View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by LBS Research Online

LBS Research Online

B Stroube

Economic consequences and the motive to discriminate

Article

This version is available in the LBS Research Online repository: http://lbsresearch.london.edu/id/eprint/1739/

Stroube, B

(2021)

Economic consequences and the motive to discriminate.

Administrative Science Quarterly.

ISSN 0001-8392

(Accepted)

SAGE Publications (UK and US)

Users may download and/or print one copy of any article(s) in LBS Research Online for purposes of research and/or private study. Further distribution of the material, or use for any commercial gain, is not permitted.

Economic consequences and the motive to discriminate

Bryan K. Stroube
bstroube@london.edu
London Business School
Regent's Park, London, NW1 4SA, United Kingdom

Last updated: March 23, 2021

Economic consequences and the motive to discriminate

ABSTRACT

Past research indicates that increasing the economic consequences of evaluations should theoretically discourage discrimination by making it more costly. In this paper I theorize that such consequences should also encourage discrimination in settings where evaluators may be motivated by performance expectations (e.g., stereotypes). I explore this theory using data from an online lending platform where a loan guarantee policy reduced the potential economic consequences of using borrower demographics during lending decisions. I find evidence that lenders evaluated female borrowers less favorably than male borrowers after the policy. This is consistent with the theory that the policy discouraged performance-motivated discrimination, while simultaneously encouraged consumption-motivated discrimination. Because I theorize about underlying motives for discrimination, the insights developed here should apply to a wide range of specific *types* of discrimination that vary according to these motives, including classic taste-based discrimination, homophily-driven discrimination, statistical discrimination, and status-based discrimination. Economic consequences may therefore represent an important dynamic link between different types of discrimination.

INTRODUCTION

Effective responses to discrimination and inequality remain elusive, and policies designed to directly address peoples' biases have generally been found ineffective (Dobbin and Kalev, 2017; Kalev, Dobbin, and Kelly, 2006; Paluck and Green, 2009). This lack of effectiveness can be particularly striking, such as the evidence that attempts to promote a meritocratic organizational culture may paradoxically increase unequal evaluations within an organization (Castilla and Benard, 2010). Economic consequences, however, appear a fruitful starting point for discouraging discrimination. This is because many types of discrimination are essentially forms of costly consumption (e.g., Becker, 1957). If the economic consequences of discriminating are high enough, then overall discrimination should go down.

This logic, however, overlooks the full range of motives that can trigger discrimination. In this paper I develop an argument for why economic consequences may not only be ineffective, but should even *encourage* discrimination in many situations. This is because evaluators are often motivated to discriminate based on performance expectations, such as what type of person they believe will be a good employee or likely to repay a loan. Consequently, increasing the economic consequences of such evaluations should encourage the use of performance stereotypes. This dynamic relationship between different types of discrimination has been largely overlooked in the literature, but has potentially important implications for understanding what drives overall levels of discrimination.

In this paper I develop the logic for why economic consequences should theoretically suppress forms of *consumption-motivated* discrimination but encourage forms of *performance-motivated* discrimination. For example, if evaluators really believe that men and women are different quality on average, then they should discriminate more, not less, when the consequence of choosing between the two genders is increased. This is important because a range of environmental characteristics at both the market and organizational level may encourage or discourage discrimination, including performance pay systems, insurance programs, and other mechanisms that alter the consequences of decisions for the people making them (Gibbons, 1998). When decisions involve the evaluation of other people, these economic consequences could theoretically motivate discriminatory behavior. As the economic consequences of such evaluations increase, discrimination motivated by performance stereotypes should rise, but discrimination motivated by consumption preferences should fall. As the potential consequences decrease, discrimination motivated by performance stereotypes should fall, but discrimination motivated by consumption preferences should rise.

I apply this theory to empirically investigate how lenders from an online peer-to-peer lending platform evaluated borrowers based on their gender. Existing research in the context indicates that women may be perceived as more reliable borrowers—that is, lenders hold positive stereotypes about women's quality—yet still be penalized by various society-level taste preferences (Armendariz and Morduch, 2010). To examine how economic consequences may encourage the expression of these during the evaluation of borrowers, I leverage the historical implementation of a loan guarantee policy on the platform—what amounted to an insurance policy for lenders—which should reduce the perceived economic consequences of lending decisions. This should theoretically motivate less discrimination driven by performance expectations, because it makes underlying beliefs about performance ability less valuable. However, it should theoretically motivate more discrimination driven by tastes, which can now be expressed with less consequence. As a result, in this context the overall gap in how men and women are evaluated should theoretically widen in favor of men as a result of the policy.

I compare 25,440 lending decisions (i.e., the amount a lender provided to a borrower) from before and after the policy to measure this change in how gender is evaluated. I find evidence that the average size of these decisions in the pre-policy period was larger to female borrowers than male borrowers, but that this effect flipped in the post-policy period. Thus, lowering the economic consequences of decisions in this market appears to have resulted in women being evaluated less favorably. However, the evidence that a fixed set of lenders changed their behavior across periods is less compelling than the evidence that the effect was between-lenders. This indicates that, similar to studies of discrimination in competitive markets, the dynamic between these two motives can function across evaluators.

By acknowledging an underlying typology of the motives that drive discriminatory behavior, this paper draws attention to a more complex dynamic between economic consequences and discrimination: although altering consequences may discourage one type of discrimination, it may encourage another. Therefore, the paper may provide guidance on how economic consequences help or hinder current approaches to anti-discrimination and diversity programs (e.g., Kalev, Dobbin, and Kelly, 2006; Paluck and Green, 2009). The study also contributes to the broader literature on evaluation processes—both the research specifically focused on discrimination (e.g., Botelho and Abraham, 2017) and the research focused on why audiences are motivated to respond to other markers of status (e.g., Malter, 2014; Simcoe and Waguespack, 2010; Kovács and Sharkey, 2014)—where one might expect economic consequences to also influence evaluations.

THE MOTIVE TO DISCRIMINATE

Discrimination is defined as the unequal treatment of otherwise equal individuals based on an observable characteristic such as race, gender, class, or other demographic trait. Pager and Shepherd (2008: 182) noted that "A key feature of any definition of discrimination is its focus on behavior... the definition of discrimination does not presume any unique underlying cause". Because discrimination is a behavior (unequal treatment), there are many potential causes of such behavior. As a result, a proliferation of theories has been proposed to describe and explain the wide range of discrimination that has been observed in practice.[1]

Each discrimination theory focuses on evaluators who make judgments about people. In the organizational literature, these evaluators include both internal managers as well as external resource providers. Examples include how discrimination may be caused by people responsible for employee hiring (Fernandez-Mateo and Fernandez, 2016; Fernandez and Greenberg, 2013), by managers responsible for evaluating employee performance and rewards (Castilla, 2008 2011), by external funders of entrepreneurs (Greenberg and Mollick, 2017; Thébaud and Sharkey, 2016; Brooks et al., 2014), or by upper management (Dahl, Dezső, and Ross, 2012; Carnahan and Greenwood, 2018).

Each discrimination theory also rests on an underlying motive for why these evaluators would want to discriminate. These motives take one of two general forms. The first is a consumption motive: discrimination for the sake of discriminating. The second is a performance motive: discrimination in the pursuit of a performance goal. The advantage of conceptualizing motives for discrimination at this level is that it allows one to abstract away from some of the theoretical baggage associated with specific theories of discrimination. This will ultimately help facilitate predictions that apply to a broader range of discrimination theory.

Consumption-motivated discrimination is driven by a direct like or dislike of others because of their demographic traits. The most general version of this motive is found in

[1] Like other work, this paper focuses on discrimination as a demand-side explanation of inequality that can arise during evaluation processes. Discrimination is a sufficient but not necessary condition for inequality. Supply-side processes can also lead to inequality (e.g., Thébaud, 2010).

theories of taste-based discrimination, where the perceived cost of a decision is shifted from p to $p(1 + d_k)$ based on how someone feels about specific demographic traits such as race or gender (Becker, 1957). Given a person's discrimination coefficient, d_k, he or she faces perceived prices that are either higher or lower than the prices faced by actors that do not hold such tastes. Thus, these types of discriminators are willing to pay to discriminate.

Specific versions of consumption-motivated discrimination include nepotism (Bennedsen et al., 2007; Goldberg, 1982), which leads to a more favorable evaluation of someone based solely on a personal or kinship tie. More recent theories that explain where and why this consumption motive exists include the theory of activist choice homophily developed by Greenberg and Mollick (2017), where women have a preference to help women in underrepresented categories succeed (i.e., "help someone penetrate barriers she can sympathize or empathize with", p. 343). A taste for demographic traits is also the underlying premise of other forms of homophily-driven discrimination, and as McPherson, Smith-Lovin, and Cook (2001: 416) noted, has been theorized since ancient times: "we love those who are like ourselves" (Aristotle, 1934: 453). While these theories differ in specific cause, consequence, and context, they are unified by their reliance on a consumption motive to explain behavior: evaluators treat people differently because of a direct like or dislike for a demographic trait itself. This consumption-motivated discrimination stands in contrast to the second basic motive for discrimination, where evaluators discriminate because of performance expectations.

Performance-motivated discrimination is premised on explicit or implicit stereotypes about someone's level of competence or performance ability. When evaluators want to make "good" decisions—for example, hire competent employees, provide bonuses to the best workers, or make loans that will be repaid—then they may discriminate if demographic traits influence their perceptions of quality. Many specific theories of discrimination in the organizational literature fall into the category of performance-motivated discrimination, because evaluators in markets are often tasked with making quality judgments.

A performance motive underlies the statistical discrimination models of Phelps (1972) and Arrow (1973), where evaluators act on their beliefs about correlations between demographic categories and other outcomes. For example, car insurance companies in the United States often charge younger male drivers more than older female drivers, because accidents have historically occurred at higher rates for the former group, and the insurer believes the correlation will exist into the future.[2] Car insurance companies do not gain personal satisfaction from their unequal treatment of men and women, but rather are motivated to do so because they believe it helps them make better decisions.[3]

In other types of performance-motivated discrimination, however, the source of performance expectations may have little or no representation in reality. For example, status-based theories of discrimination are premised on the idea that evaluators will infer that someone is lower quality because they hold a trait that society collectively believes is associated with low ability. For example, "employers prefer men because cultural beliefs about the relative performance capacity of men and women bias cognition" (Correll and Benard, 2006: 111). In practice, this type of status-based discrimination can lead to the application of double standards (Botelho and Abraham, 2017). Other types of discrimination that are performance-motivated include theories of attributional augmenting that explain how gender can change the weight given to other information (Baron, Markman, and Hirsa, 2001) and more generally how traits such as gender "frame" decisions that should rationally not directly involve such traits (Ridgeway, 2011). Performance-motivated discrimination does not even need to be conscious, as in the case of implicit biases that shape quality beliefs even when the person doing the discrimination might not be aware their beliefs are biased. These lead to performance expectations that can be "implicit, often unconscious, anticipations of the relative quality of individual members' future performance" (Correll and Ridgeway, 2003: 31). Therefore, a separate range of discrimination theories posit that discrimination

[2]For example, the website of Allstate Corporation's Esurance car insurance brand has even explicitly communicated this in the parlance of counterfactuals: "If you're a guy, all this really means is that a female clone of yourself would likely pay less for car insurance" (Esurance, 2014).

[3]Variance in the perceived ethicality of this type of discrimination is found in the fact that using gender to determine car insurance rates is illegal in the European Union.

is motivated by performance expectations rather than a consumption motive.

While the consumption motive and the performance motive are theoretically separate, it is of course possible they exist at the same time or even influence each other across time. The question of where underlying tastes and performance expectations come from is beyond the scope of this paper. However, both are typically assumed to be quite stable in the short term. For example, statistical discrimination explains performance expectations as a response to imperfect information, so behavior should be consistent absent a change in information.[4] Status construction theory provides an explanation for how status-based competency beliefs develop, where structural factors such as the historically unequal distribution of resources can lead some groups to be evaluated as higher quality than others (e.g., Ridgeway, 1991). Therefore, without significant changes to an environment one would not expect performance expectations themselves to quickly change. Tastes are often trickier to explain, as hinted at in the maxim "de gustibus non est disputandum", or "there's no accounting for taste" (Stigler and Becker, 1977). However, culture provides one obvious source of tastes: people like or dislike cultural objects such as specific music simply because others do (Salganik, Dodds, and Watts, 2006). Therefore, one would also expect consumption-motivated discrimination to be consistent absent outside social interference.

To summarize, the distinction between consumption-motivated and performance-motivated discrimination is a typology that classifies the motives underlying different types of discrimination found in the literature. This typology of motives is a useful way to conceptualize discrimination for two main reasons. First, it helps to abstract away from otherwise impor-

[4] A small literature has focused explicitly on the role of information availability at the time of decisions, typically using it as a diagnostic tool to measure the existence of taste-based versus rational statistical discrimination in specific contexts. The premise of that work is that providing evaluators more data should lead to less statistical discrimination, because when evaluators have more individual-level information they do not need to employ group-level stereotypes. This information-addition approach has been employed in studies of visa examiners (Rissing and Castilla, 2014), job applicants (Fernandez and Greenberg, 2013), medical students (Rubineau and Kang, 2012), and physician wages (Siniver, 2011). One challenge of this work is that other discrimination theories such as status-based discrimination are also based on performance expectations, but posit that peoples' cognition is biased so that new information may have little impact on their performance expectations (Correll and Benard, 2006). This means the relationship between information and discrimination may not be as straightforward as theories of statistical discrimination might indicate.

tant nuances of specific discrimination theories. For example, “statistical discrimination” has theoretical baggage related to whether people are accurate or inaccurate, which is unrelated to whether that type of discrimination is performance-motivated.[5] Second, the typology of consumption-motivated and performance-motivated discrimination helps to highlight similarities between otherwise disparate theories. For example, although theories of statistical discrimination and status-based discrimination have developed separately and have many important differences (Correll and Benard, 2006), both are driven by the performance-motive outlined here. The typology therefore allows one to theorize about what encourages a broad range of discrimination. The next section will develop an argument for why performance-motivated discrimination should respond one way with respect to economic consequences, while forms of consumption-motivated discrimination should respond the opposite.

ECONOMIC CONSEQUENCES AND THE MOTIVE TO DISCRIMINATE

The motives outlined in the previous section are a necessary but not sufficient condition for discrimination to exist in a market. Additional considerations are necessary. This includes a consideration of what might encourage or discourage an evaluator to actually enact these motives, as well as potential variance across evaluators in their propensity to enact these motives. This means that discrimination may increase or decrease as a function of changes to the behavior of specific evaluators, as well as changes to the set of evaluators. One prominent insight from past work is that economic consequences can make discrimination costlier, and thus reduce its likelihood.

This argument is built on the logic that consumption-motivated discrimination has a cost, and changes to the environment can make it costlier. For example, in the 1940’s and 1950’s the set of baseball teams that delayed racial integration won fewer games and had lower audience attendance than teams that integrated sooner (Gwartney and Haworth,

[5]For example, Pope and Sydnor (2011: 90) noted, “once one allows for the possibility of inaccurate beliefs, results from other studies that find evidence of taste-based or accurate statistical discrimination come into question.”

1974). Overall discrimination slowly decreased as firms changed their behavior to become less discriminatory. In this same vein, Siegel, Pyun, and Cheon (2018) found that in the Korean market, local firms that were reluctant to hire women faced worse performance than multinational firms that hired more women. There, total discrimination slowly decreased as new firms entered the market. Both of these examples indicate that discrimination can be reduced by increasing the competitiveness of a market. In such environments, discriminators must either reduce their discrimination or bear the cost of their tastes and risk survival. For example, Pager (2016) showed that firms that had previously discriminated in an experimental audit study (Pager, Bonikowski, and Western, 2009) were more likely to go out of business within the following six years. In extreme cases, discrimination should go down as discriminatory evaluators are replaced by non-discriminatory ones. This means that when the consequences of decisions are high, there will be both an incentive for discriminators to eventually reduce their discrimination (e.g., Gwartney and Haworth, 1974) and an incentive for non-discriminators to enter the market (e.g., Siegel, Pyun, and Cheon, 2018). In either case, overall discrimination in a market should fall as economic consequences rise.

There is also some evidence for this dynamic within organizational settings, where the economic consequences of decisions are typically a function of incentive policies that determine how tightly or loosely evaluators bear the consequences of their decisions. Incentives have traditionally been studied from the standpoint of "principles" who calibrate how closely their "agents" should bear the consequences of decisions in order to maximize output (for an organizational theory perspective, see Eisenhardt, 1989). Gibbons (1998: 116) described this fundamental tension as "The Classic Agency Model: Incentives versus Insurance". For example, paying workers an hourly rate versus a piecemeal rate has been shown to influence worker output in both theory and practice (e.g., Lazear, 2000). When these types of compensation systems are applied to managers they have been shown to impact how subordinates are evaluated. Using a field experiment, Bandiera, Barankay, and Rasul (2007) found that the implementation of a performance pay system caused managers to treat their high- and low-ability workers differently than they had when the managers were compensated with fixed wages. Along these lines, Ayres and Waldfogel (1994) conducted a "market test for discrimination" in the court system by comparing the bail amounts set by the court (which were high for minorities) to the subsequent rates charged by private bail bond dealers. The behavior of the bail bond dealers and the judges was theorized as a function of the economic consequences they faced for making accurate or inaccurate decisions. Using the assumption that the bail bond dealers had a stronger incentive to set fair rates than the judges, the authors inferred underlying levels of discrimination on the part of judges. These studies imply that discrimination will be a function of environmental characteristics, both because decision makers may alter their discriminatory behavior based on an environment, but also because an environment may encourage different types of decision makers to enter a market.

However, the expectation from the literature that increasing the economic consequences of an evaluation should discourage discrimination is derived from viewing discrimination as primarily consumption-motivated. Acknowledging the possibility of performance-motivated discrimination complicates that prediction. This is because although increasing consequences should discourage consumption-motivated discrimination, it should simultaneously *encourage* performance-motivated discrimination. Performance expectations—regardless of their source or accuracy—should be employed more frequently when the stakes of a decision are high. For example, if a basketball coach believes tall players are more skilled than short players, then tall players should be given more playing time during championship games compared to the regular season, precisely because the cost of losing (or benefit of winning) such games is greater.

This dynamic should also function in the opposite direction. Consumption-motivated discrimination will be encouraged by environments where the consequences of decisions are low, because tastes will be cheap to express. However, like above, performance-motivated discrimination will be discouraged by such environments, because the value of performance expectations in general will be lowered. This makes economic consequences a particularly important environmental characteristic as it should simultaneously influence consumption-motivated and performance-motivated discrimination in opposite directions. These predictions are agnostic to the sources of evaluators' specific beliefs and preferences, which is particularly useful in the case of performance expectations. For example, both conscious and unconscious performance expectations as well as "correct" and "incorrect" performance expectations should all respond in the same way to changes in economic consequences.

However, because consumption-motivated and performance-motivated discrimination should respond in opposite directions to a change in consequences, additional assumptions are required to predict ex-ante whether total overall discrimination will go up or down given a change in consequences. Two assumptions must come from the specific context under study: do evaluators likely 1) hold positive or negative performance expectations about the trait of interest, and 2) hold positive or negative consumption preferences about the trait of interest?

These two assumptions are necessary because the overall change in discrimination will be a function of the direction and strength of evaluators' underlying taste preferences and performance expectations, which will depend on the context. These may sometimes be correlated, but not always. Take for example the case of evaluators who hold positive performance expectations about a trait but a taste-based prejudice against it, potentially found in some forms of "model minority" discrimination. In such cases, decreasing the potential consequences of a decision should increase overall discrimination via a decrease in (positive) performance-motivated discrimination and an increase in (negative) consumption-motivated discrimination.

However, if evaluators hold both beliefs and tastes that are in the same direction, then the overall change is more difficult to predict. This is because one motive will be encouraged while the other is discouraged. In such cases the relative ex-ante strength of one motive over the other will determine whether total discrimination will increase or decrease. Despite this, a specific shift in each motive may still be useful for practical purposes, as other policies may be useful for addressing whichever motive has been encouraged, an issue that will be considered in the Discussion section. In the next section I outline the assumptions required to make ex-ante predictions in the specific context of this study.

The evaluation of gender in microfinance

The theory developed above should function independent of the specific traits being evaluated. However, a focus on gender in this study is useful for three reasons. First, gender is what psychologists refer to as a "primary category", that is, one of the handful of universal classification criterion that people employ when evaluating others (Ridgeway, 2011: 40). The result is that social relations are fundamentally "framed" by gender (Ridgeway, 2011). This makes gender a useful trait to examine for the purposes of testing a general theory about discrimination like the one in this paper. It also ties the paper to recent discrimination research in the management literature, which has focused on gender (e.g., Dahl, Dezső, and Ross, 2012; Carnahan and Greenwood, 2018; Greenberg and Mollick, 2017).

Second, there is existing research on gender in empirical contexts similar to this study. This is useful because it can provide the basis for assumptions about how evaluators in this context are likely to evaluate women in terms of both performance and consumption motives. The starting point for these assumptions is the traditional microfinance narrative that women are economically superior borrowers to men despite the cultural discrimination they face (Armendariz and Morduch, 2010; Roodman, 2012). These assumptions can in turn provide directionality for how each of the two foundational motives will respond to changes in economic consequences and what will happen to overall discrimination. I begin with assumptions about performance expectations in this context.

The positive performance expectations about female borrowers compared to male borrowers in microfinance contexts may be derived from a number of sources, including beliefs that on average women may have a higher incentive to repay loans because of fewer outside options, are more responsible with money because of general risk-aversion, and have more limited geographic mobility, which makes collection easier (Armendariz and Morduch, 2010). This general conclusion appears consistent with research in contexts even closer to this study. On Proposer.com, one of the major for-profit P2P lending platforms in the United States, Pope and Sydnor (2011) found that after controlling for available observables, such as credit score, women were more likely to be funded than men. Chen, Li, and Lai (2017) examined data from a similar Chinese lending platform and concluded that female borrowers were more likely to receive loans and less likely to default, but were charged higher interest rates (a feature of that platform); they interpreted this as evidence of positive statistical discrimination and negative taste discrimination against women. As I will discuss later, the default rates on the platform I examine in this study exhibit a similar trend. Of the loans that had already matured by the start of the study period, none of the defaults were by female borrowers, and of the loans that were actually invested in during the study period, only one of the twenty-five that eventually defaulted was a woman. Therefore, it seems most plausible that lenders hold positive stereotypes about women's performance abilities in this specific context.

Despite the likely positive performance expectations about women in this context, it also seems probable that negative taste preferences exist from the broader social context. China ranks similarly to the United States on the United Nation's Gender Inequality Index.[6] However, a number of sociological phenomena indicate the salience of gender attitudes is high. This includes profound gender disparities, such as the "missing women problem" where the ratio between the actual number of women and men in China is not what one would naturally expect (Sen, 1992; Qian, 2008). One explanation for this "son preference"—which is not exclusive to China—is the particular cultural family system that exists in countries that exhibit it (Das Gupta et al., 2003). This potentially disparate treatment continues later in life, for example the derogatory term "leftover women" ("shèngnǚ", 剩女) used to refer to women as young as twenty-five who have not yet married (Hong Fincher, 2016).

[6] In the 2015 United Nations Gender Inequality ranking of 159 countries, China is actually ranked slightly higher (37th) compared to the United States (43rd), despite its much lower ranking on the overall Human Development Index (90th for China vs 10th for the United States).

Thus, broad cultural "tastes" may be developed early and derived from widespread cultural preferences for men relative to women, even absent specific performance evaluations where quality expectations might be important. Thus, if one were to make general assumptions about lenders in this context it seems most plausible they hold 1) favorable performance expectations about women as borrowers, and 2) negative taste preferences toward women.

This means the third and final advantage of a focus on gender in this specific context is that the above assumptions are opposite of each other. The directionality of the assumptions about performance beliefs (positive) and tastes (negative) allow for clearer predictions about how the *overall* evaluation of women should be related to the economic consequences faced by those that evaluate them: increasing the economic consequences should lead to overall more favorable treatment of women, and decreasing the economic consequences should lead to overall less favorable treatment of women. This is because raising the economic consequences of lending decisions should theoretically encourage more performance-motivated discrimination (more favorable treatment of women) and less consumption-motivated discrimination (less unfavorable treatment of women); both changes will lead women to be treated more favorably than they were before. Likewise, lowering the economic consequences of lending decisions should theoretically encourage less performance-motivated discrimination (less favorable treatment of women because the positive performance expectations are made less valuable) and more consumption-motivated discrimination (more unfavorable treatment of women because negative taste preferences are now easier to express); both changes will lead women to be treated less favorably than they were before.

EMPIRICAL SETTING

The empirical context for this study was an online peer-to-peer lending platform in China. In a stylized version of peer-to-peer lending, a mediating "platform" accepts applications from potential borrowers, screens them, posts them on a website for lenders to evaluate, and then facilitates the transfer of money from lenders to borrowers and borrowers back to lenders. Loan requests are fulfilled in a piecemeal fashion, where many lenders each choose to contribute a portion of a given borrower's total loan request. Once the full loan request is met, the loan is closed and the platform facilitates the transfer of funds from the lenders to the borrower. The platform then facilitates the collection of loans and periodic borrower repayments.

The prototypical and earliest peer-to-peer lending platforms in the United States were Prosper.com, Lending Club, and the 501(c)(3) platform Kiva.org, which was designed for non-profit lending (Government Accountability Office, 2011). A handful of other platforms catered to niche markets such as student loans or medical procedures. Regulatory constraints typically limited the operations of such firms to national borders, meaning there was no international competition.

Online lending is a particularly useful context for studying evaluation processes, because researchers have access to the same information used by the lenders to evaluate the borrowers. This allows researchers to strengthen the assumption that evaluations are not being driven by omitted variables that the evaluators can see but the researcher cannot. For example, Leung and Sharkey (2013) employed the context to study perceptual factors related to the classic category-spanning discount. From the standpoint of discrimination, Pope and Sydnor (2011) and a working paper by Ravina (2012) both examined discrimination in the context of the Prosper.com marketplace. Both examined the likelihood of borrowers receiving a loan, the favorability of loan terms (a feature of the Prosper.com marketplace), and the average financial performance of different demographic groups. However, such studies have focused primarily on identifying the existence of different types of discrimination rather than what triggered that discrimination.

Online lending in China

This study took place in the context of an online for-profit peer-to-peer lending platform in China. In 2012 at the time of the study, Chinese peer-to-peer ("个人对个人" or "individual to individual") lending differed from the American context in a number of ways. This was partly due to differences in the historical development of the financial services industry in China and partly from looser regulatory constraints at the time. In the United States, peer-to-peer lending companies relied on existing third-party credit scores to screen potential borrowers.[7] China lacked an extensive national credit scoring system such as the FICO score, so the role of peer-to-peer lending companies was broader than in the United States and included more intensive verification of borrower backgrounds.

American peer-to-peer lenders technically invested in promissory notes sold by the peer-to-peer platforms, which were tied to the repayment of specific loans issued by a bank.[8] In China, however, the platforms more directly facilitated the transfer of money between lender and borrower. In theory, this resulted in less regulation since it amounted to activity outside of traditional banking institutions. This was observable in company names. For example, Ppdai, one of the first peer-to-peer lending platforms in China, was registered as Shanghai Ppdai Financial Information Service Co., Ltd.The business name of the company in this study included "commercial advising." Therefore, at the time of the study the industry still occupied an uncertain space in the broader scope of Chinese financial services with most platforms functioning as some form of financial information or advising company. Lenders had to trust that platforms were honest and competent, because it was impractical to collect individual repayments without the assistance of the platform.

The number of online Chinese peer-to-peer platforms increased rapidly from just nine in 2009 to 132 in the first quarter of 2013 (Li, 2013). By 2018 media reports estimated the number of platforms in the thousands with hundreds of billions of dollars in transactions (Feng, 2018). This growth can be viewed in light of broader economic trends, where a tradition of heavy state involvement in the Chinese financial system increased the attractiveness

[7]For example, as of June 2014, "A new Prosper borrower must be a U.S. resident in a state where Prosper loans are available, and must have a bank account, a Social Security number, and a credit score of at least 640. Prosper uses Experian to obtain credit scores." Source: `https://www.prosper.com`

[8]Both Prosper and LendingClub used WebBank, an FDIC-insured institution. Kiva loans were distributed and collected through partner microfinance institutions.

of these companies to both lenders and borrowers (Huang, 2018). State-owned banks both offered low interest rates to investors and preferential lending to state-owned enterprises. This made it difficult for individuals and small businesses to procure traditional bank loans, and drove overall demand for financial innovations such as peer-to-peer lending. It is against this institutional backdrop that this study took place.

Data for this study was collected from a platform that began offering peer-to-peer lending services in 2010. At its founding, loans were not guaranteed and functioned similar to US-based platforms such as Prosper.com, although to my knowledge never featured competitive bidding on interest rates. It first implemented a loan repayment guarantee policy in the first half of 2011 when total loan volume was still low. In early 2012 the company updated its loan guarantee policy to cover loans of all credit rating levels. These types of loan guarantee policies were common in the industry as a way to assuage fears about repayment and attract new lenders. For example, competitor Ppdai began offering a principle guarantee in July of 2011.These guarantees were featured prominently in the marketing material of such companies, and can be viewed as a stepping stone to the eventual introduction of packaged financial products whereby companies would simply aggregate individual loans together so individual choice was no longer necessary.[9] Because of the lack of a comprehensive national credit scoring system, lenders had always had to trust in the peer-to-peer companies to perform proper due diligence on potential borrowers. Therefore, the guarantees acted as a mechanism to demonstrate that the companies' incentive to perform adequate due diligence on borrowers was aligned with the interests of lenders.

Because borrower screening was more involved for Chinese companies than their US counterparts, this created the potential for mismatch between the supply and demand of money on the platform.[10] For example, during the time period examined in this study, the platform attracted much more money for lending than borrowers available to accept

[9]Huang (2018: 72) noted that a 2016 regulatory measure impacted the range of potential business models within the industry, including the ability to offer these types of guarantees.

[10]A U.S. company could simply employ well established third-party credit scores as their primary criterion for allowing a borrower on the platform.

it. This meant that all the loan requests that the platform allowed to be posted were fulfilled, sometimes in just a matter of hours. I next discuss how these features of the context influenced the empirical design.

EMPIRICAL STRATEGY

Data

The data for the study consisted of all realized evaluations of borrowers, that is, every decision made by lenders on the platform. Each row of these data indicated how much a specific lender decided to provide a specific borrower and when they made the evaluation.

Dependent variable. The main outcome of interest was how lenders altered their evaluation of male and female borrowers based on economic consequences. That is, did their relative evaluations of female borrowers become more or less favorable. To measure this, I focused on the amount of money a lender decided to contribute to a specific borrower's loan request. This empirical choice is similar to other discrimination studies that have focused on later-stage evaluations, such as the allocation of employee bonuses between men and women (Castilla and Benard, 2010), even though it is possible that inequality also existed at earlier stages such as the promotion or hiring process. If a lender invested in the same loan more than once, the total amount invested in that loan by the lender was aggregated for purposes of analysis.[11]

By using a lender's loan size decision, it was largely possible to sidestep the selection and timing effects that are difficult to account for in non-experimental data. For example, the speed at which borrowers were funded in this particular context meant it was unrealistic to assume that all lenders viewed all borrowers (i.e., to assume that all 3,087 lenders active at

[11] This occurred with some frequency in the data. An employee of the company indicated that because loans were sometimes funded almost instantaneously, a prospective lender might have first attempted a smaller sized investment than his or her actual target and then repeated the process until they had either invested their desired amount in the loan or other lenders had already fully satisfied the loan request. Lenders might even have used third-party software tools for this purpose, although this was discouraged by the company.

least once during the period viewed all 558 loans, which would have resulted in 1,722,546 pairs). This is because at any given time only a handful of active loans were available, ranging from between zero other borrowers and fifteen other borrowers at any given time (Figure A1). Further, all borrowers during the period under study had their loan requests fulfilled. For this reason I focus on the 25,440 realized decisions, and then include control variables that account for the availability of other loans at the exact time of each of these decisions.

Independent variables. The main counterfactual of interest is how the relative evaluation of women is different when economic consequences of a decision are higher or lower. This requires an interaction of two variables. First, a binary variable that captures the gender of the borrower. Second, a binary variable that indicates whether the decision was made in an environment with relatively higher or lower economic consequences as compared to a separate baseline environment. The interaction between these two variables captures whether economic consequences influenced the evaluation of gender.

Control variables. The most important control variables consisted of other observable traits of the borrower. Chief among these were the direct financial characteristics of the loan itself: i.e., the company-assigned credit score category, the interest rate, the loan term, and the amount of money requested. I also controlled for a set of other available borrower characteristics. These included the purpose of the loan use and a range of other information about the personal and professional situation of the borrower; these are detailed in the regression table notes.

In addition to these borrower characteristics, I created time-dependent variables that helped capture lender choice. Each of these variables was calculated based on the timestamp of the focal lending decision. These included the count of other loans that were available at the same time (mean 3.7 loans), and the number of those loans that were women (mean 0.55 loans), as well as a measure of how long the focal loan had been posted on the website based on the difference between the time of the focal lending decision and the first decision from any lender to that specific loan (mean 14.7 hours).

Finally, I also created a set of lender variables to potentially control for variance in each lender's performance expectations about borrowers. This included whether the lender and borrower shared a geographic location (8.2% of decisions), as well as measures of a lender's past experience on the platform. Again, these lender experience variables were calculated based on the timestamp of each focal decision by a specific lender. These lender experience variables included the number of previous lending decisions the lender had made, the number of those decisions that were to female borrowers, the amount of total money lent, and the previous number of loans that were likely to already be known to have defaulted. A lender's personal experience with defaulted loans was inferred from a combination of the loan terms and loan outcomes of each lender's past decisions. For example, if a lender had made a loan in January that had a six-month term, then knowledge of the outcome of that specific loan would be available in July. Any decisions made after July would then be made with the knowledge of whether that specific previous loan had defaulted. Less than 5% of the sample of lending decisions were made at a time when the lender had already experienced a default from one of their prior lending decisions.

Research design and methods

The goal of this study is to understand how shifts in economic consequences may alter how gender is evaluated. Variance in consequences is required to test this theory, where the potential economic consequences of decisions is shifted to be higher or lower than they were previously. Evaluations can then be compared before and after the change.

This study employed a setting where the potential economic consequences of using gender during lending decisions was reduced. In early 2012 the company updated its loan principle repayment guarantee policy. For practical purposes, this guarantee amounted to an insurance policy for lenders. Unlike an existing policy implemented in the first half of 2011, this new policy covered every loan on the platform (the previous policy did not cover HR loans). It was funded by assessing service fees on loans ranging from zero to five percent depending on the company-assigned credit rating of the loan. The policy would be expected to reduce—if not remove—the economic consequence of lending to one particular borrower over another.[12]

Therefore, performance-motivated discrimination should decrease from the pre-policy period to the post-policy period given that the value of performance expectations about gender was reduced. At the same time, consumption-motivated discrimination should increase from the pre-policy period to the post-policy period because the consequences of expressing those tastes were also lowered. Combining these predictions with the context-specific assumptions outlined previously (i.e., the existence of positive underlying performance expectations and negative tastes) leads to an overall empirical expectation that the change will cause women to be evaluated less favorably than they were previously.

To test this, I used 30-day windows of data from either side of the policy change. The difference of interest is not in whether men and women were evaluated differently, but whether the policy itself altered these relative evaluations. Therefore, I employed a difference-in-differences design using each of these two periods to measure the potential shift in relative evaluations. In practice this takes the form of an interaction between the policy period variable ($guarantee = true$) and the gender of the borrower ($gender = female$), so that a positive coefficient on the interaction term would indicate women were subsequently treated more favorable, and a negative coefficient less favorable.

Employing limited 30-day windows on either side of the policy change strengthened the identification in two ways. First, it limited the potential influence of other concurrent events in the firm, industry, or broader economy. This is important given how fast the industry, as well as business models of the firms, was evolving (Huang, 2018). Second, it limited the potential for lender learning (e.g., Altonji and Pierret, 2001; Freedman and Jin,

[12] As noted earlier, loan collection outside of the platform was never practically feasible. Therefore, even in the pre-policy period lenders had to trust that the company was honest and would be able to collect repayments over the life of the loan.

2011), which could theoretically influence performance expectations. The lender experience variables outlined above also help control for differences in lender experience.

Assuming that the women in the post-period were not significantly different from the women in the pre-period—and the same was true for men—then this basic difference-in-differences should provide a valid measure of how the policy altered lenders' evaluation of gender. This is because the main assumption of this approach is not that men and women are identical, but rather that the characteristics of borrowers of each gender is similar across the study windows. To relax this assumption, however, I ran a set of ordinary least squares regression models that included the control variables previously reviewed. Equation 1 represents the general form of these models, where $DecisionSize_{i,j}$ is the amount of money that lender i loaned to borrower j (contingent on making a loan), $Gender_j$ is whether borrower j was female, $Policy_{\{0,1\}}$ is a binary variable for whether the decision was made during the post-policy period, and X_j is the primary financial characteristics of the pertinent loan: the interest rate, term of the loan, size of the requested loan, and credit rating category. X_j is then expanded in stages to include the full range of borrower control variables, followed by the lender experience variables.

$$DecisionSize_{i,j} = Policy_{\{0,1\}} + Gender_j + Policy_{\{0,1\}} * Gender_j + X_j \tag{1}$$

Descriptive statistics

Lenders made 25,440 decisions to lend across 558 loans during the study window. The decisions during the two 30-day periods are summarized in Table 1, with 10,975 in the pre-policy period and 14,465 in the post-policy period. In total, approximately 15 percent of these decisions were to women. The average investment size was 882 RMB in the 30 days before compared to 865 RMB in the 30 days after. However, the absolute number of decisions was greater in the second period and highlights the volatility in loan supply, with some days where a shortage of borrowers meant no loans were made (Figure A2).

[Insert Table 1 about here]

The 558 individual loans in the window (317 in the pre-policy period and 241 in the post-policy period) are summarized in Table 2. The overall loan interest rates varied from 6.1 to 24.4 percent with the most common categories of 13 and 15 percent. Loan sizes also varied from 3,000 RMB (approximately $475) to 500,000 RMB, with the majority of loans in the 3,000 or 5,000 RMB categories. The loan repayment terms ranged from three to 24 months, with the majority being three or six month loans. All loans were fully funded, and most very quickly, with the average taking just over eight hours. During this period, the demand from lenders therefore outstripped available borrowers.

[Insert Table 2 about here]

Additional information on the 558 loans that were open for lending at some point during the window is presented in the supplementary material. Only 25 (of which one was a woman) of the total 558 loans ended up as "bad debt" (Table A1). The majority of borrowers indicated they would use their loans for short-term turnover (Table A2). The distribution of the company-assigned credit rating ranged from AA (highest quality) to HR (high risk). The pool of loans grew progressively larger as the quality decreased (Table A3). About half of loans were rated the two highest risk categories of E or HR, with only five loans rated AA or A.

There was a total of 3,087 unique lenders who made at least one lending decision during the study period. 1,524 of these lenders were active in both periods, 566 were active in only the pre-period, and 997 in only the post-period. The higher number of unique lenders after the policy change may be a function of the platform growth process and supply of loans.

RESULTS

A simple difference-in-differences between the two periods with respect to gender provides preliminary evidence of how the policy altered discrimination. This calculation is produced in Model 1 of Table 3. The coefficient on the interaction of borrower sex and the policy is

-422.3 RMB ($p = 0.00009$), which represents the change in the relative evaluation of female borrowers. This negative coefficient indicates that the policy led women to be evaluated less favorably. Women were actually evaluated more favorably than men before the policy, but after the policy, men were evaluated more favorably than women. Using the regression model described earlier, I then introduced controls to relax the assumption that the borrowers within each gender category were the same across periods (i.e., the men were similar before and after the change, and the women were similar before and after the change). Model 2 in Table 3 adds controls for the most important "hard" economic traits of a loan: the loan's credit rating, interest rate, term, and size. The directionality on the interaction term was the same, although the magnitude of the coefficient decreased to -221.9 RMB ($p = 0.045$).

[Insert Table 3 about here]

A fundamental theoretical challenge in studies of demographic disparities is understanding what exactly a demographic attribute represents. Even demographic variables such as race present serious taxonomical challenges, where definitions change over time and it is not always apparent what category membership specifically entails or how the information is interpreted (Charles and Guryan, 2011). To further complicate matters, many individual characteristics are confounded with other characteristics. Therefore, it is impossible to ever be fully certain of how people interpret demographic information. These issues are important because they influenced what an "ideal" model specification should look like.

In light of this I next ran a model with a full range of borrower information and controls for the availability of other loans at the time the decision was made. Including these additional variables slightly reduced the sample size because of missing data. These results are presented in Model 3 of Table 3. The results remain consistent with previous models; the coefficient of the interaction term was -323.5 RMB ($p = 0.026$). Finally, I added the controls related to individual lender experience at the time the lender made each specific decision. These results are presented in Model 4 of Table 3. The coefficient on the interaction term was again consistent with previous models: -383.9 RMB ($p = 0.005$).

Further empirical tests

The previous section tested the main effect of how evaluations of gender changed across the two periods. The following sections further explore this effect from three complementary dimensions: the set of evaluators, the heterogeneity of the effect, and the nature of the policy.

Evaluator behavior. The preceding analyses indicated that the policy led borrowers to be evaluated differently based on their gender. Two basic pathways could have contributed to this effect. The first is a within-lender channel where specific lenders changed their behavior as a result of the policy. The second is a between-lender channel where different individuals behaved differently across the two periods. Because discrimination is often conceptualized as a market-level outcome, these channels are not mutually exclusive, and it is plausible that both could have contributed to the effect.

I used a lender fixed effects model to investigate these two channels. The main challenge to employing lender fixed effects in this context was the extent to which it restricted the sample. Although about half of all lenders that were active during the study period made loans both pre- and post-policy, a lender must have made at least one decision to a male borrower and one decision to a female borrower in both the pre- and post-policy periods to possess the minimum required amount of variance. Subsetting the data to such lenders resulted in 405 lenders who made a total of 11,874 decisions across the span of the two windows (this represents 13.1% of lenders and 46.7% of decisions from the original sample; Table A5 includes descriptive statistics). The final model described in the previous regressions was then rerun using the subsample eligible for lender fixed effects and the subsample ineligible for lender fixed effects (Models 1 and 2 in Table A6). For the subsample that was ineligible for fixed effects, the results were similar to the main analyses. For the subsample that was eligible for the fixed effect model, the coefficient was in the same direction but quite noisy.

These results arguably provide stronger evidence for the between-lender channel than the within-lender channel, even though the evidence is not inconsistent with the later. This relative importance of between-lender effects reflects other discrimination research that indicates that within-evaluator behavior may be slower to change (Siegel, Pyun, and Cheon, 2018). One explanation is the "imprinting" that can occur from an experience of specific environments (Marquis and Tilcsik, 2013). However, I found evidence that the effect was stronger (more negative) when interacted with either the total number of previous loans a lender had made or the total number of defaults a lender had experienced (Table A7). One caveat of prior experience in this setting is that more experienced lenders may also pay more attention to platform changes simply because they are more active.

In order to better understand the nature of potential within-lender changes, I expanded the time frame by an additional thirty days before and after the main sample window. This approach has natural tradeoffs, as such a wide window (60 days on either side of the policy) is more likely to overlap with other changes in such a rapidly developing industry. However, the benefit is that a larger number of lenders (702) can meet the criterion for a fixed effects estimation. I therefore ran the model using this sample as a robustness test (Model 3 of Table A6). The magnitude of the coefficient is very similar to the original fixed effects model and much more precise. This strengthens the evidence for within-lender effects, although they do appear somewhat weaker in magnitude than the between-lender effects, implications that will be explored in the Discussion section.

Heterogeneity of the effect. It is possible the measured effect is not observed equally across the range of decisions. For example, if a lender considered a sum of money to be too small to be worth an active decision, then one would expect the decision to be as good as random no matter the potential consequences. This indicates there may be some lower threshold for this mechanism to function. To test for this, I excluded small decisions from the analysis. Limiting the decisions to those equal to or greater than 200 RMB resulted in around 14,000 decisions. These data were then used in the regression model described earlier. As expected, the magnitude of the previously observed effect was larger for this subset (Table A8).

A second form of heterogeneity may also exist if male and female lenders hold different tastes or beliefs about quality. Some studies have found evidence for how the gender of evaluators may influence discrimination (e.g., Srivastava and Sherman, 2015; Greenberg and Mollick, 2017). However, Heilman (2012: 129) concluded that "In the vast majority of studies conducted on gender stereotypes, no differences have been found in the reactions of male and female respondents." To test whether an effect exists in this context, I reran the final previously specified model and included the gender of the lender interacted with both the borrower gender and the policy (Table A9; also Table A10 for crosstabs). Neither the two-way interaction between lender and borrower gender, nor the three-way interaction with the policy variable was statistically significant at conventional levels, meaning I did not find evidence that women and men evaluated gender differently in this context. This finding therefore appears consistent with the majority of studies on gender stereotypes.

Policy treatment specification. Given that the updated guarantee policy covered loans with an HR ("high risk") credit rating while the previous policy did not, it seems plausible the effect would be strongest on that subset of borrowers. A blanket guarantee policy may have also reduced the perceived economic consequences for all borrower choice on the platform. If this is the case, then loans that were previously covered, yet flagged as more risky (e.g., "E" credit ratings), may have been impacted differently than less risky categories (e.g., "B" credit ratings). To test for these effects, I split the sample into three separate subsamples: 1) loans with a credit rating of "HR", 2) loans with a credit rating of "AA" through "C", and 3) loans with a credit rating of "D" or "E." Despite the overall prevalence of HR-rated loans (Table A3), they were on average smaller than other categories, so represented only 14.8% of the original sample of lending decisions. The majority of decisions were to "D" and "E" loans (56.3%).

I then ran the same model specification across the three subsamples (Table A11). The directionality of the coefficients in the three models mirrored earlier results, however, the coefficients were imprecise for both the HR and the AA/A/B/C samples. The much smaller sample size of the HR subsample may have contributed to the imprecision of its coefficient estimate. However, the larger (negative) magnitude and more precise estimate for the D/E subsample provides some support that the policy may have cued lenders to consider financial risk more broadly. This indicates that salience of economic consequences may be enough to change how borrowers are evaluated. One reason for this may be that the policy represented a fundamental shift that removed all economic consequences related to choice between borrowers. At the platform-level, this represents a shift from "some" risk when choosing borrowers to "zero risk" involved in the choice. Even if such a gap is small it may be qualitatively important. In short, economic consequences may be perceived in terms of the broader environment (i.e., the platform) as well as the more narrow decision. These findings may also further support evidence of the between-lender processes explored in the lender fixed effects analysis. This is because platform-level policies are likely more salient to new lenders who join the platform as a result of advertising or other communication that may highlight such policies.

A second consideration is related to changes in the purposes of loans. The main models control for loan purpose, however, the policy appears to have coincided with shifts in loan usage categories (Table A2). This could be due to changing incentives at the platform level to prioritize certain types of loans to post on the platform, or other dynamics related to how loans are classified. Given that the most salient shift appears to be a decrease in personal consumption loans and increase in short-term turnover loans, I re-ran the final model from Table 3 first with only loans from those two categories, and then again with only the short-term turnover loans (Table A12). The original effect was present in both of these cases. I interpret this as evidence that even if the policy altered what types of borrowers were allowed on the platform or how they were classified, the effects do not appear contingent on such a shift occurring.

Finally, because policy changes such as the one in this study are not experimentally exogenous, I also constructed and tested two "placebo" policies. The first placebo test set the treatment date one month before the actual date and compared the 30-day windows on either side, so that this sample had not experienced the actual policy change. The second placebo treatment employed the same approach but moved the treatment date to one month after the actual date, so that this sample had already experienced the policy change. The analyses were then replicated for these two samples (Table A13). The coefficients on the interaction terms were insignificant in the primary financial control models. However, the coefficient was positive in the second placebo sample with the full set of controls, although this was contingent on specifically including the control for the number of other concurrent female loans. This post-treatment placebo may not be an ideal test, however, because in reality the full sample had already been treated. Further, in settings such as this where the most salient goal of evaluators is to make money, the lack of any economic incentive at all to discriminate between individual borrowers naturally leads the industry itself to change. Indeed, firms in the industry soon began to offer automatic investment options and financial products based on bundles of individual loans, so that individual choice between loans no longer occurred. The second placebo test may indicate that some borrowers had already begun to mimic this behavior by simply randomly choosing borrowers, which would lead women to be treated more favorably again as the difference between genders is ultimately equalized.

DISCUSSION

In this study I examined the role of economic consequences in motivating discriminatory evaluations. I theorized that consequences affect discrimination by simultaneously encouraging and discouraging two separate motives for discrimination: a consumption motive driven by taste preferences for a specific trait, and a performance motive driven by performance expectations about that same trait. Reducing the economic consequences of evaluations should discourage performance-motivated discrimination and encourage consumption-motivated discrimination. Increasing the consequences should encourage performance-motivated discrimination and discourage consumption-motivated discrimination. This is because economic consequences alter the value of enacting tastes and acting on performance expectations. They therefore create a dynamic relationship between otherwise very different types of discrimination.

The first step required to apply this theory in a specific context was to develop priors regarding the performance expectations and tastes of evaluators in that context. In this paper, these priors consisted of positive performance expectations but negative taste preferences. The second step was to analyze a situation where the potential economic consequences of evaluations had been altered. I found evidence that reducing economic consequences in this context led women to be evaluated less favorably. When interpreted in conjunction with the priors about evaluators, this result was consistent with expectations derived from the theory.

Organizational implications

Although the study took place in the context of a peer-to-peer platform, the insights about economic consequences may translate to more traditional organizational contexts when one considers the role of incentive policies. Many incentive policies have been documented in the compensation literature, including "piece rates, options, discretionary bonuses, promotions, profit sharing, efficiency wages, deferred compensation," and related approaches (Prendergast, 1999). All of these should increase or decrease the consequences of decisions for the people impacted by such policies. Even tools that are not explicitly economic in nature, such as non-monetary employee awards (Gallus and Frey, 2016), may increase the perceived consequences of evaluations and thus produce similar effects. This is because such tools link the performance outcome of an evaluation to the perceived rewards or punishments of evaluators. For example, changing the financial compensation of managers to be more or less closely tied to the performance of their employees should alter the levels at which managers will be motivated to discriminate. Therefore, consumption-motivated discrimination can be discouraged by introducing higher-powered incentives. Performance-motivated discrimination can be discouraged by shielding evaluators from the consequences of their decision.

However, organizations must also be cognizant that using economic incentives to reduce one form of discrimination may simultaneously motivate other forms of discrimination. This means generic prescriptive recommendations are not possible. The theory requires assumptions about the nature of people's ex-ante performance expectations and tastes in order to predict overall changes in discrimination. It was possible to establish these assumptions for the context of this study. However, organizations will need to turn to research on specific types of discrimination, prejudice, and biases to understand how altering incentives in a specific context is most likely to impact overall levels of discrimination in that context. Luckily, this task should be made easier by the diverse body of research on discrimination in specific contexts.

For example, Kalev, Dobbin, and Kelly (2006) found that many corporate diversity programs have mixed effectiveness and generally observed "We know a lot about the disease of workplace inequality, but not much about the cure" (p. 590). One potential reason for this is that managers may not truly understand the underlying motives of evaluators, and reexamining policies in light of the distinction between consumption-motivated and performance-motivated discrimination may be fruitful. Paluck and Green (2009: 341) reviewed 985 reports of particular prejudice-reduction interventions that included "multicultural education, antibias instruction more generally, workplace diversity initiatives, dialogue groups, cooperative learning, moral and values education, intergroup contact, peace education, media interventions, reading interventions, intercultural and sensitivity training, cognitive training, and a host of miscellaneous techniques and interventions." Some of these interventions, such as multicultural education, might be expected to impact only one motive for discrimination; e.g., decrease consumption-motivated discrimination but have no impact on performance-motivated discrimination. This means that it might still be useful to use incentives to decrease one type of discrimination even if it increases another type of discrimination, assuming complementary interventions can then be subsequently employed.

The within-lender fixed effects analyses in this paper also warrant additional interpretation in light of the potential organizational implications. This is because overall discrimination can decrease via multiple channels—both selection and treatment effects—sometimes occurring at the same time. This is most clear in traditional markets. For example, women have historically been undervalued in the Korean managerial labor market (Siegel, Pyun, and Cheon, 2018). As such a market grows (i.e., the economic consequences of entering it increase), taste-motivated discrimination should decrease via both selection and treatment effects: non-discriminators such as multinational firms will be encouraged to enter to access under-utilized talent, and local firms will be encouraged to reduce their own taste-motivated discrimination. Both these avenues will lead to less overall discrimination, although the former process may typically be faster than the later. In organizational settings, however, selection processes may be more constrained. This means that if organizations are limited to a fixed set of evaluators, the overall effect of economic consequences on discrimination may be more muted than if the pool of evaluators turns over more freely. Indeed, the fixed effects results from this study are consistent with but less clear than the results from the sample ineligible for fixed effects. This may be partly due to that fact that evaluators may be slow to change—an interpretation consistent with the dynamics of traditional markets (i.e., Siegel, Pyun, and Cheon (2018: 18): "the market is moving toward a new equilibrium free of discrimination, but very slowly."). The organizational implication of this would be that economic consequences will be more important when evaluator turnover is higher, which echos the importance of selection effects that has been identified in other evaluation processes (Kovács and Sharkey, 2014).

Future directions

While this study shed light on the dynamic relationship between economic consequences and two different motives for discrimination, it provided only general guidance about which motive will dominate in a static setting. In settings where evaluators are not constrained by strong economic penalties, one would expect tastes to play a larger role than would otherwise be expected, as tastes will already be relatively cheap to exert. In settings where decisions already lead to significant economic consequences, tastes will be less pronounced because they are already prohibitively expensive. Likewise, performance-motivated discrimination should be most prevalent when consequences are already high. This insight may be useful for interpreting existing studies of discrimination: those conducted in settings such as a laboratory (where economic consequences of decisions are low) may be more prone to measure consumption-motivated discrimination, whereas those conducted in the field (where real economic consequences of decisions are higher) may be more likely to capture performance-motivated discrimination.

Caution should also be taken when attempting to generalize these specific empirical results to other settings, as they represent just one particular context. Context-specific assumptions are required before directional predictions can be made. The general approach to encourage or suppress their expression, however, should apply to a wide range of settings. For example, Thébaud and Sharkey (2016) concluded that women-led small businesses had more difficulty securing loans following the financial crisis of 2007–2008. In addition to the role of uncertainty during such macroeconomic periods, one might also consider how the economic incentives within lending institutions might have changed. If recessions increase the perceived economic penalty for making bad decisions—for example, if layoffs within banks had increased—then loan officers should be more motivated to employ their beliefs about quality during such periods and less motivated by their tastes.

Future work may also benefit from attempting to directly situate these motives within the managerial research on anti-prejudice and diversity programs. In addition to economic incentives, managers have a range of tools that can moderate the perceived or real consequences of decisions. This includes "social" incentives such as awards (Gallus and Frey, 2016), but also policies that foster organizational cultures that might trigger taste or performance motives. For example, Castilla and Benard (2010) found that actively promoting a meritocratic organizational culture can lead to more discrimination. One interpretation of this finding is that participants actually did believe that men were more deserving of the bonus than women, but only rewarded them at higher rates when they were told that rewarding people based on quality was critical to their own job performance. Without such a nudge people may actually have a taste preference to treat people equally. Therefore, it is possible that performance-motivated and consumption-motivated discrimination can be triggered and suppressed via a range of different interventions, with economic incentives simply the most obvious starting point.

Finally, the construct of potential economic consequences is different than "accountability" as it is typically studied in discrimination research. For example, when studying pay disparities in organizations, Castilla (2015: 315) defined organizational accountability as "a set of procedures making certain individuals (or a group of individuals) responsible for ensuring the fair compensation and distribution of rewards among employees." This falls within a more general social psychology definition of accountability as "pressures to justify one's causal interpretations of behavior of others" (Tetlock, 1985: 227). Of utmost importance is "accountability for what?" Accountability for diversity in-and-of-itself should be most useful for achieving diversity, but vague accountability for the quality of a decision (thus increasing the economic consequences) might encourage more discrimination.

Beyond discrimination research, the study also speaks to the broader literature on evaluation processes and organizations. The distinction between consumption-motivated and performance-motivated discrimination presented in this paper may be useful to explain why people respond to generic forms of status: people may believe status markers help them make better decisions (i.e., facilitate a performance motive), but may also derive direct utility from interacting with high-status actors (i.e., facilitate a consumption motive). For example, Malter (2014) attempted to separate the returns to organizational status in the wine industry into two underlying components: quality signals versus conspicuous consumption. He noted of the Podolny (1993) view of status that "audiences would not have to rely on status to infer quality if quality were perfectly observable" (Malter, 2014: 276), but empirically demonstrated that conspicuous consumption matters (a consumption motive), and status—at least in the wine industry—matters in its own right independent of quality concerns. Similar results regarding how audiences separately evaluate the symbolic versus objective value of traits has been found in other contexts (Frake, 2016). I complement this and similar research by highlighting that the ultimate importance of such status traits to evaluators should also depend on the economic consequences under which evaluations are made. For example, consumption of high-status objects may be motivated by tastes if there are limited economic consequences for doing so. However, one would not expect the same to be as true if the consequences were increased. In such cases, performance-motives should come to dominate the evaluation process.

References

Altonji, J. G. and Pierret, C. R.
2001 "Employer learning and statistical discrimination." The Quarterly Journal of Economics, 116:313–350.

Aristotle
1934 The Nicomachean Ethics. London, Great Britain: Harvard University Press, William Heinemann Ltd.

Armendariz, B. and Morduch, J.
2010 The Economics of Microfinance. MIT Press.

Arrow, K.
1973 "The theory of discrimination." In O. Ashenfelter and A. Rees (eds.), "Discrimination in Labor Markets," 3–33. Princeton, NJ: Princeton University Press.

Ayres, I. and Waldfogel, J.
1994 "A market test for race discrimination in bail setting." Stanford Law Review, 987–1047.

Bandiera, O., Barankay, I., and Rasul, I.
2007 "Incentives for Managers and Inequality Among Workers: Evidence from a Firm-Level Experiment." Quarterly Journal of Economics, 122:729–773.

Baron, R. A., Markman, G. D., and Hirsa, A.
2001 "Perceptions of women and men as entrepreneurs: Evidence for differential effects of attributional augmenting." Journal of Applied Psychology, 86:923–929.

Becker, G. S.
1957 The Economics of Discrimination. Chicago: University of Chicago Press.

Bennedsen, M., Nielsen, K. M., Perez-Gonzalez, F., and Wolfenzon, D.
2007 "Inside the Family Firm: The Role of Families in Succession Decisions and Performance." The Quarterly Journal of Economics, 122:647–691.

Botelho, T. L. and Abraham, M.
2017 "Pursuing Quality: How Search Costs and Uncertainty Magnify Gender-based Double Standards in a Multistage Evaluation Process." Administrative Science Quarterly, 1–33.

Brooks, A. W., Huang, L., Kearney, S. W., and Murray, F. E.
2014 "Investors prefer entrepreneurial ventures pitched by attractive men." Proceedings of the National Academy of Sciences, 111:4427–4431.

Carnahan, S. and Greenwood, B. N.
2018 "Managers' Political Beliefs and Gender Inequality among Subordinates: Does His Ideology Matter More Than Hers?" Administrative Science Quarterly, 63:287–232. Publisher: Administrative Science Quarterly.

Castilla, E.
2008 “Gender, Race, and Meritocracy in Organizational Careers.” American Journal of Sociology, 113:1479–1526.

Castilla, E. J.
2011 “Bringing Managers Back In: Managerial Influences on Workplace Inequality.” American Sociological Review, 76:667–694.

Castilla, E. J.
2015 “Accounting for the Gap: A Firm Study Manipulating Organizational Accountability and Transparency in Pay Decisions.” Organization Science, 26:311–333.

Castilla, E. J. and Benard, S.
2010 “The Paradox of Meritocracy in Organizations.” Administrative Science Quarterly, 55:543–676.

Charles, K. K. and Guryan, J.
2011 “Studying discrimination: Fundamental challenges and recent progress.” Annual Review of Economics, 3:479–511.

Chen, D., Li, X., and Lai, F.
2017 “Gender discrimination in online peer-to-peer credit lending: evidence from a lending platform in China.” Electronic Commerce Research, 17:553–583.

Correll, S. J. and Benard, S.
2006 “Biased estimators? Comparing status and statistical theories of gender discrimination.” Advances in Group Processes, 23:89–116.

Correll, S. J. and Ridgeway, C. L.
2003 “Expectation states theory.” In J. Delamater (ed.), “Handbook of Social Psychology,” 29–51. New York: Kluwer Academic/Plenum.

Dahl, M. S., Dezső, C. L., and Ross, D. G.
2012 “Fatherhood and Managerial Style: How a Male CEO’s Children Affect the Wages of His Employees.” Administrative Science Quarterly, 57:669–693.

Das Gupta, M., Jiang Zhenghua, Li Bohua, Xie Zhenming, Woojin Chung, and Bae Hwa-Ok
2003 “Why is Son Preference so Persistent in East and South Asia? A Cross-Country Study of China, India and the Republic of Korea.” Journal of Development Studies, 40:153–187.

Dobbin, F. and Kalev, A.
2017 “Are Diversity Programs Merely Ceremonial? Evidence-Free Institutionalization.” In R. Greenwood, C. Oliver, T. B. Lawrence, and R. E. Meyer (eds.), “The Sage Handbook of Organizational Institutionalism,” 808–828. London: Sage.

Eisenhardt, K. M.
1989 “Agency Theory: An Assessment and Review.” The Academy of Management Review, 14:57–74.

Esurance
2014 “Why Women Pay Less for Car Insurance | Esurance.”

Feng, E.
2018 “China’s P2P lenders brace for renewed regulatory crackdown.” Financial Times.

Fernandez, R. M. and Greenberg, J.
2013 “Race, network hiring, and statistical discrimination.” Research in the Sociology of Work, 24:81–102.

Fernandez-Mateo, I. and Fernandez, R. M.
2016 “Bending the Pipeline? Executive Search and Gender Inequality in Hiring for Top Management Jobs.” Management Science, 62:3636–3655.

Frake, J.
2016 “Selling Out: The Inauthenticity Discount in the Craft Beer Industry.” Management Science.

Freedman, S. M. and Jin, G. Z.
2011 “Learning by doing with asymmetric information: Evidence from Prosper.com.” Working Paper 16855, National Bureau of Economic Research.

Gallus, J. and Frey, B. S.
2016 “Awards: A strategic management perspective.” Strategic Management Journal, 37:1699–1714.

Gibbons, R.
1998 “Incentives in Organizations.” The Journal of Economic Perspectives, 12:115–132.

Goldberg, M. S.
1982 “Discrimination, Nepotism, and Long-Run Wage Differentials.” The Quarterly Journal of Economics, 97:307–319.

Government Accountability Office
2011 “Person-to-person lending: New regulatory challenges could emerge as the industry grows.” Report to Congressional Committees GAO-11-613, United States Government Accountability Office.

Greenberg, J. and Mollick, E.
2017 “Activist Choice Homophily and the Crowdfunding of Female Founders.” Administrative Science Quarterly, 62:341–374. Publisher: SAGE Publications Inc.

Gwartney, J. and Haworth, C.
1974 “Employer Costs and Discrimination: The Case of Baseball.” Journal of Political Economy, 82:873–881.

Heilman, M. E.
2012 “Gender stereotypes and workplace bias.” Research in Organizational Behavior, 32:113–135.

Hong Fincher, L.
2016 Leftover Women: The Resurgence of Gender Inequality in China. London: Zed Books, 2nd edn.

Huang, R. H.
2018 “Online P2P Lending and Regulatory Responses in China: Opportunities and Challenges.” European Business Organization Law Review, 19:63–92.

Kalev, A., Dobbin, F., and Kelly, E.
2006 “Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies.” American Sociological Review, 71:589–617.

Kovács, B. and Sharkey, A. J.
2014 “The Paradox of Publicity How Awards Can Negatively Affect the Evaluation of Quality.” Administrative Science Quarterly, 59:1–33.

Lazear, E. P.
2000 “Performance Pay and Productivity.” American Economic Review, 90:1346–1361.

Leung, M. D. and Sharkey, A. J.
2013 “Out of Sight, Out of Mind? Evidence of Perceptual Factors in the Multiple-Category Discount.” Organization Science, 25:171–184.

Li, J.
2013 China P2P lending service industry whitepaper (中国 P2P 借贷服务行业白皮书）. China Business Network New Finance Research Center (第一财经). Beijing: China Economic Publishing House.

Malter, D.
2014 “On the causality and cause of returns to organizational status: Evidence from the grands crus classés of the Médoc.” Administrative Science Quarterly, 59:271–300.

Marquis, C. and Tilcsik, A.
2013 “Imprinting: Toward a Multilevel Theory.” Academy of Management Annals, 7:195–245.

McPherson, M., Smith-Lovin, L., and Cook, J. M.
2001 “Birds of a feather: Homophily in social networks.” Annual Review of Sociology, 27:415–444.

Pager, D.
2016 “Are Firms That Discriminate More Likely to Go Out of Business?” Sociological Science, 3:849–859.

Pager, D., Bonikowski, B., and Western, B.
2009 “Discrimination in a Low-Wage Labor Market: A Field Experiment.” American Sociological Review, 74:777–799.

Pager, D. and Shepherd, H.
2008 “The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets.” Annual Review of Sociology, 34:181–209.

Paluck, E. L. and Green, D. P.
2009 “Prejudice Reduction: What Works? A Review and Assessment of Research and Practice.” Annual Review of Psychology, 60:339–367.

Phelps, E. S.
1972 “The statistical theory of racism and sexism.” The American Economic Review, 62:659–661.

Podolny, J. M.
1993 “A status-based model of market competition.” American Journal of Sociology, 829–872.

Pope, D. G. and Sydnor, J. R.
2011 “What’s in a picture? Evidence of discrimination from Prosper.com.” Journal of Human Resources, 46:53–92.

Prendergast, C.
1999 “The Provision of Incentives in Firms.” Journal of Economic Literature, 37:7–63.

Qian, N.
2008 “Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance.” The Quarterly Journal of Economics, 123:1251–1285.

Ravina, E.
2012 “Love & loans: The effect of beauty and personal characteristics in credit markets.” SSRN Scholarly Paper ID 1101647, Social Science Research Network, Rochester, NY.

Ridgeway, C.
1991 “The Social Construction of Status Value: Gender and Other Nominal Characteristics.” Social Forces, 70:367–386.

Ridgeway, C. L.
2011 Framed by Gender: How Gender Inequality Persists in the Modern World. Oxford University Press. Google-Books-ID: NqoWmrAoQn0C.

Rissing, B. A. and Castilla, E. J.
2014 “House of Green Cards Statistical or Preference-Based Inequality in the Employment of Foreign Nationals.” American Sociological Review, 79:1226–1255.

Roodman, D.
2012 Due Diligence: An Impertinent Inquiry Into Microfinance. Washington, DC: Center For Global Development.

Rubineau, B. and Kang, Y.
2012 “Bias in White: A Longitudinal Natural Experiment Measuring Changes in Discrimination.” Management Science, 58:660–677.

Salganik, M. J., Dodds, P. S., and Watts, D. J.
2006 “Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market.” Science, 311:854–856.

Sen, A.
1992 “Missing women.” British Medical Journal, 304:587–588.

Siegel, J., Pyun, L., and Cheon, B. Y.
2018 “Multinational Firms, Labor Market Discrimination, and the Capture of Outsider’s Advantage by Exploiting the Social Divide.” Administrative Science Quarterly, 1–28.

Simcoe, T. S. and Waguespack, D. M.
2010 “Status, Quality, and Attention: What’s in a (Missing) Name?” Management Science, 57:274–290.

Siniver, E.
2011 “Testing for statistical discrimination: The case of immigrant physicians in Israel.” LABOUR, 25:155–166.

Srivastava, S. B. and Sherman, E. L.
2015 “Agents of Change or Cogs in the Machine? Reexamining the Influence of Female Managers on the Gender Wage Gap.” American Journal of Sociology, 120:1778–1808.

Stigler, G. J. and Becker, G. S.
1977 “De Gustibus Non Est Disputandum.” The American Economic Review, 67:76–90.

Tetlock, P. E.
1985 “Accountability: A Social Check on the Fundamental Attribution Error.” Social Psychology Quarterly, 48:227–236.

Thébaud, S.
2010 “Gender and Entrepreneurship as a Career Choice: Do Self-assessments of Ability Matter?” Social Psychology Quarterly, 73:288–304.

Thébaud, S. and Sharkey, A. J.
2016 “Unequal hard times: The influence of the great recession on gender bias in entrepreneurial financing.” Sociological Science, 3:1–31.

Table 1: Summary of lender decisions in the pre- and post-policy windows.

	Pre-policy period	Post-policy period	Diff	p-value	Both periods
Number of decisions	10975	14465	3490		25440
Avg investment size (RMB)	882	865	-17	0.65	872
Max investment size	100000	80000	-20000		100000
Min investment size	50	50	0		50
Stdev investment size	3220	2979	-241		3085
Unique lenders in window	2090	2521	431		3087
Unique loans in window	319	241	-78		558
Avg interest rate	13.9	14.3	0.4	0.00	14.1
Avg loan term (months)	8.1	9.3	1.1	0.00	8.8
% decisions same geography	9.0	7.6	-1.3	0.00	8.2
% decisions to women	18.8	12.7	-6.0	0.00	15.3
Avg # concurrent loans	2.7	4.5	1.8	0.00	3.7
Avg # concurrent loans, women	0.4	0.7	0.3	0.00	0.6
Avg time elapsed since first decision (h)	13.7	15.4	1.7	0.00	14.7

Table 2: Overall loan-level summaries of the pre- and post-policy windows.

	Pre-policy period	Post-policy period	Diff	p-value	Both periods
Number of loans	317	241	-76		558
Number of unique borrowers	242	214	-28		420
% loans male	86.4	85.9	-0.5	0.85	86.2
% loans female	13.6	14.1	0.5	0.85	13.8
Avg interest rate	14.3	14.3	0.0	0.92	14.3
Avg loan term (months)	6.7	7.6	0.9	0.03	7.0
Avg total loan size (RMB)	30542	51918	21376	0.00	39774
Max total loan size (RMB)	500000	500000	0		500000
Min total loan size (RMB)	3000	3000	0		3000
Avg borrower age	32.3	35.1	2.9	0.00	33.5
Avg fund time (hours)	6.4	10.5	4.0	0.06	8.2
Avg unique lenders / loan	34.6	60.0	25.4	0.00	45.6

Table 3: Linear regression of loan policy and borrower gender on investment size decision. Model 1 represents the most basic crosstab, Model 2 adds controls for the primary financial characteristics of the loan, Model 3 adds additional borrower controls, and Model 4 adds lender experience controls. The slight sample size differences in later models is the result of incomplete data for some of the added demographic variables.

	DV: decision size (Chinese RMB)			
	(1)	(2)	(3)	(4)
Guarantee policy (true)	52.98	−20.79	−22.80	7.11
	(42.67)	(44.97)	(67.10)	(63.30)
	$p = 0.22$	$p = 0.65$	$p = 0.74$	$p = 0.92$
Borrower sex (female)	271.84***	184.95**	188.30**	216.04**
	(75.43)	(76.46)	(90.51)	(85.38)
	$p = 0.0004$	$p = 0.02$	$p = 0.04$	$p = 0.02$
Borrower sex * policy	−422.26***	−221.91**	−323.50**	−383.91***
	(107.72)	(110.56)	(145.14)	(136.87)
	$p = 0.0001$	$p = 0.05$	$p = 0.03$	$p = 0.01$
Loan interest rate		42.06***	49.32***	23.12
		(12.07)	(15.41)	(14.72)
		$p = 0.0005$	$p = 0.002$	$p = 0.12$
Loan term		13.63**	1.28	−5.74
		(5.95)	(7.69)	(7.27)
		$p = 0.03$	$p = 0.87$	$p = 0.43$
Loan size		0.002***	0.002***	0.002***
		(0.0002)	(0.0002)	(0.0002)
		$p = 0.00$	$p = 0.00$	$p = 0.00$
Elapsed time posted			3.75***	3.46***
			(0.90)	(0.85)
			$p = 0.00004$	$p = 0.0001$
Count of concurrent active loans			−19.20**	−18.78**
			(9.76)	(9.21)
			$p = 0.05$	$p = 0.05$
...count of above, women			−27.15	−17.06
			(44.11)	(41.61)
			$p = 0.54$	$p = 0.69$
Geographic overlap (true)				207.01***
				(67.95)
				$p = 0.003$
Count of past decisions				−5.97***
				(1.38)
				$p = 0.00002$
...count of above, to women				0.01***
				(0.0002)

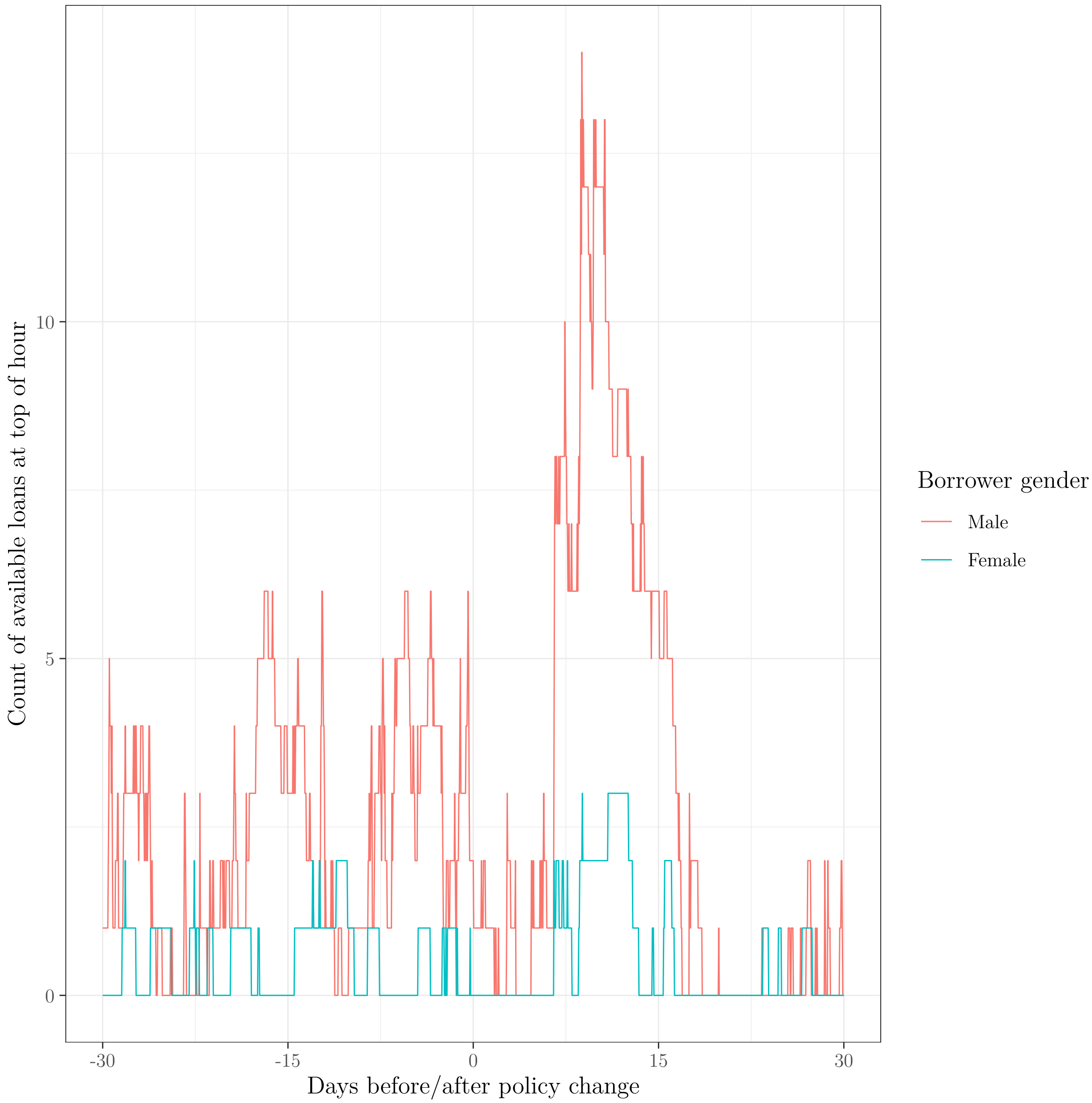
				$p = 0.00$
Total volume of previous decisions				7.99
				(8.02)
				$p = 0.32$
Count of previous defaults				-708.27^{***}
				(37.80)
				$p = 0.00$
Constant	831.17^{***}	-19.69	366.78	715.56
	(32.67)	(480.20)	(1,089.34)	(1,027.29)
	$p = 0.00$	$p = 0.97$	$p = 0.74$	$p = 0.49$
Controls for credit rating	No	Yes	Yes	Yes
Other borrower characteristics†	No	No	Yes	Yes
Observations	25,440	25,440	24,212	24,212
R^2	0.001	0.01	0.02	0.13
Adjusted R^2	0.001	0.01	0.01	0.12

Note: *p<0.1; **p<0.05; ***p<0.01

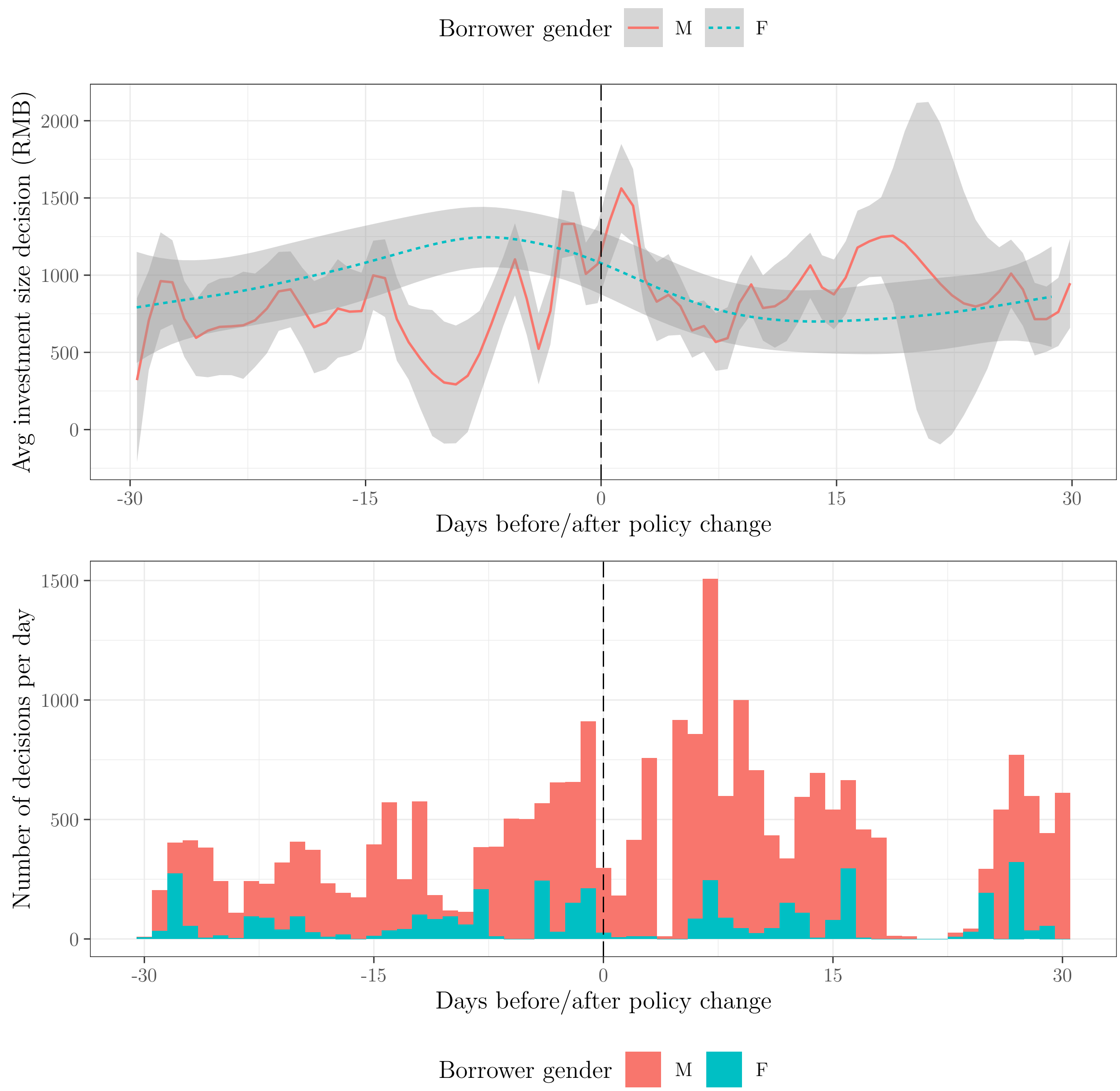
†In addition to gender, borrower characteristics included loan purpose, province, age, level of academic degree, salary range, office characteristics (type, size, and industry), and the existence of a car, house, spouse, or children.

APPENDIX

Figures and tables to supplement the main analyses and provide additional context.



Appendix Figure A1: Count of the number of loans at the top of each hour that had already received their first investment but had not yet received their last investment. This approximates the amount of between-borrower choice available to a lender at any given point in time. For a summary of loan-level characteristics, see Table 2.



Appendix Figure A2: Temporal descriptive statistics for the 30 days +/- the policy change. Smoothed with generalized additive models using basis dimension $k = 75$. The difference in the smoothness of the two lines is the result of the disproportionate volume of decisions to men; see Table 1. Histogram using $binwidth = 1$ day.

Appendix Table A1: Loan outcomes as of August 2013.

Category	Loan count	Male %	Female %
Already complete (success)	524	85.9	14.1
Bad debt	25	96.0	4.0
In repayment	8	75.0	25.0
Overdue	1	100.0	0.0

Appendix Table A2: Distribution of loan use categories.

Category	Loan count	Pre-policy	Post-policy
Short-term turnover	259	97	162
Personal consumption	155	145	10
Other	75	47	28
Startup investment	55	23	32
Redecoration	6	2	4
Wedding preparations	3	2	1
Education/training	2	1	1
Automobile	2	0	2
Housing	1	0	1

Appendix Table A3: Averages of credit rating categories.

Credit rating	Number of loans	Interest rate %	Loan term	Loan size
AA	3	6.9	3.0	4,333
A	2	13.0	9.0	42,500
B	65	11.4	7.4	31,708
C	87	13.3	7.4	43,393
D	119	13.7	7.5	67,159
E	135	14.9	7.4	40,273
HR	147	16.2	6.0	19,259

Appendix Table A4: Basic exposition of borrower gender effects.

Borrower	Post-policy	Avg investment	Decision count
M	FALSE	831.2	8,917
M	TRUE	884.1	12,621
F	FALSE	1,103.0	2,058
F	TRUE	733.7	1,844

Appendix Table A5: Summary of decision counts at the lender level for the lender fixed effects subsample analyses. To be included in the sample, a lender must have made at least one loan to a man and one to a woman in both periods; 405 lenders meet this criteria, making a total of 11,874 total decisions.

Statistic	Mean	St. Dev.	Min	Max
Count of decisions to women, pre-policy	2.9	2.3	1	16
Count of decisions to women, post-policy	2.3	2.0	1	16
Count of decisions to men, pre-policy	11.5	11.5	1	83
Count of decisions to men, post-policy	12.7	11.6	1	96

Appendix Table A6: Lender fixed effects OLS regression of loan policy and gender on investment size decision. Model 1 is the subsample ineligible for fixed effects, Model 2 is the fixed effects eligible subsample with lender fixed effects included. See table Table A5 for additional details about the sample. Model 3 is a fixed effects robustness test where the sample is created from lending decisions from sixty days before and after the policy.

	DV: decision size (Chinese RMB)		
	FE ineligible	Main FE	Expanded FE
	(1)	(2)	(3)
Guarantee policy (true)	231.95^{***}	-104.62	-95.56
	$p = 0.01$	$p = 0.35$	$p = 0.16$
Borrower sex (female)	204.77^{*}	221.06^{*}	106.93
	$p = 0.10$	$p = 0.07$	$p = 0.16$
Borrower sex * policy	-389.23^{**}	-190.63	-199.74^{*}
	$p = 0.05$	$p = 0.32$	$p = 0.06$
Loan interest rate	35.54^{*}	42.64^{**}	38.16^{***}
	$p = 0.09$	$p = 0.05$	$p = 0.003$
Loan term	0.93	-22.05^{**}	-14.75^{**}
	$p = 0.93$	$p = 0.04$	$p = 0.03$
Loan size	0.001^{***}	0.004^{***}	0.003^{***}
	$p = 0.0001$	$p = 0.00$	$p = 0.00$
Elapsed time posted	3.98^{***}	1.40	-0.23
	$p = 0.0003$	$p = 0.31$	$p = 0.78$
Count of concurrent active loans	-11.69	-33.27^{**}	-16.49^{**}
	$p = 0.34$	$p = 0.02$	$p = 0.04$
...count of above, women	29.89	-33.08	-52.89^{*}
	$p = 0.59$	$p = 0.60$	$p = 0.09$
Geographic overlap (true)	198.57^{**}	48.80	55.45
	$p = 0.03$	$p = 0.64$	$p = 0.38$
Count of past decisions	-19.18^{***}	-1.31	5.49^{**}
	$p = 0.00$	$p = 0.82$	$p = 0.03$
...count of above, to women	0.02^{***}	0.001	-0.003^{***}
	$p = 0.00$	$p = 0.23$	$p = 0.00$
Total volume of previous decisions	40.77^{**}	33.91	-20.47
	$p = 0.02$	$p = 0.32$	$p = 0.21$
Count of previous defaults	-274.76^{***}	-306.27	-4.87
	$p = 0.005$	$p = 0.27$	$p = 0.95$
Constant	723.23	-186.59	$9,927.90^{***}$
	$p = 0.62$	$p = 0.92$	$p = 0.00$
Controls for credit rating	Yes	Yes	Yes
Borrower controls†	Yes	Yes	Yes
Lender fixed effects	No	Yes	Yes

Observations	12,980	11,232	27,359
R^2	0.09	0.28	0.26
Adjusted R^2	0.09	0.25	0.24
Note:	*p<0.1; **p<0.05; ***p<0.01		
	†See Table 3 for list of variables.		

Appendix Table A7: Moderating effect of the number of past decisions made by a lender (Model 1) and the number of previous defaults a lender experienced (Model 2).

	DV: decision size (Chinese RMB)	
	(1)	(2)
Guarantee policy (true)	123.03^{*}	13.29
	$p = 0.09$	$p = 0.84$
Borrower sex (female)	32.71	139.75
	$p = 0.75$	$p = 0.11$
Borrower sex * policy	-241.78	-333.74^{**}
	$p = 0.13$	$p = 0.02$
Loan interest rate	24.19	24.56^{*}
	$p = 0.11$	$p = 0.10$
Loan term	-3.34	-5.47
	$p = 0.65$	$p = 0.46$
Loan size	0.002^{***}	0.002^{***}
	$p = 0.00$	$p = 0.00$
Elapsed time posted	3.53^{***}	3.47^{***}
	$p = 0.00004$	$p = 0.00005$
Count of concurrent active loans	-18.64^{**}	-18.83^{**}
	$p = 0.05$	$p = 0.05$
...count of above, women	-24.46	-18.18
	$p = 0.56$	$p = 0.67$
Geographic overlap (true)	206.47^{***}	207.37^{***}
	$p = 0.003$	$p = 0.003$
Count of past decisions	-4.55^{***}	-5.67^{***}
	$p = 0.002$	$p = 0.00004$
...count of above, to women	5.49	5.33
	$p = 0.50$	$p = 0.51$
Total volume of previous decisions	0.01^{***}	0.01^{***}
	$p = 0.00$	$p = 0.00$
Count of previous defaults	-718.68^{***}	-725.05^{***}
	$p = 0.00$	$p = 0.00$
Policy * previous loan count	-2.60^{***}	
	$p = 0.0002$	
Borrower sex * previous loan count	4.35^{***}	
	$p = 0.0004$	
Policy * borrower sex * previous loan count	-3.50^{**}	
	$p = 0.05$	
Policy * previous defaults		-76.25
		$p = 0.18$
Borrower sex * previous defaults		736.77^{***}
		$p = 0.00$

Policy * borrower sex * previous defaults		-528.30^{***}
		$p = 0.001$
Constant	717.59	714.05
	$p = 0.49$	$p = 0.49$
Controls for credit rating	Yes	Yes
Other borrower characteristics†	Yes	Yes
Observations	24,212	24,212
R^2	0.13	0.13
Adjusted R^2	0.12	0.12
Note:	*p<0.1; **p<0.05; ***p<0.01	
	†In addition to gender, borrower characteristics included loan purpose, province, age, level of academic degree, salary range, office characteristics (type, size, and industry), and the existence of a car, house, spouse, or children.	

Appendix Table A8: Tests for sample heterogeneity: OLS model regressing policy and demographics on investment size decisions equal to or greater than 200 RMB, using final model from Table 3. This reduces the sample roughly in half.

	DV: decision size (Chinese RMB)
Guarantee policy (true)	-29.3
	$p = 0.8$
Borrower sex (female)	350.1^{**}
	$p = 0.02$
Borrower sex * policy	-636.3^{***}
	$p = 0.01$
Loan interest rate	41.8
	$p = 0.2$
Loan term	-11.9
	$p = 0.4$
Loan size	0.004^{***}
	$p = 0.0$
Elapsed time posted	7.8^{***}
	$p = 0.000$
Count of concurrent active loans	-25.6
	$p = 0.2$
...count of above, women	-23.6
	$p = 0.8$
Geographic overlap (true)	287.7^{**}
	$p = 0.02$
Count of past decisions	-6.4^{***}
	$p = 0.003$
...count of above, to women	0.01^{***}
	$p = 0.0$
Total volume of previous decisions	-0.8
	$p = 1.0$
Count of previous defaults	-702.5^{***}
	$p = 0.0$
Constant	$1,036.9$
	$p = 0.6$
Controls for credit rating	Yes
Borrower controls†	Yes
Observations	13,878
R^2	0.1
Adjusted R^2	0.1
Note:	*p<0.1; **p<0.05; ***p<0.01
	†See Table 3 for list of variables.

Appendix Table A9: Test for lender gender effects. More fully specified OLS model regressing policy and demographics on investment size decision.

	DV: decision size (Chinese RMB)
Guarantee policy (true)	63.04
	$p = 0.35$
Borrower sex (female)	265.12^{***}
	$p = 0.005$
Borrower sex * policy	-444.06^{***}
	$p = 0.003$
Lender sex (female)	138.38^{*}
	$p = 0.09$
Lender sex * policy	-254.93^{**}
	$p = 0.02$
Lender sex * borrower sex	-226.25
	$p = 0.20$
Lender sex * borrower sex * policy	278.70
	$p = 0.27$
Loan interest rate	23.41
	$p = 0.12$
Loan term	-5.80
	$p = 0.43$
Loan size	0.002^{***}
	$p = 0.00$
Elapsed time posted	3.46^{***}
	$p = 0.0001$
Count of concurrent active loans	-18.75^{**}
	$p = 0.05$
...count of above, women	-16.54
	$p = 0.70$
Geographic overlap (true)	207.68^{***}
	$p = 0.003$
Count of past decisions	-6.09^{***}
	$p = 0.0000$
...count of above, to women	0.01^{***}
	$p = 0.00$
Total volume of previous decisions	8.53
	$p = 0.29$
Count of previous defaults	-706.30^{***}
	$p = 0.00$
Constant	670.72
	$p = 0.52$
Controls for credit rating	Yes

Other borrower characteristics[†]	Yes
Observations	24,212
R^2	0.13
Adjusted R^2	0.12
Note:	*p<0.1; **p<0.05; ***p<0.01 [†]See Table 3 for list of variables.

Appendix Table A10: Basic exposition of gender effects accounting for lender gender.

Lender	Borrower	Post-policy	Avg investment	Decision count
M	M	FALSE	815.4	7,042
M	M	TRUE	939.4	9,596
M	F	FALSE	1,127.2	1,613
M	F	TRUE	803.2	1,398
F	M	FALSE	890.5	1,875
F	M	TRUE	708.9	3,025
F	F	FALSE	1,015.2	445
F	F	TRUE	515.9	446

Appendix Table A11: Full sample divided into three separate samples based on credit ratings: 1) HR only, 2) AA, A, B, C, and 3) D and E. The final model from Table 3 was then run on each sample, exluding credit rating controls so that cross-model comparisons are meaningful.

	DV: decision size (Chinese RMB)		
	HR	AA,A,B,C	D,E
	(1)	(2)	(3)
Guarantee policy (true)	−42.71	784.96**	93.53
	p = 0.88	p = 0.02	p = 0.32
Borrower sex (female)	370.49	−142.51	−55.48
	p = 0.20	p = 0.79	p = 0.70
Borrower sex * policy	−262.64	−157.98	−406.62*
	p = 0.56	p = 0.83	p = 0.07
Loan interest rate	−25.13	126.12**	28.23
	p = 0.51	p = 0.03	p = 0.28
Loan term	−3.75	10.71	−6.24
	p = 0.91	p = 0.63	p = 0.64
Loan size	0.0001	0.001	0.003***
	p = 0.94	p = 0.38	p = 0.00
Elapsed time posted	7.23**	6.48**	4.84***
	p = 0.02	p = 0.04	p = 0.0000
Count of concurrent active loans	19.01	−4.43	−11.17
	p = 0.59	p = 0.87	p = 0.40
...count of above, women	−60.62	1.27	−77.01
	p = 0.57	p = 0.99	p = 0.21
Geographic overlap (true)	−19.41	18.68	352.54***
	p = 0.92	p = 0.89	p = 0.0002
Count of past decisions	−3.73	−0.14	−8.31***
	p = 0.25	p = 0.96	p = 0.0000
...count of above, to women	0.01***	0.01***	0.01***
	p = 0.00	p = 0.00	p = 0.00
Total volume of previous decisions	−10.02	−24.55	23.16**
	p = 0.59	p = 0.13	p = 0.04
Count of previous defaults	−454.71***	−952.98***	−683.88***
	p = 0.0000	p = 0.00	p = 0.00
Constant	−363.21	569.52	248.64
	p = 0.84	p = 0.64	p = 0.74
Other borrower characteristics†	Yes	Yes	Yes
Observations	3,568	6,917	13,727
R^2	0.09	0.16	0.13
Adjusted R^2	0.07	0.15	0.13

Note: *p<0.1; **p<0.05; ***p<0.01
†See Table 3 for list of variables.

Appendix Table A12: Recreation of final Table 3 model, limited by loan use category. Model 1 is restricted to short-term turnover and personal consumption loans, and Model 2 is restricted to only short-term turnover loans.

	DV: decision size (Chinese RMB)	
	(1)	(2)
Guarantee policy (true)	60.43	33.98
	$p = 0.49$	$p = 0.77$
Borrower sex (female)	216.89*	535.38***
	$p = 0.07$	$p = 0.01$
Borrower sex * policy	-402.94^{**}	-671.14^{**}
	$p = 0.05$	$p = 0.02$
Loan interest rate	8.04	1.11
	$p = 0.65$	$p = 0.97$
Loan term	0.78	11.08
	$p = 0.93$	$p = 0.44$
Loan size	0.002***	0.002***
	$p = 0.00$	$p = 0.000002$
Elapsed time posted	6.45***	7.50***
	$p = 0.0000002$	$p = 0.0000002$
Count of concurrent active loans	-10.95	-25.90^{*}
	$p = 0.38$	$p = 0.08$
...count of above, women	14.85	44.96
	$p = 0.80$	$p = 0.51$
Geographic overlap (true)	356.94***	84.08
	$p = 0.00005$	$p = 0.42$
Count of past decisions	-5.92^{***}	-7.79^{***}
	$p = 0.001$	$p = 0.0002$
...count of above, to women	0.01***	0.01***
	$p = 0.00$	$p = 0.00$
Total volume of previous decisions	7.12	17.66
	$p = 0.48$	$p = 0.15$
Count of previous defaults	-660.74^{***}	-703.65^{***}
	$p = 0.00$	$p = 0.00$
Constant	-239.07	489.38
	$p = 0.76$	$p = 0.59$
Controls for credit rating	Yes	Yes
Other borrower characteristics†	Yes	Yes
Observations	14,508	10,436
R^2	0.13	0.13
Adjusted R^2	0.12	0.13
Note:	*p<0.1; **p<0.05; ***p<0.01	

[†]See Table 3 for list of variables.

Appendix Table A13: OLS models using placebo samples. Note: Models 1 and 2 use a sample constructed by shifting the policy date 30-days prior to the actual date, and Models 2 and 3 do the same for a sample 30-days after. Models 2 and 4 mirror Table 3 Model 4.

	DV: decision size (Chinese RMB)			
	(1)	(2)	(3)	(4)
Placebo policy 1 (true)	-221.1^{***}	-145.0^{**}		
	$p = 0.000$	$p = 0.04$		
Placebo policy 2 (true)			-203.9^{***}	-230.9^{***}
			$p = 0.0$	$p = 0.0$
Borrower sex (female)	10.4	59.9	-111.6^{*}	-219.6^{***}
	$p = 1.0$	$p = 0.7$	$p = 0.1$	$p = 0.01$
Borrower sex * placebo policy 1	88.0	14.3		
	$p = 0.6$	$p = 1.0$		
Borrower sex * placebo policy 2			98.0	195.4^{**}
			$p = 0.3$	$p = 0.05$
Loan interest rate	42.7^{***}	7.2	34.1^{***}	51.4^{***}
	$p = 0.001$	$p = 0.7$	$p = 0.002$	$p = 0.000$
Loan term	9.6	8.4	4.5	-10.9
	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.2$
Loan size	0.003^{***}	0.004^{***}	0.001^{***}	0.001^{***}
	$p = 0.0$	$p = 0.0$	$p = 0.0$	$p = 0.001$
Elapsed time posted		2.6^{*}		4.6^{***}
		$p = 0.1$		$p = 0.0$
Count of concurrent active loans		-32.2		-0.7
		$p = 0.2$		$p = 0.9$
...count of above, women		-34.8		-60.7^{***}
		$p = 0.6$		$p = 0.005$
Geographic overlap (true)		366.9^{***}		30.8
		$p = 0.000$		$p = 0.6$
Count of past decisions		-5.2^{***}		-5.0^{***}
		$p = 0.002$		$p = 0.000$
...count of above, to women		0.01^{***}		0.01^{***}
		$p = 0.0$		$p = 0.0$
Total volume of previous decisions		3.3		-5.1
		$p = 0.8$		$p = 0.4$
Count of previous defaults		-700.2^{***}		-470.1^{***}
		$p = 0.0$		$p = 0.0$
Constant	88.6	$1,007.0$	410.2	-9.5
	$p = 0.9$	$p = 0.4$	$p = 0.4$	$p = 1.0$
Controls for credit rating	Yes	Yes	Yes	Yes
Other borrower characteristics†	No	Yes	No	Yes

Observations	17,946	16,908	35,576	34,548
R^2	0.03	0.1	0.01	0.1
Adjusted R^2	0.02	0.1	0.01	0.1
Note:	*p<0.1; **p<0.05; ***p<0.01			
	†See Table 3 for list of variables.			