THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Why does regional information matter? Evidence from peer-to-peer lending

Citation for published version:
Wang, T, Zhao, S & Shen, X 2021, 'Why does regional information matter? Evidence from peer-to-peer lending', *The European Journal of Finance*, vol. 27, no. 4-5, pp. 346-366.
https://doi.org/10.1080/1351847X.2020.1720262

Digital Object Identifier (DOI):
10.1080/1351847X.2020.1720262

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
The European Journal of Finance

Publisher Rights Statement:
This is an Accepted Manuscript of an article published by Taylor & Francis in The European Journal of Finance on 28 January 2020, available online: https://www.tandfonline.com/doi/full/10.1080/1351847X.2020.1720262

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 23. Feb. 2022

Why does Regional Information Matter? Evidence from Peer-to-Peer Lending

Abstract

In this paper, we study regional discrimination in a peer-to-peer lending scenario and provide novel empirical evidence for theories of soft information collection and information cost. We find that the regional information matters for borrowers' funding probabilities and that discrimination is profit-oriented or taste-oriented depending on the specific region. Moreover, using borrowers' birthplace as an instrumental variable, we find no evidence of genuine discrimination based purely on region in the peer-to-peer lending market.
Keywords: Peer-to-Peer Lending, Regional Discrimination, Information Cost Theory

February 10, 2020

1. Introduction

Information acquisition plays a central role in a lending relationship (Lin et al., 2013; Liberti and Petersen, 2018). Lenders often allocate a considerable amount of time and effort to the information about potential borrowers so as to reduce the information asymmetry and gain a strategic advantage in the ensuing financial transaction (Agarwal and Hauswald, 2010). Based on collected information, borrower groups with certain characteristics may be preferred by lenders, while others may be discriminated against. In the financial sector, one of the main obstacles to identifying the discrimination source is quantifying the interaction process between borrowers and lenders. Individuals usually have complicated social interactions, which are not perfectly observable by researchers, while endogeneity, in this case, becomes a notable concern (Dell'Ariccia et al., 2012). Turner (1999) raises information access as one of the most significant challenges for empirical studies on discrimination. In their statement, omitted variables may mislead researchers to conclude that discrimination exists when none actually does. Guryan and Charles (2013) have a similar concern with omitted variables, which would overestimate the discrimination's magnitude.

The online lending market for individuals, often referred to as the peer-to-peer (P2P) lending market, is particularly suited for inquiries into the subjective aspects of individual decision making in a lending relationship. P2P lending refers to the unsecured loans generated by lenders to borrowers through lending platforms (Funk et al., 2011). It is an online service that directly matches lenders and borrowers and provides the chance for open and transparent micro-credit transactions between individuals, integrating Internet technology with micro-finance. Depending on different scenarios, P2P lending markets could play as either a substitute for or a complement of the traditional banking system (Tang, 2019). Most P2P lending companies operate and provide services entirely online, reducing loan costs by forgoing the expensive intermediaries (Klafft, 2008) and facilitating the matching between borrowers and lenders. These platforms disclose various types of borrower information, including credit history, as well as various personal statements, such as education, salary levels, and the borrowers' residential area. Furthermore, they usually earn a profit by charging a fee as the cost of information provision. Compared with traditional financial institutions, the trading pattern is more transparent. More importantly, the lenders can only access borrower profile data via the P2P platforms; therefore, researchers are able to collect the same information as lenders.

In this paper, we study the source of regional discrimination in a P2P lending scenario. Observed region-oriented differences in funding probabilities can be due to taste-related factors, profit-related factors or a combination of both. For instance, given the lower economic development and information diffusion level within certain regions, the information of loan requests from these regions can be less reliable, making access to credit at fair prices more difficult for qualified borrowers. As a result of these adverse selection effects, the average default rate of funded projects from certain regions will increase, along with worsening beliefs among lenders about the average quality of the borrowers. In this case, regional information is a useful signal of some characteristics that cannot be directly observed but that are related to the borrowers' capability of paying back their debt. Subsequently, a self-reinforcing Arrowian profit-oriented discrimination naturally arises (Arrow, 1998). On the other hand, regions where the borrowers reside may have no pecuniary implications on the lenders, which can affect their decisions from a non-economic perspective, in which case a Beckerian taste-oriented discrimination (Becker, 2010) situation arises.

We use data from Renrendai, a leading Chinese P2P platform, in order to empirically examine the existence and sources of regional discrimination in the online lending market. We define regional discrimination as the phenomenon in which after controlling for all of the explicit information shared on the platform, the information about the economic region where the loan applicants reside is still significantly connected to the success rate of loans. To motivate our empirical strategy, we use signaling theory (Spence, 1978), information cost theory (Meyer, 1967) and soft information theory(Liberti and Petersen, 2018) to develop testable statements about the sources of regional discrimination.

In the lending market, taste-oriented and profit-oriented discrimination behave differently. If the discrimination is taste-oriented, then differentiated groups have to offer better terms in order to qualify themselves, such as providing more detailed and complete personal information, offering higher interest rates and reducing uncertainty (Larrimore et al., 2011). The increased interest income and reduced credit risk will make compensate for the discrepancy caused by prejudice and meet the business demand (Guo et al., 2016). In this case, the average financial performance is better when the borrower belongs to the discriminated group. However, if the existing discrimination is mainly profit-oriented, then from the lenders' perspective, the performance of borrowers is highly likely to be correlated with their group membership (Turner, 1999). According to the theory of information cost (Meyer, 1967), lenders are inclined to use group memberships as proxy variables and to have preferred choices over disparate groups of people, especially when the acquisition cost for the borrowers' exact information is prohibitive. In this case, the financial performance of the differentiated borrowers should be lower than average.

In this paper, we study the lenders' regional discrimination against online loan applicants from different regions of China, respectively. Based on our analysis, we find strong evidence of regional discrimination: with all other factors being equal, borrowers from the Eastern region of China have a higher success rate than the average, while the loan success rate is significantly lower from the Western region and Northeastern region of China. When checking the actual default rate of both regions, we find that discrimination against applicants from the Eastern and Northeastern regions is profit-oriented, while for their Western region counterparts it is mainly taste-oriented. We also find that the discrimination against Western region applicants can be explained by the two regional indicators of economic development and financial risk, while the discrimination against Eastern and Northeastern borrowers still exists after controlling for the regional information. Moreover, using borrowers' birthplace as the instrumental variable, we find no evidence of the existence of genuinely region-based discrimination. Based on these results, we provide an interpretation consistent with the theories of home bias and information loss.[1] For applicants who reside near the lenders, lenders tend to increase the weight of soft information in their decision-making process and the discrimination is more likely to be profit-oriented. For the remotely-residing applicants, lenders prefer to use hard information about the region to make a decision, and the discrimination tends to be taste-oriented.

Our paper contributes to the literature in three main aspects. First, it adds to the stream of regional analysis on P2P lending. For instance, Lin et al. (2013) use Prosper, an online lending platform based in the U.S., as a sample and find that loan applications are more likely to succeed among people within the same region than those across different regions. Burtch et al. (2014) empirically examine the impacts of cultural differences and geographic distance between borrowers and lenders on lending. Our paper takes a step further by exploring the reason for such regional discrimination and empirically tests our explanations.

Secondly, our work provides novel empirical evidence for information cost theory and soft information usage in financial transactions. Although P2P platforms have attempted to reduce the information gap by examining qualification validating data (Serrano-Cinca et al., 2015; Weiss et al., 2010) or by designing appropriate mechanisms (Wei and Lin, 2016), the credit risk was still high on average for P2P loans due to the low investment threshold and borrowing requirements (Pope and Sydnor, 2011). For individual lenders, collecting

[1] as stated in Agarwal and Ben-David (2018)

information is costly and time consuming. Therefore, these theories predict that lenders tend to make use of some observable information as a noisy signal for the unobservables. Our results are aligned with Liao et al. (2014a), who state that lenders can infer extra information from borrowers' profiles. In particular, lenders are likely to use information about the regions where borrowers reside as a proxy for the unobservable repayment ability of borrowers.

Finally, in view of the unbalanced development in provinces throughout China, many scholars have found evidence for market discrimination. According to Wang and Zheng (2017), loan success rates with diverse aims in various categories are impacted by the regional economic development to varying degrees. The basic credit and identity information of borrowers in underdeveloped cities is of greater concern than that of borrowers in developed cities, explaining the disparity between borrowing success rates in less-developed compared to more-developed cities. Jiang and Zhou (2016) also found the same economic effect. Liao et al. (2014b) show that regional discrimination at the provincial level exists in the peer- to-peer lending market, and that this taste-oriented discrimination is a kind of irrational behavior. In contrast, our paper focuses on the regional level, and we investigate how lenders' decision making is affected by information on the borrowers' regions of residence. Because of discrimination, people in high-income regions are inclined to decrease their financing costs, while low-income region residents need to increase interest rates to attract lenders. This discrimination imposes additional difficulties on borrowers from less-developed regions and prevents some from obtaining proper funding opportunities. Our paper also has important implications for regulators from the perspective of balancing the levels of regional development and promoting an equal opportunity policy.

The remainder of the paper is organized as follows. In Section 2, we review related literature and briefly introduce regional discrimination in China. In Section 3, we show the dataset and the settings of variables. Section 4 examines the existence of regional discrimination and regional differences in loan default rates. We also include two provincial economic factors and apply the applicants' birthplace as an instrumental variable to identify the reason for regional discrimination. Section 5 concludes.

2. Background

2.1. Regional Discrimination in Financial Transactions

Regional discrimination in a lending relationship, sometimes rephrased as regional redlining, refers to the phenomenon of lenders discriminating against borrowers from certain allegedly redlined areas. The current literature on regional redlining focuses mainly on institutional lenders. The majority of studies (Benston and Horsky, 1992; Schafer and Ladd, 1981; Munnell et al., 1993) have found little evidence of the differences between households in the redlined areas compared with the controlled areas in terms of their ability to secure lending offers. However, this conclusion is restricted to banks and financial institutions in the U.S. since they are highly regulated by the Community Reinvestment Act, which imposes an affirmative action on lenders. For individual lenders who are not constrained by such regulatory policies, evidence shows that lenders do have some preferences based on borrowers' regional information (Burtch et al., 2014).

In China, regional discrimination and its influence on financial decisions have been widely discussed. Historically, internal migration in China was tightly controlled for management reasons, and many barriers to free mobility, such as the *Hukou* system (Afridi et al., 2015), have not been entirely eliminated. Under this system, every Chinese citizen was legally bound to register his or her single permanent place of residence, and strict controls were imposed on the mobility of *Hukou* holders. Due to this immobility, regions have strong subcultures and have developed different lifestyles, which may affect lenders' beliefs and, subsequently, their decisions in a potential lending relationship (Liao et al., 2014a,b; Jiang and Zhou, 2016; Peng et al., 2016). Moreover, from a macroeconomics perspective, regions across China have unbalanced economic development, regulatory maturity and business environments. Therefore, the regional location of a Chinese citizen may impact on their employment opportunities, health and education services and benefits. It is also possible for lenders to use regional information as a proxy variable for some unobservable characteristics of the borrowers.

According to the economic region classification issued by the National Bureau of Statistics of China, China can be divided into four regions: Eastern, Western, Central and Northeastern. There are a considerable number of differences between each of these regions, such as in economic development (Kanbur and Zhang, 1999), labor markets (Cai et al., 2002) and in lifestyles (Feng et al., 2009) among those regions. Provincial-level studies (Peng et al., 2016) have shown statistical evidence that lenders prefer borrowers from high-income provinces. Focusing on the regional level, our research will explore the underlying reasons for such discrimination.

2.2. Related Literature

Broadly speaking, discrimination refers to the different treatment of a specific group of people. These groups must vary according to some characteristics valued in the market. The reason for discrimination has thus been under debate, and two main branches of explanations exist: taste-oriented and profit-oriented discrimination. Becker (2010) states that if individuals wish to exercise a discriminatory preference, they must put prejudice ahead of profits and behave as if they are willing to pay something, either directly or by forgoing income, to avoid interaction with those individuals. Therefore, the discrimination is based on individual taste and results from personal preference towards that group. Depending on the situation, taste-oriented discrimination could be competed away by the market (Turner, 1999; Han, 2011) or be sustainable (Pęski and Szentes, 2013).

Phelps (1972) and Arnott (1972) define discrimination in an alternative way. They emphasize the role of information asymmetry in the market. In their work, group memberships may act as proxy variables for some important characteristics that are relevant to production and profit, but may not be directly observable or are cost prohibitive for gathering information. Therefore, this kind of statistical discrimination is classified as profit-oriented (Guryan and Charles, 2013), and is based on rational optimizing behavior and imperfect information. Researchers have already conducted several tests in order to measure and identify discrimination in the contexts of the labor market (Turner, 1999; Altonji and Pierret, 2001) and police investigations (Knowles et al., 2001).

There is both experimental and empirical evidence of the existence of discrimination. For instance, Fershtman and Gneezy (2001) designed games to measure the trust between people from different ethnic groups. Bertrand and Mullainathan (2004) sent fictitious CVs in response to Help-Wanted advertisements, with applicants' names randomly assigned for majority/minority groups and check whether the call-back rate is different for 'applicants' with different name-implied ethnic characteristics. Muravyev et al. (2009), Bellucci et al. (2010), Cheng et al. (2015) and Guariglia and Mateut (2016) use empirical data to analyse the gaps between the rejection rates of projects (loan, mortgage, etc.) among individuals in the different groups.

Regarding the criteria upon which the discrimination takes place, scholars have surveyed a number of areas, such as gender, age, culture, and race (Schafer and Ladd, 1981; Goering, 1996; Blanchflower et al., 2003; Ravina, 2007; Calomiris et al., 1994; Deku et al., 2016). In addition, regional factors are also discussed in the context of mortgage lending(Ladd, 1998), banking industry (Edie and Riefler, 1931), institutional development (Huyghebaert and Wang, 2016) and property market (Schafer and Ladd, 1981).

Factors affecting P2P investors' trading decisions are widely studied in the current literature and are generally divided into hard and soft information. Hard information is composed of personal data (Pope and Sydnor, 2011; Claessens et al., 2018), loan terms (Klafft, 2008), and proposed interest rates (Puro, 2010). Conversely, soft information consists mainly of the information that borrowers voluntarily publish on the P2P platform (Iyer, 2009; Han, 2018; Jiang et al., 2018), such as narrative statements (Caldieraro et al., 2018; Netzer et al., 2019), individuals' appearance (Duarte et al., 2012), public observable communications between borrowers and lenders (Xu and Chau, 2018), social ties (Freedman and Jin, 2017), and linguistic features (Larrimore et al., 2011). Both hard and soft data are important factors which affect lender decisions (Dorfleitner et al., 2016; Agarwal and Ben-David, 2018; Thompson and Cowton, 2004; Feller et al., 2017). Based on the data from Prosper, a P2P lending platform in the US, Pope and Sydnor (2011) find that African Americans are less likely to succeed in borrowing than white people with similar credit ratings, while there is no significant evidence that gender affects borrower funding success (Barasinska and Schäfer, 2010).

3. Data and Variables

We obtained the listing and loan data from Renrendai, one of the largest Chinese peer-to-peer lending platforms. The listing includes the amount of funds that a borrower wishes to raise and the interest rate that he or she is willing to pay. Similar to the mechanism of Prosper Inc. (Duarte et al., 2012), the loan-soliciting process is as follows. First, borrowers submit their funding requests online, along with necessary and voluntary information to invite bids, before the platform investigates the credit documents, reviews the quality and creates a credit rating for each borrower. Then, lenders can decide whether or not to place bids on the listed loan and how much to invest on the basis of the available information.[2] If the loan request has been open for seven days but does not receive enough funds, the platform will automatically cancel the request.

The sample period is June 2015 to July 2016. This 14-month period includes an external policy shock in the middle. On 28 December 2015, the Chinese government published a white paper indicating that tightened regulations would be placed on the P2P lending market. At the beginning of 2016, Renrendai officially announced cooperation with China Minsheng Bank on funds' depositing. The users' account information and the capital flows began to be supervised by the bank. As a result, the 14 months were divided into two parts: the pre-regulation period from June 2015 to December 2015 and the post-regulation period from January 2016 to July 2016. We created a time dummy, *post*_2015 to indicate whether a loan application was published on the website before or after the regulation.

We identified and downloaded all 265,041 loan applications from this sampling period. Following prior studies (Ding et al., 2018; Chen et al., 2018), we use unsecured loans for our next studies.[3] After data cleaning, we validated 77,992 closed unsecured loan application records, of which 6,557 applications were successfully funded by the lenders.

A complete list of all variables derived from Renrendai can be found in Table A.3.

For exposition purposes, we divided the variables into four groups. The first group concerns the performance of loan applications, including three variables: a funding indicator, which is equal to 1 if the application was successfully funded and 0 otherwise; and a default indicator, which is equal to 1 if the application was funded but the borrower fails to pay the loan and 0 otherwise. We also created a categorical variable named *status* to represent

[2]The lenders can also delegate the investment decisions to the platform. In this case, the investment choices are determined by algorithms and any algorithmic bias may also lead to a discriminatory result (Hajian et al., 2016; Chander, 2016).

[3]There are three types of loan applications in total: unsecured, joint liability and field certification. Joint liability and field certification loans have a significant lower default risk and a higher likelihood of receiving loans compared with the unsecured loans.

the overall funding status, which has three possible values: if the loan application succeeds, status is "regular" when the loan is paid back on time and is "default" if not. If the loan application does not succeed, status is "rejected".

The second group contains the variables derived from loan characteristics, including the application time, the loan amount, the interest rate and the loan duration.

The third group mainly concerns the borrowers' profile, including their age, gender, marital status, and revenue. It also includes a credit rating issued by the Renrendai platform. This rating has the same 7 levels as the Prosper data used in Duarte et al. (2012), i.e., AA, A, B, C, D, E and HR where HR means high risk. As shown in Table A.1, most of the applications are rated HR. Since the success rate has an obvious gap between applications with an HR rating and those with a non-HR rating, we created a rating indicator, which equals to 1 when the rating is not HR and 0 when it is HR.

Insert Table A.1

The third group also contains the regional information of each loan application. Due to reasons such as culture, habits and economic development, the numbers of applications in each province vary greatly. Following the classification method published by the Chinese Statistics Bureau,[4], we created a dummy *is_east* for the applications from the Eastern region and the dummy *is_west* for those from the Western region. All of the average disposable income data and the region classifications are from the website of the National Bureau of Statistics of China[5]. The provincial information is shown in Table A.2.

Insert Table A.2

[4] http://www.stats.gov.cn/ztjc/zthd/sjtjr/dejtjkfr/tjkp/201106/t20110613_71947.htm.
[5] The data portal of National Bureau of Statistics of China is http://data.stats.gov.cn

The fourth group contains two supplemental variables obtained from the Chinese National Bureau of Statistics. It includes provincial data of annual GDP and the bad debt rate.

Details of the variables are shown in Table A.3. We also took into account possible effects of the regulatory policy change initiated in December 2015, when the Chinese government acted to supervise and regulate online loans. The China Banking Regulatory Commission (CBRC) publicly solicits opinions to regulate the business activities of online lending information intermediaries. After the Chinese government's action, P2P platforms became responsible for examining and verifying the credit quality of borrowers before putting their loan applications online. Therefore, this regulation could be regarded as an external shock to borrowers' average credit quality.

Table A.4 reports the summary statistics before and after the government regulatory policy took place. From Table A.4, it is clear that after the regulation, the average loan amount and interest rate become lower and the rejection rate sharply declined, which implies that the platform possibly made substantial efforts to improve the criteria for loan applications to be published. Table A.5 compares the full sample, the Eastern region and the Western region. According to the t-tests, it seems that the proportions of default/regular/rejected loan applications were not significantly different between the Eastern and Western regions; while at the same time, the average monetary value of loan requests from the Eastern region was 9.43% higher than that from the Western region. If we believe that lenders generally tend to be more cautious about larger loans, then the descriptive statistics in A.5 may imply lenders' general preference for loan applications from the Eastern region over those from the Western region, which we will formally discuss in the next section.

Insert Tables A.3, A.4, and A.5

4. Empirical Analysis

The empirical study was conducted in four steps. First, we study whether the regional information of a loan application is associated with its likelihood of success or default. Second, we add two indicators (GDP and non-performing loan ratio; i.e., bad debt ratio) into our statistical model to see whether these dummies help to explain the regional discrimination. Third, we used an instrumental variable to identify whether or not the regional discrimination was genuine. Finally, we conducted two robustness checks by 1) using the Northeastern region as a potentially-discriminated region to test the consistency of our empirical results and theoretical explanation and 2) fixing the benchmark group to test the co-existence of positive and negative discrimination.

4.1. The Existence of Regional Discrimination

We started our analysis by relating the probability of a loan being funded to our regional dummies. As we can see from Table A.4, the mean funding success rate was 99.38% after the new regulatory policy was implemented [6]. Therefore, we ran two groups of regressions to examine the relationship between the regional information and the likelihood of success of loan applications, respectively. The first group only used the data before the regulation took place, while the second group used the full sample with the regulation dummy and cross-terms. The model specifications of the first group of regressions were as follows:[7]

$$is\_success = is\_east/is\_west + LoanSpecControls + BorrowerProfileControls + \epsilon$$

[6] Actually, only 11 of 1767 applications failed after 31/12/2015.

[7] Loan specification control variable include: $log(amount)$, $log(interest)$ and $log(duration)$; and borrower profile control variables include: $rating$, $is\_verified$, $log(age)$, $gender$, $education$, $is\_married$, $is\_divorced$, has_house, has_car, is_middle_income and is_high_income.

The second group is similar to the first one, except that the regulation dummy $post\_2015$ and its cross-term $post\_2015 \times is\_east(is\_west)$ were added to the model.

$$
\begin{aligned}
is\_success = & is\_east/is\_west + post\_2015 + post\_2015 \times is\_east/is\_west \\
& + LoanSpecControls + BorrowerProfileControls + \epsilon
\end{aligned}
$$

The results are shown in Table A.6. Specifications (1)-(4) are related to the Western region, and Specifications (5)-(8) are related to the Eastern region. Obviously, all of the coefficients for the Western region dummy are negative at a significance level lower than 0.05. Correspondingly, all of the coefficients for the Eastern region dummy are positive at a level of significance lower than 0.01. These results confirm that there is indeed regional discrimination in the loan success rate against borrowers from the Western region. Such discrimination remains even when we use the full sample to include the regulatory shock in the beginning of year 2015. [8]

Insert Table A.6

Verification status ($is\_verified$) and credit rating level ($rating$) represent the platform's effort to mitigate the information asymmetry between lenders and borrowers. Unsurprisingly, these two variables have a positive association with the probability of successful loan applications, which implies that these steps taken by the platform are informative and that they do have some substantial impact in closing the information gap between lenders and borrowers. However, the significance of regional dummies implies that regional discrimination still exists. Besides the main results, most of the coefficients of the other variables also

[8]We admit that, to some extent, the robustness of post-regulation results is limited due to an unbalanced sample size (there are 76,225 pre-regulation samples of unsecured loan applications, while only 1,767 samples after the regulatory shock).

fit our expectations and current literature. For instance, in terms of loan characteristics, a borrowing with a larger amount, a higher interest rate and a shorter period is less likely to be funded because lenders need to bear a higher risk for higher returns (Iyer, 2009; Dorfleitner et al., 2016).

Although Table A.6 shows the probabilistic difference in the success rates of loan applications initiated by borrowers from different regions, it would be interesting to see whether or not similar relationships exist between the likelihood of default and the regional dummies. If so, the discrimination would be more likely to be profit-oriented or taste-oriented otherwise. The statistical models are as follows:

$$is\_default = is\_east/is\_west + LoanSpecControls + BorrowerProfileControls + \epsilon$$

As before, if we take into account the policy shock, the model becomes:

$$\begin{aligned} is\_default = & is\_east/is\_west + post\_2015 + post\_2015 \times is\_east/is\_west \\ & + LoanSpecControls + BorrowerProfileControls + \epsilon \end{aligned}$$

Since only a successful application can default, we restricted our sample to all of the successful loans.

Insert Table A.7

Table A.7 shows the associations between the likelihood of default and regional information.

For the Eastern region, the implication of this result together with Table A.6 is straightforward: all other things equal, lenders favor loan applications from the Eastern region (a higher likelihood of being funded), and those funded loans from the Eastern region do have lower likelihood of default, which confirms the lenders' judgment. From this perspective, lenders' preferences towards applications from the Eastern region are aligned with the definition of profit-oriented discrimination.

However, the situation of loan applications from the Western region is quite different. Tables A.6 and A.7 show that the likelihood of being funded and defaulting are both negatively associated with the Western region dummy. These results imply that, compared with an average loan application, lenders are more demanding regarding those from the Western region: even the quality of these loans is *de facto* higher than average. In this vein, lenders' discrimination against Western-region loan applications is not likely to be driven by the potential profit; that is, it is taste-oriented.

For robustness, we also ran multi-nominal regressions to include all listing application records in our sample, regressing a three-outcome dependent variable with region indicators along with loan characteristics and borrower profile. The regression design is as follows:

$$Status = is\_east/is\_west + LoanSpecControls + BorrowerProfileControls + \epsilon$$

The dependent variable, *status*, has three outcomes:

- Outcome 1: Regular (successful application with normal repayment);
- Outcome 2: Default (successful application with default);
- Outcome 3: Failed (unsuccessful application).

We also include policy shock in our analysis.

$$status = is\_east/is\_west + post\_2015 + post\_2015 \times is\_east/is\_west$$
$$+ LoanSpecControls + BorrowerProfileControls + \epsilon$$

Table A.8 reports the corresponding results. Specifications (1) and (2) show the results of the regressions for Western-region borrowers, and Specifications (3) and (4) show those of the Eastern-region borrowers. Specifications (1) and (3) show the results of regressions that only include pre-policy observations, and Specifications (2) and (4) show those that include both pre- and post-policy samples.

Insert Table A.8

We have two remarks concluded from the results.

Remark 1. *For Eastern region loans, both failed and default indicators are statistically significant with reasonable signs, which provides additional evidence to support the hypothesis that lenders' favoring of the Eastern region is profit-oriented.*

Loan applications from the Eastern region have a lower application failure rate (i.e., a higher success rate, consistent with Table A.7) and a lower default rate. The significance also exists if we extend our analysis to the full sample by adding the post-policy dummy and related cross-terms. In general, loan applications from the Eastern region have a higher success rate and a lower default rate. These results fit the hypothesis of profit-oriented discrimination; that is, lenders' discriminatory behavior is likely to be based on the consideration of profit motivation (a higher probability of receiving repayment with interest).

Remark 2. *For Western region loans, both failed and default indicators are statistically insignificant, which implies that the discriminatory results in Table A.6 are not likely to be driven by profit motivation.*

The insignificant coefficients (p=0.05) of the Western region dummy in regressions (1) and (2) indicate that conditional on a successful application, if a borrower resides in the Western region, he or she does not necessarily have a higher probability of default. Therefore, we obtained some asymmetry here: for the Eastern region, the discrimination seems to be profit-oriented, since the region dummy is significant in the probabilities of both funding and default; while for the Western region, the discrimination is taste-oriented, and there is no solid connection between the default rate and the region dummy. However, some important issues remain: What are the factors behind the regional dummy? Is the discrimination genuinely region-based? In the next subsection, we add two explanatory variables and apply an instrumental variable (IV) approach to further explore these questions.

4.2. An IV Test with Regional Economic Attributes

In this section, we firstly added two regional economic factors into our regression designs from Section 4.1 to test whether the significance of regional estimators disappear. Next, we extracted the information of borrowers' original birthplace from ID numbers as an instrumental variable to bridge the causality between regional attributes and the success rate of loan applications. We introduced two possible factors related to the borrowers' capability to repay the loans into our statistical model – the provincial GDP and bad debt rate (non-performing loan rate reported in banking lending) – to represent the levels of economic development and provincial financial risks, respectively. After adding these two variables into the regression model of Table A.6, we report the updated results in Table A.9.

Insert Table A.9

For the two newly-added variables, the provincial GDP was marginally and positively related to the success rate of applicants (but not significant for Western province) and the provincial bad debt rate was significantly and negatively associated with the success rate.

These findings are aligned with the explanation of Peng et al. (2016); Jiang and Zhou (2016), who state that when making loan-granting decisions, lenders – perhaps subliminally – link economic factors with geographic information, although such information does not explicitly appear in the application document.

Regarding the coefficients for the regional dummies, we obtained asymmetric findings. After adding in provincial GDP and bad debt rate, the estimated coefficients for the Western region dummies are insignificant even at a $p = 0.1$ level while those for the eastern-province dummies were significant at a $p = 0.01$ level. This implies that the discrimination can be fully explained by these two economic factors for the Western region, however, for positively discriminated regions, there are two possible explanations for the significance of regional dummies: first, the discrimination is triggered by genuine region-based prejudice; or second, lenders might derive some soft information from the regional information (Liberti and Petersen, 2018). This explanation is also consistent with the information cost theory proposed by Meyer (1967), who states that such discrimination is motivated by reduced information-searching costs for avoiding the potential losses. Lenders thus rely on both the application document and their established notions derived from their private knowledge, such as provincial GDP and bad debt rate, and are more inclined to invest in Eastern province borrowers.

We employed an instrumental variable approach to check whether or not the regional discrimination was genuine. Fortunately, we had the first three digits of the borrowers' ID cards in the application document, the first two of which represents their birthplace information. As shown in the correlation matrix of all of these relevant variables in Figure A.1, a strong correlation exists between the place where the borrower resides [9] and the place where the borrower was born.

Insert Figure A.1

[9] This is also the address shown in the application document.

On the other hand, the borrower's ID number is simply regarded as a proof of ID verification. Lenders seldom use this as an informative piece of borrowers' profile. Moreover, the birthplace is hardly to be manipulated; hence, it is exogenous in our analysis. Therefore, the borrower's birthplace is a valid instrumental variable for us to determine what the discrimination stands for. If the regional discrimination is genuine and the discrimination is really about the region in which the borrower resides, then the IV regression would show the significance of the fitted value of the region indicator.

Since the dependent variable is binary, the standard method of IV regression is likely to lead to inefficient estimators. We thus used two variations of IV technologies: a two-stage least square (2SLS) estimation for binary variables, as developed by Newey (1987), and a two-stage residual inclusion (2SRI) estimation developed by Terza et al. (2008). The first method is very close to the standard IV regression, with the only change is to replace the second-stage OLS with a Probit regression. Taking is_west as an example, the two stages are:

$$OLS : is\_west = birth\_west + bad\_debt + GDP + v$$

$$\begin{aligned} Probit : is\_success &= \widehat{is\_west} + bad\_debt + GDP + LoanSpecControls \\ &\quad + BorrowerProfileControls + \epsilon \end{aligned}$$

The second method is mainly for non-linear regressions. It performs the same OLS regression in the first stage but takes the residual as a new variable in the second stage, with a Probit regression conducted together with the endogenous explanatory variable itself:

$$
\begin{aligned}
OLS: is\_west &= birth\_west + bad\_debt + GDP + v
\end{aligned}
$$

$$
\begin{aligned}
Probit: is\_success &= \widehat{v} + is\_west + bad\_debt + GDP + \epsilon \\
&\quad + LoanSpecControls + BorrowerProfileControls
\end{aligned}
$$

All of the relevant results are reported in Tables A.10 and A.11.

Insert Tables A.10 and A.11

The coefficients of the IV regional indicators in Tables A.10 and A.11 reveal intriguing results: after using birthplace as the instrumental variable, the estimates for both Western and Eastern region dummies become insignificant; that is, the discrimination is not genuinely region-based.

The overall results in Tables A.9 A.10 and A.11 can be interpreted as follows. For applicants who reside in the Western region, the lender's main concern seems to be on the economic side, represented by the fact that after adding GDP and bad debt rate, both the original and IV-case coefficients for the regional dummies were no more significant even at a p = 0.1 level. However, for the Eastern region, the lender's preference could not be fully explained by the selected variables indicating economic development and average capability of repayment (provincial GDP and bad rate). Therefore, region dummies may be used to proxy different information across different regions. We will further discuss possible explanations later.

4.3. Discrimination against the North-eastern Region

According to Lin and Fu (2017), the North-Eastern region also suffers from investment dis-crimination, as shown by a widely circulated saying: “no investment outside Shan-haiguan”.[10]. The post-2015 period also saw a large number of negative news reports about all three provinces in this region. As evidence, in 2016, Dr. Keqiang Li, the Premier of China, even used the saying to warn officials present at a formal government meeting (Zhao and Xu, 2016).

Therefore, it is worth employing our research method to check whether the Northeastern region is in fact negatively discriminated against by P2P lenders and, if so, which kind of discrimination applies. The situation is different between the Northeastern region and the Western region. The latter region is generally less-developed (except for Chongqing and Inner-Mongolia, the GDP per capita of all the other nine Western provinces is below national average)[11]. Nevertheless, the Northeastern region is not traditionally economically weak (even in 2015, Liaoning’s and Jilin’s GDP per capita is above the national average)[12]. Therefore, we expect that the discrimination against the Northeastern region, if it does in fact exist, should not be fully explained by economic factors as well.

Insert Tables A.12

We replicated our analysis with a region dummy representing loan applications from the Northeastern region. For simple illustration, we included the robustness checks.[13]

The results are summarized as follows.

[10]Shanhaiguan is an important pass in ancient China, and all three northeastern provinces are outside Shanhaiguan. This saying *de facto* means "Do not invest in the north-eastern region".

[11]Data is from Statistics Bureau of China

[12]Data are from the Statistics Bureau of China

[13]We conducted a full analysis as we did with the eastern region. The results had no significant difference with the results in Table A.12.

1) A lower success rate and a higher default rate of loan application were associated with North-Eastern applicants;

2) The region dummy remained significant at a p = 0.05 level, even after introducing GDP and bad debt ratio as additional control variables;

3) Using birthplace as the IV, the regional dummy became insignificant.

Compared with the Western region, a crucial difference is that the likelihood of default in the Northeastern region was significantly positively associated with the region dummy. It implies that lenders' discrimination against applications from the Northeastern region may be profit-driven. In addition, similar to those of the Eastern region, the factors behind the regional dummy go beyond economic development and the average capability of repayment.

4.4. Co-existence of Positive and Negative Discrimination

In the above analysis, one might wonder whether positive discrimination towards the Eastern region is just an alternative manifestation of negative discrimination. Typically, this concern could be solved by choosing an appropriate control group and keep it consistent in all regression functions. In our studies, the only possible control group is the Central region if we have to keep the benchmark group consistent. However, the Central region is not "neutral" for discrimination since provincial evidence shows that loan applicants from the Central region suffer from similarly negative discrimination as the Western region does (Jiang and Zhou, 2016).

Therefore, in this subsection, we conduct a robustness check of the co-existence of both positive and negative discrimination. We use two sub-samples and keep the rest of the observations as the control group consistently throughout the analysis. By doing so, we construct an enough large control group that is close to an "average" applicant in China. The Eastern sub-sample includes Beijing and Shanghai (is_sub_east), the two most significant economic centres of China; the Western sub-sample consists of several provinces in the Southwest and Northwest of China (is_sub_west), including Sichuan, Chongqing, Qinghai, Ningxia, Gansu, Xinjiang and Tibet. [14] We report the regression results in Table A.13.

Insert Table A.13

Our findings show that the positive and negative regional discrimination can co-exist with a large and consistent control group (Column 1). The default rate is lower for applicants from sub-east region which shows that the discrimination against this economically-developed region is profit-oriented (Column 2). After controlling for provincial economic factors, discrimination against sub-west region borrowers disappears (Column 3) which strengthens our explanation that lenders are inclined to use hard information to judge borrowers from remote areas. After adding birthplace IV, we find no evidence of discrimination (Column 4) which shows that the discrimination shown in Column 1 is not genuinely region-based.

Overall, our sub-region analysis results are consistent with those in Sections 4.1 and 4.2.

4.5. Discussion of the Empirical Results

To summarize, regional discrimination widely exists in the Chinese online lending market. We found strong evidence of positive discrimination against applicants residing in the Eastern region of China and negative discrimination against those residing in the Western and Northeastern regions. None of the discrimination was genuinely region-based.

The explanation that our results support is one of home bias and information loss due to lenders' limited capability to collect soft information (Liberti, 2017; Agarwal and Ben-David, 2018; Baltzer et al., 2015). Most of the lenders resided in the Eastern region,[15] which is also

[14] We rule out those provinces that are closer to the Eastern region.

[15] See the report published by Yingcan Consulting https://www.sohu.com/a/124127840_530780

physically close to the Northeastern region.[16] Therefore, lenders can collect more soft information regarding these two regions, while such capability is undermined when borrowers are from the Western region, making the physical distance to the majority of lenders considerably longer. Information cost theory (Meyer, 1967) then predicts that lenders will rely more on region-related hard information, such as provincial GDP and non-performing loan ratio, to decide on corresponding loan applications.[17] Following the same logic, the weight of soft information will increase in lenders' decision-making process when the borrowers are physically close to the lenders. Our empirical results also provide convincing evidence for this explanation. On the one hand, the region dummies are still significant when GDP and the non-performing loan ratio were controlled for the North-Eastern and Eastern regions; on the other hand, when using the birthplace as the instrumental variable, we rule out the possibility of genuinely region-based discrimination in the whole market. Therefore, the part of the region dummy that remains unexplained is, by definition of Liberti and Petersen (2018), soft information, and thus the weight of soft information does vary along with the distance, as shown by the difference of significance level of regional dummy in Table A.9.

5. Conclusion

This research focuses on regional discrimination in the Chinese P2P lending market and explores the potential reasons behind it. Based on a dataset from one of the largest and well-developed P2P platforms in China, we find evidence that favors the existence of regional discrimination. We also use the instrument variable method to rule out the possibility of genuinely region-based discrimination.

As for the reasons for regional discrimination, we provide novel empirical evidence of the information cost theory and soft information acquisition in the P2P lending market. In

[16]See the region distribution map in Figure A.2.

[17]According to Table A.7, the decision process is not always aligned with the profit maximising objective.

particular, our findings suggest that lenders might use the region where borrowers reside as a proxy variable with which to infer more information. This information could contain both hard and soft information, and the lender's capability to collect soft information diminishes when borrowers are physically far from the lenders.

From the policy perspective, some lenders employ algorithms to help them make lending decisions, and these algorithms, together with human beings themselves, have shown an intention towards regional discrimination. This is aligned with the theory of Chander (2016), who argued that even a transparent, facially-neutral algorithm can still produce discriminatory results and that this discrimination could potentially be self-enforcing and contribute to the polarization of the social economy. Therefore, our research may motivate further study regarding whether an algorithm helps mitigate or alleviate discriminatory behavior in the financial decision-making process.

We acknowledge some limitations of our research. First, although all of our empirical evidence is aligned with the soft information and distance trade-off, as well as the information cost theory, we cannot claim that our explanation is unique by ruling out alternative explanations. An example is that lenders may simply have different expectations regarding the future development of institutions and economic development of certain regions, due to a lack of information about lenders' residential regions. Second, the text in the description of loans is another source of soft information that lenders can extract from the application webpage. It may increase our explanatory powers if we employed machine learning techniques to reveal some useful information or patterns and add them into our statistical model. Finally, GDP and bad debt rate are introduced as additional control variables, however, these two variables are far from an exhaustive list of lenders' potential concerns. Nevertheless, we leave the discussion of other variables (e.g., variables related to regional culture and social security) open for future research.

References

Afridi, F., Li, S.X., Ren, Y., 2015. Social identity and inequality: The impact of china's hukou system. Journal of Public Economics 123, 17–29.

Agarwal, S., Ben-David, I., 2018. Loan prospecting and the loss of soft information. Journal of Financial Economics 129, 608–628.

Agarwal, S., Hauswald, R., 2010. Distance and private information in lending. The Review of Financial Studies 23, 2757–2788.

Altonji, J., Pierret, C., 2001. Employer learning and statistical discrimination. Quarterly Journal of Economics 464, 313–350.

Arnott, S., 1972. Some mathematical models of race discrimination in the labor market. Racial Discrimination in Economic Life , 187–204.

Arrow, K.J., 1998. What has economics to say about racial discrimination? Journal of economic perspectives 12, 91–100.

Baltzer, M., Stolper, O., Walter, A., 2015. Home-field advantage or a matter of ambiguity aversion? local bias among german individual investors. The European Journal of Finance 21, 734–754.

Barasinska, N., Schäfer, D., 2010. Does gender affect funding success at the peer-to-peer credit markets? evidence from the largest german lending platform .

Becker, G.S., 2010. The economics of discrimination. University of Chicago press.

Bellucci, A., Borisov, A., Zazzaro, A., 2010. Does gender matter in bank–firm relationships? evidence from small business lending. Journal of Banking & Finance 34, 2968–2984.

Benston, G.J., Horsky, D., 1992. The relationship between the demand and supply of home financing and neighborhood characteristics: an empirical study of mortgage redlining. Journal of Financial Services Research 5, 235–260.

Bertrand, M., Mullainathan, S., 2004. Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. American economic review 94, 991–1013.

Blanchflower, D.G., Levine, P.B., Zimmerman, D.J., 2003. Discrimination in the small-business credit market. Review of Economics and Statistics 85, 930–943.

Burtch, G., Ghose, A., Wattal, S., 2014. Cultural differences and geography as determinants of online prosocial lending. MIS Quarterly 38.

Cai, F., Wang, D., Du, Y., 2002. Regional disparity and economic growth in china: The impact of labor market distortions. China Economic Review 13, 197–212.

Caldieraro, F., Zhang, J.Z., Cunha Jr, M., Shulman, J.D., 2018. Strategic information transmission in peer-to-peer lending markets. Journal of Marketing 82, 42–63.

Calomiris, C., Kahn, C., Longhofer, S., 1994. Housing finance intervention and private incentives: Helping minorities and the poor. Journal of Money, Credit, and Banking 26, 634–74.

Chander, A., 2016. The racist algorithm. Mich. L. Rev. 115, 1023.

Chen, X., Ye, D., Deng, J., 2018. Can readability of loan description promote lending success rate of online. China Industrial Economics 3, 174–192.

Cheng, P., Lin, Z., Liu, Y., 2015. Racial discrepancy in mortgage interest rates. The Journal of Real Estate Finance and Economics 51, 101–120.

Claessens, S., Frost, J., Turner, G., Zhu, F., 2018. Fintech credit markets around the world: size, drivers and policy issues. BIS Quarterly Review September .

Deku, S.Y., Kara, A., Molyneux, P., 2016. Access to consumer credit in the uk. The European Journal of Finance 22, 941–964.

Dell'Ariccia, G., Igan, D., Laeven, L.U., 2012. Credit booms and lending standards: Evidence from the subprime mortgage market. Journal of Money, Credit and Banking 44, 367–384.

Ding, J., Li, Y., Zeng, Y., Li, Z., 2018. Double information value and information transfer of two-way traders in p2p online lending. Nankai management review 2, 4–15.

Dorfleitner, G., Priberny, C., Schuster, S., Stoiber, J., Weber, M., de Castro, I., Kammler, J., 2016. Description-text related soft information in peer-to-peer lending–evidence from two leading european platforms. Journal of Banking & Finance 64, 169–187.

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

Edie, L., Riefler, W., 1931. Money rates and money markets in the united states. Journal of the American Statistical Association 26, 112.

Feller, J., Gleasure, R., Treacy, S., 2017. Information sharing and user behavior in internet-enabled peer-to-peer lending systems: an empirical study. Journal of Information Technology 32, 127–146.

Feng, K., Hubacek, K., Guan, D., 2009. Lifestyles, technology and co2 emissions in china: A regional comparative analysis. Ecological Economics 69, 145–154.

Fershtman, C., Gneezy, U., 2001. Discrimination in a segmented society: An experimental approach. The Quarterly Journal of Economics 116, 351–377.

Freedman, S., Jin, G.Z., 2017. The information value of online social networks: lessons from peer-to-peer lending. International Journal of Industrial Organization 51, 185–222.

Funk, B., Buerckner, D., Hilker, M., Kock, F., Lehmann, M., Tiburtius, P., et al., 2011. Online peer-to-peer lending: A literature review. The Journal of Internet Banking and Commerce 16, 1–18.

Goering, J.M., 1996. Mortgage Lending, Racial Discrimination, and Federal Policy. The Urban Institute Press, Washington, DC.

Guariglia, A., Mateut, S., 2016. External finance and trade credit extension in china: does political affiliation make a difference? The European Journal of Finance 22, 319–344.

Guo, Y., Zhou, W., Luo, C., Liu, C., Xiong, H., 2016. Instance-based credit risk assessment for investment decisions in p2p lending. European Journal of Operational Research 249, 417–426.

Guryan, J., Charles, K.K., 2013. Taste-based or statistical discrimination: The economics of discrimination returns to its roots. The Economic Journal 123, F417–F432.

Hajian, S., Bonchi, F., Castillo, C., 2016. Algorithmic bias: From discrimination discovery to fairness-aware data mining, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, ACM. pp. 2125–2126.

Han, J.T., 2018. The persuasion of borrowers' voluntary information in peer-to-peer lending: An empirical study based on elaboration likelihood model. Computers in Human Behavior 78, 200–214.

Han, S., 2011. Creditor learning and discrimination in lending. Journal of Financial Services Research 40, 1–27.

Huyghebaert, N., Wang, L., 2016. Institutional development and financing decisions: evidence from a cross-regional study on chinese listed firms. The European Journal of Finance 22, 288–318.

Iyer, R., 2009. Screening in new credit markets: Can individual lenders infer borrower creditworthiness in peer-to-peer lending? Management Science 59, 17–35.

Jiang, C., Wang, Z., Wang, R., Ding, Y., 2018. Loan default prediction by combining soft information extracted from descriptive text in online peer-to-peer lending. Annals of Operations Research 266, 511–529.

Jiang, Y., Zhou, A., 2016. Is there regional discrimination in p2p lending? empirical data from renrendai. Journal of Central University of Finance and Economics 2016, 29–39.

Kanbur, R., Zhang, X., 1999. Which regional inequality? the evolution of rural–urban and inland–coastal inequality in china from 1983 to 1995. Journal of comparative economics 27, 686–701.

Klafft, M., 2008. Online peer-to-peer lending: A lenders' perspective. SSRN Electronic Journal .

Knowles, J., Persico, N., Todd, P., 2001. Racial bias in motor vehicle searches: Theory and evidence. Journal of Political Economy 109, 203–29.

Ladd, H., 1998. Evidence on discrimination in mortgage lending. Journal of Economic Perspectives 12, 41–62.

Larrimore, L., Jiang, L., Larrimore, J., Markowitz, D., Gorski, S., 2011. Peer to peer lending: The relationship between language features, trustworthiness, and persuasion success. Journal of Applied Communication Research 39, 19–37.

Liao, L., Li, M., Wang, Z., 2014a. The intelligent investor:not-fully-marketized interest rate and risk identify——evidence from p2p lending. Economic Research Journal 2014, 125–137.

Liao, L., Li, M., Wang, Z., 2014b. Regional discrimination in chinese internet finance. The Journal of Quantitative & Technical Economics 31, 54–70.

Liberti, J.M., 2017. Initiative, incentives, and soft information. Management Science 64, 3714–3734.

Liberti, J.M., Petersen, M.A., 2018. Information: Hard and soft. Review of Corporate Finance Studies 8, 1–41.

Lin, M., Prabhala, N.R., Viswanathan, S., 2013. Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. Management Science 59, 17–35.

Lin, Y., Fu, C., 2017. Jilin revitalization from the perspective of new structural economics:analysis, tools and policy plans in jilin report. Social Science Journal 6, 5–20.

Meyer, P.A., 1967. Price discrimination, regional loan rates, and the structure of the banking industry. The Journal of Finance 22, 37–48.

Munnell, A.H., Tootell, G.M., Browne, L.E., McEneaney, J., 1993. Is discrimination racial or geographic? Manuscript. Federal Reserve Bank of Boston, undated .

Muravyev, A., Talavera, O., Schäfer, D., 2009. Entrepreneurs' gender and financial constraints: Evidence from international data. Journal of Comparative Economics 37, 270–286.

Netzer, O., Lemaire, A., Herzenstein, M., 2019. When words sweat: Identifying signals for loan default in the text of loan applications. Journal of Marketing Research 56, 960–980.

Newey, W.K., 1987. Efficient estimation of limited dependent variable models with endogenous explanatory variables. Journal of Econometrics 36, 231–250.

Peng, H., Yang, L., Tan, X., 2016. How regional differences affect p2p lending behavior——based on empirical evidence of renrendai. Journal of Contemporary Economics 38, 21–34.

Pęski, M., Szentes, B., 2013. Spontaneous discrimination. American Economic Review 103, 2412–36.

Phelps, E., 1972. The statistical theory of racism and sexism. The American Economic Review 62, 659.

Pope, D., Sydnor, J., 2011. What's in a picture? evidence of discrimination from prosper. com. Journal of Human resources 46, 53–92.

Puro, L., 2010. Borrower decision aid for people-to-people lending. Decision Support Systems 49, 52–60.

Ravina, E., 2007. Beauty, personal characteristics, and trust in credit markets. SSRN Electronic Journal .

Schafer, R., Ladd, H., 1981. Discrimination in mortgage lending. MIT Press, Cambridge, Mass.

Serrano-Cinca, C., Gutierrez-Nieto, B., Lopez-Palacios, L., 2015. Determinants of default in p2p lending. PLOS ONE 10, 0139427.

Spence, M., 1978. Job market signaling, in: Uncertainty in economics. Elsevier, pp. 281–306.

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

Terza, J.V., Basu, A., Rathouz, P.J., 2008. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. Journal of health economics 27, 531–543.

Thompson, P., Cowton, C.J., 2004. Bringing the environment into bank lending: implications for environmental reporting. The British Accounting Review 36, 197–218.

Turner, M., 1999. Mortgage lending discrimination: A review of existing evidence.

Wang, T., Zheng, T., 2017. External environmental influencing factors of regional differences in china's p2p online loan industry. Technical Economy 36, 98–105.

Wei, Z., Lin, M., 2016. Market mechanisms in online peer-to-peer lending. Management Science 63, 4236–4257.

Weiss, G.N., Pelger, K., Horsch, A., 2010. Mitigating adverse selection in p2p lending–empirical evidence from prosper. com .

Xu, J.J., Chau, M., 2018. Cheap talk? the impact of lender-borrower communication on peer-to-peer lending outcomes. Journal of Management Information Systems 35, 53–85.

Zhao, S., Xu, T., 2016. Why keqiang li is worried about "no investment outside shanhaiguang". http://www.bjnews.com.cn/news/2016/10/20/420551.html. [Online; accessed 19-July-2008].

Appendix A. Figures and Tables

Table A.1: Credit Ratings

Credit Rating	Number	Percentage	Success Rate
AA	18	0.02%	55.62%
A	40	0.05%	55.00%
B	65	0.08%	47.70%
C	209	0.27%	60.78%
D	2286	2.93%	70.69%
E	2946	3.78%	65.23%
HR	72428	92.87%	3.90%

Table A.2: Applications in targeted regions

Variables	Province	Number	Percentage(%)
Eastern Provinces	Beijing	3305	52.0
	Fujian	4198	
	Guangdong	12335	
	Hainan	591	
	Hebei	2233	
	Jiangsu	4492	
	Shandong	4130	
	Shanghai	2827	
	Tianjin	634	
	Zhejiang	5683	
Western Provinces	Chongqing	1504	21.9
	Gansu	880	
	Guangxi	2440	
	Guizhou	1520	
	Inner mongolia	1336	
	Ningxia	376	
	Qinghai	151	
	Shaanxi	1764	
	Sichuan	4170	
	Tibet	116	
	Xinjiang	1002	
	Yunnan	1790	
	Anhui	2277	26.1
	Heilongjiang	1643	
	Henan	3684	
	Hubei	3075	
	Hunan	3159	
	Jiangxi	1820	
	Jilin	1023	
	Liaoning	1922	
	Shanxi	1812	

Table A.3: Variables Constructed from Renrendai Data

	Variable Name	Variable Definition
Performance Indicator:	*is_ successful*	An indicator that equals one if an application is fully funded and becomes a loan and is zero otherwise.
	in_ default	An indicator that equals one if a funded project fails to pay back the loans and is zero otherwise.
	status	A categorical variable, which has three possible values: if the loan application succeeds, status is "regular" when the loan is paid back on time and is "default" if not. If the loan application does not succeed, status is "rejected".
Loan Characteristics:	*amount*	The requested loan amount in 1000 CNY.
	interest	The rate the borrower pays on the loan.
	duration	In how many months the loan matures.
	post-2015	Whether the loan application is listed after the new regulatory policy being effective.
Borrower's Profile:	*age*	Borrower's age, range from 21 to 62.
	gender	Borrower's gender, 1 if the borrower is male, 0 otherwise.
	education	Borrower's education level, 1-5 represents high school or lower, junior college, undergraduate and postgraduate or higher respectively.
	is_ east	Whether the borrower resides in any of the eastern provinces.
	is_ west	Whether the borrower resides in any of the western provinces.
	is_ married	1 if the borrower is married, 0 otherwise.
	is_ divorced	1 if the borrower is divorced, 0 otherwise.
	has_ house	1 if the borrower is a house owner, 0 otherwise.
	has_ car	1 if the borrower is a car owner, 0 otherwise.
	rating	Borrower's credit rating issued by the website. 1 for Non-HR ratings, 0 for HR rating.
	is_ verified	1 if the borrower has uploaded a valid document and verified by the website, 0 otherwise.
	middle_ income	1 if the borrower's monthly revenue is between 5000 and 10000, 0 otherwise.
	high_ income	1 if the borrower's monthly revenue is above 10000, 0 otherwise.
	birth_ east	1 if the borrower was born in the 10 provinces in the eastern region, 0 otherwise.
	birth_ west	1 if the borrower was born in the 10 provinces in the western region, 0 otherwise.
Supplemental Variables:	*bad_ rate*	The provincial non-performing loan ratio.
	gdp	The annual provincial GDP.

Table A.4: Summary Statistics: Pre- and Post-Regulation

Variables	Total			Pre-Regulation			Post-Regulation		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Default	77992	0.01	0.12	76225	0.01	0.10	1767	0.17	0.38
Regular	77992	0.07	0.25	76225	0.05	0.22	1767	0.82	0.38
Rejected	77992	0.92	0.28	76225	0.94	0.24	1767	0.01	0.08
Amount	77992	56.11	82.27	76225	57.01	82.99	1767	17.23	10.51
Interest	77992	12.22	1.00	76225	12.23	0.99	1767	11.67	1.12
Duration	77992	16.69	7.82	76225	16.75	7.83	1767	14.09	7.07
Age	77992	30.65	6.24	76225	30.62	6.24	1767	31.78	6.22
Gender	77992	0.86	0.35	76225	0.86	0.35	1767	0.86	0.35
Education	77992	1.94	0.80	76225	1.93	0.80	1767	2.35	0.78
Is_married	77992	0.49	0.50	76225	0.49	0.50	1767	0.57	0.49
Is_Divorced	77992	0.04	0.21	76225	0.04	0.21	1767	0.04	0.20
Has_House	77992	0.44	0.50	76225	0.44	0.50	1767	0.56	0.50
Has_Car	77992	0.26	0.44	76225	0.26	0.44	1767	0.29	0.46
Rating	77992	0.07	0.26	76225	0.06	0.24	1767	0.60	0.49
Is_verified	77992	0.45	0.50	76225	0.44	0.50	1767	1.00	0.00
Middle_income	77992	0.35	0.48	76225	0.35	0.48	1767	0.42	0.49
High_income	77992	0.23	0.42	76225	0.23	0.42	1767	0.15	0.36
Bad_debt	77992	1.77	0.59	76225	1.77	0.59	1767	1.74	0.59
GDP	77992	40.30	24.60	76225	40.27	24.56	1767	41.67	26.16

Table A.5: Summary Statistics: Full Sample, Eastern Provinces, Western Provinces

Variables	Total			Eastern Provinces			Western Provinces			T-test	
	N	Mean	SD	N	Mean	SD	N	Mean	SD	Mean Dif	SD
Default	77992	0.01	0.12	40528	0.01	0.12	17049	0.01	0.12	0.00	0.12
Regular	77992	0.07	0.25	40528	0.07	0.26	17049	0.07	0.26	0.00	0.26
Rejected	77992	0.92	0.28	40528	0.91	0.28	17049	0.91	0.28	0.00	0.28
Amount	77992	56.11	82.27	40528	58.20	84.06	17049	52.92	80.13	5.29***	82.92
Interest	77992	12.22	1.00	40528	12.22	0.99	17049	12.23	0.98	0.00	0.99
Duration	77992	16.69	7.82	40528	16.72	7.81	17049	16.72	7.79	0.00	7.81
Age	77992	30.65	6.24	40528	30.26	5.94	17049	30.68	6.23	-0.24***	6.03
Gender	77992	0.86	0.35	40528	0.87	0.34	17049	0.84	0.37	0.03***	0.35
Education	77992	1.94	0.80	40528	1.92	0.81	17049	1.99	0.79	-0.07***	0.80
Is_married	77992	0.49	0.50	40528	0.48	0.50	17049	0.48	0.50	0.01	0.50
Is_Divorced	77992	0.04	0.21	40528	0.03	0.18	17049	0.06	0.23	-0.02***	0.20
Has_House	77992	0.44	0.50	40528	0.37	0.48	17049	0.48	0.50	-0.11***	0.49
Has_Car	77992	0.26	0.44	40528	0.25	0.43	17049	0.26	0.44	-0.01*	0.43
Rating	77992	0.07	0.26	40528	0.08	0.27	17049	0.07	0.26	0.005***	0.26
Is_verified	77992	0.45	0.50	40528	0.47	0.50	17049	0.45	0.50	0.01***	0.50
Middle_income	77992	0.35	0.48	40528	0.38	0.48	17049	0.34	0.47	0.04***	0.48
High_income	77992	0.23	0.42	40528	0.27	0.44	17049	0.18	0.38	0.09***	0.43
Bad_debt	77992	1.77	0.59	40528	1.72	0.63	17049	1.95	0.73	-0.22***	0.66
GDP	77992	40.30	24.60	40528	55.97	23.77	17049	19.31	7.33	36.66***	20.50

Table A.6: Regional discrimination

	Dependent variable:							
	is_success							
	Logit	Probit	Logit	Probit	Logit	Probit	Logit	Probit
is_west	−0.119**	−0.060**	−0.118**	−0.059**				
	(0.048)	(0.025)	(0.048)	(0.025)				
is_east					0.195***	0.106***	0.194***	0.105***
					(0.041)	(0.022)	(0.041)	(0.022)
post_2015			6.169***	3.134***			7.212***	3.552***
			(0.321)	(0.116)			(0.711)	(0.223)
is_west × post_2015			1.035	0.500				
			(1.055)	(0.364)				
is_east × post_2015							−1.288	−0.520**
							(0.788)	(0.257)
rating	2.744***	1.556***	2.739***	1.550***	2.733***	1.550***	2.728***	1.544***
	(0.046)	(0.026)	(0.046)	(0.026)	(0.046)	(0.026)	(0.046)	(0.026)
is_verified	7.087***	2.834***	7.086***	2.829***	7.084***	2.834***	7.083***	2.829***
	(0.500)	(0.149)	(0.500)	(0.149)	(0.500)	(0.149)	(0.500)	(0.149)
Constant	18.717***	10.562***	18.429***	10.270***	18.636***	10.494***	18.357***	10.206***
	(1.713)	(0.923)	(1.707)	(0.916)	(1.714)	(0.923)	(1.708)	(0.917)
Borrowers' Profile	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	76,225	76,225	77,992	77,992	76,225	76,225	77,992	77,992
McFadden's R^2	0.4973	0.4994	0.5963	0.5972	0.4978	0.4999	0.5967	0.5976

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

Table A.7: Default Rate of Borrowers from Different Regions

	Dependent variable:							
	is_default							
	logistic	*probit*	*logistic*	*probit*	*logistic*	*probit*	*logistic*	*probit*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
is_west	-0.243^{**}	-0.131^{**}	-0.238^{**}	-0.128^{**}				
	(0.115)	(0.065)	(0.114)	(0.065)				
is_east					-0.354^{***}	-0.223^{***}	-0.335^{***}	-0.213^{***}
					(0.098)	(0.055)	(0.096)	(0.054)
post2015			0.134	0.076			0.003	0.006
			(0.103)	(0.058)			(0.132)	(0.074)
is_west × post2015			0.262	0.174				
			(0.215)	(0.120)				
is_east × post2015							0.374^{**}	0.219^{**}
							(0.180)	(0.101)
Constant	3.376	2.730	3.840	2.978^{*}	2.990	2.474	3.611	2.850^{*}
	(3.523)	(1.925)	(3.007)	(1.655)	(3.544)	(1.939)	(3.023)	(1.663)
Borrowers' Profile	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,801	4,801	6,557	6,557	4,801	4,801	6,557	6,557
McFadden's Pseudo R^2			0.564	0.566			0.564	0.565
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$							

Table A.8: Multinomial Regression on Loan Status

	Multinomial Regressions							
	(1)		(2)		(3)		(4)	
	default	failed	default	failed	default	failed	default	failed
is_west	−0.198*	0.072	−0.147	0.082				
	(0.102)	(0.054)	(0.102)	(0.054)				
is_east					−0.287***	−0.276***	−0.305***	−0.274***
					(0.084)	(0.046)	(0.084)	(0.046)
post_2015			0.140	−6.128***			−0.082	−7.276***
			(0.100)	(0.323)			(0.129)	(0.713)
is_west×post_2015			0.086	−1.057				
			(0.206)	(1.057)				
is_east×post_2015							0.460***	1.475*
							(0.173)	(0.790)
Constant	6.068*	−17.698***	13.675***	−16.126***	−6.507***	−21.272***	14.068***	−16.123***
	(3.683)	(1.799)	(2.864)	(1.782)	(1.786)	(1.790)	(2.871)	(1.785)
Borrowers' Profile	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	76225	76225	77992	77992	76225	76225	77992	77992
Akaike Inf. Crit.	21,071.930	21,071.930	22,365.920	22,365.920	21,072.860	21,072.860	22,334.870	22,334.870

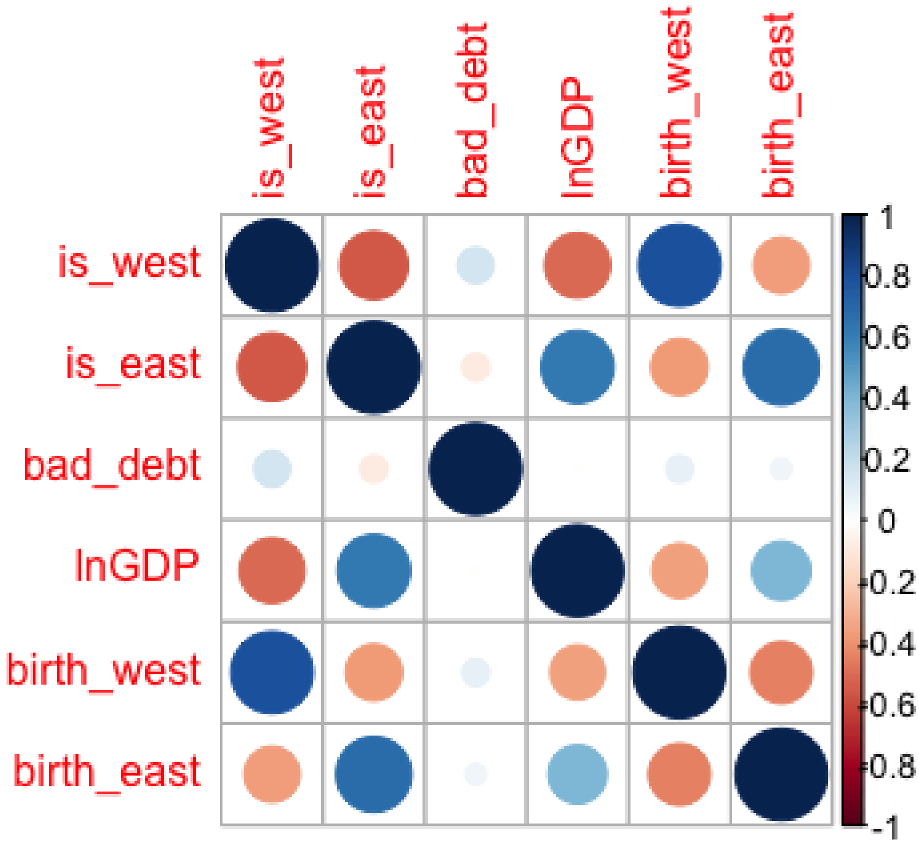
Note: *p<0.1; **p<0.05; ***p<0.01

Table A.9: Effects of Provincial GDP and Bad Debt Ratio

	Dependent variable:							
	is_success							
	logistic	*probit*	*logistic*	*probit*	*logistic*	*probit*	*logistic*	*probit*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
is_west	−0.048	−0.019	−0.048	−0.018				
	(0.056)	(0.030)	(0.056)	(0.030)				
is_east					0.197***	0.108***	0.194***	0.105***
					(0.052)	(0.028)	(0.052)	(0.028)
ln_GDP	0.055	0.032*	0.055*	0.033*	−0.016	−0.009	−0.014	−0.007
	(0.033)	(0.018)	(0.033)	(0.018)	(0.036)	(0.019)	(0.036)	(0.019)
bd_rate	−0.098***	−0.054***	−0.096***	−0.051***	−0.086**	−0.046**	−0.084**	−0.044**
	(0.035)	(0.018)	(0.035)	(0.018)	(0.034)	(0.018)	(0.034)	(0.018)
post_2015			6.152***	3.124***			7.214***	3.554***
			(0.321)	(0.116)			(0.711)	(0.224)
is_west×post_2015			1.055	0.513				
			(1.055)	(0.364)				
is_east×post_2015							−1.299*	−0.527**
							(0.788)	(0.257)
Constant	18.464***	10.385***	18.166***	10.080***	19.061***	10.736***	18.753***	10.420***
	(1.748)	(0.942)	(1.742)	(0.935)	(1.752)	(0.944)	(1.746)	(0.937)
Borrowers' Profile	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	76,225	76,225	77,992	77,992	76,225	76,225	77,992	77,992
McFadden's R^2	0.4976	0.4997	0.5965	0.5974	0.4980	0.5001	0.5968	0.5977

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

Figure A.1: Correlation Matrix



As shown in the legend on the right of the figure, deeper colours implies stronger correlations.

Table A.10: IV 2SLS Regression

	Dependent variable:			
	success			
	(1)	(2)	(3)	(4)
$\widehat{is\_west}$	−0.004	−0.002		
	(0.039)	(0.039)		
$\widehat{is\_east}$			0.017	0.009
			(0.045)	(0.045)
$\widehat{is\_west}$ × post_2015		0.696^{*}		
		(0.396)		
$\widehat{is\_east}$ × post_2015				-0.617^{**}
				(0.285)
post_2015		3.103^{***}		3.600^{***}
		(0.115)		(0.233)
bd_rate	-0.055^{***}	-0.054^{***}	-0.054^{***}	-0.053^{***}
	(0.019)	(0.019)	(0.018)	(0.018)
ln_GDP	0.037^{*}	0.038^{**}	0.030	0.035
	(0.019)	(0.019)	(0.025)	(0.025)
Constant	10.344^{***}	10.035^{***}	10.400^{***}	10.059^{***}
	(0.945)	(0.938)	(0.955)	(0.948)
Borrowers' Profile	Yes	Yes	Yes	Yes
Loan Characteristics	Yes	Yes	Yes	Yes
Observations	76,225	77,992	76,225	77,992
McFadden's R^2	0.4976	0.5974	0.4997	0.5974
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table A.11: IV 2SRI Regression

	Dependent variable:			
	success			
	(1)	(2)	(3)	(4)
is_west	−0.004	−0.001		
	(0.039)	(0.039)		
is_east			0.026	0.017
			(0.045)	(0.045)
is_east × post_2015				-0.527^{**}
				(0.257)
is_west × post_2015		0.512		
		(0.364)		
$\hat{v}_{west}$	−0.035	−0.042		
	(0.060)	(0.060)		
$\hat{v}_{east}$			0.127^{**}	0.138^{**}
			(0.055)	(0.054)
post_2015		3.123^{***}		3.553^{***}
		(0.116)		(0.224)
Constant	10.336^{***}	10.021^{***}	10.400^{***}	10.057^{***}
	(0.945)	(0.938)	(0.955)	(0.948)
Borrowers' Profile	Yes	Yes	Yes	Yes
Loan Characteristics	Yes	Yes	Yes	Yes
Observations	76,225	77,992	76,225	77,992
McFadden's R^2	0.4997	0.5974	0.5003	0.5979
Note:				*p<0.1; **p<0.05; ***p<0.01

Table A.12: Northeastern Region

	Dependent variable:			
	is_success	is_default	is_success	
	logistic	*logistic*	*logistic*	*probit*
	(1)	(2)	(3)	(4)
is_ne	-0.211^{**}	0.943^{***}	-0.193^{**}	
	(0.090)	(0.196)	(0.092)	
$\widehat{is\_ne}$				0.014
				(0.058)
post2015	6.288^{***}	0.257^{***}	6.274^{***}	3.190^{***}
	(0.306)	(0.093)	(0.306)	(0.109)
is_ne × post2015	9.728	-1.189^{***}	9.736	
	(136.025)	(0.443)	(136.049)	
$\widehat{is\_ne}$ × post2015				2.490
				(21.682)
ln_GDP		0.104^{*}	0.056^{*}	0.038^{**}
		(0.061)	(0.029)	(0.016)
bd_rate		0.120^{*}	-0.105^{***}	-0.053^{***}
		(0.070)	(0.034)	(0.018)
Constant	18.412^{***}	2.057	18.175^{***}	10.037^{***}
	(1.707)	(3.084)	(1.736)	(0.932)
Borrowers' Profile	Yes	Yes	Yes	Yes
Loan Characteristics	Yes	Yes	Yes	Yes
Observations	77,992	6,557	77,992	77,992
McFadden's R^2	0.5963	0.3232	0.5966	0.5974
Note:				*p<0.1; **p<0.05; ***p<0.01

Table A.13: Sub-Sample Results

	Dependent variable:			
	is_success	is_default	is_success	
	logistic	*logistic*	*logistic*	*probit*
	(1)	(2)	(3)	(4)
is_sub_west	-0.112^{*}	-0.165	-0.117	
	(0.067)	(0.158)	(0.072)	
is_sub_east	0.202^{***}	-0.539^{***}	0.160^{**}	
	(0.071)	(0.188)	(0.081)	
$is\_\widehat{sub}\_west$				-0.061
				(0.047)
$is\_\widehat{sub}\_east$				-0.090
				(0.107)
post2015	6.648^{***}	0.186^{*}	6.641^{***}	3.486^{***}
	(0.382)	(0.099)	(0.382)	(0.146)
Constant	18.698^{***}	4.057	18.668^{***}	10.756^{***}
	(1.702)	(3.024)	(1.735)	(0.939)
Macro Factors	No	No	Yes	Yes
Borrowers' Profile	Yes	Yes	Yes	Yes
Loan Characteristics	Yes	Yes	Yes	Yes
Observations	77,992	6,557	77,992	77,992
McFadden's R^2	0.6022	0.3205	0.6022	0.6033
Note:		*p<0.1; **p<0.05; ***p<0.01		

Figure A.2: Region Map