslim borrowers), and vice versa. Second, we studied stereotypical discrimination by considering the patterns of lending by caste, and found that *all* lenders, including Shudra lenders, tended to discriminate against Shudra borrowers before their lending decisions were automated through an algorithm that does not consider castes as a dimension for assigning credit. In this section, we move on to assess the financial implications of cultural biases in lending, that is, we ask: "Does the cultural debiasing associated with the use of a robo-advising tool improve or worsen lenders' performance?"

6.1. In Which Direction Should Performance Change?

Intuitively, if lenders have biases in favor of or against a certain religion or caste, they might need to reach deeper into the pool of borrowers of the religion (or caste) they favor when choosing to whom they lend their money. As a result, they should lend to less creditworthy borrowers of the preferred religion (or caste), whereas they should reject more creditworthy borrowers of the religion (or caste) against whom they are discriminating.

Our setting displays two features that make it appropriate for testing for the presence and effects of cultural biases. First, it abstracts from the potential screening and monitoring roles of in-group lending, which are instead documented and studied extensively in Fisman, Paravisini, and Vig (2017) and Fisman, Sarkar, Skrastins, and Vig (2020). Our platform connects lenders and borrowers all over India and does not involve a form of localized relationship lending, in which the loan officer of a local branch of a commercial bank screens and monitors local borrowers on behalf of the lender (the bank) when making her lending decisions. This screening/monitoring channel is a defining feature of local-branch lending by loan officers to local borrowers, who live and operate in the same small social environment as the loan officers. In that setting, this channel is so compelling that, quantitatively, Fisman, Paravisini, and Vig (2017) and Fisman, Sarkar, Skrastins, and Vig (2020) show it prevails relative to the negative effects of discrimination in terms of borrower quality in the context of localized relationship lending. The fact that, in our setting, the screening/monitoring channel is limited makes it an ideal setting in which to test for the effects and value of cultural discrimination in the field.

Because the screening/monitoring channel is inactive in our case, our prediction is the *opposite* relative to the result in Fisman, Paravisini, and Vig (2017) and Fisman, Paravisini, and Vig (2017). We conjecture that cultural debiasing improves rather than worsens lending performance.

The second unique feature of our setting is that the lending decision does not simultaneously include an assessment of borrowers' risk profile and a potential discriminatory choice: the risk levels of borrowers and their loan requests are assessed ex ante by the Faircent algorithm when the borrowers are admitted to the P2P platform. When making decisions, lenders see this risk assessment of potential borrowers, which is unrelated to borrowers' religion or caste. By studying lending decisions under the robo-advising tool, we can ex-post verify directly that the Faircent algorithms does not use religion or caste when assessing borrowers' risk, by checking that the shares of borrowers by religion or caste are equalized across lenders of different types (see Figure 1 and Figure 5). Decoupling the risk assessment from the lending decision, which are conjoint decisions for loan officers of a commercial bank's local branch, allows us to consider deviations from the robo-advising-determined outcomes as driven by active discrimination on the part of lenders rather than by systematic heterogeneity in risk levels that might correlate with borrowers' religion or caste.

The fact that both the lender and the econometrician observe the same risk assessment for each borrower that the platform performs is a third important feature of our setting because it allows assessing the channels through which the performance of lenders change based on when they choose borrowers relative to when the Auto Invest chooses them. Indeed, we can test directly whether Auto Invest improves lenders' performance by reallocating the risk profiles of the borrowers in their portfolios. In particular, we can test whether lenders performed worse when making biased choices because they tended to choose riskier in-group borrowers when choosing alone and whether the Auto Invest, by reducing the riskiness of the in-group borrowers in lenders' portfolios, improves lenders' performance.

6.2. Performance after Cultural Debiasing: Sign, Size, and Channels

We first consider a measure of the extensive margin of lending performance—whether borrowers default on their loans. Loan defaults are a commonly used measures of performance and have also been studied as an outcome for moral financial decision-making (e.g., see Guiso et al. (2013)). In our case, loan default allow testing whether in-group borrowers are less likely to default on in-group lenders relative to out-group borrowers due to stigma from the in-group community, which is the opposite of what cultural biases predict.

We will then move on to assess the effects of debiasing on actual loan returns, which are the ultimate measure of lenders' performance as they invest in each borrower by extending them a loan.

Finally, we will assess the channels through which the debiasing induced by Auto Invest affects lenders' performance.

_6.2.1_ Loan Default (Extensive Margin) To assess loans' default, we consider all the loans in our sample that were closed by the last month we have available—March 2020—and categorize as defaulted those loans that had been delinquent for more than 90 days at the time of closure. This definition is close to the regulatory definition by Reserve Bank of India (RBI) as well as to the economic definition of a defaulted loan, whereby the borrower did not pay back in full the loan's future value to the lender at the time the platform stopped monitoring the borrower.[29]

We start by considering in-group vs. out-group discrimination. For Hindu lenders, we estimate the following specification by OLS:

$$\begin{aligned} Delinquent\ Loan_{ij} = \alpha + \gamma\ Muslim\ Borrower_j \\ + \delta\ Muslim\ Borrower_j \times Auto\ Invest_j \\ + \theta\ Hindu\ Borrower_j \times Auto\ Invest_j + \zeta\ \boldsymbol{x}_i + \epsilon_{i,j}, \end{aligned} \tag{3}$$

where $Delinquent\_Loan_{ij}$ is equal to 1 if the loan associated with borrower i and lender j is closed as delinquent, and all other variables are defined as discussed above.

This specification allows us to assess three implications of the discrimination hypothesis in terms of performance. First, we should observe that, on average, Muslim borrowers are less likely to default relative to Hindu borrowers in Hindu lenders' portfolios *before* Auto Invest is used ($\gamma < 0$), because Hindu lenders should dig deeper into the Hindu pool and hence choose Hindu borrowers of worse quality when making unassisted decisions. We report evidence for this result both graphically and in table format. In panel A of Figure 8, we plot the estimated coefficient $\hat{\gamma}$ for the full sample of Hindu lenders ("All") as well as when estimating equation 3 separately for the subsamples based on the heterogeneity of Hindu-Muslim conflict to lenders we discussed in the previous section. We find that not only Muslim borrowers are on average 4.6 percentage-point less likely to default than Hindu borrowers before lenders access Auto Invest, but the size of this estimate is higher for the groups of Hindu lenders for whom the conflict should be more salient—lenders in states with a higher incidence of Hindu-Muslim riots, those in states with a higher support for the BJP, and younger lenders. Columns (1)-(2) of Table 3 show that the baseline lower likelihood of default by Muslim borrowers before lenders access robo-advising is statistically different from zero.

The second implication of biased discrimination is that, before controlling directly for the risk profile of borrowers, we should observe that after Hindu lenders access Auto Invest Hindu borrowers reduce their likelihood of default by more than Muslim borrowers.[30] That is, we should observe that $\theta < \delta$ when estimating equation 3. Results in column (2) of Table 3 are consistent with this implication: Muslim borrowers reduce their likelihood of default by 7.3 percentage points whereas Hindu borrowers by 11.2 percentage points—an effect that is 53% larger than for Muslim borrowers.

The third implication we bring to the data relates to the channels that should explain the difference in

[29]Note that we do not observe whether lenders engaged in litigation to collect borrowers' debentures after the loan was closed. If lenders were ultimately able to obtain a higher repayment than what is registered in the company's accounts, unfortunately we cannot know.

[30]Note that Auto Invest might provide additional benefits in terms of performance than those due to cultural debiasing. For instance, it might provide lenders with a more diversified portfolio of borrowers. For this reason, the likelihood of default by Muslim borrowers might also decline with robo-advising. Discrimination though undoubtedly predicts that the likelihood of default should decline by more for Hindu than Muslim borrowers.

performance by Hindu and Muslim borrowers before and after robo-advising. If Hindu lenders were willing to dig deeper in the pool of Hindu borrowers and pick riskier borrowers in that group, because on our platform the highest-risk borrowers are disproportionally more likely to default than other borrowers (see Figure A.4 for the relation between ex-post default probabilities and interest rates), we should observe that once we control for the observable proxies for risk to lenders—interest rate, maturity, and loan amount—the differences in defaults across Hindu and Muslim borrowers should disappear. Moreover, the estimated improvement after accessing robo-advising should converge for both groups to the improvement we estimated for Muslim borrowers without controls for risk, which under a discrimination story should not be driven by borrower-level risk characteristics. Column (3) of Table 3 shows results that align with these predictions. Once we control for borrower-level risk, neither Muslim borrowers appear to default less than Hindu borrowers before robo-advising nor do Hindu borrowers appear as improving their performance by more than Muslim lenders after robo-advising. Also, the two point estimates for the decline in default rates—7.2 and 7 percentage points—are indistinguishable from the estimated improvement of Muslim borrowers in column (2), which is 7.3 percentage points.

Moving on to stereotypical bias, we assess the same three implications of cultural debiasing in the data—that Shudra borrowers were less likely to default than other borrowers before robo-advising, that the improvement in default after robo-advising was larger for non-Shudra borrowers, and that controlling for borrower-level risk characteristics should shrink the differences in defaults before and after lenders accessed robo-advising if these differences were driven by lenders choosing riskier (which, on our platform, means disproportionally more likely to default) borrowers before robo-advising.

Graphically, Panel B of Figure 8 shows that not only Shudra borrowers were less likely to default than other borrowers before lenders accessed Auto Invest, but this difference in default rates increased with borrowers' recognizability as members of the Shudra caste. The left plot shows that Shudra borrowers whose probability of being recognized was high (above 80%) were about 13 percentage points less likely to default than other borrowers, whereas then we consider the full sample of Shudra borrowers this lower likelihood of default was 4.5 percentage points. Moreover, the right plot shows that this result is also starker for lenders who reside in states with a higher incidence of crime against Shudras, which we interpreted as higher salience of stereotypical discrimination against members of the lower caste in the previous section. Indeed, among highly recognizable Shudras, those picked by lenders in high-crime states were about twice less likely to default than other Shudra borrowers.

We also find evidence consistent with the second implication in column (5) of Table 3, where we find that non-Shudra borrowers' performance improved by more that Shudra borowers' performance, although the difference between these two effects (16 and 14.8 percentage points) is economically small and not statistically significant.

Finally, columns (6) shows that, once we control for borrower-level risk, both estimated effects decline and are aligned to about 10 percentage points. Also, before robo-advising, the lower likelihood of default of Shudra borrowers declines even though it does not become economically and statistically insignificant as we found for the case of Muslim borrowers in column (3).

6.2.2 Fraction of Loan Repaid (Intensive Margin)

We move to study the intensive margin of performance by considering two variables. First, the share of the overall amount due by the borrower (including principal and interest) that the borrower pays back to the lender, which allows us to assess the extent to which borrowers are willing to fully repay loans. Second, the actual return that lenders earned on each loan by the time of maturity, which is the ultimate measure of performance at the lender level.

The share of repaid amount is capped at 1 for borrowers who repay their loan in full, because no borrowers pay more than what is due to the lenders. In principle, the share can be as low as 0 if a borrower does not repay anything, but in our data the number of borrowers who repay less than 20% of their amount due is minimal, because the platform expels borrowers who pay less than 20% of the amount due on any outstanding loans.

We study the shares of repaid amount across groups and before and after access to Auto Invest in Figure 9. In Panel A, we focus on in-group vs. out-group debiasing. The graph to the left plots the cumulative distribution functions (CDFs) for the share of the amount due paid by Hindu borrowers (solid green line) and by Muslim borrowers (orange dashed line) before Hindu lenders use Auto Invest, and hence at a time when Hindu lenders made all their lending choices autonomously. The cumulative distributions display evidence consistent with our conjecture that Hindu lenders might dig deeper into the pool of Hindu borrowers than into the pool of Muslim borrowers. Muslim borrowers who receive loans from Hindu lenders are substantially more likely to repay larger shares of their amount due before the loan servicing is closed. In fact, no Muslim borrowers picked by Hindu lenders pay less than 80% of the amount due. To the contrary, the repayment behavior of Hindu lenders is more volatile: about 20% of them repay less than 40% of the amounts due, and even when considering those who pay at least 80%, the CDF of Hindu borrowers is flatter than that of Muslim borrowers.[31]

The right graph of Figure 9 plots the CDFs for the shares of amount due repaid by Hindu and Muslim borrowers to Hindu lenders after Hindu lenders start to use Auto Invest. Hindu borrowers appear to improve disproportionally more than Muslim borrowers along the distribution. For instance, the share of Hindu borrowers who pay back more than 90% of their loans among those picked by Hindu lenders under Auto Invest increases to about 40% from 30% before Auto Invest.

We then consider the intensive margin of performance for stereotypical discrimination in Panel B of Figure 9. The left graph reports the CDFs of the percentage of amounts due repaid by borrowers whose probability of being Shudra is below the median (solid green line) or above the median (orange dashed line) before lenders accessed the robo-advising tool. Throughout the support of the probability of being Shudra, and especially in the left part of the distribution, we find that before the introduction of Auto Invest, borrowers who are likely to be Shudra tend to repay a higher fraction of their amount due, relative to other borrowers.

[31] Note also that the share of Hindu borrowers who repay intermediate amounts between 50% and 80% is negligible.

Once lenders move to Auto Invest in the right graph of Panel B of Figure 9, the distance between the CDFs of the two types of borrowers decreases, and the drop is largely driven by an improvement of the low-probability Shudra borrowers' repayment behavior—the solid green line shifts to the right. Even for the case of stereotypical discrimination, the results for the pre-Auto Invest period are consistent with lenders imposing higher standards for the selection of high-probability Shudra borrowers than for the selection of other borrowers when they make unassisted choices. Once the robo-advising tool debiases lenders' stereotypical discrimination, the standards applied to Shudra and non-Shudra borrowers converge.

6.3. Overall Loan Returns

The analysis of the share of loans repaid is qualitative, so we move on to a more quantitative assessment by considering the overall return each loan provides to the lender by the time of maturity. We consider two types of specifications. First, we assess the changes in the average lender's return through standard OLS specifications. Second, because the effects could be driven by specific parts of the loan return distribution (e.g., very risky loans or very safe loans), we follow D'Acunto and Rossi (ming) and also estimate a set of quantile regressions of the following form:

$$Q_\tau(Returns_{i,j,t}) = \alpha(\tau) + \beta(\tau)\ Auto\ Invest_{j,t} +\ X'_{i,j,t}\ \zeta(\tau) + \epsilon_{i,j,t}, \tag{4}$$

whose outcome variable is quantile Q_τ of the distribution of the return associated with borrower i and j throughout the sample period. All other variables are defined as in equation (3).

To interpret the estimates of equation (4), consider the special case of the median, which is the 50^{th} percentile of the distribution. The coefficient $\hat{\beta}(50)$ estimates that the median return was $\hat{\beta}(50)$ higher after lender j moved to Auto Invest relative to before. A positive $\hat{\beta}(50)$ would suggest the median of the distribution has shifted to the right. The advantage of estimating quantile regressions is that we can assess how the whole intensive margin (distribution) has changed rather than focusing on specific moments, such as the conditional mean.

We report the results for estimating the baseline OLS specification in columns (1)-(2) of Panel A of Table 4 and the quantile regression estimates in columns (3)-(8).

In the first line of each column, we focus on specifications that *do not* control for the risk characteristics of loans, such as interest rates and amount, so that we allow for the possibility that difference between the returns before and after access to the robo-advising tool are driven by a different distribution of the risk of loans in each lender's portfolio.

Based on the hypothesis that Hindu lenders picked in-group borrowers of worse quality before moving to Auto Invest, we should find that they improve their performance after accessing to Auto Invest. The same should be true for Muslim lenders, who, under cultural biases, should have picked in-group borrowers of worse quality than otherwise available out-group borrowers. And, indeed, the first line of columns (1)-(2) of Panel A of Table 4 reveals that both Hindu and Muslim lenders improve their performance in terms of average loan returns.

Moving on to the quantile regression results (columns (3)-(8)), they reveal that most of the improvement in returns is driven by substantially higher returns in the left tail of the return distribution, as can be seen by the fact that the size of the estimated coefficients is larger for the 25^{th} and 50^{th} percentiles of the return distribution and declines as we move towards the right (the 75^{th} percentile) of the return distribution.

In the second line of each column, instead, we add borrower-level loan characteristics as controls to assess the extent to which any changes before and after Auto Invest might indeed be driven by systematically different choices in terms of borrower risk—for instance, whether lenders tended to choose systematically riskier in-group borrowers and Auto Invest instead chose borrowers with lower levels of risk, irrespective of their religion.

The evidence is consistent with a risk channel that explains the performance improvement lenders enjoy when moving to Auto Invest. Once we absorb differences in the riskiness of loans, conditional returns do not differ when lenders make choices on their own or when the robo-advising tool makes choices on their behalf, either on average or in terms of the different parts of the distribution of loan returns, including the left tail of the distribution.

We perform the same analysis for the case of stereotypical discrimination. The first line of Panel B of Table 4 estimates the relationship for Shudra borrowers without controlling for the loans' risk characteristics, whereas the second line estimates the relationship conditional on borrower-level risk proxies.

The patterns we uncover for stereotypical discrimination are the same as we found in the case of in-group vs. out-group discrimination. In particular, the bulk of the improvement in returns is driven by the elimination of a left tail of low returns lenders earned when choosing borrowers on their own. Moreover, once we estimate the differential returns conditional on loans' risk measures, we detect no systematic differences in returns with and without Auto Invest, which is consistent with the possibility that the return improvement under Auto Invest is due to the selection of a less risky set of borrowers and especially avoiding a left tail of high-risk low-return borrowers.

Ultimately, our analysis of performance suggests that eliminating cultural biases improves lenders' performance and this improvement is driven by a change in the composition of the borrower pool that reduces lenders' risk exposure. This reduction is largely driven by eliminating a left tail of low-return loans and is consistent with the prediction of in-group vs. out-group discrimination that lenders affected by this cultural bias might tend to dig deeper into the pool of in-group borrowers and hence select riskier borrowers who end up providing them with low returns.

7 Quantifying the Cost of Cultural Biases at the Lender Level

We conclude our analysis by proposing an exercise to quantify the aggregate effects of cultural biases on the earnings of the lenders, who are the discriminating agents in our setting. To this aim, we depart from the analysis at the lender-borrower-loan level proposed thus far and focus on lender-level performance measurement.

We start by computing the change in the returns each lender made on their investment before and after accessing the automated robo-advising tool, both in general as well as separately for the amounts lenders disbursed to in-group vs. out group borrowers as well as to Shudra vs. non-Shudra borrowers. For each lender, we define the total return on the investment before and after Auto Invest as follows:

$$Lender\ Return_{i,t} = 100 \times \frac{\sum_j Amount\ Disbursed_{i,j,t} \times Loan\ Return_{j,t}}{\sum_j Amount\ Disbursed_{i,j,t}}, \tag{5}$$

where $Return_{i,t}$ is the overall return on the aggregate investment made by lender i earned either before access to Auto Invest ($t = PRE$) or after access to Auto Invest ($t = POST$); $Amount\ Disbursed_{i,j,t}$ is the amount (in rupees) lender i disbursed to loan j, which was issued either before or after access to Auto Invest (t); $Loan\ Return_{j,t}$ is the return of loan j which lender i contributed to.

The quantities defined by equation (5) thus capture the total returns to invested amounts lenders obtained before and after using Auto Invest. We then compute the lender-level change in total return across the two conditions:

$$Change\ Lender\ Return_i = Lender\ Return_{i,POST} - Lender\ Return_{i,PRE}, \tag{6}$$

where a positive value indicates that lender i earned a higher total return on their investment after accessing Auto Invest relative to before, and a negative value the opposite.

In Figure 10, we plot the density of the distributions of $Change\ Lender\ Return_i$ for lenders in the in-group vs. out-group discrimination sample (panel A) and for those in the stereotypical discrimination sample (panel B). For each distribution, we indicate the mean of the distribution with a solid vertical line and compare it to a dashed vertical line that indicates no change in returns. We find that the average lender in the in-group vs. out-group discrimination sample earned a 4.5-percentage-point higher total return after accessing Auto Invest relative to before, whereas the average lender in the stereotypical discrimination sample earn a 7.3-percentage-point higher return.

Our regression results at the lender-borrower-loan level suggested that most of the improvement in terms of loan defaults and repayment behavior derived from borrowers belonging to demographics that lenders tended to favor before moving to Auto Invest. As expected, this pattern holds in terms of lender-level total returns. For instance, if we compute the change in lender returns defined in equation 5 separately for Hindu lenders based on the amounts they lent to Hindu borrowers or to Muslim borrowers before and after Auto Invest, we find Hindu lenders on average gained a 4.3-percentage-point higher return on the amounts disbursed to Hindu borrowers, whereas they actually on average "lost" (an insignificant) 0.5 percentage points in returns to the amounts disbursed to Muslim borrowers. Virtually the whole improvement in the average lender-level returns derive from higher returns earned on the loans disbursed to favored demographic groups.

To capture the rupee-level change in performance and hence the lender-level and aggregate value of cultural biases, we need a measure of lender-level performance in which returns are value-weighted—they are weighted by the rupee amounts each lender disburses to borrowers on the platform. A challenge to define such measure is that the amounts lenders disbursed before and after accessing Auto Invest might differ for many reasons, which are potentially unrelated to cultural biases, lender characteristics, or platform characteristics. We therefore compute the following:

$$\begin{aligned}
&Change\ Lender\ Value_i \\
&\qquad = [Amount\ Disbursed_{i,POST} \times Return_{i,POST} - Amount\ Disbursed_{i,PRE} \times Return_{i,PRE}] \\
&\qquad - [Amount\ Disbursed_{i,POST} \times Return_{i,POST} - Amount\ Disbursed_{i,PRE} \times Return_{i,POST}]. \qquad (7)
\end{aligned}$$

The expression defined in equation (7) allows us to purge the difference in total rupee-value earnings at the lender level merely due to the fact that lenders might disburse different amounts before and after accessing Auto Invest, irrespective of the returns they earn in the two periods.

Note that equation (7) is equivalent to $Amount\ Disbursed_{i,PRE} \times Change\ Lender\ Return_i$, and hence this measure can be interpreted as the change in lender-level return after accessing Auto Invest relative to before weighted by the rupee amount the lender disbursed on the platform before accessing Auto Invest, which cannot have been determined by the returns the lender earned since after starting to use the tool. Ultimately, this value captures the incremental rupee amount each lender would have earned in the period before accessing Auto Invest had they enjoyed the same return they did enjoy after moving to Auto Invest rather than the actual return they earned over that period.

We find that the average change in lenders' rupee value for lenders in the in-group vs. out-group discrimination sample is ₹457, which is about 6% of the average amount of resources disbursed by each lender in the period before accessing Auto Invest (₹7,543).[32] When we consider the stereotypical discrimination sample, we find that the average change is higher: ₹862, which represents about 12% of the average amount lenders in this sample disbursed before accessing Auto Invest (₹7,091). Overall, the estimated cost of cultural biases appears to be sizable relative to the amounts lenders invested for both types of biases we study.

Note that the calculation proposed in equation (7) does not account for the possibility that the rupee amounts disbursed to each demographic groups would have been different had cultural biases not influenced

[32] Note that the average size of loans on the platform is substantially larger (₹90,523), because, as we discussed when introducing our setting, each loan borrowers received is financed by multiple lenders.

lenders' choices in the period before accessing Auto Invest.

To assess whether accounting for this difference could influence our estimates of the costs of cultural biases substantially, we thus propose a modified version of the computation in equation (7). In this case, we separate the amounts disbursed and returns earned in the pre- and post-periods by Hindu lenders coming from Hindu and Muslim borrowers for the case of in-group vs. out-group discrimination and from Shudra and non-Shudra borrowers for the case of stereotypical discrimination.

To obtain a counterfactual for the first period, we split the amounts Hindu lenders disbursed through the platform in the pre-period among the two borrower groups (Hindus vs. Muslims and Shudra vs. non-Shudra) based on the shares of the post-period funds lenders attribute to each group rather than the true shares they attributed to them in the pre-period. In this way, we keep the true total amounts lenders disbursed through the platform in the pre-period fixed, but we assume that, if lenders faced no cultural biases in the pre-period, they would have split such amounts between borrowers based on the post-period shares.

Economically, the change in the shares of Hindu borrowers in Hindu lenders' portfolios cannot be large given that the majority of borrowers on the platform are Hindu. Indeed, the share of Hindu borrowers in Hindu lenders' portfolio moves from 89.7% in the pre-period to 88.7% in the post-period. For this reason, we would not expect that the correction we propose in this second method will deliver estimates for the cost of in-group vs. out-group bias that are substantially different from those discussed above. And, indeed, we find that the average lender-level change in earnings for Hindu lenders based on this correction is ₹382, which is quite similar to the value estimated above.

The same correction, instead, is likely to imply a larger estimated cost of the bias for the case of stereotypical discrimination, because the share of Shudra borrowers in Hindu lenders' portfolios goes from 41.8% in the pre-period to 46.2% in the post-period. When accounting for this composition change in the pre-period, we obtain an average cost of bias of ₹2,254 at the lender level, which is more than twice as large as the estimate that does not use this correction.

8 Conclusions

We propose a unique setting in which to test for and quantify the extent and effects of cultural biases in a large-stake consumer-finance setting. We detect evidence that both in-group vs. out-group discrimination and stereotypical discrimination are prevalent and economically sizable. These forms of discrimination make discriminating individuals—in our case, lenders—worse off in terms of consumption utility, because by discriminating they finance loans by borrowers who perform worse than other (discriminated) borrowers available on the lending platform.

Our test and results suggest a new role that robo-advising tools might have for future research in economics—

they can provide a benchmark to assess the types and sizes of agents' biases in decision-making. For instance, by coding robo-advising tools that embed different forms of biases or rules of thumb detected in the literature and by comparing decision-makers' unassisted choices with those they make after accessing such tools, one could disentangle the role of alternative biases and quantify them.

Moreover, future research should study whether exposure to robo-advising suggestions lets decision-makers learn about optimal choices and develop rules of thumb that can assist them also when a robo-advisor is not available. For instance, by using interactive robo-advising tools that teach borrowers how to concrete goal-setting saving strategies (Gargano and Rossi (2020)) or provide just-in-time simple financial literacy contents (Burke et al. (2021)).

More broadly, our results beget additional work across several fields on understanding how human and machine-based decision-making interact and complement or substitute each other in a world in which the combination of the two forms of decision-making is becoming ubiquitous in all daily economic decision-making problems agents face.

References

Adams, P. D., S. Hunt, C. Palmer, and R. Zaliauskas (2019). Testing the effectiveness of consumer financial disclosure: Experimental evidence from savings accounts. Technical report, National Bureau of Economic Research.

Agarwal, S., S. Alok, P. Ghosh, S. Ghosh, T. Piskorski, and A. Seru (2017). Banking the unbanked: What do 255 million new bank accounts reveal about financial access? *Columbia Business School Research Paper* (17-12).

Agarwal, S., S. Alok, P. Ghosh, and S. Gupta (2019). Fintech and credit scoring for the millennials: evidence using mobile and social footprints. *Available at SSRN 3507827*.

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *The quarterly journal of economics 115*(3), 715–753.

Alesina, A., P. Giuliano, and N. Nunn (2013). On the origins of gender roles: Women and the plough. *Quarterly Journal of Economics 2*(128), 469–530.

Ambedkar, B. R. (1947). *Who were the Shudras?*, Volume 1. Ssoft Group, INDIA.

Avenancio-León, C. and T. Howard (2019). The assessment gap: Racial inequalities in property taxation. *Available at SSRN 3465010*.

Balyuk, T. (2019). Financial innovation and borrowers: Evidence from peer-to-peer lending. *Rotman School of Management Working Paper* (2802220).

Balyuk, T. and S. Davydenko (2019). Reintermediation in fintech: Evidence from online lending. *Michael J. Brennan Irish Finance Working Paper Series Research Paper* (18-17).

Bapuji, H. and S. Chrispal (2020). Understanding economic inequality through the lens of caste. *Journal of Business Ethics 162*(3), 533–551.

Bartlett, R., A. Morse, R. Stanton, and N. Wallace (2019). Consumer-lending discrimination in the fintech era. Technical report, National Bureau of Economic Research.

Becker, G. (1957). The economics of discrimination. *University of Chicago Press*.

Berglund, H. (2004). Religion and nationalism: Politics of bjp. *Economic and Political Weekly*, 1064–1070.

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

Bhagavatula, S., M. Bhalla, M. Goel, and B. Vissa (2017). The business of religion and caste in india. *Working Paper*.

Bhagavatula, S., M. Bhalla, M. Goel, and B. Vissa (2018). Cultural homophily in corporate boards and firm. *Working Paper*.

Bhat, M. M. A. (2020). Hate crimes in india.

Bhattacharya, U., A. Hackethal, S. Kaesler, B. Loos, and S. Meyer (2012). Is unbiased financial advice to retail investors sufficient? answers from a large field study. *The Review of Financial Studies 25*(4), 975–1032.

Bhavnani, R. R. (2014). India National and State Election Dataset.

Bhutta, N., A. Hizmo, and D. Ringo (2021). How much does racial bias affect mortgage lending? evidence from human and algorithmic credit decisions. *Evidence from Human and Algorithmic Credit Decisions (July 15, 2021)*.

Bohren, J. A., K. Haggag, A. Imas, and D. G. Pope (2019). Inaccurate statistical discrimination. Technical report, National Bureau of Economic Research.

Borjas, G. J. and M. S. Goldberg (1978). Biased screening and discrimination in the labor market. *The american economic review 68*(5), 918–922.

Brass, P. R. (2011). *The production of Hindu-Muslim violence in contemporary India*. University of Washington Press.

Breza, E. (2019). Peer effects and loan repayment: Evidence from the krishna default crisis. *Working Paper*.

Brownstein, M. (2015). Implicit bias.

Burke, C., I. Chak, K. Croxson, F. D'Acunto, R. Jonathan, A. Rossi, and J. Shaw (2021). Improving households' debt management with robo-advising. *Working Paper*.

Bursztyn, L., B. Ferman, S. Fiorin, M. Kanz, and G. Rao (2018). Status goods: Experimental evidence from platinum credit cards. *The Quarterly Journal of Economics 133*(3), 1561–1595.

Bursztyn, L., S. Fiorin, D. Gottlieb, and M. Kanz (2019). Moral incentives in credit card debt repayment: evidence from a field experiment. *Journal of Political Economy 127*(4), 1641–1683.

Chhibber, P. and R. Verma (2014). The bjp's 2014'modi wave': An ideological consolidation of the right. *Economic and Political Weekly*, 50–56.

Chhibber, P. K. and R. Verma (2018). *Ideology and identity: The changing party systems of India*. Oxford University Press.

Chidambaram, S. et al. (2020). The civil society roots of bjp's majoritarian nationalism.

Chiu, L., B. Wolfe, and W. Yoo (2018). Do fintech lenders fairly allocate loans among investors? quid pro quo and regulatory scrutiny in marketplace lending. *Quid Pro Quo and Regulatory Scrutiny in Marketplace Lending (September 30, 2018)*.

Cowgill, B. and C. E. Tucker (2020). Algorithmic fairness and economics. *The Journal of Economic Perspectives*.

Crouzet, N., A. Gupta, and F. Mezzanotti (2019). Shocks and technology adoption: Evidence from electronic payment systems.

D'Acunto, F. (2019a). Identity and choice under risk. *Available at SSRN 3466626*.

D'Acunto, F. (2019b). Tear down this wall street: The effect of anti-market ideology on investment decisions. *Available at SSRN 2705545*.

D'Acunto, F., A. Fuster, and M. Weber (2020). Diverse policy committees are more effective. *Working Paper*.

D'Acunto, F., D. Hoang, M. Paloviita, and M. Weber (2019a). Human frictions in the transmission of economic policy. *Working Paper* (2019-07).

D'Acunto, F., D. Hoang, M. Paloviita, and M. Weber (2019b). Iq, expectations, and choice. Technical report, National Bureau of Economic Research.

D'Acunto, F., N. Prabhala, and A. G. Rossi (2019). The promises and pitfalls of robo-advising. *The Review of Financial Studies 32*(5), 1983–2020.

D'Acunto, F., M. Prokopczuk, and M. Weber (2019). Historical antisemitism, ethnic specialization, and financial development. *The Review of Economic Studies 86*(3), 1170–1206.

DAcunto, F., T. Rauter, C. K. Scheuch, and M. Weber (2020). Perceived precautionary savings motives: Evidence from fintech. Technical report, National Bureau of Economic Research.

D'Acunto, F., A. Rossi, and M. Weber (2019). Crowdsourcing peer information to change spending behavior. *Chicago Booth Research Paper* (19-09).

D'Acunto, F. and A. G. Rossi (2020). Robo-advising. *Handbook of Technological Finance*.

D'Acunto, F. and A. G. Rossi (forthcoming). Regressive mortgage credit redistribution in the post-crisis era. *Review of Financial Studies*.

D'Acunto, F., M. Weber, and J. Xie (2019). Punish one, teach a hundred: The sobering effect of punishment on the unpunished. *Fama-Miller Working Paper*, 19–06.

D'Acunto, F., J. Xie, and J. Yao (2020). Trust and contracts: Empirical evidence. *Working Paper*.

D'Andrea, A. and N. Limodio (2019). High-speed internet, financial technology and banking in africa. *BAFFI CAREFIN Centre Research Paper* (2019-124).

Dhattiwala, R. and M. Biggs (2012). The political logic of ethnic violence: The anti-muslim pogrom in gujarat, 2002. *Politics & Society 40*(4), 483–516.

Diep-Nguyen, H. and H. Dang (2019). Social collateral. Technical report, Working Paper.

Dobbie, W., A. Liberman, D. Paravisini, and V. S. Pathania (2020). Measuring bias in consumer lending. *Review of Economic Studies*.

Drexler, A. and A. Schoar (2014). Do relationships matter? evidence from loan officer turnover. *Management Science 60*(11), 2722–2736.

Duarte, J., S. Siegel, and L. Young (2012). Trust and credit: The role of appearance in peer-to-peer lending. *The Review of Financial Studies 25*(8), 2455–2484.

Dumont, L. (1980). *Homo hierarchicus: The caste system and its implications*. University of Chicago Press.

Engineer, A. (1997). *Communal riots in post-independence India*. Universities Press.

Fisman, R., D. Paravisini, and V. Vig (2017). Cultural proximity and loan outcomes. *American Economic Review 107*(2), 457–92.

Fisman, R., A. Sarkar, J. Skrastins, and V. Vig (2020). Experience of communal conflicts and intergroup lending. *Journal of Political Economy 128*(9), 3346–3375.

Fox, R. G. (1969). Varna schemes and ideological integration in indian society. *Comparative Studies in Society and History 11*(1), 27–45.

Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther (2017). Predictably unequal? the effects of machine learning on credit markets. Technical report, CEPR Discussion Papers.

Ganguly, S. (2003). The crisis of indian secularism. *Journal of Democracy 14*(4), 11–25.

Gargano, A. and A. G. Rossi (2020). There's an app for that: Goal-setting and saving in the fintech era. *Available at SSRN*.

Gargano, A. and A. G. Rossi (2021). Correcting present bias in saving decisions with fintech.

Ghassem-Fachandi, P. (2012). *Pogrom in Gujarat: Hindu nationalism and anti-Muslim violence in India*. Princeton University Press.

Graff, V., J. Galonnier, et al. (2012). Hindu-muslim communal riots in india ii (1986-2011). *Online Encyclopedia of Mass Violence, Paris, Sciences Po, CERI*.

Guiso, L., P. Sapienza, and L. Zingales (2006). Does culture affect economic outcomes? *Journal of Economic perspectives 20*(2), 23–48.

Guiso, L., P. Sapienza, and L. Zingales (2009). Cultural biases in economic exchange? *The quarterly journal of economics 124*(3), 1095–1131.

Guiso, L., P. Sapienza, and L. Zingales (2013). The determinants of attitudes toward strategic default on mortgages. *The Journal of Finance 68*(4), 1473–1515.

Hewstone, M., M. Rubin, and H. Willis (2002). Intergroup bias. *Annual review of psychology 53*(1), 575–604.

Higgins, S. (2019). Financial technology adoption. *Working Paper*.

Hjort, J. (2014). Ethnic divisions and production in firms. *The Quarterly Journal of Economics 129*(4), 1899–1946.

Hjort, J., C. Song, and C. Yenkey (2021). Ethnic investing and the value of firms.

Iyer, R., A. I. Khwaja, E. F. Luttmer, and K. Shue (2016). Screening peers softly: Inferring the quality of small borrowers. *Management Science 62*(6), 1554–1577.

Jenkins, L. D. (2008). Legal limits on religious conversion in india. *Law and contemporary problems 71*(2), 109–127.

Jenkins, R. (2014). *Social identity*. Routledge.

Jha, S. (2014). 'unfinished business': Historic complementarities, political competition and ethnic violence in gujarat. *Journal of Economic Behavior & Organization 104*, 18–36.

Jost, J. T. and M. R. Banaji (1994). The role of stereotyping in system-justification and the production of false consciousness. *British journal of social psychology 33*(1), 1–27.

Karlan, D., M. Mobius, T. Rosenblat, and A. Szeidl (2009). Trust and social collateral. *The Quarterly Journal of Economics 124*(3), 1307–1361.

Kaul, N. (2017). Rise of the political right in india: Hindutva-development mix, modi myth, and dualities. *Journal of Labor and Society 20*(4), 523–548.

Lorenzen, D. N. (1999). Who invented hinduism? *Comparative studies in society and history 41*(4), 630–659.

Maikkēl, E. E. (1999). *Untouchable: Dalits in Modern India*. Lynne Rienner Publishers.

Malmendier, U. and S. Nagel (2011). Depression babies: do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics 126*(1), 373–416.

Menon, N. and A. Nigam (2007). *Power and contestation: India since 1989*. Zed Books.

Mitra, A. and D. Ray (2014). Implications of an economic theory of conflict: Hindu-muslim violence in india. *Journal of Political Economy 122*(4), 719–765.

Mobius, M., T. Rosenblat, and Q. Wang (2016). Ethnic discrimination: Evidence from china. *European Economic Review 90*, 165–177.

Muthukumar, A. (2020). Casteism camouflaged as culture. *Harvard Political Review*.

Naaraayanan, S. L. (2019). Women's inheritance rights and entrepreneurship gender gap. Technical report.

NCRB (2019). Crime in india—statistics, volume ii. *National Crime Records Bureau of India*.

Nosek, B. A., M. R. Banaji, and A. G. Greenwald (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice 6*(1), 101.

Oza, R. (2007). The geography of hindu right-wing violence in india. *Violent geographies: Fear, terror, and political violence*, 153–74.

Paravisini, D., V. Rappoport, and E. Ravina (2017). Risk aversion and wealth: Evidence from person-to-person lending portfolios. *Management Science 63*(2), 279–297.

Parsons, C. A., J. Sulaeman, M. C. Yates, and D. S. Hamermesh (2011). Strike three: Discrimination, incentives, and evaluation. *American Economic Review 101*(4), 1410–35.

Payne, B. K., H. A. Vuletich, and J. L. Brown-Iannuzzi (2019). Historical roots of implicit bias in slavery. *Proceedings of the National Academy of Sciences 116*(24), 11693–11698.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The american economic review 62*(4), 659–661.

Prakash, G. (2007). Secular nationalism, hindutva, and the minority. *The crisis of secularism in India*, 177–188.

Pritlove, C., C. Juando-Prats, K. Ala-Leppilampi, and J. A. Parsons (2019). The good, the bad, and the ugly of implicit bias. *The Lancet 393*(10171), 502–504.

Pursiainen, V. (2020). Cultural biases in equity analysis. *Journal of Finance, Forthcoming*.

Rambachan, A., J. Kleinberg, J. Ludwig, and S. Mullainathan (2020). An economic perspective on algorithmic fairness. In *AEA Papers and Proceedings*, Volume 110, pp. 91–95.

Rao, G. (2019). Familiarity does not breed contempt: Generosity, discrimination, and diversity in delhi schools. *American Economic Review 109*(3), 774–809.

Ravina, E. (2019). Love & loans: The effect of beauty and personal characteristics in credit markets. *Available at SSRN 1107307*.

Rossi, A. G. and S. P. Utkus (2020). Who benefits from robo-advising? evidence from machine learning. *Evidence from Machine Learning (March 10, 2020)*.

Schoar, A. (2012). The personal side of relationship banking. *Available at SSRN 2024653*.

Sharma, S. (2015). Caste-based crimes and economic status: Evidence from india. *Journal of comparative economics 43*(1), 204–226.

Simon, H. A. (1993). Altruism and economics. *The American Economic Review 83*(2), 156–161.

Sinha, G. S. and R. C. Sinha (1967). Exploration in caste stereotypes. *Social Forces 46*(1), 42–47.

Tajfel, H., J. C. Turner, W. G. Austin, and S. Worchel (1979). An integrative theory of intergroup conflict. *Organizational identity: A reader 56*, 65.

Tang, H. (2019). Peer-to-peer lenders versus banks: substitutes or complements? *The Review of Financial Studies 32*(5), 1900–1938.

Tantri, P. (2021). Fintech for the poor: Financial intermediation without discrimination. *Review of Finance 25*(2), 561–593.

Ticku, R. (2015). Riot rewards? study of bjp's electoral performance and hindu muslim riots. Technical report, The Graduate Institute of International and Development Studies.

Vallee, B. and Y. Zeng (2019). Marketplace lending: A new banking paradigm? *The Review of Financial Studies 32*(5), 1939–1982.

Verma, V. (2016). Secularism in india. *The Oxford Handbook of Secularism*, 214.

Ziegfeld, A. (2020). A new dominant party in india? putting the 2019 bjp victory into comparative and historical perspective. *India Review 19*(2), 136–152.

Figure 2: Lending to In-Group vs. Out-Group Borrowers: Extent of Debiasing (Intensive Margin)

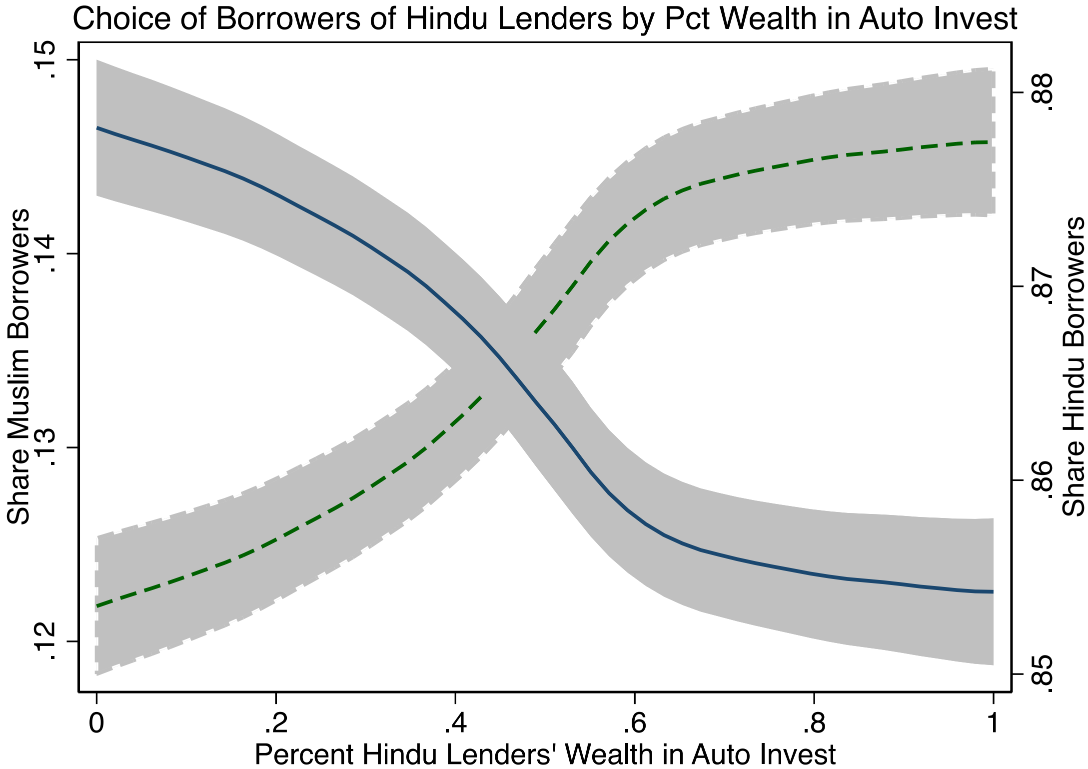


Figure 2 plots the coefficient estimates of kernel-weighted local mean smoothing regressions for whether borrowers are Hindu (blue, solid line, measured on the right y-axis) and whether borrowers are Muslim (green, dashed line, measured on the left y-axis) on the share of their available funds Hindu borrowers who moved to the robo-advising lending tool (Auto Invest) allocated to such tool. This share represents the intensive margin of usage of Auto Invest by Hindu borrowers. Grey bandwidths correspond to 95% confidence intervals around the estimated coefficients. We use an Epanechnikov kernel and evaluate the smooth at 50 points.

Figure 3: Spatial Heterogeneity of In-group vs. Out-group Conflict

Panel A. Hindu-Muslim Riots, 1980-2000

Panel B. Average Vote Shares for the Bharatiya Janata Party (BJP) 1977-2015

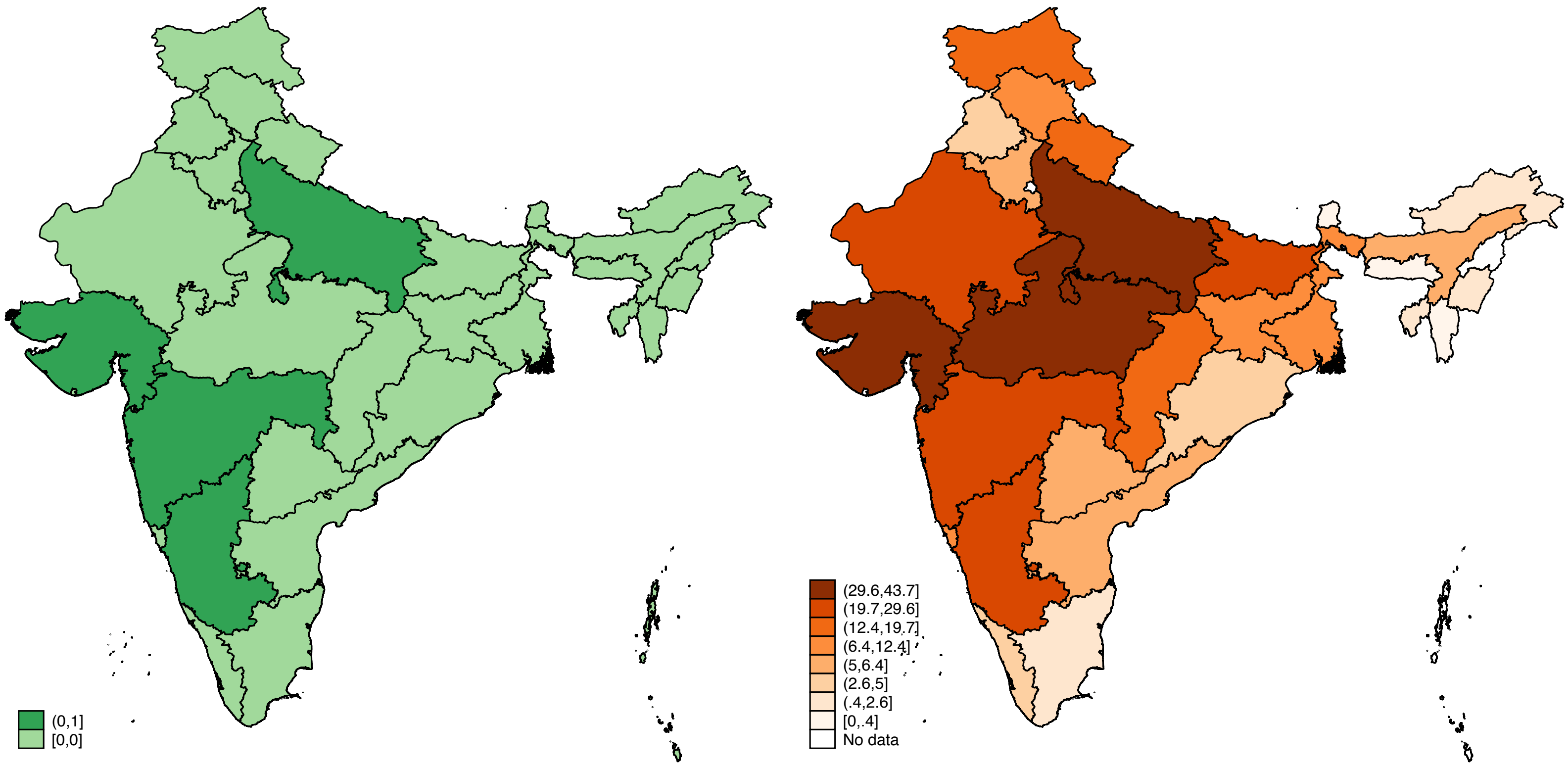
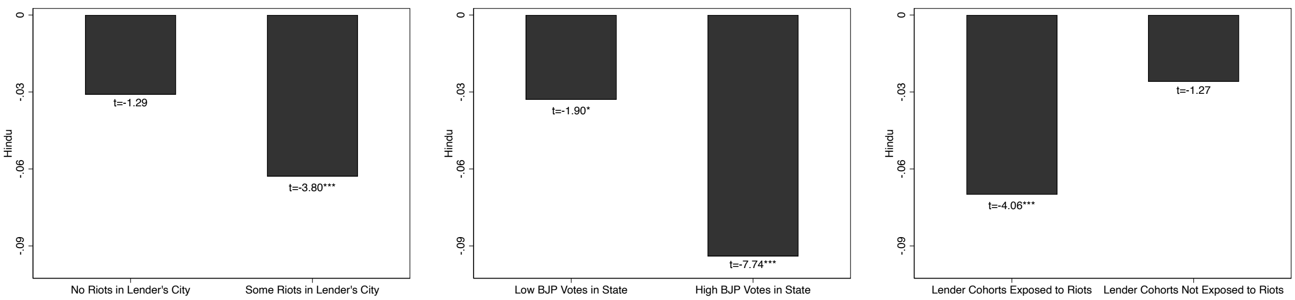


Figure 3 depicts the spatial variation of proxies for the vividness of Hindu-Muslim conflict across Indian states. Panel A compares states in which large-scale riots between Hindus and Muslims and/or pogroms against the Muslim minority happened between 1980 and 2000. Dark green states (Gujarat, Uttar Pradesh, Delhi, Marahashtra, and Karnataka) are states in which such events happened based on Ticku (2015). Panel B compares states based on the average vote share of BJP candidates to national and local elections between 1977 and 2015. We obtain candidate-level election results from 1977 to 2015 from Bhavnani (2014). We first compute the voting shares for each election in each state and then average these shares within states. The darker is a state, the higher is the average BJP vote share.

Figure 4: Change in Lending to Out-group Borrowers: Salience of Hindu-Muslim Conflict

Panel A. Bias Before Auto Invest ($\hat{\gamma}$ Coefficient)



Panel B. De-Biasing After Auto Invest ($\hat{\delta}$ Coefficient)

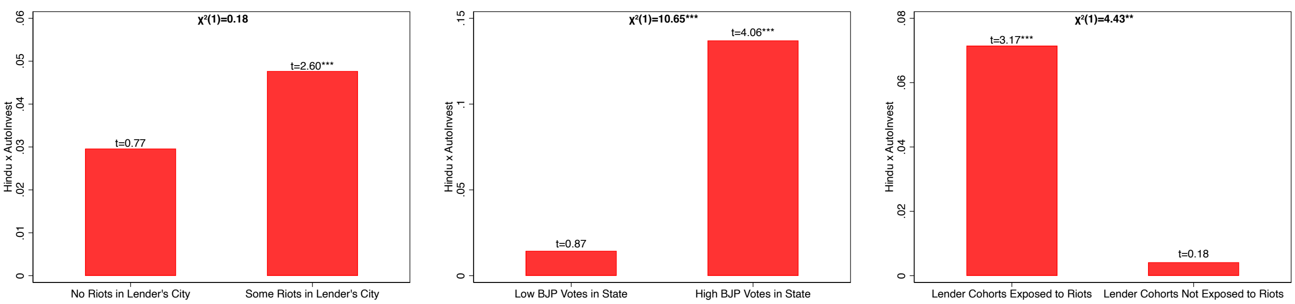


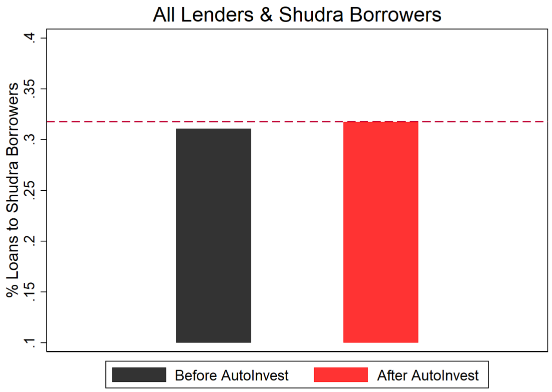
Figure 4 reports the results of estimating the following specification by ordinary least squares across different subsamples reported on top of each column:

$$
\begin{aligned}
Muslim\ Borrower_{i,j} = \alpha + \beta\ Auto\ Invest_j + \gamma\ Hindu\ Lender_j \\
+ \delta\ Hindu\ Lender_j \times Auto\ Invest_j + \zeta\ \boldsymbol{x}_i + \epsilon_{i,j}
\end{aligned}
$$

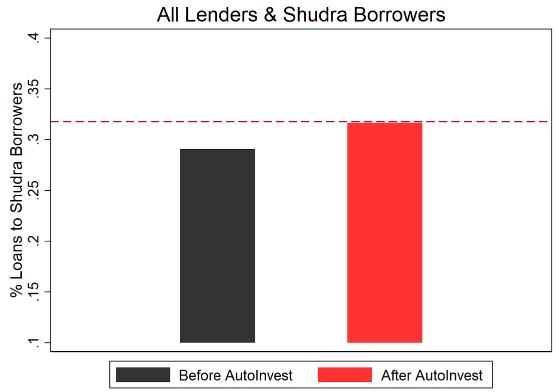
where $Muslim\ Borrower_{i,j}$ is equal to 1 if the borrower i who receives funding from lender j is Muslim, and zero otherwise; $Auto\ Invest_j$ is equal to 1 if the lender made the loans after activating Auto Invest and 0 otherwise; $Hindu\ Lender_j$ is equal to 1 if lender j is Hindu; and $\boldsymbol{x}_i$ is a vector of loan-level characteristics that are direct proxies for the risk profiles of the loans lenders extend to borrowers—loan maturity (measured in months), loan amount, and the annual interest rate associated with the loan. These loan-level characteristics are assigned to borrowers by the platform's algorithm when the loan requests are vetted before borrowers access the borrower pool. We cluster standard errors at the lender level. Panel A plots estimated coefficient $\hat{\gamma}$, which captures the extent of lender bias before accessing Auto Invest. Panel B plots estimated coefficient $\hat{\delta}$, which captures the de-biasing effect of Auto Invest.

Figure 5: Lending to Discriminated Borrowers:
Shudra Caste Borrowers Before and After Debiasing

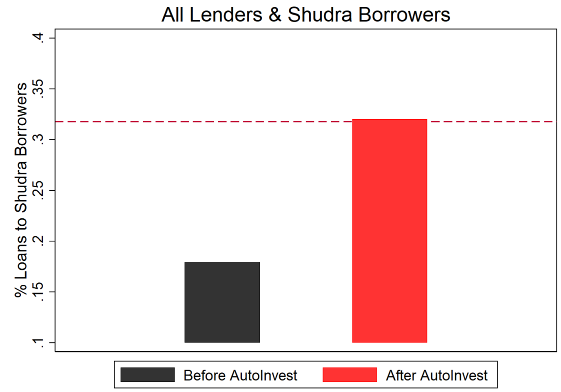
Panel A. All Lenders



(a) Caste Barely Recognizable (Prob≥0%)

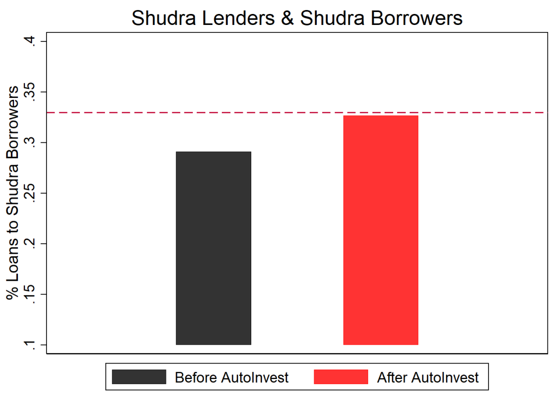


(b) Caste Somewhat Recognizable (Prob>50%)

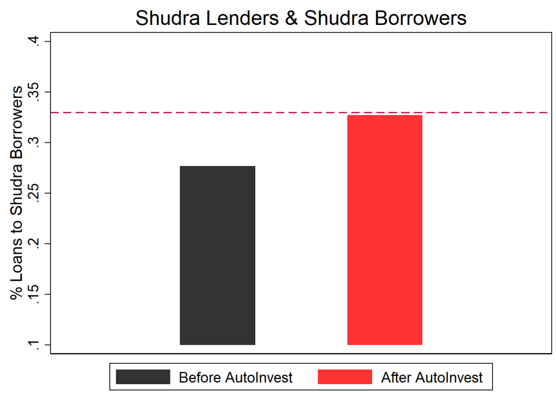


(c) Caste Easily Recognizable (Prob>70%)

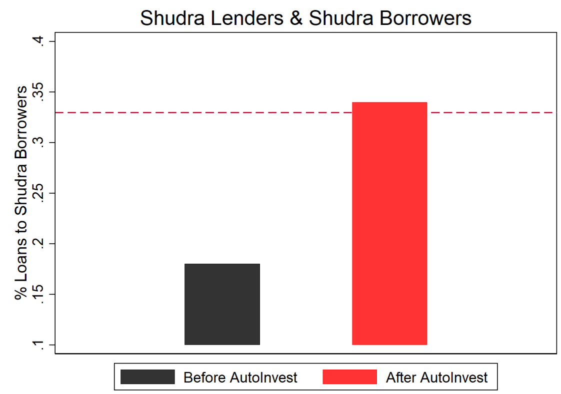
Panel B. Only Shudra Lenders (Easier to Recognize Own Caste)



(d) Caste Barely Recognizable (Prob≥0%)



(e) Caste Somewhat Recognizable (Prob>50%)

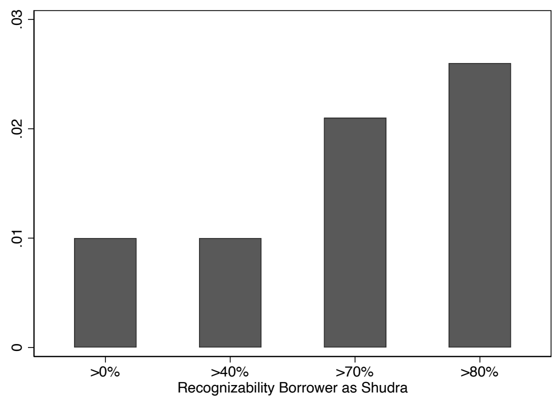


(f) Caste Easily Recognizable (Prob>70%)

Figure 5 plots the average share of borrowers in Hindu lenders' portfolios who are Shudra before lenders moved to the robo-advising tool (Auto Invest, black bars) and after lending decisions are made by Auto Invest (red bars). Panel A consider all Hindu lenders on the platform whereas Panel B only includes Shudra Hindu lenders, for whom recognizing the caste of Shudra borrowers might be weakly easier. In each panel, the left graph considers all borrowers in lenders' portfolios; the middle graph only considers borrowers whose caste can be recognized by a human with a probability above 50% as defined by the algorithm designed by Bhagavatula et al. (2018); the right graph only considers borrowers whose caste can be recognized with a probability above 70% based on the same algorithm.

Figure 6: Change in Lending to Discriminated Borrowers—Shudra Caste Borrowers

Panel A. De-Biasing After AutoInvest by Recognizability of Shudra Borrowers



Panel B. Heterogeneous De-biasing After AutoInvest by Salience of Shudra Discrimination

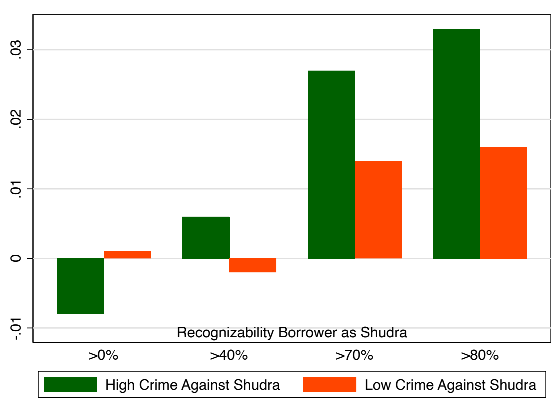


Figure 6 reports the results of estimating the following specification by ordinary least squares:

$$Shudra\ Borrower_{i,j} = \alpha + \beta\ Auto\ Invest_j + \zeta\ \boldsymbol{x}_i + \epsilon_{i,j}$$

where $Shudra\ Borrower_{i,j}$ is equal to 1 if the borrower i who receives funding from lender j is a Shudra, and zero otherwise; $Auto\ Invest_j$ is equal to 1 if the lender made the loans after activating Auto Invest and 0 otherwise; and $\boldsymbol{x}_i$ is a vector of loan-level characteristics that are direct proxies for the risk profiles of the loans lenders extend to borrowers—loan maturity (measured in months), loan amount, and the annual interest rate associated with the loan. These loan-level characteristics are assigned to borrowers by the platform's algorithm when the loan requests are vetted before borrowers access the borrower pool. Panel A reports the $\hat{\beta}$ coefficients across Shudra borrowers with different recognizability. Panel B further divides the estimates across states with high and low crimes against Shudra. We cluster standard errors at the lender level.

Figure 7: Spatial Heterogeneity of Salience of Stereotypical Discrimination

Crimes Against Scheduled Castes per inhabitant (2018)

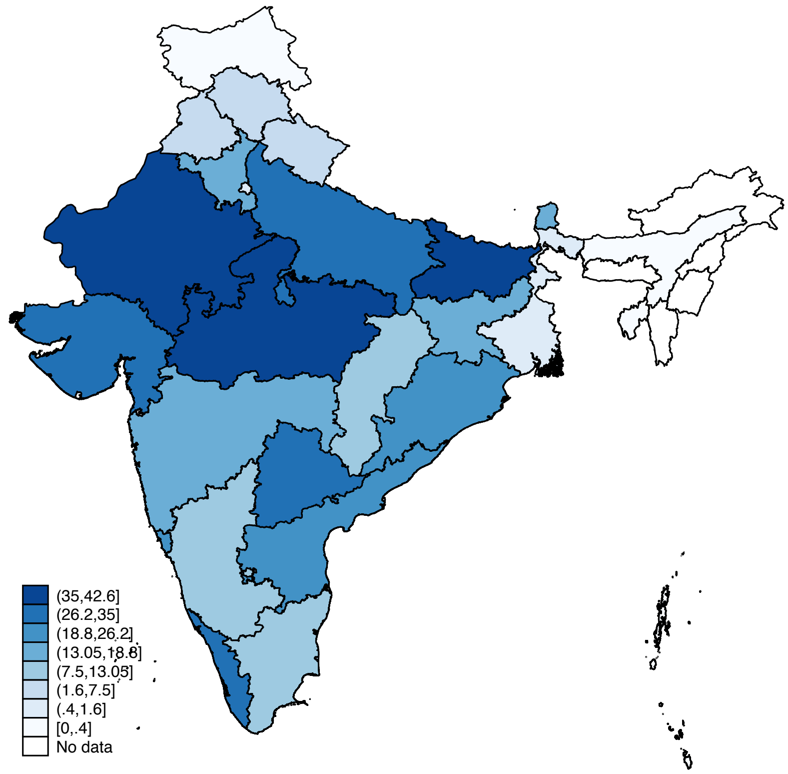
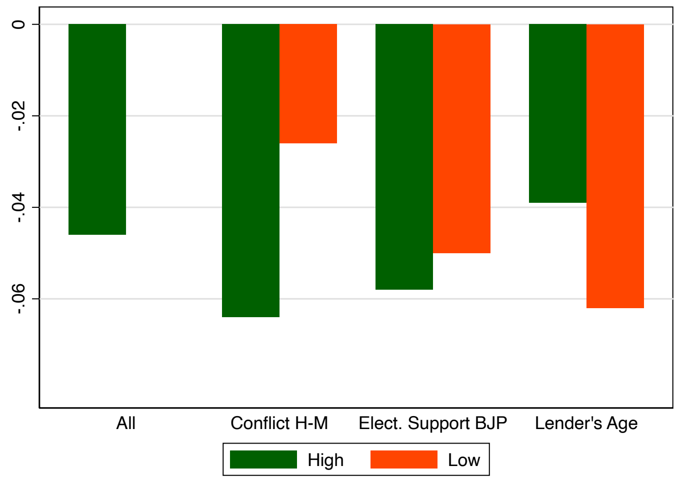


Figure 7 depicts the spatial variation of a proxy for the salience of discrimination against lower castes by Hindus across Indian states, that is, the number of crimes against Scheduled Castes (which includes members of the Shudra varna as well as those belonging to lower castes) per 100,000 inhabitants in 2018 based on the official data from the Indian National Crime Records Bureau (NCRB (2019)). The darker is a state, the higher is the number of crimes against Schedules Classes per inhabitant in the state.

Figure 8: Lower Default of Discriminated Borrowers Relative to Others Before Auto Invest

Panel A. In-group vs. Out-group Discrimination



Panel B. Stereotypical Discrimination

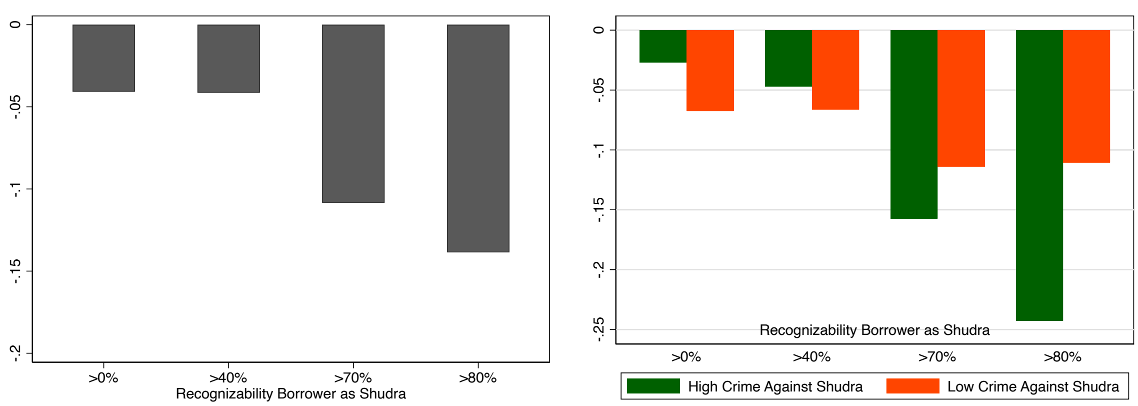
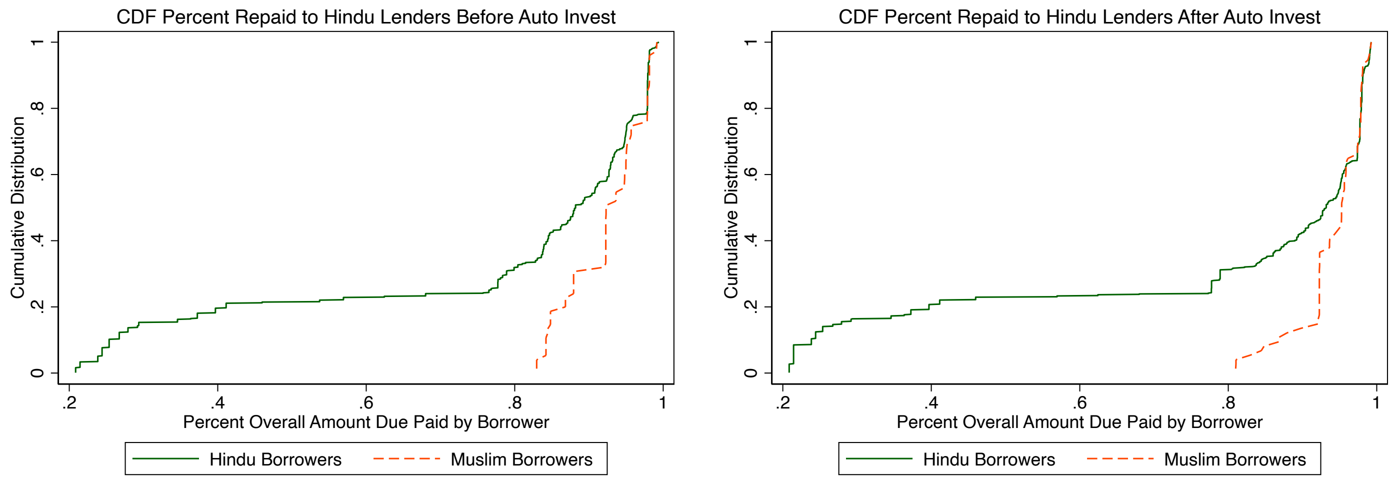


Figure 8 plots the relative default of Muslim borrowers relative to Hindu borrowers in Hindu lenders' portfolios (Panel A) and of Shudra borrowers relative to other Hindu borrowers in all Hindu lenders' portfolios before lenders moved to the robo-advising tool (Auto Invest) and across different subsamples. Panel A includes borrowers in Hindu lenders' portfolios. In Panel B, the probability of caste recognition of Shudra borrowers is based on the algorithm developed by Bhagavatula et al. (2018).

Figure 9: Fraction of Loan Repaid Before and After Debiasing

Panel A. In-group vs. Out-group Discrimination



Panel B. Stereotypical Discrimination

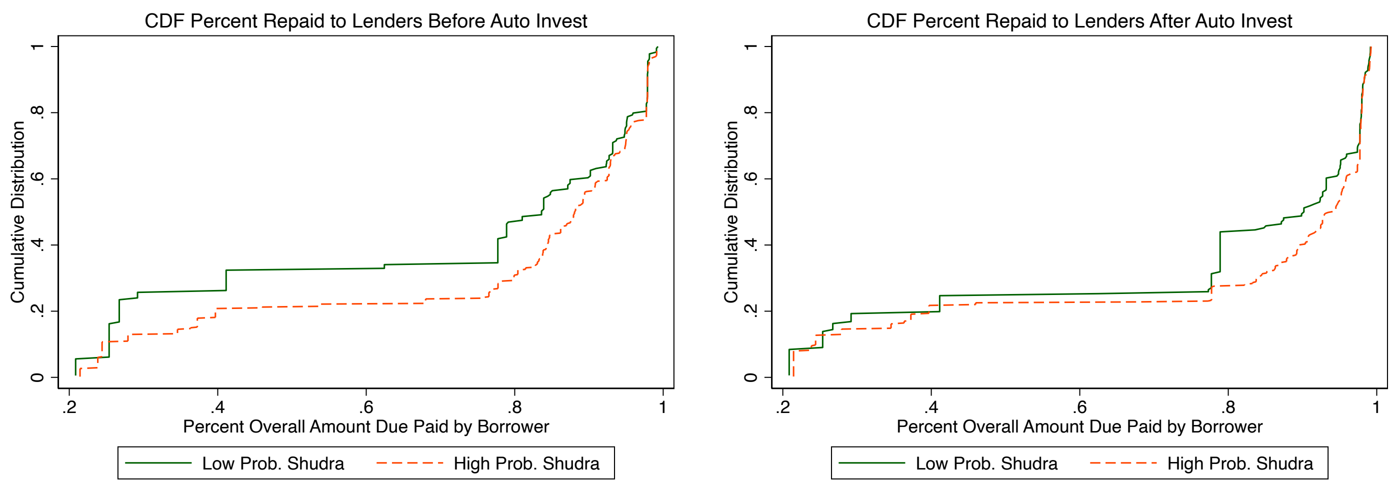
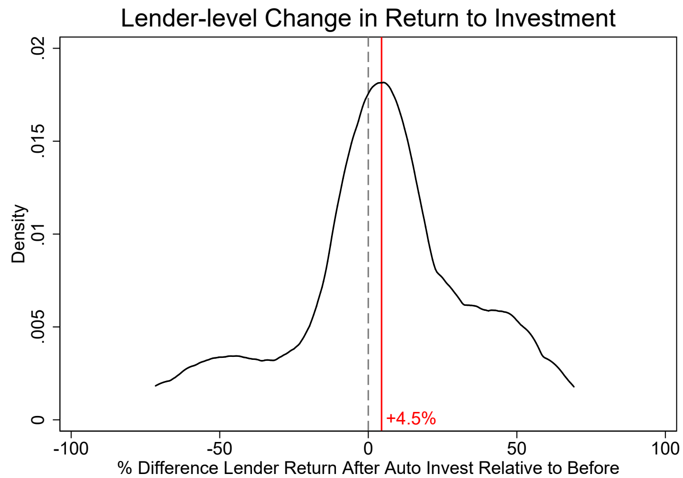


Figure 9 plots a set of cumulative distribution functions (CDFs) for different groups of borrowers over the share of the overall amounts due (principal plus interest) that borrowers repaid by the time their loan account was closed. In all Panels, the left graph refers to the CDFs of borrowers in lenders' portfolios before moving to the robo-advising tool (Auto Invest), whereas the right graph refers to the CDFs after moving to Auto Invest. Panel A includes borrowers in Hindu lenders' portfolios. The green solid lines are the CDFs for Hindu borrowers and the orange dashed lines for Muslim borrowers. Panel B only includes Shudra borrowers in Hindu lenders' portfolios. The green solid lines are the CDFs for Shudra borrowers whose probability of caste recognition is below 15% based on the algorithm developed by Bhagavatula et al. (2018), and the orange dashed lines for other Shudra borrowers.

Figure 10: Lender-level Change in Returns After Debiasing

Panel A. In-group vs. Out-group Discrimination



Panel B. Stereotypical Discrimination

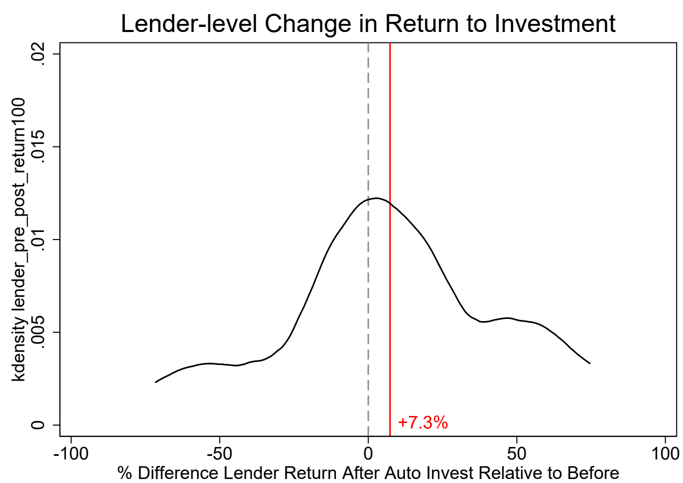


Figure 10 plots the density of the lender-level average change in the returns to loans originated on the platform after using Auto Invest relative to before. To compute the change, we first compute the overall return to the amounts invested by each lender separately before and after using Auto Invest across all loans they originated. We then subtract the two returns and average these changes across lenders. Loan-level returns are thus value weighted within lenders and lender-level returns are equally weighted across lenders. Panel A refers to Hindu lenders' loans to Hindu borrowers (in-group vs. out-group discrimination), whereas Panel B refers to all Hindu lenders' loans to Shudra borrowers (stereotypical discrimination).

Table 1. Summary Statistics

Panel A. In-group vs. Out-group Discrimination Sample

	N. obs.	Mean	St. dev.	25*th* perc.	Median	75*th* perc.
Muslim Borrower	113,283	0.13	0.34	0.00	0.00	0.00
Hindu Lender	113,283	0.99	0.11	1.00	1.00	1.00
Auto Invest	113,283	0.45	0.50	0.00	0.00	1.00
Auto Invest allocation (%)	113,283	0.59	0.37	0.22	0.60	1.00
Tenure (months)	113,283	22.08	8.98	15.00	24.00	24.00
Loan Amount (rupees)	113,283	131,074	102,575	50,000	100,000	188,000
Interest Rate	113,283	0.24	0.07	0.20	0.23	0.27
Delinquent	113,127	0.29	0.45	0.00	0.00	1.00

Panel B. Stereotypical Discrimination Sample

	N. obs.	Mean	St. dev.	25*th* perc.	Median	75*th* perc.
Shudra Borrower	62,831	0.39	0.49	0.00	0.00	1.00
Auto Invest	62,831	0.43	0.49	0.00	0.00	1.00
Auto Invest allocation (%)	62,831	0.58	0.37	0.22	0.57	1.00
Tenure (months)	62,831	22.03	9.02	12.00	24.00	30.00
Loan Amount (rupees)	62,831	131,797	105,994	50,000	100,000	200,000
Interest Rate	62,831	0.24	0.07	0.20	0.23	0.26
Delinquent	62,736	0.28	0.45	0.00	0.00	1.00

Table 1 reports summary statistics for the main variables in the analysis across the two datasets used in the analysis of in-group vs. out-group discrimination (Panel A) and stereotypical discrimination (Panel B). In both panels, the unit of observation is a lender-borrower-loan triad. Borrower-lender characteristics include the religion/caste of borrowers and lenders. *Auto Invest* is a dummy variable that equals 1 if the lender uses the robo-advising lending tool, whereas *Auto Invest allocation* is the share of funds lenders have available on the P2P platform that they allocate to the robo-advising tool. Loan-level characteristics include the loans' tenure, size, and annual interest rate, as well as a dummy variable that equals 1 if the loan was delinquent at the time it was closed and zero otherwise.

Table 2. Change in Lending to Out-group Borrowers: Hindu vs. Muslim

Dependent variable: Muslim Borrower					Low Use Auto Invest	High Use Auto Invest
	(1)	(2)	(3)	(4)	(5)	(6)
Hindu Lender × Auto Invest	0.045**	0.046**	0.045**	0.043**	0.009	0.052*
	(2.51)	(2.51)	(2.07)	(1.96)	(0.23)	(1.94)
Hindu Lender	-0.058***	-0.058***				
	(-3.52)	(-3.54)				
Auto Invest	-0.026	-0.025	-0.030	-0.033	0.011	-0.048*
	(-1.45)	(-1.40)	(-1.43)	(-1.53)	(0.29)	(-1.83)
Maturity		0.012***	0.009***	0.010***	0.010**	0.009**
		(4.78)	(3.20)	(3.35)	(2.31)	(2.31)
Loan Amount (₹000)		-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
		(-10.34)	(-10.81)	(-10.92)	(-5.47)	(-9.27)
Interest Rate		-0.009	-0.015	-0.012	0.013	-0.035
		(-0.47)	(-0.74)	(-0.62)	(0.47)	(-1.25)
Constant	0.181***	0.163***	0.119***	0.120***	0.099***	0.137***
	(11.14)	(8.74)	(11.72)	(11.77)	(6.77)	(9.60)
Lender Fixed Effects			✓	✓	✓	✓
Year Fixed Effects				✓	✓	✓
N. obs.	113,284	113,283	113,283	113,283	39,366	72,104

Table 2 reports the results of estimating the following specification by ordinary least squares:

$$\begin{aligned} Muslim\ Borrower_{i,j} &= \alpha + \beta\ Auto\ Invest_j + \gamma\ Hindu\ Lender_j \\ &+ \delta\ Hindu\ Lender_j \times Auto\ Invest_j + \zeta\ \boldsymbol{x}_i + \epsilon_{i,j} \end{aligned}$$

where $Muslim\ Borrower_{i,j}$ is equal to 1 if the borrower i who receives funding from lender j is Muslim, and zero otherwise; $Auto\ Invest_j$ is equal to 1 if the lender made the loans after activating Auto Invest and 0 otherwise; $Hindu\ Lender_j$ is equal to 1 if lender j is Hindu; and $\boldsymbol{x}_i$ is a vector of loan-level characteristics that are direct proxies for the risk profiles of the loans lenders extend to borrowers—loan maturity (measured in months), loan amount, and the annual interest rate associated with the loan. These loan-level characteristics are assigned to borrowers by the platform's algorithm when the loan requests are vetted before borrowers access the borrower pool. We cluster standard errors at the lender level.

Table 3. Loan Defaults Before and After Debiasing

Dependent variable:	**In-group vs. Out-group Discrimination**			**Stereotypical Discrimination**		
Delinquent Loan	Hindu Lenders			All Lenders		
	(1)	(2)	(3)	(4)	(5)	(6)
Auto Invest	-0.108***			-0.157***		
	(-5.08)			(-6.25)		
Hindu Borrower× Auto Invest		-0.112***	-0.072***			
		(-5.21)	(-4.55)			
Muslim Borrower× Auto Invest		-0.073**	-0.070***			
		(-2.49)	(-2.82)			
Muslim Borrower	-0.024**	-0.046**	-0.002			
	(-2.02)	(-2.58)	(-0.10)			
Non-Shudra Borrower× Auto Invest					-0.160***	-0.100***
					(-6.04)	(-4.39)
Shudra Borrower× Auto Invest					-0.148***	-0.104***
					(-4.54)	(-3.32)
Shudra Borrower				-0.038***	-0.041***	-0.030**
				(-3.03)	(-2.89)	(-2.14)
Maturity			-0.025***			-0.020
			(-2.72)			(-1.26)
Loan Amount (₹000)			-0.002***			-0.002***
			(-4.07)			(-4.19)
Interest Rate			0.492***			0.479***
			(7.79)			(5.16)
Constant	0.499***	0.501***	0.403***	0.490***	0.491***	0.421***
	(30.95)	(30.84)	(17.69)	(32.13)	(32.25)	(7.77)
Lender Fixed Effects			✓			✓
Time Fixed Effects			✓			✓
N. obs.	16,985	16,985	16,985	6,821	6,821	6,821
R-Square	0.012	0.012	0.263	0.020	0.020	0.357

Table 3 reports the results of estimating variations of the following specification by ordinary least squares:

$$\begin{aligned} Delinquent\ Loan_{ij} &= \alpha + \beta\ Auto\ Invest_j + \gamma\ Muslim\ Borrower_j \\ &+ \delta\ Muslim\ Borrower_j \times Auto\ Invest_j \\ &+ \theta\ Hindu\ Borrower_j \times Auto\ Invest_j + \zeta\ \boldsymbol{x}_i + \epsilon_{i,j}, \end{aligned}$$

where $Delinquent\ Loan_{ij}$ is equal to 1 if the loan associated with borrower i and lender j is closed as delinquent; $Auto\ Invest_j$ is equal to 1 if the lender made the loans after activating Auto Invest and 0 otherwise; $Muslim\ Borrower_{i,j}$ ($Hindu\ Borrower_{i,j}$) is equal to 1 if the borrower i who receives funding from lender j is Muslim (Hindu), and zero otherwise; and $\boldsymbol{x}_i$ is a vector of loan-level characteristics that are direct proxies for the risk profiles of the loans lenders extend to borrowers—loan maturity (measured in months), loan amount, and the annual interest rate associated with the loan. These loan-level characteristics are assigned to borrowers by the platform's algorithm when the loan requests are vetted before borrowers access the borrower pool. We cluster standard errors at the lender level.

Table 4. Loan Returns Before and After Debiasing: Quantile Regressions

Panel A. In-group vs. Out-group Discrimination

Dependent variable:	**OLS**		**25th** *percentile*		**Median**		**75th** *percentile*	
Loan's Return	Hindu (1)	Muslim (2)	Hindu (3)	Muslim (4)	Hindu (5)	Muslim (6)	Hindu (7)	Muslim (8)
Auto-Invest	0.217***	0.065**	0.250*	0.173**	0.274***	0.121***	0.109***	-0.004
(Without Risk Controls)	(4.92)	(2.25)	(1.82)	(2.23)	(11.22)	(3.84)	(11.66)	(-0.29)
Auto-Invest	-0.063**	-0.006	-0.008	-0.002	-0.035	-0.000	0.000	0.008
(With Risk Controls)	(-2.20)	(-0.30)	(-1.01)	(-0.38)	(-0.04)	(-0.04)	(0.02)	(1.04)
N. obs.	2,326	220	2,326	220	2,326	220	2,326	220

Panel B. Stereotypical Discrimination

Dependent variable:	**OLS**		**25th** *percentile*		**Median**		**75th** *percentile*	
Loan's Return	Low Prob Shudra (1)	High Prob Shudra (2)	Low Prob Shudra (3)	High Prob Shudra (4)	Low Prob Shudra (5)	High Prob Shudra (6)	Low Prob Shudra (7)	High Prob Shudra (8)
Auto-Invest	0.079***	0.068***	0.485***	0.080*	0.088***	0.072***	0.030***	0.025***
(Without Risk Controls)	(2.89)	(4.54)	(4.61)	(1.73)	(3.02)	(8.13)	(5.64)	(9.29)
Auto-Invest	-0.053	-0.039	-0.001	0.000	-0.011	0.014**	0.020	0.020***
(With Risk Controls)	(-0.82)	(-1.04)	(-0.07)	(0.05)	(-0.55)	(1.99)	(0.55)	(4.83)
N. obs.	462	1,158	462	1,158	462	1,158	462	1,158

Table 4 reports the results of estimating the following set of quantile regressions:

$$Q_\tau(Loan's\ Return_{i,j,t}) = \alpha(\tau) + \beta(\tau)\ Auto\ Invest_{j,t} +\ X'_{i,j,t}\ \zeta(\tau) + \epsilon_{i,j,t},$$

whose outcome variable is quantile Q_τ of the distribution of the (standardized) loan return associated with borrower i and lender j throughout the sample period; $Auto\ Invest_j$ is equal to 1 if the lender made the loans after activating Auto Invest and 0 otherwise; and $X_{i,j,t}$ is a vector of loan-level characteristics that are direct proxies for the risk profiles of the loans lenders extend to borrowers—loan amount, and the annual interest rate associated with the loan. These loan-level characteristics are assigned to borrowers by the platform's algorithm when the loan requests are vetted before borrowers access the borrower pool. In each panel, the first row reports the estimates of $\beta(\tau)$ *without* controlling for loan's risk characteristics. The second row reports the estimates of the same specifications when risk controls are included. We cluster standard errors at the lender level.

Online Appendix:

How Costly Are Cultural Biases? Evidence from FinTech

Francesco D'Acunto, Pulak Ghosh, Rajiv Jain, Alberto G. Rossi

Not for Publication

Figure A.1: Platform's Screening of Borrowers 1: Credit Scores of Approved and Rejected Borrowers

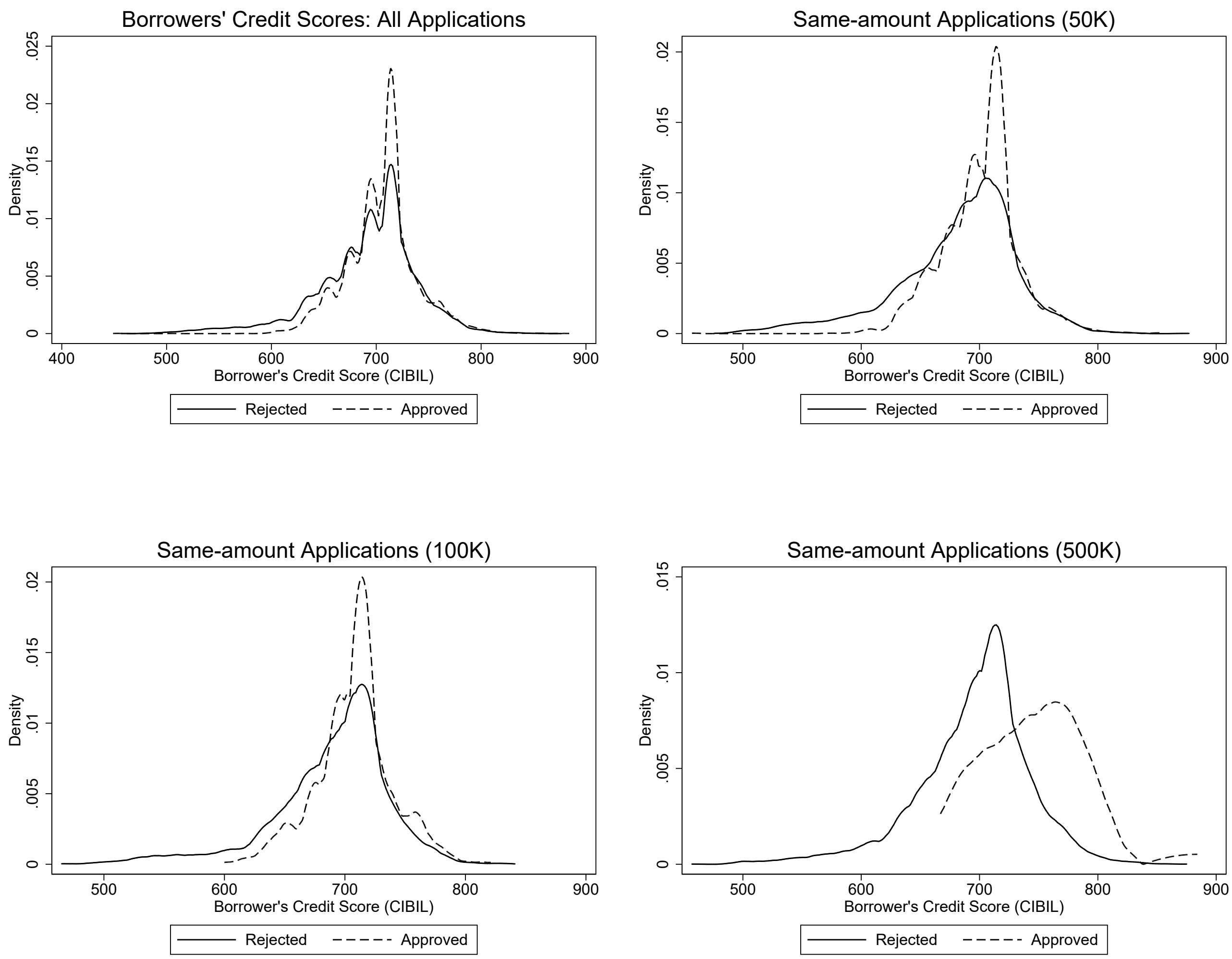


Figure A.2: Platform's Screening of Borrowers 2:
Interest Rates, Maturities, and Loan Amounts by Credit Score

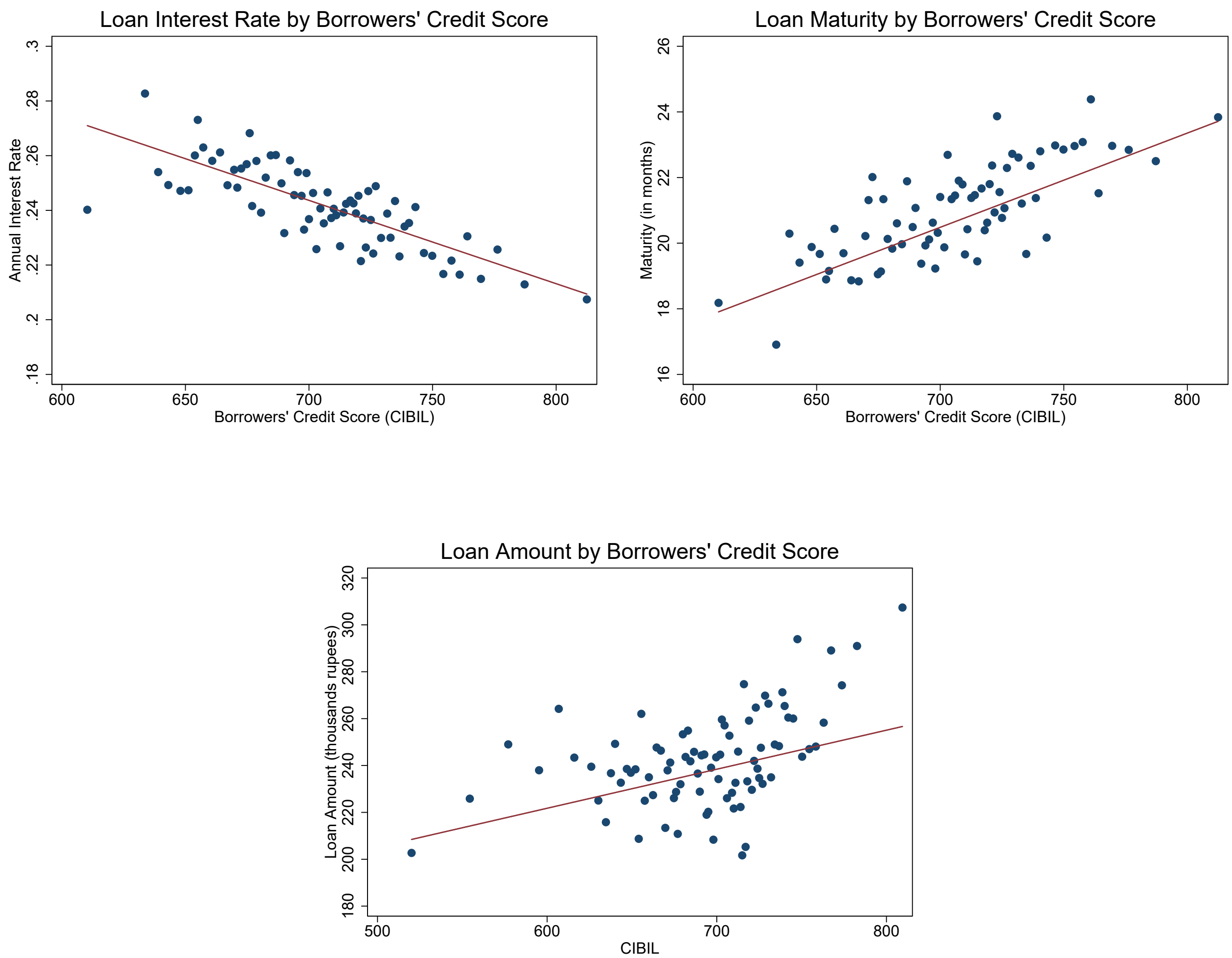


**Figure A.3: Geographic Distribution of Lending:
Number of Indian States in which Each Lender Disburses Funds**

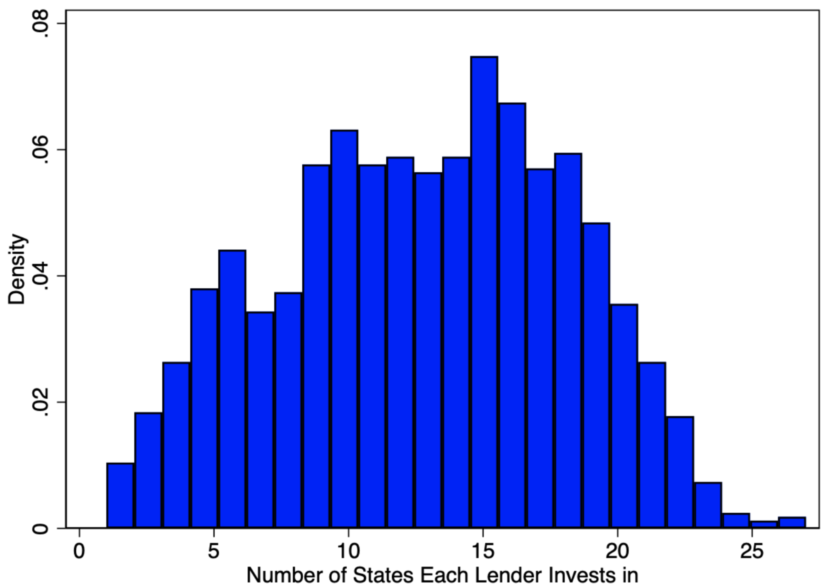


Figure A.4: Ex-post Likelihood of Default and Interest Rates

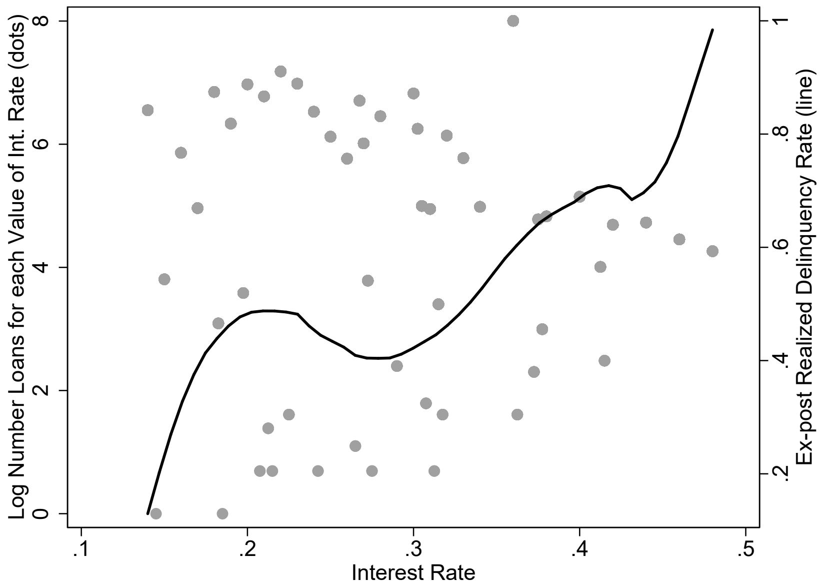


Figure A.5: Distribution of Probabilities that Borrowers are Shudra

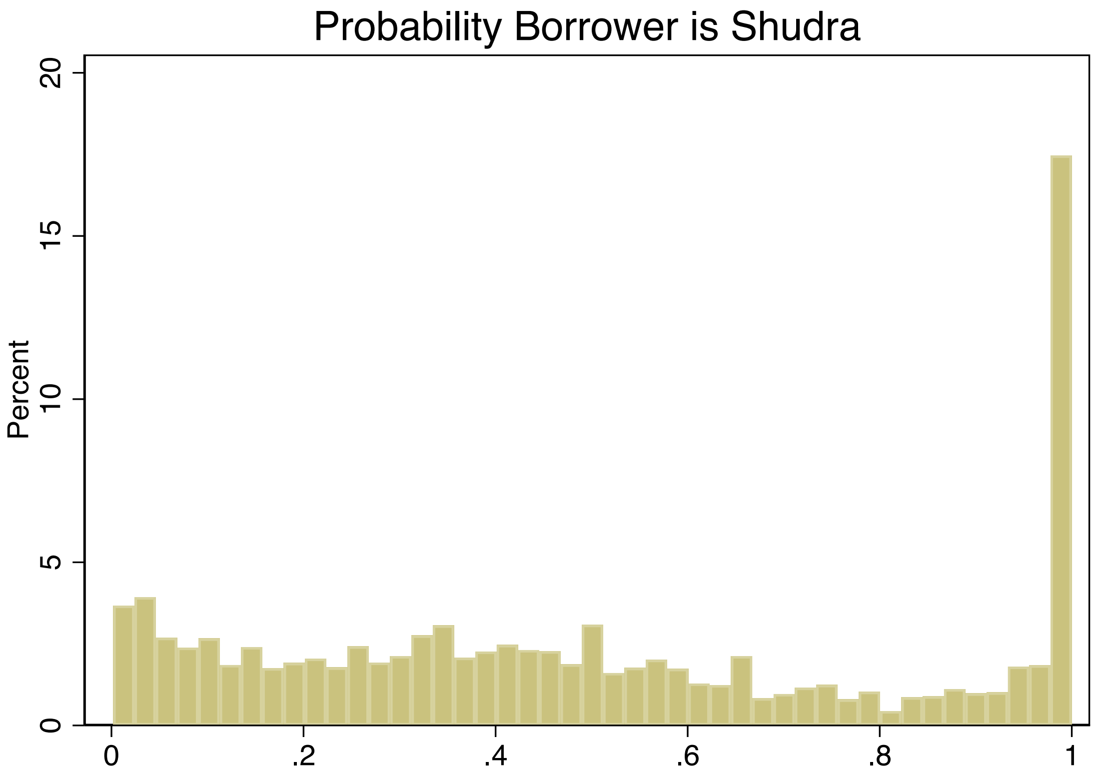


Figure A.6: Intensive Margin Performance Before and After Auto Invest: Full Sample

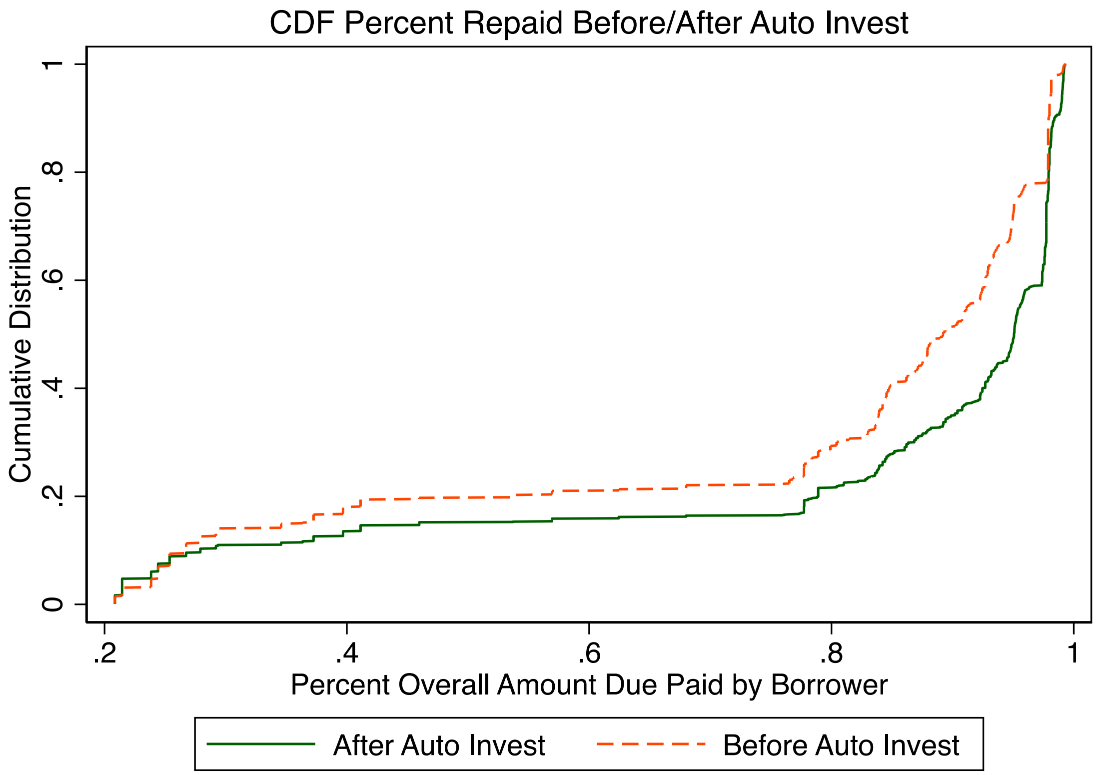


Table A.1. Change in Lending to Out-group Borrowers: Robustness, Statistical Inference

Dependent variable: Muslim Borrower	(1)	(2)	(3)	(4)	Low Use Auto Invest (5)	High Use Auto Invest (6)
Hindu Lender × Auto Invest	0.045	0.046	0.045	0.043	0.009	0.052
By Lender	(2.51)**	(2.51)**	(2.07)**	(1.96)**	(0.23)	(1.94)*
By Lender and Borrower	(2.24)**	(2.24)**	(2.00)**	(1.90)*	(0.23)	(1.93)*
By Lender, Borrower, and Month	(2.13)**	(2.14)**	(2.51)**	(2.19)**	(0.21)	(2.71)***
By Lender Surname, Borrower Surname, and Month	(2.16)**	(2.18)**	(2.01)**	(2.03)**	(0.22)	(1.94)*
Lender Fixed Effects			✓	✓	✓	✓
Year Fixed Effects				✓	✓	✓
N. obs.	113,284	113,283	113,283	113,283	39,366	72,104

Table A.1 reports the results of estimating the following specification by ordinary least squares:

$$\begin{aligned} Muslim\ Borrower_{i,j} &= \alpha + \beta\ Auto\ Invest_j + \gamma\ Hindu\ Lender_j \\ &+ \delta\ Hindu\ Lender_j \times Auto\ Invest_j + \zeta\ \boldsymbol{x}_i + \epsilon_{i,j} \end{aligned}$$

where $Muslim\ Borrower_{i,j}$ is equal to 1 if the borrower i who receives funding from lender j is Muslim, and zero otherwise; $Auto\ Invest_j$ is equal to 1 if the lender made the loans after activating Auto Invest and 0 otherwise; $Hindu\ Lender_j$ is equal to 1 if lender j is Hindu; and $\boldsymbol{x}_i$ is a vector of loan-level characteristics that are direct proxies for the risk profiles of the loans lenders extend to borrowers—loan maturity (measured in months), loan amount, and the annual interest rate associated with the loan. These loan-level characteristics are assigned to borrowers by the platform's algorithm when the loan requests are vetted before borrowers access the borrower pool. In each line, we report the t-statistics estimated with the indicated level of clustering of standard errors.