Cultural Diversity and Borrowers' Behavior: Evidence from Peer-to-Peer Lending

Zhongfei Chen[1], Ming Jin[1], Athanasios Andrikopoulos[2], Youwei Li[2]

1. School of Economics, Jinan University, No.601 Huangpu West Road, Guangzhou, Guangdong 510632, China.
2. Hull University Business School, Hull HU67RX, UK. Email: A.Andrikopoulos@hull.ac.uk. Corresponding Author. Athanasios Andrikopoulos. ORCID: 0000-0003-3998-5751.

Abstract

We study cultural diversity and borrowers' behavior using data of peer-to-peer lending platform Renrendai. We proxy cultural diversity with the Linguistic Diversity Index, measured by the population weighted number of dialects spoken in a region, and we show that it has a negative (positive) effect on the loan amount (default rate) of the borrowers. We address endogeneity using two novel instruments, the river length and land slope of Chinese cities, a Heckman two-stage model, and an IV-Heckit model. We also study areas where financial institutions' loan balances are higher (lower) than average. In areas with low (high) loan balances, the amount borrowed (the default rate) is affected more (less). We argue that lenders' behavior is a reason that borrowers in diverse cultures apply for smaller loans. Our results pass a number of robustness tests. Finally, we offer suggestions for improving risk management and inclusive financial development.

Keywords: Culture; Cultural diversity; Dialect diversity; Peer-to-peer lending.

JEL Classification Numbers: G21; G41; G20.

1. Introduction

Peer-to-peer (P2P) lending offers a new opportunity for lenders and investors (Dorfleitner et al, 2016). Based on the Internet, it lowers transaction costs compared to traditional banking (Wei and Lin, 2016). P2P lending complements the shortcomings of traditional banking for small loans and replaces banking for infra-marginal bank borrowers (Tang, 2019).

China exemplifies the significant effect FinTech can have on the financial industry. Its mobile-based connectivity ecosystem, the scarcity of consumer-targeted bank offerings and the innovation-friendly regulatory framework have allowed P2P lending firms to seize large market shares. China is the largest market for FinTech credit, with about 2,500 FinTech credit platforms, and their credit volumes have been steadily increasing, with cumulative lending reaching CNY 1.36 trillion (around USD 210 billion) (Vives, 2019).

Culture is a global concept of social relationship and an important question is whether borrowers from different cultural backgrounds have distinct borrowing behaviors. However, its quantitative measurement is difficult due to domestic cultural diversity. Researchers generally regard country of origin, religion and ethnicity as proxy variables for different cultures (Fisman et al., 2017), but the feasibility of applying it to China is relatively low. It is obviously not realistic to use country type as a proxy variable. Even between cities, cultures are too extraordinary to measure: some cities share the same culture with nearby cities, while others are culturally distinct from close cities and have cultural similarities with cities that are geographically distant, not to mention that there are numerous multicultural cities.

Religion is also not a good choice of variable for Chinese cities since there is a variety of religious beliefs in China and it is difficult to determine the religion of P2P borrowers through historical borrowing information. China has 56 ethnic groups, so ethnic groups might appear to be a good proxy for culture. However, the distribution of ethnic groups in China is geographically uneven: the Han nationality has the widest distribution and its population occupies the largest proportion, while the other 55 ethnic minorities are small and live with their own tribe or together with other tribes. The distribution of settlements and large mixed habitats makes it difficult to obtain their specific distribution status, so the relevant data is difficult to obtain. Besides, the subtle cultural differences of the ethnic group blends make it even more difficult to use regional variables. Thus, using racial data is also infeasible.

Dialect is the cultural symbol of a region. People who use the same dialect often have a stronger sense of cultural identity. According to the *Chinese Dialect Dictionary*, there are 17 major dialects in China, and these dialects are divided into 105 subdialects (Xu and Ichiro, 1999). However, it is hard to identify the dialect that a borrower uses. Luckily, studies show that diversity is also an important factor, which changes not only economics, but also personal utility functions (Easterly and Levine, 1997; Akerlof and Kranton, 2000). Moreover, it is possible to calculate the dialect diversity of the city where a borrower lives.

So, is personal financial behavior also affected by the diversity of dialects? People using different dialects have distinct living habits and cultural conventions, almost as if they are from different countries (Erbaugh, 1995). Thus, if there are multiple dialects in an area, this will lead to communication barriers in the region and, ultimately, borrowers with different dialects will be affected. The Ethnolinguistic Fractionalization Index (ELF) has been used by Mauro (1995), Easterly and Levine (1997), and Alesina et al. (2003) showing that linguistic diversity makes a difference in many areas. Bian et al. (2019) examine the value of language as an important dimension of culture to banks in China. We use dialect diversity in China as a proxy for cultural diversity. Cultural diversity leads to communication barriers, interest bias, and prejudice which changes local trust level. Moreover, although the popularization of Mandarin has reduced some communication barriers, dialect differences still reflect persistent cultural differences and the effect of dialects on culture identity would still exist (Falck et al., 2012). In other words, cultural diversity may affect social capital and cultural identity, and as a result can change personal financial behavior. Accordingly, borrowers who live in an area where there is a variety of dialects may be influenced in their borrowing behavior due to lack of trust, communication barriers, or different financial concepts. This diverse cultural background is also one of personal soft information.

So, we use dialect diversity as a proxy for cultural diversity. Although Fisman et al. (2017) argue that local cultural diversity may affect borrowers' behavior; our paper differs from their study, which discusses culture and loan outcomes, while we try to understand the effect of cultural diversity on personal financial behavior. Culture and cultural diversity are different concepts. Moreover, they regard country of origin, religion, and ethnicity as proxy variables for different cultures, but the feasibility of using these variables for China is low. Hence, we offer a different proxy for cultural diversity, which is more reasonable for the Chinese P2P market, the largest FinTech market for credit (Vives, 2019).

Still, it might not be obvious whether dialect diversity is an appropriate proxy for cultural diversity. For example, it may be argued that the popularization of Mandarin would make dialects to no longer be a symbol of culture. Mandarin has been popularized for many years, a fact which may help people with different dialects to communicate in commercial affairs. However, Erbaugh (1995) argues that dialect has a long tradition and it has not significantly been weakened by the popularization of Mandarin. After field investigation in Hebei, Jiangsu, and Guangxi in 2011, the State Language Commission of China also concluded that although Mandarin has been the official language in China for years, the dialects remain the basic means of communication for local residents. Moreover, a rather small proportion of residents uses Mandarin as their mother language in provinces like Hebei. In the mega cities like Beijing and Shanghai, there are many foreigners. However, most of them can speak Mandarin in addition to their own dialects. The ability to speak a common language can play an important role. However, the same dialect indicates people from the same background, which may also play a different and important role in economic activities. Falck et al. (2012) also argue that although standard German is now much more widespread, dialect differences still reflect persistent cultural differences across German regions that have developed over centuries. We believe that dialects still play a vital role in Chinese daily life and a borrower's financial behavior is affected by dialect diversity.

Even if the language barriers or foreign accents are not an issue for a borrower, Fisman et al. (2017) show that same culture between lenders and borrowers could improve both loan quality and quantity. Large cultural diversity means it is unlikely that a borrower shares the same dialect with the credit auditor of a local bank, which may decrease their success rate of borrowing money from that bank. Moreover, our study is not the only one that treats dialects as a proxy for culture. Falck et al. (2012) also argue that dialects can be used as a measure of cultural identity. Gong et al. (2011) also use dialect diversity as a proxy for cultural diversity. Finally, it is worth mentioning cultural differences and cultural discrimination among different groups of dialect users (Falck et al., 2012; Liu et al., 2020).

Culture-related research mainly involves corporate management, macroeconomics and finance; the extant literature on the impact of culture on personal financial behavior is growing (Tan et al., 2019).[1] Research on P2P lending also focuses more on the personal information listed on websites, but other perspectives should be considered as well. Moreover, P2P lending company bankruptcies occur frequently. We offer the criteria to classify a P2P lending company in the high-risk category in Table A1. This information suggests what kind of borrowers of online P2P lending present higher risk.

The main premise of our study is that culture creates barriers in the lending markets. Borrowers ashamed to borrow or who cannot borrow from local banks due to language barriers or cultural discrimination would choose P2P lending for borrowing (Falck et al., 2012; Fisman et al., 2017), since nobody would know they have borrowed money, or they would not need to communicate with anyone. We also consider migration in robustness checks, and the results remain robust.

We offer a number of significant contributions. To the best of our knowledge, we are the first to examine the link between cultural diversity and borrowers' behavior in peer-to-peer lending. We also propose two novel instruments, the river length and land slope of cities in China, to address endogeneity issues. As an example, cultural conventions may affect both dialect diversity and an individual's financial behavior.

Moreover, in conjunction with the dialect diversity database, we construct an unbalanced panel with 210,841 historical loan records covering the period 2013–2018 and examine the impact of cultural diversity on lending behavior. In this way, we expand the literature on P2P lending and contribute to a better understanding of microfinance from the perspective of cultural diversity. Finally, we address the gap in the literature on how culture affects borrowing behavior and offer policy implications for risk management and government development of inclusive finance.

In the next section, we review the relevant literatures on P2P lending and culture. In Section 3, we formulate our research hypotheses. Section 4 provides our sample and summary statistics and the methodology used. Section 5 provides the empirical results and accounts for endogeneity. Section 6 tests the robustness of our results. Section 7 offers a number of policy implications and the final section concludes.

2. Related Literature

P2P lending is a new and important product for the modern finance industry. Analysts follow factors influencing borrowers' financial actions on P2P lending firms and classify them into soft and hard information (Bachmann et al., 2011). Hard information refers to personal information that can be collected in a standardized way, such as gender, age, income, education, and credit rating (Lin et al., 2017). The advantages of hard information are that it is easy to gather from the P2P lending platform's website and is objective.

However, hard information alone is not enough. Soft information is not as easy to analyze, but it is helpful in matching the borrowers with the lenders more efficiently. Many studies on soft information try to find connections between the words borrowers use in their lending description and the borrowers' mood. Research into loan descriptions is abundant. For example, Larrimore et al. (2011) claim that appropriate language use in borrowers' lending purpose description would improve their persuasiveness. Herzenstein et al. (2011) argue that identity affects the decision of investors.

Moreover, previous studies find a connection between P2P lending and other soft information, such as pictures of borrowers and even punctuation marks in lending descriptions (Duarte et al., 2012). Dorfleitner et al. (2016) study the relation of soft factors derived from the description texts to the probability of successful funding and to the probability of default in P2P lending. Spelling errors, punctuation, text length and the mentioning of positive emotion-inducing keywords predict the funding and default probability on P2P lending platforms under study (Jiang et al., 2017; Chen et al., 2018; Xia et al., 2020; Zhang et al., 2020).

Social interactions also reduce the information asymmetry and enhance the success rate of borrowers in P2P lending (Lin et al., 2013). Cultural dimensions are popular in business and finance research (Karolyi, 2016). Social trust can affect finance and transaction costs (Aggarwal and Goodell, 2014). Moreover, financial decisions can be affected by culture at the individual, firm and country level (Anderson et al., 2011).

Hahn (2013) suggests that common cultural heritage between countries promotes cross-border bank loans. According to Fisman et al. (2017), cultural similarities between lenders and borrowers have been shown to improve the quality of loan and reduce default rate.

Tan et al. (2019) offer evidence that cultural differences affect trading behavior. More specifically, they capture how culture can affect an individual's propensity to trade, and the way the person trades. Moreover, lenders tend to lend money to borrowers with fewer differences in culture or residence (Burtch et al., 2014). Anderson et al. (2011) investigate the impact of cross-cultural differences in behavior on international portfolio allocations.

However, it is not easy to quantify culture diversity in China since there is a large number of different cultures with a complex distribution. Luckily, we can use cultural diversity. Factors like gender (Chen et al., 2020), linguistics (Desmet et al., 2020) and culture (Frijns et al., 2016) make diversity a complex topic. Chen et al. (2020), using data from Renrendai, show that lending to female borrowers is linked to better loan performance, compared to their male peers. However, female borrowers must compensate lenders by having higher profitability to have a similar funding likelihood. Thus, they document a gender gap against female borrowers.

Earlier research suggests that the different dimensions of diversity include geographic location and socioeconomic factors. Easterly and Levine (1997) argue that ethnic diversity explains most of Africa's underdeveloped characteristics, including undeveloped financial systems and distorted foreign exchange markets.

Social or ethnic diversity, similarly to cultural diversity, has an effect on economic decisions. For instance, Alesina and Ferrara (2005) argue that cities in America are creative and prosperous while having significant ethnic and cultural diversity. Studies on China also support the view that diversity matters; for example, Herrmann-Pillath et al. (2014) find that ethnic diversity has a negative impact on economic growth across Chinese provinces.

As ethnic diversity has an impact on economics and may even lead to undeveloped financial systems, does cultural diversity also play a role in financial systems and individual financial behavior? Cultural diversity may lead to financial behavior diversity.

Chinese cities are differentiated culturally because of the vast area of the country and the migration of people from different regions that has taken place over the centuries. Normally, the dialects of immigrants would be different from those of the locals, such as the Hakka (or "the guest people") have the same dialect and living habits. Erbaugh (1995) argues that people using different dialects are like foreigners. Norman (1988) claims that the Chinese language family is more like related language groups rather than dialects of the same language. Falck et al. (2012) argue that dialect is an important aspect of cultural identity, which creates intangible cultural borders within a country and impedes economic exchange.

Previous research provides strong evidence that linguistics can affect economics and institutions. For instance, Alesina et al. (2003) use historical data from 190 countries to measure the effects of ethnic, linguistic, and religious heterogeneity, and find they have an impact on the quality of institutions and growth. Akerlof and Kranton (2000) also use the ELF and find that it has an impact on economics in Africa. Several recent studies support the view that cultural diversity, which is proxied by dialect diversity, affects a firm's decision, migration, etc. (Bian et al., 2019; Liu et al., 2020; Jiang et al., 2020). However, very few studies have considered individual differences in diversified cultures like those in China.

We analyze how cultural diversity, which we utilize as a proxy for soft information, affects borrowers' actions using a unique cultural diversity database built by Liu et al. (2020) and Renrendai's historical transaction data. Our contribution is not only meant to help lenders acquire more information to reduce the default rate, but also to narrow the gap in the literature on how cultural diversity influences personal financial decisions.

3. Hypotheses

As we discussed in the literature review section, soft information, such as spelling errors, punctuation, text length and the mentioning of positive emotion-inducing keywords, affects the borrowers' funding and default probability on P2P lending (Jiang et al., 2017; Chen et al., 2018; Xia et al., 2020; Zhang et al., 2020). Same culture means a higher level of trust and lower communication and transaction costs, which help to improve the quality of loans (Fisman et al., 2017).

Living in a city with a high level of cultural diversity means that chances are high that other people are from different cultural backgrounds, which in turn means high transaction costs or low trust levels within the city. As a result, the financial decisions of borrowers may be affected by cultural diversity. Cultural diversity is also soft information, which affects the behavior of borrowers in P2P lending markets. Therefore, we argue that individuals' borrowing behavior from multidialectal regions are influenced by cultural diversity. Thus, we explore the impact of cultural diversity on both the amount of borrowing and the default rate of borrowers.

Fisman et al. (2017) argue that same culture between lenders and borrowers helps to increase loan amounts. The amount of the loan that borrowers apply for should reflect their personal borrowing requirement to a large extent. One might expect that the loan amount may not be affected by cultural diversity in P2P lending markets since the lenders and borrowers do not know each other. However, there are three types of loans in Renrendai platform, including pure online credit loans, institution guarantee loans, and field certification loans. Borrowers of the first type are totally isolated from lenders and indeed do not know each other; however in the latter two, borrowers must get a certificate offline, which means they must contact someone who represents the Renrendai platform or other financial institutions.

Actually, the loan amounts in the Renrendai platform are not entirely determined by borrowers' demand. In all the loan types, borrowers apply for a certain loan at first, and Renrendai or local institutions with partnerships review their information and decide the final loan amount and interest rate. Borrowers of institution guarantee loans and field certification loans must pass the offline personal qualification reviews.

Before November 30, 2017, Renrendai had a special investor protection policy, according to which if borrowers fail to keep their promises, institution or Renrendai would repay the loan principal using their reserves. After 2018, Renrendai changed the investor protection policy and stopped drawing reserves from loans. They announced that after borrowers defaulting, and if a third party is willing to pay the borrower overdue payments, the creditor's rights are transferred to the third party. Sometimes, the P2P lending platform and the institutions with partnerships will also play the role of a third party. Therefore, no matter which kind of policy is, they do not want borrowers to default. As a result, Renrendai and institutions with partnerships also treat loan applications with caution as banks do, which makes the situation very similar to the situation of Fisman et al. (2017).

The applications of pure online credit loans are reviewed online. While for borrowers of institution guarantee loans and field certification loans, it is necessary for them to conduct qualification review offline. Therefore, in institution guarantee loans and field certification loans, accent bias, communication problems and cultural discrimination still exist and affect loan amounts.

Based on the previous discussion, we develop the following main Hypothesis about how cultural diversity affects borrowing demand from P2P lending companies.

Hypothesis 1: In cities with higher cultural diversity, the amounts of institution guarantee loans and field certification loans will be smaller.

Aggarwal and Goodell (2014) argue that culture affects social trust and financial transaction costs. Although the dialects diversity data is different from other cultural data, there is existing evidence, which illustrates that diversity has an impact on financial development and an individual's utility function (Easterly and Levine, 1997; Akerlof and Kranton, 2000).

Fisman et al. (2017) argue that cultural proximity will increase the quantity of credit and reduce default. Although our paper focuses on online loans, it cannot be denied that in societies with high cultural diversity individuals have less sense of cooperation due to strategic uncertainty (see Kets and Sandroni, 2021). Moreover, the P2P lending in Renrendai includes the loan application, qualification review, and collection of arrears, which makes the online cases very similar to the off-line cases.

Dilling (2011) finds that cultural diversity has a negative impact on Corporate Social Responsibility (CSR) perception. Dinesen and Sønderskov (2015) emphasize that cultural diversity reduces social trust. Chen et al. (2016) argue that a low likelihood of default and high financial reporting quality are the channels through which social trust promotes bank finance to Non-state-owned enterprises (NSOEs) in China. Low social trust means that locals are less likely to keep their contracts, and this phenomenon will also be reflected in online transactions.

Therefore, we assume that borrowers will be affected by dialect diversity, and people coming from cities with higher dialect diversity will have a higher default rate. Thus, we develop the following main hypothesis about how cultural diversity affects default rates.

Hypothesis 2: The default rate of P2P borrowers who live in cities with higher cultural diversity should be higher as compared to the default rate of P2P borrowers who live in cities with lower cultural diversity.

It is more difficult for people from different cultural backgrounds to trust each other (Fisman, et al., 2017; Bian, et al., 2019). As a result, the loan amounts are not only affected by offline reviews, they might be affected by investors' decisions as well. In order to get a loan from investors, borrowers must be consistent with investors' requirements and meet their demands. Hence, we develop the following hypothesis:

Hypothesis 3: Borrowers in multicultural cities apply for small amounts of money in order to increase the probability of getting approved.

4. Sample and Methodology

4.1 Sample and summary statistics

Our sample is obtained from Renrendai's historical loan records and correspondences to the Chinese market. We construct an unbalanced panel with 210,841 historical loan records covering the period from 2013 to 2018. Renrendai is a leading P2P online loan platform and one of the oldest of its kind in China (Chen et al., 2020). Compared with other regional P2P platforms, this platform has a good reputation, so the time span is longer and its users are from all over the country. Thus, the characteristics of borrowers in different regions can be obtained from their historical transaction data and are highly representative. Hence, we can analyze the relationship between borrowers' loan amount and default rate, and the dialect diversity index, measured by the population weighted number of dialects spoken in a region.

Our aim is to study the impact of cultural diversity on the behavior of borrowers. More specifically, we study the impact of cultural diversity on borrowing requirements and default risk. We proxy borrowing requirements using the amount borrowed and the default risk from the borrowing default rate of Renrendai's historical transaction data. The borrowing amount is obtained directly from the historical data, and the default rate is reflected by the default of the borrower. It should be noted that in 2016 the Renrendai loan platform adjusted the contract number of loose-standard loans in order to make them untraceable. Renrendai also increased the standard of their borrowers' review conditions, so the default rate was greatly reduced. Since defaults occur more often in earlier years, recent years' loan data are less likely to show default. If only defaults and related data are considered, there will be selection bias. Hence, we only use historical data before 2017 when analyzing the default rate.

Our main explanatory variable is cultural diversity as proxied by dialect diversity. We use the dialect diversity index for 276 prefecture-level cities in China. Our dialect data is from a 1980s survey (Xu and Ichiro, 1999; Liu et al., 2020). The dialects covered 2113 counties of China. Fig.1 offers a picture of the number of different dialects across China, and Table 2 offers the summary statistics of loan status, cultural diversity, personal information and the macroeconomic environment in each prefecture-level city (denoted as city for short).[2] As we can see from Fig.1 and Table 2, the number of dialects varies from region to region and is unevenly distributed - from one to five dialects in a city. Since the indicator "default" does not include new borrowing data after 2017 or borrowing failure data for 2017, the number of this indicator is significantly less than other indicators.

[Insert Figure 1]

Instead of considering only the number of dialects spoken in a city, it is better to take a weighted number of dialects. The reason can be explained with an example. Suppose there are two dialects, a and b, spoken both in city A and city B, the percentage of people speaking dialect a (b) is 10% (90%) and 50% (50%) in city A and city B, respectively. If we calculate the number of dialects, there are two dialects both in city A and city B, but only 10% of the population speaks dialect a in city A, while 50% of the population speaks dialect a in city B.

Moreover, Montalvo and Reynal-Querol (2005) argue that it is better to use an index of discrete polarization to measure potential conflicts. Therefore, to solve this problem, Xu et al. (2015) count the population of each dialect users by one minus the Herfindahl index of the ethnolinguistic group. Eq. (1) is the dialect diversity index, marked as *Diver*.[3]

$$Diver_m = 1 - \sum_{j=1}^{n} S_{mj}^2, \tag{1}$$

where n is the total number of dialects other than the main one used in the city, and S_{mj} represents the population weight of the *j-th* sub-dialect in city m. Following Xu et al. (2015) and Liu et al. (2020), the population weights are based on the fifth population census in China carried out in 2000. When the dialect in a city is concentrated, S_{mj}^2 is larger and $Diver_m$ is lower; conversely, when the dialect within a city is more dispersed, S_{mj}^2 is lower and $Diver_m$ is higher. In China, there are 17 main dialects and 105 sub-dialects, hence n takes values from

0 to 104; there are 276 prefecture-level cities with available data used in the current study, hence *m* takes values from 1 to 278. Eq. (1) has also been used by Mauro (1995), Easterly and Levine (1997) and Alesina et al. (2003) to calculate the ELF, which is similar to our dialect diversity index. As a robustness test (see Section 5.5), we use the number of dialects (*Dn*) as an explanatory variable.

Apart from cultural diversity, the borrowing amount and borrower's default rate are influenced by the borrowers' characteristics. Thus, we use the users' own information influencing factors in all regressions, including gender, marital status, work experience, education, age, house property, car property, and income level as control variables.

Age, house property, and car property data can be obtained directly from historical transaction data on the Renrendai website. Gender, marital status, work experience, education, and income levels need to be handled afterwards. Gender is a dummy variable that takes the value of one for female borrowers and zero for male borrowers. The different categories of marital status available on the website are "unmarried", "divorced", "widowed" and "married". We create a dummy variable *Marital status*, which takes the value of one if the borrower is married and zero otherwise. Education level is a variable that takes values on a scale of one for borrowers with "high school or lower", two for "college education" or "junior college", three for "Undergraduate degree" and four for borrowers with "Graduate degree or above".

In order to protect the privacy of the borrower, work experience and income level are only given a range in the Renrendai loan website, so we averaged the ranges.[4] We take this value when the range has only the maximum or minimum value. For instance, when the work experience of the borrower is "within one year" and the income level is "5000-10,000 CNY", then we define the work experience as one year and the income level as 7,500 CNY.

In addition to the borrower's personal features, their default rate will also be affected by the amount borrowed, the borrowing interest rate and the duration of the loan. Therefore, we use the borrower's repayment cycle (such as monthly repayment), the type of borrowing (such as equal principal and interest) and other information to measure the single-period repayment amount (*Lppay*) indicator, which controls the content contained in the above-mentioned order attributes, and reduces the loss of freedom caused by too many control variables.

It should be noted that in the past, credit scores or credit ratings were important control variables. However, the personal data of Renrendai's loan platform is updated in real time, hence, whenever the website is accessed, the borrowers' information is updated. Because of this up-to-date information, the borrower's credit level should not be used as a control variable for the amount borrowed or the default rate since it would decrease if they defaulted.

[Insert Table 1]

Considering that more developed cities attract more immigrants from all over the country, resulting in greater dialect diversity, we control the impact of a number of macroeconomic indicators. The economic situation and geographical location of different regions affect the financial behavior of local residents. Hence, the Gross Domestic Product (GDP) of the borrower's area is used as a control variable, and the data is from *City Statistical Yearbook.* The location of the city is also an important factor.[5] Defaulters in the provincial capital are more likely to be collected by the platform. Therefore, we control for the impact of this attribute on the default rate.

The regional information matters for borrowers' funding probability and default probability (Wang et al. (2021). Therefore, we also control for province fixed effects to measure geographical influence, and these dummy variables can also make up for some differences in regional financial levels that have not been considered. At the same time, in order to control the effects of different years of policies and economic levels on borrowers' financial behavior, we also add year fixed effects in the regressions. GDP per capita is a better variable to measure the economic environment since we study personal financial behavior. However, the population data in China are collected using surveys, and hence contain statistical errors. Moreover, there are data that are more available for GDP compared to GDP per capita. These reasons make GDP a better choice. However, we test the robustness of our results using GDP per capita and average income.

Considering that the amount borrowed, the income level of the borrower, and the sum of the regional income of the borrower's location are large, we use natural logarithms when using these indicators. Table 1 defines the variables.

[Insert Table 2]

Table 2 offers the summary statistics of the variables under study. 28.74% of the borrowers are women and the average work experience is three years. 55.33% of the borrowers are married. The average age of borrowers is 36 years. 47.40% of the borrowers own a house and 31.94% own a car. The region has an average of 1.76 dialects and an average differentiation index of 0.24. Borrowers in provincial capitals account for 42.9% of the sample.

4.2 Methodology

To test our first hypothesis, the estimation equation is:

$$L\_amount_{i,t} = c_p + \tau_t + \beta_0 + \beta_1 Diver_m + \sum_{k=2}^{K} \beta_k \, Control_{i,t} + \varepsilon_{i,t} \qquad (2)$$

where $L\_amount_{i,t}$ is the logarithm of the borrowing amount that loan $i = 1,2 \ldots N$, was issued at time $t = 1,2 \ldots T$, where t is measured in years.[6] c_p is a time-invariant provincial-specific unobservable effect. There are 34 provinces (municipalities/autonomous regions/special administrative regions) in China, of which 30 with available data.[7] τ_t is a common unobservable year-specific effect and $\varepsilon_{i,t}$ is the time-varying individual-specific idiosyncratic error. $Diver_m$ is the explanatory variable as proxied by the dialect diversity of the borrower's location (see Figure 1 and Table A2., describing the number of dialects in each city). *Control* includes 11 variables, nine personal information variables and two city's macroeconomic variables. The control variables are presented in Table 1.

By studying the sign of the coefficient β_1, we can obtain the influence of cultural diversity on the dependent variable (loan amount). We can obtain the magnitude of the influence by comparing the β_1 coefficients in different subsamples. According to Hypothesis 1, the sign of the borrowing amount and the cultural diversity index coefficient tested using model (2) should be negative. For Hypothesis 2, we use a Probit model as follows:

$$Probit\left(Bdebt_{i,t}\right) = d_p + \mu_t + \gamma_0 + \gamma_1 Diver_m + \sum_{k=2}^{K} \gamma_k Control_{i,t} + u_{i,t} \quad (3)$$

where $Bdebt_{i,t}$ is a dummy variable that takes the value of one if the loan has been defaulted and zero otherwise. $u_{i,t}$ is the time-varying individual-specific idiosyncratic error. The other variables are defined analogously as in Equation (2). According to our hypothesis, the sign of the cultural diversity index coefficient tested by model (3) should be positive.

5. Empirical results

5.1 Loan amount and default rate

Table 3 reports the impact of cultural diversity on the natural logarithm of the loan amount. Columns (1)-(3) present the findings without control variables (1), with borrowers' attribute variables (2) (including gender, work experience, marriage, education, age, house ownership, car ownership and log of income), with city characteristics (log of GDP), fixed province and time effects (3), respectively. The results show that whether we include control variables or not, the coefficient on cultural diversity is significantly negative. On average, when dialect diversity increases by 1%, the natural logarithm of loan amount decreases by

0.05%. Moreover, cultural diversity of borrowers' city reduces the amount borrowers applied for, which is consistent with our first hypothesis.

[Insert Table 3]

According to our first hypothesis, in cities with higher cultural diversity, the loan amounts are smaller. Our results support this hypothesis. Hence, borrowers from cities with lower cultural diversity may have more borrowing opportunities, so when they demand a small loan, other borrowing channels will help them out. For the same reason, in cities with high cultural diversity borrowers have fewer borrowing opportunities and resort to P2P lending for even small loan amounts. It is also likely that borrowers from cities with lower cultural diversity may lack lending experience, so they tend to try micro loans out of caution.

Turn to control variables, female, married and well-educated borrowers tend to apply for higher amounts of money ceteris paribus. The same is true for older, high-income borrowers and those with many years of work experience. Borrowers who own a house or a car apply for larger amounts compared to those who do not own a house or a car. Higher GDP of the city where the borrower lives also increases the amount borrowed at any one time.

The default rate of borrowers is one of the key variables in this study since it can indicate characteristics of defaulters, which, in turn, help lenders and P2P lending companies identify high-risk borrowers. In this section, we explore the impact of cultural diversity on borrowers' default rate. Table 4 offers the results. Columns (1)-(3) present the regression results without control variables (1), with personal attributes (2) with the characteristics of the borrower's city of residence, province and time fixed effects (3). In columns (1)-(3) the coefficients of cultural diversity are all significantly positive. Hence, the higher the cultural diversity borrowers have, the higher their default rate is, which means both lenders and companies should take into account this soft information.

The coefficients of the other variables indicate that personal and city-specific variables have an impact on borrowers' default rate. More specifically, low repayment amount per period, being male, long work experience, being unmarried, low education level, being younger, owning a car, and high-income increase the likelihood of default.

It seems puzzling that lower repayment amount and higher income levels make borrowers tend to renege on their debt. P2P lending companies would spend less effort collecting the overdue loans for the sake of costs. Borrowers of small amounts are also more likely to default on their loans. As information on income level is not reviewed, high default risk borrowers often fake a high income in order to get money from a lender.

We use *pcapital* and *L_GDP* to measure if a borrower lives in a provincial capital, and the income level of the city, respectively. The results show that borrowers from capital cities and high GDP cities are linked with low default. Borrowers from capital cities and high GDP cities may have higher income and job opportunities, so they have better repayment abilities. Moreover, provincial capital cities and high GDP cities have better transportation, which makes debt collection easier and defaulter prosecution more likely.

The results of Tables 3 and 4 suggest that cultural diversity has an impact on an individual's financial behavior. More specifically, borrowers in multicultural cities generally apply for smaller loans at any given time and have higher default rates.

[Insert Table 4]

5.2 Accounting for Endogeneity

5.2.1 Instrumental variables

Endogeneity problems could be caused by measurement errors, omitted important explanatory variables, or simultaneity bias (Greene, 2012). In this study, measurement errors or unobservable variables may affect dialect diversity and financial development at the same time and cause endogeneity problems when we use dialect diversity as an explanatory variable. For example, cultural conventions may affect both dialect diversity and an individual's financial behavior. According to unbalanced development, there is also the risk of minority languages being endangered. People choose to learn and use Mandarin, which offers a communication advantage as it is the most widely spoken language in China. We address endogeneity by using two instrumental variables for dialect diversity.

Dialect formation is influenced by a number of geographical factors. Rivers have been significant geographic barriers causing a lack of communication between residents of different areas. However, rivers have also been important paths for the immigrants' long migration by boat. Moreover, humans have historically lived along rivers (Gani and Gani, 2011) and seafaring people often settled near the rivers after migrating there. Therefore, we expect riverside areas to have more dialects, not only because it was easier to form new cultures, but also because it was more likely that those areas would accept new migrants and new dialects. Hence, we expect the total river length of cities to be positively correlated with dialect diversity (or cultural diversity).

Moreover, previous studies (Michalopulous, 2012; Bian et al, 2019) show that land quality and elevation are strongly associated with linguistic diversity. Slope is the relief or terrain degree of land surface, which is a comprehensive representation of regional altitude and surface roughness and could stand for land quality and elevation. Areas with more hillsides would have higher land slopes and larger dialect diversity.

Although in the past urban centers were often concentrated near rivers, modern transportation made rivers no longer an important factor for financial development, so this variable meets the exogeneity requirement of an instrument. Land quality and elevation no longer affect financial markets, due to the modern construction industry, so this variable should meet the exogeneity requirement of an instrument, as well. Hence, we address endogeneity by using river length and land slope as instruments for dialect diversity.

Both river length and land slope are at the city level. We use ArcMap to calculate the sum of all rivers in the 1:40000 vector map of China National Basic Geographic Information Center as the length of rivers in each city (Rl). We take logarithms of the variable river length and thus reduce the influence of outliers. The data for land slope is at a 1 km resolution and from You et al. (2018) with the re-sampled Digital Elevation Model data.

[Insert Table 5]

Table 5 presents the results after adding river length and land slope as instruments. The results remain the same with the baseline results. We performed a weak instrument variable test on the instruments. From the first stage F-statistics, AR statistics, and Wald statistics, we see that both river length and land slope passed the relevance and exogeneity conditions to be considered valid instruments. Hence, the conclusion that dialect diversity affects personal financial behavior remains robust after dealing with endogeneity.

5.2.2 Heckman two-stage model

The sample of P2P loans does not include applicants who have not received a loan. One potential problem is that unobservable features, which influence the applicant's ability to receive a loan and are correlated with the dialect diversity, could drive the results. For example, only higher quality applicants could receive a loan.

To assess the extent to which this potential selection bias affects the results, we estimate a Heckman (1979) two-stage sample selection model for the baseline Model (3). A selection equation estimates the probability of funding success using a Probit model.

Table 6 offers the results, where there are 57,355 failed loan applications (i.e. the difference between the number of observations (108,552) minus the number of censored observations (51,197), namely 57,355 unfunded loans). After correcting for the sample selection, the dialect diversity variable remains statistically significant in explaining the probability of default.

[Insert Table 6]

We also, use IV-Heckit model (Wooldridge, 2001) to deal with endogeneity in the identification of the equation (3). The results are presented in Table 7. We estimate a Probit model for funding success rate as the selection indicator in step 1. Then we obtain the inverse Mills ratio from step 1. Finally, we estimate the structural equation using an IV-probit. The impact of dialect diversity on the funding success rate remains significant when we control for endogeneity using the IV-Heckit model.

[Insert Table 7]

5.3 Loan types

The borrowers in the Renrendai platform can apply for pure credit loans, institution guarantee loans, or field certification loans. The first one is an online application, which means there is no offline communication in the transaction. Institution guarantee loans are guaranteed by financial companies, so the borrowers must communicate with the financial companies. In field certification loans, the borrowers' characteristics are confirmed by Ucredit (www.ucredit.cn), which is a subsidiary of Renrendai, specializing in offline services.

As we discussed in Hypothesis 1, the loan amount may be affected by the accent bias, communication problems or cultural discrimination. Columns (1) and (2) of Table 8 present the results that correspond to Hypothesis 1 when we divide the entire sample by whether the borrower has offline communication with others. For the pure online loans, there is no offline communication, dialect diversity is insignificant to the loan amounts, as shown in Table 8. For this kind of loan, there is no need for face-to-face audit, indicating no communication. In contrast, in columns (2) and (3), which need offline auditing and screening, the results are quite different. The dialect diversity reduces the loan amount. This result confirms the role of language barriers.

Moreover, we provide the default rate results when the borrower does not have offline communication with others in column (3) of Table 8, and the coefficient is still significantly negative. As all the successful funded offline loans (i.e. 134,897 in 157,704) have been repaid, there are no regression results for the offline subsample.

[Insert Table 8]

5.4 Local financial market development

The development of a local financial market is closely associated with individuals' financial behavior, which means that borrowers under different financial conditions are affected by cultural diversity to varying degrees. Therefore, we divide our sample into two groups according to the loan balance of the local bank in the borrower's city. Samples are classified into the "high loan balance" group if the loan balance is higher than average, while the other samples are classified into the "low loan balance" group.

Table 9 reports the impact of cultural diversity on borrower behavior when we control for financial development, measured by local banks' loan balance. In both kinds of cities with a higher and a lower loan balance, cultural diversity has a negative impact on the loan amount of borrowers and the impact is significant at 1% level. Moreover, in cities with a lower loan balance, the absolute value of the coefficient of cultural diversity is higher compared to cities with a higher loan balance, and the difference in the coefficients between the two groups is significant at 1% level of confidence based on the Seemingly unrelated estimation (Suest) test (Table 9). Hence, the loan amounts of borrowers from more culturally diverse cities is smaller than that of borrowers from less culturally diverse cities, and the loan amount decreases significantly in cities with a lower loan balance.

[Insert Table 9]

The analysis in Section 5.1 leads to two possible reasons why the coefficient of the loan amount is negative. One reason is that borrowers in multicultural cities have limited sources of borrowing, even for small loans, so they are forced to borrow from P2P companies. This reason is confirmed here when we observe that borrowers from lower loan balance cities are more sensitive; the absolute value of the coefficients of *Diversity* is much larger than the one for the high loan balance group. Another reason, as we discuss in the robustness checks (5.5.2) section, might be the migrants arriving into the city that do not have a financial track record or collateral, so tend to rely on P2P lending, which is easier for them to access.

The default risk is also positively linked with cultural diversity in both kinds of cities. The absolute value of cultural diversity's coefficient is higher in multicultural cities, and the difference in the coefficients between the two groups is significant at 1% level of confidence based on the Suest test (Table 9). We believe this is because borrowers in cities with a higher loan balance, despite having more loan options, still need to borrow money from a P2P with a higher interest rate, so the loan they apply for may be a high-risk loan that has been rejected by the local financial market. Since there may be peer review, it is not strange that borrowers from financially developed cities have a higher default risk.

In conclusion, in both types of cities, the baseline results are still unchanged, but there are differences between them, which not only indicate the reason for the results in Section 5.1, but also claim a high-risk factor.

5.5 The lenders' decision

According to our results, borrowers tend to apply for a small loan in a single application if they live in multicultural cities. In this section, we explore whether the baseline results change if we use only the success in getting the loan approved sample or the failure sample. We divide the full sample into the success and failure group and re-run the regressions.[8]

The results are presented in Table 10. As we discussed before, for the full sample the coefficient of cultural diversity in column (1) suggests that cultural diversity negatively affects the amount of the loan in a statistically significant way. The negative impact of cultural diversity and loan amount still exists in the success sample, and the absolute value of the coefficient is even larger than in the baseline model. However, the coefficient of cultural diversity in column (3) suggests that it positively affects the amount of the loan in the failure sample, but the coefficient is not statistically significant.

[Insert Table 10]

6. Robustness checks

6.1. An alternative definition of dialect diversity

In order to measure cultural diversity, we use dialect diversity as the key independent variable in the baseline model. Although the number of dialects has a weight asymmetry problem, this indicator has less information loss than dialect diversity. Thus, most cultural diversity studies in China use the number of dialects as a proxy for cultural diversity.

To test the robustness of dialect diversity, we use the number of dialects (*Dn*) as another proxy for cultural diversity. Table 11 offers the results when the main independent variable is the number of dialects. Yet again, cultural diversity has a negative impact on the amount of the loan; the positive impact on the default rate remains unchanged.

[Insert Table 11]

6.2. Interprovincial migration

Our third robustness check refers to migration and its effects on Chinese cultural diversity. Today, migration is still ongoing, as urbanization has accelerated in China during recent decades (Chan, 2013). People still move to other cities to find a better job, to pursue higher education, or to get married. Both the local and their native culture affect migrants. The borrower's city in the historical P2P lending data that we used is based on the city they live in. Therefore, there may be biases since some migrants are not affected by the local culture as much as the natives are. In addition, the fact that most migrants in the mega cities do not use the local dialects, make the problem of measuring cultural diversity more difficult. In order to deal with these problems, we use the ID number provided by Renrendai to discover which province a borrower was born in since interprovincial culture has greater influence.[9] Table 12 offers the results without interprovincial migrants. Our findings are still unchanged when we keep samples without interprovincial migration.

[Insert Table 12]

6.3. Macroeconomic Environment

Then, we use different proxies for the macroeconomic environment. Personal financial behavior is affected by personal income; we have already controlled borrowers' income level as personal information. Moreover, the local economic environment matters as well, so we have also added GDP since the higher the GDP, the more loans are given in an economy. However, to test the robustness of our results, we replace GDP with GDP per capita and average income as proxies for macro economy. We still choose GDP because the population data in China is not reliable and there is more data available on GDP. Table 13 offers the results when we replace GDP with GDP per capita, average salary, and one-year lag of GDP; our results remain robust.[10]

[Insert Table 13]

7. Policy Implications

Our focus is to understand the impact of cultural diversity on lending behavior using a sample of 210,841 historical borrowing data for the period 2013-2018. Towards this objective, we provide a battery of statistical tests. In this section, we discuss the implications of this study for P2P lending companies and traditional financial markets and institutions.

We suggest that online P2P lending companies, when reviewing borrowers, should take note not only of hard information, but also of soft information like cultural background and the financial development of the borrower's location. Although the influence of cultural belonging on personal financial behavior is inevitable, reasonable guidance is still necessary. When formulating the financial policies for different regions, fairness is not the only thing to consider. Proper consideration of the adaptability of the local economy is also necessary, as well as strengthening the residents' financial knowledge, trying to reduce the obstacles to the implementation of policies caused by culture, and improving the effectiveness of policies.

We note that dialect or cultural diversity affect not only borrowers of P2P lending, but also traditional financial markets and institutions. In addition, the behavioral differences of individuals from multicultural societies warn us that culture has a mind of its own, so some policies that promote population concentration should be implemented gradually.

8. Conclusions

We proxy cultural diversity in China using dialect diversity to study the impact of cultural diversity on lending behavior. The empirical results reveal that the amount borrowers apply for in a single application decreases with rising local dialect diversity of the city where the borrower lives, while cultural diversity has a positive impact on default rate. The results remain robust after controlling for loan feature, borrower information, and the characteristics of the borrower's resident city, province, and time fixed effects.

We note that barriers created by dialect diversity are stronger for older generations in China, since the Chinese government implemented compulsory education in 1986.[11] 50% of the borrowers in our sample are older than 35 (Table 2, median age is 36). Another argument is that an online platform like Renrendai expands the loan market beyond geographic boundaries. However, the borrowers of P2P lending are still affected by local cultural diversity as traditional banks do. Borrowers ashamed or unable to borrow from local banks due to language barriers would choose P2P lending since nobody would know they have borrowed money, or they would not need to communicate with anyone.

To address possible endogeneity issues in the baseline model, we use a two-step regression with the river length and the land slope of cities as alternative instruments. The sample of P2P loans does not include applicants that have not received a loan. To assess the extent to which this potential selection bias affects the results, we estimate a Heckman (1979) two-stage sample selection model and an IV-Heckit model to deal with endogeneity. Our results do not alter the baseline results.

In order to study the effect of the development of regional financial markets on the relationship between individuals and culture, we explore the heterogeneity of areas where financial institutions' loan balances are higher or lower than the average loan balance. Both types of regions are affected by cultural diversity in the same way, but the impact is different. In areas with low loan balances, the loan amount of borrowers is affected to a greater degree, while in areas with high loan balances the default rate of borrowers is affected more.

We also conduct regression analysis on the "successful borrowing" and "failed borrowing" groups. The coefficient of cultural diversity on borrowing amount in a single application of the successful borrowing group is still significantly negative and higher than the full sample’s, while the coefficient of the failed borrowing group is positive but insignificant. We use the number of dialects instead of the Linguistic Diversity Index for robustness check. The regression results remain the same.

We consider migration and its effects. By removing the data of population migration between provinces, the regression results are still robust in the non-migrated data. Hence, cultural diversity affects borrowers' borrowing behavior, including borrowing demand and default rates. Finally, we use a Heckman two-stage selection model to deal with potential selection biases in model (3), and the results remain robust.

Our work is not without its limitations, and these limitations themselves suggest interesting questions for future research. One such limitation is the development of a theoretical framework to analyze how cultural diversity affects borrowers' behavior. Moreover, Renrendai is a typical representative P2P lending platform in China; however, the loan amount in one platform does not represent all the P2P lending platforms in China. An interesting future direction is to find data for all the P2P lending platforms in China and link the total loan amount in each city to the number of dialects in this city.

Previous research claims that diversity affects individuals' utility functions, institutions and economics. According to our results, the behavior of residents in different regions on the credit market is affected by local cultural diversity; high levels of cultural diversity result in small loan amounts per single application and high risk of default. The level of mainstream financial development in the borrower's location is also a key factor; when the bank in the borrower's location can meet their loan requirements, borrowers from more culturally diverse cities are more likely to default.

References

Aggarwal, R., Goodell, J. W. 2014. "National cultural dimensions in finance and accounting scholarship: An important gap in the literatures?" Journal of Behavioral and Experimental Finance 1: 1–12.

Akerlof, G. A., Kranton, R. E. 2000. "Economics and identity." Quarterly Journal of Economics 115: 715–753.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., Wacziarg, R. 2003. "Fractionalization." Journal of Economic Growth 8: 155–194.

Alesina, A., Ferrara, E. L. 2005. "Ethnic diversity and economic performance." Journal of

Economic Literature 43: 762–800.

Anderson, C. W., Fedenia, M., Hirschey, M., Skiba, H. 2011. “Cultural influences on home bias and international diversification by institutional investors.” Journal of Banking and Finance 35: 916–934.

Bachmann, A., Becker, A., Buerckner, D., Hilker, M., Kock, F., Lehmann, M., Tiburtius, P. 2011. “Online peer-to-peer lending-a literature review.” Journal of Internet Banking and Commerce 16: 2–18.

Bian, W., Ji, Y., Zhang, H. 2019. “Does dialect similarity add value to banks? Evidence from China.” Journal of Banking and Finance 101: 226–241.

Burtch, G., Ghose, A., Wattal, S. 2014. “Cultural Differences and Geography as Determinants of Online Pro-Social Lending.” Management Information Systems Quarterly 38: 773–794.

Chan, K.W. 2013. China: internal migration. The Encyclopedia of global human migration.

Chen, X., Huang, B., Ye, D. 2018. “The role of punctuation in P2P lending: Evidence from China.” Economic Modelling 68: 634-643.

Chen, X., Huang, B., Ye, D. 2020. “Gender gap in peer-to-peer lending: evidence from China.” Journal of Banking and Finance 112: 105633.

Chui, A., Kwok, C., Zhou, G. 2016. “National culture and the cost of debt.” Journal of Banking and Finance 69: 1–19.

Desmet, K., Gomes, J. F., Ortuño-Ortín, I. 2020. “The geography of linguistic diversity and the provision of public goods.” Journal of Development Economics 143: 102384.

Dilling, P F A. 2011. “Stakeholder perception of corporate social responsibility.” International Journal of Management and Marketing Research, 4: 23–34.

Dinesen, P T, Sønderskov, K M. 2015. “Ethnic diversity and social trust: Evidence from the micro-context.” American Sociological Review, 80: 550–573.

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 and Finance 64: 169–187.

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

Easterly, W., Levine, R. 1997. "Africa's growth tragedy: policies and ethnic divisions." Quarterly Journal of Economics 112: 1203–1250.

Erbaugh, M. S. 1995. "Southern Chinese dialects as a medium for reconciliation within Greater China." Language in Society 24: 79–94.

Falck, O., Heblich, S., Lameli, A., Südekum, J. 2012. "Dialects, cultural identity, and economic exchange." Journal of Urban Economics 72: 225–239.

Fisman, R., Paravisini, D., Vig, V. 2017. "Cultural proximity and loan outcomes." American Economic Review 107: 457–92.

Frijns B, Dodd O, Cimerova H. 2016. "The impact of cultural diversity in corporate boards on firm performance." Journal of Corporate Finance 41: 521–541.

Gani, M. R., Gani, N. D. 2011. "River-margin habitat of Ardipithecus ramidus at Aramis, Ethiopia 4.4 million years ago." Nature communications 2: 602.

Gong, Y., Chow, I. H. S., Ahlstrom, D. 2011. "Cultural diversity in China: Dialect, job embeddedness, and turnover." Asia Pacific Journal of Management 28: 221–238.

Greene, W. H., 2012. Econometric Analysis (7th Edition). Pearson, New York.

Hahn, F. R. 2013. "Culture, Geography and Institutions: Empirical Evidence from Small-scale Banking." Economic Journal 124: 859–886.

Heckman, J. J. 1979. "Sample selection bias as a specification error." Econometrica 47: 153–161.

Herzenstein, M., Sonenshein, S., Dholakia, U. M. 2011. "Tell me a good story and I may lend you money: The role of narratives in peer-to-peer lending decisions.” Journal of Marketing Research 48: 138–149.

Herrmann-Pillath, C., Libman, A., Yu, X., 2014. “Economic integration in China: Politics and culture.” Journal of Comparative Economics 42: 470–492.

Karolyi, G. A., 2016. “The gravity of culture for finance.” Journal of Corporate Finance 41: 610–625.

Kets W, Sandroni A. A theory of strategic uncertainty and cultural diversity. The Review of Economic Studies, 2021, 88(1): 287-333.

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. J., Liu, Y J., Lu, R., 2020. “Social heterogeneity and local bias in peer-to-peer lending–evidence from China.” Journal of Comparative Economics 48: 302-324.

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.

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, X., Li, X., Zheng, Z. 2017. “Evaluating borrower’s default risk in peer-to-peer lending: evidence from a lending platform in China.” Applied Economics 49: 3538–3545.

Liu, Y., Jiao, Y., Xu, X. 2020. “Promoting or preventing labor migration? Revisiting the role of language.” China Economic Review 60: 101407.

Mauro, P. 1995. “Corruption and growth.” Quarterly Journal of Economics 110: 681–712.

Montalvo, J. G., Reynal-Querol, M. 2005. “Ethnic diversity and economic development.”

Journal of Development Economics 76: 293–323.

Norman, J. 1988. Chinese. Cambridge University Press. Cambridge.

Tan, G., Cheong, C., S, Zurbruegg R. 2019. “National culture and individual trading behavior.” Journal of Banking and Finance 106: 357–370.

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

Vives, X. 2019. “Digital Disruption in Banking.” Annual Review of Financial Economics 11: 243–272.

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

Wooldridge, J. 2001. Econometric Analysis of Cross Section and Panel Data. The MIT Press, Massachusetts.

Xia, Y., He, L., Li, Y., Liu, N., Ding, Y. 2020. “Predicting loan default in peer-to-peer lending using narrative data.” Journal of Forecasting 39: 260–280.

Xu, B., Ichiro, I. 1999. Large dictionary of Chinese Dialects. Zhonghua Book Company, Beijing. (In Chinese).

You, Z., Feng. Z M., Yang Y Z. 2018. “Relief Degree of Land Surface Dataset of China (1 km).” Journal of Global Change Data and Discovery 2: 151–155.

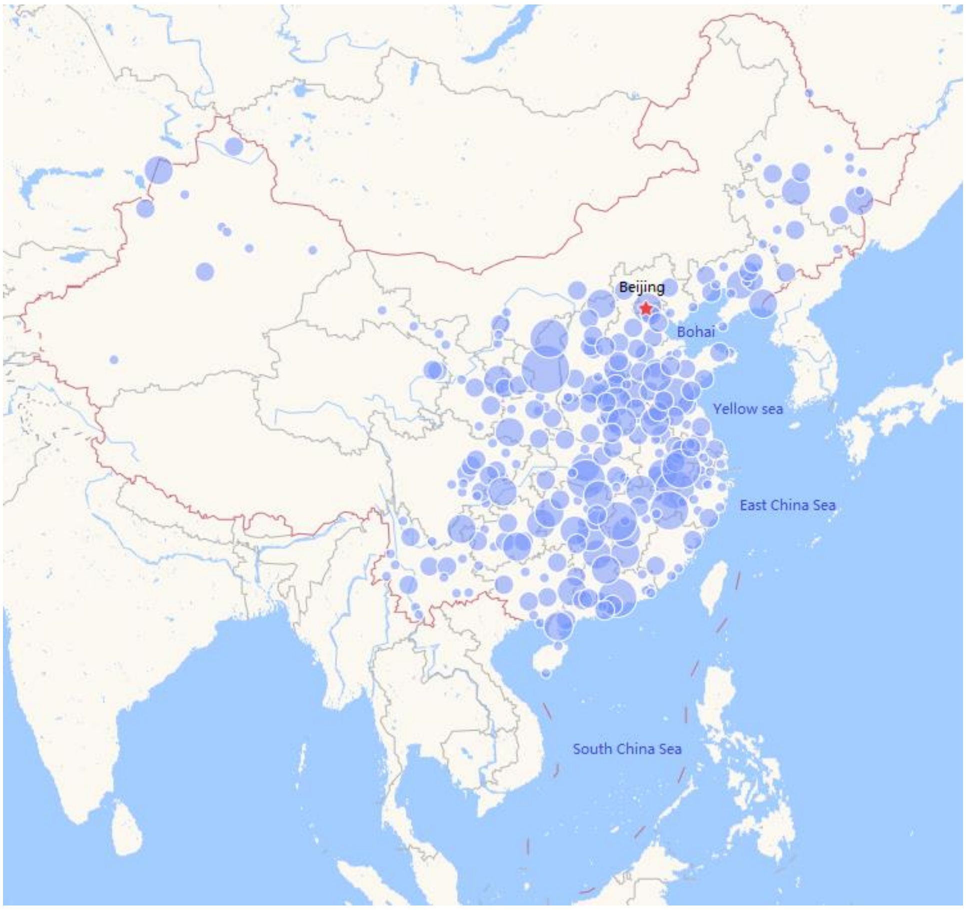
Zhang, W., Wang, C., Zhang, Y., Wang, J. 2020. "Credit risk evaluation model with textual features from loan descriptions for P2P lending." Electronic Commerce Research and Applications 42: 100989.

Acknowledgements

We wish to thank the participants of the internal seminar series at University of Hull and Prof. Mahbub Zaman for their valuable comments. The usual disclaimer applies.

Figures

Fig. 1 Number of dialects.



Data sources: *Chinese Dialect Dictionary*. This figure was produced using Amap API. Larger circles represent more dialects in the area. The source of data is from Liu et al. (2020). The details of the data for different regions are further disclosed in Table A2 in Appendix.

Tables

Table 1. Variables Definition.

This Table presents the variables used in the current study.

Type	Variable	Symbol	Definition
Dependent variables	Individual financial Behavior	L_amount	The logarithm of loan amount
		Bdebt	A dummy variable that takes the value of one if the loan has been defaulted and zero otherwise
Independent variables	Cultural diversity	*Diver*	Dialect diversity takes values in the interval from zero to one with one being the maximum
		Dn	Number of dialects
Control variables	Personal information variables	Gender	A dummy variable that takes the value of one for female borrowers and zero for male borrowers
		Worktime	Work experience, ranges from zero to five, with five being the maximum
		Mar	Marital status. A dummy variable that takes the value of one if the borrower is married and zero otherwise
		Edu	Education level
		Age	Age
		House	A dummy variable that takes the value of one if the borrower owns a house and zero otherwise
		Car	A dummy variable that takes the value of one if the borrower owns a car and zero otherwise
		L_income	The logarithm of income
		Lppay	The logarithm of amount that the borrower needs to repay in a single term
	Macroeconomic variables	L_GDP	The logarithm of GDP
		Pcaptial	A dummy variable that takes the value of one if the borrower lives in a provincial capital and zero otherwise
		L_gdppc	The logarithm of GDP per capita
		L-asal	The logarithm of average income

Notes: As there is no precise data on the population, there is a lot of missing data on GDP per capita, especially for earlier periods. People usually move to other places to work in China.

Table 2. Summary Statistics.

This Table reports the summary statistics of loan status, cultural diversity, personal information and a city's macroeconomic situation. We collect 210,841 historical borrowing data for the period 2013-2018 to examine the impact of cultural diversity on lending behavior. Since the indicator "default" does not include new borrowing data after 2017 and borrowing failure data for 2017, the number of observations of the variable *Bdebt* is significantly less than the number of observations of other variables.

Variable Symbol	Variable Name	Average	St.d	Min	Max	No. of observations
L_amount	Ln(lamount)	10.8197	0.9561	6.9078	13.8155	210841
Bdebt	Default or not	0.0084	0.0913	0.0000	1.0000	58960
Diver	Dialdiversity	0.2408	0.2030	0.0008	0.7822	202250
Dn	Dialnumber	1.7657	0.7464	1.0000	5.0000	202250
Gender	Gender	0.2874	0.4526	0.0000	1.0000	210841
Worktime	Work exp	3.0456	1.5170	0.0000	5.0000	210712
Mar	Marriage	0.5533	0.4972	0.0000	1.0000	210841
Edu	Education	1.2649	0.7583	0.0000	3.0000	210841
Age	Age	35.4628	7.8527	20.0000	77.0000	210841
House	House	0.4740	0.4993	0.0000	1.0000	210841
Car	Car	0.3194	0.4662	0.0000	1.0000	210841
L_income	Ln(income)	9.2500	0.8243	7.3132	10.8198	210841
Lppay	Ln(repayment)	7.7644	0.7806	3.8616	12.7564	210841
L_GDP	Ln (GDP)	6.4584	1.0217	2.7305	8.2468	207840
L_gdppc	Ln (GDPpc)	11.2812	0.5139	9.0066	12.2807	206114
L_asal	Ln(asal)	11.1385	0.3107	10.1039	11.9173	188986
Pcapital	Provcapital	0.4936	0.5000	0.0000	1.0000	210841

Notes: Ln(lamount) stands for Ln (loan amount), Dialdiversity stands for Dialect diversity Dialnumber stands for Dialect number, Work exp stands for Work experience, Ln (GDPpc) stands for Ln (GDP per capita), Ln(asal) stands for Ln (average income), Provcapital stands for Provincial capital.

Table 3. Loan Amount.

This Table reports the impact of cultural diversity on the loan amount. The regression is based on model (1). Columns (1)-(3) present the result without control variables (1), with personal attributes (2) with city characteristics, province and time fixed effects (3).

	(1) L_amount	(2) L_amount	(3) L_amount
Diver	-0.1892***	-0.0513***	-0.0457***
	(0.0108)	(0.0096)	(0.0125)
Gender		0.2376***	0.1511***
		(0.0039)	(0.0038)
Worktime		0.0290***	0.0063***
		(0.0012)	(0.0013)
Mar		0.3252***	0.3115***
		(0.0040)	(0.0039)
Edu		0.1483***	0.0664***
		(0.0028)	(0.0029)
Age		0.0162***	0.0177***
		(0.0002)	(0.0002)
House		-0.0252***	-0.0245***
		(0.0040)	(0.0042)
Car		-0.1251***	-0.1300***
		(0.0045)	(0.0046)
L_income		0.3509***	0.2781***
		(0.0023)	(0.0025)
L_GDP			0.0837***
			(0.0032)
Constant	10.8812***	6.5480***	6.4094***
	(0.0032)	(0.0200)	(0.0321)
Year fixed effect	No	No	Yes
Province fixed effect	No	No	Yes
No. of observations	202250	202128	200077
Adjusted. R^2	0.0020	0.2220	0.2730

Notes: The definitions of the variables are reported in Table 1. Standard errors of estimates are reported in parenthesis. Standard errors are clustered robust. ***, **, and * indicate the 1%, 5%, and 10% significance levels, respectively.

Table 4. Default Rate.

This Table reports the impact of cultural diversity on borrower default rate. The regression is based on model (2). Columns (1)-(3) present the result without control variables (1), with personal attributes (2) with city characteristics, province and time fixed effects (3).

	(1) Bdebt	(2) Bdebt	(3) Bdebt
Diver	0.8071***	0.7420***	0.8867***
	(0.0876)	(0.0970)	(0.1210)
Lppay		-0.6606***	-0.5847***
		(0.0347)	(0.0383)
Gender		-0.3112***	-0.2355***
		(0.0527)	(0.0596)
Worktime		0.1337***	0.2093***
		(0.0127)	(0.0150)
Mar		-0.1691***	-0.2015***
		(0.0421)	(0.0477)
Edu		-0.2672***	-0.2429***
		(0.0279)	(0.0309)
Age		-0.0185***	-0.0204***
		(0.0027)	(0.0032)
House		0.1950***	0.1293***
		(0.0413)	(0.0495)
Car		0.4248***	0.3066***
		(0.0460)	(0.0522)
L_income		0.0783***	0.1204***
		(0.0257)	(0.0301)
Pcaptial			-0.1531**
			(0.0740)
L_GDP			-0.2227***
			(0.0403)
Constant	-2.6183***	2.1413***	2.7634***
	(0.0300)	(0.2986)	(0.4238)
Year fixed effect	No	No	Yes
Province fixed effect	No	No	Yes
No. of observations	57583	57583	57332
Pseudo R^2	0.0160	0.1695	0.3034

Notes: The definitions of the variables are reported in Table 1. Standard errors of estimates are reported in parenthesis. Standard errors are clustered robust. ***, **, and * indicate the 1%, 5%, and 10% significance levels, respectively.

Table 5. Endogeneity Problems (1).

This Table reports the impact of cultural diversity on borrowers' behavior while adding the river length (Rl)-Column (1) and (3) - or average land slope (Slope) as instruments to address endogeneity.

The second stage	(1) L_amount	(2) L_amount	(3) Bdebt	(4) Bdebt
Diver	-0.4020***	-0.2606**	1.2536**	2.9041***
	(0.0670)	（0.0861）	(0.5391)	(0.9637)
Control variables	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes
Province fixed effect	Yes	Yes	Yes	Yes
The first stage	*Diver*	*Diver*	*Diver*	*Diver*
Rl	0.0580***		0.0652***	
	(0.0015)		(0.0031)	
Slope		0.1348**		0.0975***
		（0.0031）		(0.0062)
The first stage F-value	1579.16	1869.12		
AR statistics	36.47***	9.17***	4.91**	13.44***
Wald statistics	36.06***	9.15***	4.91**	13.29***
No. of observations	188395	199987	55016	57328

Notes: The data for the length of rivers in each city (Rl) variable is from China National Basic Geographic Information Center and calculated by the 1:40000 vector map with ArcMap, and it is in natural logarithms. The data for the land slope of each city (Slope) variable is from Relief Degree of Land Surface Dataset of China (1km). The definitions of the other variables are reported in Table 1. The dependent variable is the natural logarithm of the loan amount in columns (1) and (2) and the dummy variable *Bdebt* in columns (3) and (4). The control variables included in the estimation are gender, work experience, marriage, education, age, real estate, car, log of income and log of GDP in column (1) and column (2), and monthly payment, gender, work experience, marriage, education, age, real estate, car, log of income, a dummy variable if the borrower lives in a provincial capital, and log of GDP in column (3) and column (4). Fixed effects are included in the estimation. Standard errors of estimates are reported in parenthesis. Standard errors are clustered robust. ***, **, and * indicate the 1%, 5%, and 10% significance levels, respectively.

Table 6. Endogeneity Problems (2).

Heckman selection model controlling for the probability of funding success

Dependent variable	*Bdebt*
Diver	0.8864***
	(0.1309)
Control variables	Yes
Year fixed effect	Yes
Province fixed effect	Yes
No. of observations	108,552
No. of censored observations	51,197

Notes: The selection indicator in the first stage is (funding success) one if the application was successful and zero otherwise. The number of *failed loan applications* is the difference between the number of observations (108,552) minus the number of censored observations (51,197), namely 57,355 unfunded loans. The control variables included in the estimation of both steps are monthly payment, gender, work experience, marriage, education, age, real estate, car, log of income, a dummy variable if the borrower lives in a provincial capital, and the natural logarithm GDP per capita. Fixed effects are also included in the estimation. The definitions of the variables are reported in Table 1. Standard errors of estimates are reported in parenthesis. Standard errors are clustered robust. ***, **, and * indicate the 1%, 5%, and 10% significance levels, respectively.

Table 7. Endogeneity Problems (3).

IV-Heckit model controlling for the probability of funding success.

	(1) Bdebt	(2) Bdebt
Diver	1.2882**	2.9951***
	(0.5481)	(1.0048)
Control variables	Yes	Yes
Year fixed effect	Yes	Yes
Province fixed effect	Yes	Yes
	Diver	*Diver*
Rl	0.0642***	
	(0.0030)	
Slope		0.0948***
		(0.0062)
AR statistics	5.01**	13.28***
Wald statistics	5.01**	13.12***
No. of observations	55016	57328

Notes: Rl and Slope are the instrumental variables. The selection indicator in the first stage is (funding success) one if the application was successful and zero otherwise. The control variables included in the estimation are monthly payment, gender, work experience, marriage, education, age, real estate, car, log of income, a dummy variable if the borrower lives in a provincial capital, and log of GDP in both stages. The instrumental variable is river length (Rl) in column (1) and the land slope (Slope) in column (2). The definitions of the other variables are reported in Table 1. The dependent variable is the dummy variable *Bdebt*. Fixed effects are included in the estimation. Standard errors of estimates are reported in parenthesis. Standard errors are clustered robust. ***, **, and * indicate the 1%, 5%, and 10% significance levels, respectively.

Table 8. Loan Types.

This Table reports the impact of cultural diversity on borrower behavior when we control for loan types. The first and third columns refer to observations of pure online transaction, and the second column refers to observations of offline transaction.

	(1) Pure online L_amount	(2) With offline L_amount	(3) Pure online Bdebt
Diver	-0.0023	-0.0365***	0.4556***
	(0.0215)	(0.0128)	(0.1570)
Control variables	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes
Province fixed effect	Yes	Yes	Yes
No. of observations	67676	132399	3414
Adjust. R^2	0.1604	0.2256	
Pseudo R^2			0.1239

Notes: In columns (1) and (2) the dependent variable is the natural logarithm of loan amount and in column (3) the dependent variable is the dummy variable Bdebt. The control variables are gender, work experience, marriage, education, age, real estate, car, log of income and log of GDP in columns (1) and (3), and monthly payment, gender, work experience, marriage, education, age, real estate, car, the natural logarithm of income, a dummy variable if the borrower lives in a provincial capital and the natural logarithm of GDP in column (3). Fixed effects are also included in the estimation. The variables definitions are reported in Table 1. Standard errors of estimates are reported in parenthesis. Standard errors are clustered robust. ***, **, and * indicate the 1%, 5%, and 10% significance levels, respectively.

Table 9. Financial Development.

This Table reports the impact of cultural diversity on borrower behavior when we control for financial development, measured by local bank's loan balance. The full sample is divided in two groups. Samples are classified into the "high loan balance" group if the loan balance is higher than average, the other samples are classified into the "low loan balance" group.

	(1)	(2)	(3)	(4)
Loan balance	High	Low	High	Low
	L_amount	L_amount	Bdebt	Bdebt
Diver	-0.0826***	-0.1589***	1.6778***	0.7822***
	(0.0147)	(0.0264)	(0.2113)	(0.1770)
Control variables	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes
Province fixed effect	Yes	Yes	Yes	Yes
No. of observations	164932	46689	35145	9167
Suest test	Chi2(1) = 17.14		Chi2(1) = 9.38	
	Prob>chi2 = 0.0000***		Prob>chi2 = 0.0022***	
Adjust. R^2	0.2830	0.2180		
Pseudo R^2			0.2883	0.3101

Notes: In columns (1) and (2) the dependent variable is the natural logarithm of loan amount and in columns (3) and (4) the dependent variable is the dummy variable Bdebt. The control variables are gender, work experience, marriage, education, age, real estate, car, log of income and log of GDP in columns (1) and (2), and monthly payment, gender, work experience, marriage, education, age, real estate, car, the natural logarithm of income, a dummy variable if the borrower lives in a provincial capital and the natural logarithm of GDP in columns (3) and (4). Fixed effects are also included in the estimation. The variables definitions are reported in Table 1. Standard errors of estimates are reported in parenthesis. Standard errors are clustered robust. ***, **, and * indicate the 1%, 5%, and 10% significance levels, respectively. Suest test stands for Seemingly unrelated estimation test.

Table 10. Lender's decision.

This Table reports the impact of cultural diversity on borrower behaviors while considering success or not. We divide the full sample into a success and a failure sample.

	(1) Full L_amount	(2) Success L_amount	(3) Failure L_amount
Diver	-0.0457***	-0.1027***	0.0336
	(0.0125)	(0.0122)	(0.0275)
Control variables	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes
Province fixed effect	Yes	Yes	Yes
No. of observations	200077	152420	47657
Adjust. R^2	0.2730	0.2670	0.1990

Notes: The dependent variable is the natural logarithm of loan amount. The control variables included in the estimation are gender, work experience, marriage, education, age, real estate, car, the natural logarithm of income and the natural logarithm of GDP. Fixed effects are also included in the estimation. The definitions of the variables are reported in Table 1. Standard errors of estimates are reported in parenthesis. Standard errors are clustered robust. ***, **, and * indicate the 1%, 5%, and 10% significance levels, respectively.

Table 11. Robustness Checks (1).

This Table reports the impact of cultural diversity on borrower behaviors while the proxy variable of cultural diversity is the number of dialects instead of dialect diversity.

	(1) L_amount	(4) Bdebt
Dn	-0.0200***	0.1988***
	(0.0039)	(0.0340)
Control variables	Yes	Yes
Year fixed effect	Yes	Yes
Province fixed effect	Yes	Yes
No. of observations	200077	57332
Adjust. R^2	0.2730	
Pseudo R^2		0.2995

Notes: The dependent variable is the natural logarithm of loan amount in the first column and the dummy variable *Bdebt* in the second column. The control variables included in the estimation are gender, work experience, marriage, education, age, real estate, car, log of income and log of GDP in column (1) and monthly payment, gender, work experience, marriage, education, age, real estate, car, log of income, a dummy variable if the borrower lives in a provincial capital and the natural logarithm GDP in column (2). Fixed effects are also included in the estimation. The definitions of the variables are reported in Table 1. Standard errors of estimates are reported in parenthesis. Standard errors are clustered robust. ***, **, and * indicate the 1%, 5%, and 10% significance levels, respectively.

Table 12. Robustness Checks (2).

This Table reports the impact of cultural diversity on borrowers' behavior without the samples of borrowers who do not live in the province where they were born.

	(1) L_amount	(2) Bdebt
Diver	-0.0698***	0.8015***
	(0.0146)	(0.1394)
Control variables	Yes	Yes
Year fixed effect	Yes	Yes
Province fixed effect	Yes	Yes
No. of observations	144333	42082
Adjusted. R^2	0.2690	
Pseudo R^2		0.3212

Notes: The dependent variable is the natural logarithm of loan amount in the first column and the dummy variable *Bdebt* in the second column. The control variables included in the estimation are gender, work experience, marriage, education, age, real estate, car, log of income and log of GDP in column (1) and monthly payment, gender, work experience, marriage, education, age, real estate, car, log of income, a dummy variable if the borrower lives in a provincial capital and the natural logarithm GDP in column (2). Fixed effects are also included in the estimation. Standard errors of estimates are reported in parenthesis. Standard errors are clustered robust. ***, **, and * indicate the 1%, 5%, and 10% significance levels, respectively.

Table 13. Robustness Checks (3).

This Table reports the impact of cultural diversity on borrowers' behavior while we proxy macroeconomic environment with GDP per capita and average salary instead of GDP.

	(1)	(2)	(3)	(4)	(5)	(6)
	GDP per capita		Average Salary		One-Year Lag of GDP	
	L_amount	Bdebt	L_amount	Bdebt	L_amount	Bdebt
Diver	-0.0336***	0.8528***	-0.0580***	0.9303***	-0.0443***	0.8843***
	(0.0126)	(0.1218)	(0.0132)	(0.1218)	(-3.53)	(7.31)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Province fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	198720	57332	181686	57330	200077	57332
Adjust. R^2	0.2720		0.2660		0.273	
Pseudo R^2		0.3028		0.3003		0.3035

Notes: The dependent variable is the natural logarithm of loan amount in the first column and the dummy variable *Bdebt* in the second column. The control variables are gender, work experience, marriage, education, age, real estate, car, log of income and log of GDP per capita in column (1) and monthly payment, gender, work experience, marriage, education, age, real estate, car, log of income, a dummy variable if the borrower lives in a provincial capital and the natural logarithm GDP per capita in column (2). The natural logarithm GDP per capita is replaced by the average salary in columns (3) and (4), and by the lag of natural logarithm of GDP in columns (5) and (6). Fixed effects are also included in the estimation. The definitions of the variables are reported in Table 1. Standard errors of estimates are reported in parenthesis. Standard errors are clustered robust. ***, **, and * indicate the 1%, 5%, and 10% significance levels, respectively.

Appendix.

Table A1. Classification Criteria for P2P lending companies in the high-risk category.

Labels	Classification criterion
Closed	1. The P2P lending company has paused bid for more than three months and the announcements have stopped. 2. The P2P lending company issued an announcement for closure.
In transition	The P2P lending company has stopped P2P service and it is turning to other business.
Pause bid	The P2P lending company has paused bid for more than three months, but the announcements are still normal.
Website closed	The website of the P2P lending company cannot be browsed, but there is no negative news regarding this company.
Escape with money	1. The website cannot be browsed. 2. Customer service cannot be reached. 3. No employees at the registered address. 4. Losing all contacts.
Cash withdraw with difficulty	Unable to withdraw or restrict withdrawal and the P2P lending company has no announcement about the future redemption.
Investigation intervention	Public Relations Agency Filed for Intervention.
Deferred payment	1. The P2P lending company issued an announcement for winding up, but the loan cannot be repaid normally and needs to be deferred (e.g., 12/24 terms). 2. The P2P lending company has paused bid and issued an announcement for deferred payment.
Dispute	The P2P lending company is still issuing bids, but the investors report that a large amount of loans are overdue or cannot be paid normally.

Notes: P2P lending companies are classified as high risk when there is a high chance of going bankrupt. This table presents the classification criteria reported by WangDaiZhiJia.

Table A2. Number of dialects.

Province	City	Dn	Province	City	Dn	Province	City	Dn
Anhui	Anqing	2	Henan	Xuchang	2	Shandong	Binzhou	1
Anhui	Bengbu	2	Henan	Zhengzhou	2	Shandong	Dezhou	2
Anhui	Bozhou	1	Henan	Zhoukou	3	Shandong	Dongying	1
Anhui	Chizhou	3	Henan	Zhumadian	2	Shandong	Heze	3
Anhui	Chuzhou	2	Heilongjiang	Daqing	2	Shandong	Jinan	2
Anhui	Fuyang	2	Heilongjiang	Harbin	3	Shandong	Jining	2
Anhui	Hefei	1	Heilongjiang	Hegang	1	Shandong	Laiwu	1
Anhui	Huaibei	2	Heilongjiang	Heihe	1	Shandong	Liaocheng	2
Anhui	Huainan	1	Heilongjiang	Jixi	3	Shandong	Linyi	4
Anhui	Huangshan	4	Heilongjiang	Jiamusi	1	Shandong	Qingdao	2
Anhui	Maanshan	1	Heilongjiang	Mudanjiang	2	Shandong	Rizhao	2
Anhui	Tongling	1	Heilongjiang	Qitaihe	1	Shandong	Tai'an	3
Anhui	Wuhu	2	Heilongjiang	Qiqihar	1	Shandong	Weihai	1
Anhui	Suzhou	3	Heilongjiang	Shuangyashan	1	Shandong	Weifang	2
Anhui	Xuancheng	4	Heilongjiang	Suihua	2	Shandong	Yantai	2
Beijing	Beijing	3	Heilongjiang	Yichun	1	Shandong	Zaozhuang	2
Fujian	Fuzhou	1	Hubei	Ezhou	1	Shandong	Zibo	2
Fujian	Longyan	2	Hubei	Enshi	2	Shanxi	Datong	3
Fujian	Ningde	2	Hubei	Huanggang	1	Shanxi	Jincheng	3
Fujian	Quanzhou	1	Hubei	Huangshi	1	Shanxi	Jinzhong	2
Fujian	Xiamen	1	Hubei	Jingmen	3	Shanxi	Linfen	2
Gansu	Dingxi	2	Hubei	Jingzhou	4	Shanxi	Shuozhou	2
Gansu	Jiayuguan	1	Hubei	Shiyan	2	Shanxi	Taiyuan	2
Gansu	Jinchang	1	Hubei	Suizhou	2	Shanxi	Xinzhou	2
Gansu	Lanzhou	1	Hubei	Wuhan	2	Shanxi	Yangquan	2
Gansu	Longnan	1	Hubei	Xianning	1	Shanxi	Yuncheng	1
Gansu	Pingliang	2	Hubei	Xiangyang	1	Shanxi	Changzhi	1
Gansu	Qingyang	2	Hubei	Yichang	1	Shaanxi	Ankang	2
Gansu	Tianshui	2	Hunan	Changde	1	Shaanxi	Baoji	1
Gansu	Wuwei	1	Hunan	Chenzhou	3	Shaanxi	Hanzhong	3
Gansu	Zhangye	1	Hunan	Hengyang	3	Shaanxi	Shangluo	2
Guangdong	Chaozhou	1	Hunan	Loudi	1	Shaanxi	Tongchuan	1
Guangdong	Dongguan	1	Hunan	Shaoyang	3	Shaanxi	Weinan	1
Guangdong	Foshan	1	Hunan	Xiangtan	2	Shaanxi	Xi'an	1
Guangdong	Guangzhou	1	Hunan	Xiangxi	3	Shaanxi	Xianyang	2
Guangdong	Huizhou	4	Hunan	Yiyang	1	Shaanxi	Yan'an	5
Guangdong	Jiangmen	1	Hunan	Yongzhou	2	Shaanxi	Yulin	4
Guangdong	Jieyang	1	Hunan	Yueyang	3	Shanghai	Shanghai	1
Guangdong	Maoming	2	Hunan	Zhangjiajie	2	Sichuan	Bazhong	1
Guangdong	Meizhou	1	Hunan	Changsha	2	Sichuan	Chengdu	2
Guangdong	Shantou	1	Jilin	Baicheng	1	Sichuan	Dazhou	1
Guangdong	Shaoguan	3	Jilin	Jilin	2	Sichuan	Deyang	2
Guangdong	Shenzhen	2	Jilin	Liaoyuan	1	Sichuan	Guang'an	1
Guangdong	Yunfu	2	Jilin	Siping	1	Sichuan	Guangyuan	1
Guangdong	Zhanjiang	3	Jilin	Songyuan	1	Sichuan	Leshan	1
Guangdong	Zhaoqing	2	Jilin	Tonghua	2	Sichuan	Luzhou	1
Guangdong	Zhongshan	1	Jilin	Yanbian	1	Sichuan	Meishan	1
Guangdong	Zhuhai	2	Jilin	Changchun	1	Sichuan	Mianyang	2
Guangxi	Baise	2	Jiangsu	Changzhou	1	Sichuan	Nanchong	2
Guangxi	Beihai	1	Jiangsu	Huai'an	1	Sichuan	Neijiang	2
Guangxi	Guigang	2	Jiangsu	Lianyungang	2	Sichuan	Panzhihua	1
Guangxi	Guilin	2	Jiangsu	Nanjing	2	Sichuan	Suining	2
Guangxi	Hechi	1	Jiangsu	Nantong	2	Sichuan	Ya'an	1
Guangxi	Hezhou	2	Jiangsu	Suzhou	1	Sichuan	Ziyang	1

Guangxi	Liuzhou	1	Jiangsu	Taizhou	2	Sichuan	Zigong	1
Guangxi	Nanning	2	Jiangsu	Wuxi	1	Tianjin	Tianjin	2
Guangxi	Qinzhou	1	Jiangsu	Suqian	2	Xinjiang	Altay	2
Guangxi	Wuzhou	3	Jiangsu	Xuzhou	2	Xinjiang	Bayingoleng	2
Guizhou	Anshun	1	Jiangsu	Yancheng	2	Xinjiang	Bortala	2
Guizhou	Bijie	1	Jiangsu	Yangzhou	1	Xinjiang	Changji	1
Guizhou	Guiyang	2	Jiangsu	Zhenjiang	2	Xinjiang	Hami	1
Guizhou	Liupanshui	2	Jiangxi	Fuzhou	1	Xinjiang	Hotan	1
Guizhou	Qiandongnan	2	Jiangxi	Ganzhou	3	Xinjiang	Karamay	1
Guizhou	Qiannan	3	Jiangxi	Ji'an	3	Xinjiang	Tacheng	3
Guizhou	Qianxinan	1	Jiangxi	Jingdezhen	1	Xinjiang	Turpan	1
Guizhou	Tongren	3	Jiangxi	Jiujiang	3	Xinjiang	Urumqi	1
Guizhou	Zunyi	2	Jiangxi	Nanchang	2	Yunnan	Baoshan	1
Hainan	Haikou	1	Jiangxi	Shangrao	4	Yunnan	Chuxiong	2
Hainan	Sanya	1	Jiangxi	Xinyu	1	Yunnan	Dehong	1
Hebei	Baoding	2	Jiangxi	Yichun	4	Yunnan	Diqing	1
Hebei	Cangzhou	2	Jiangxi	Yingtan	1	Yunnan	Honghe	1
Hebei	Chengde	2	Liaoning	Anshan	3	Yunnan	Kunming	2
Hebei	Handan	3	Liaoning	Benxi	2	Yunnan	Lijiang	2
Hebei	Hengshui	2	Liaoning	Chaoyang	2	Yunnan	Lincang	2
Hebei	Langfang	2	Liaoning	Dalian	1	Yunnan	Nujiang	1
Hebei	Qinhuangdao	1	Liaoning	Dandong	3	Yunnan	Qujing	1
Hebei	Shijiazhuang	2	Liaoning	Fushun	1	Yunnan	Puer	1
Hebei	Tangshan	1	Liaoning	Fuxin	1	Yunnan	Wenshan	1
Hebei	Xingtai	2	Liaoning	Huludao	2	Yunnan	Xishuangbanna	1
Hebei	Zhangjiakou	2	Liaoning	Jinzhou	1	Yunnan	Yuxi	1
Henan	Anyang	2	Liaoning	Liaoyang	1	Yunnan	Zhaotong	3
Henan	Hebi	2	Liaoning	Shenyang	2	Zhejiang	Hangzhou	2
Henan	Jiaozuo	2	Liaoning	Tieling	2	Zhejiang	Huzhou	1
Henan	Kaifeng	3	Liaoning	Yingkou	1	Zhejiang	Jiaxing	1
Henan	Luoyang	2	Inner Mongolia	Hohhot	2	Zhejiang	Jinhua	1
Henan	Luohe	2	Inner Mongolia	Wuhai	1	Zhejiang	Lishui	1
Henan	Nanyang	2	Ningxia	Guyuan	3	Zhejiang	Ningbo	1
Henan	Pingdingshan	1	Ningxia	Shizuishan	2	Zhejiang	Quzhou	1
Henan	Puyang	1	Ningxia	Wuzhong	1	Zhejiang	Shaoxing	1
Henan	Sanmenxia	2	Ningxia	Yinchuan	1	Zhejiang	Taizhou	1
Henan	Shangqiu	2	Qinghai	Haidong	2	Zhejiang	Wenzhou	2
Henan	Xinyang	2	Qinghai	Xining	2	Chongqing	Chongqing	3

Notes: The data is from Liu et al. (2020). In China, there are 23 provinces, four municipalities, five autonomous regions and two special administrative regions. Following the literature, we denote them as provinces for short. Two special administrative regions (Hong Kong and Macao) are not included due to their special status and restricted P2P loan markets for the residents there. We have also not considered Tibet and Taiwan since the data is missing for these two regions. So, there are 34 provinces and we have not included data for four of them. Dn is the number of dialects. There are 276 prefecture-level cities with available data in this study.

Table A3.

This Table offers the definitions of income and work experience.

Monthly income	Value assigned	Work experience	Value assigned
<1000	1000	No work experience	0
1001–2000	1500	<1 year	1
2000–5000	3500	1-3 years	2
5000–10000	7500	3-5 years	4
10000–20000	15000	>5 years	5
20000–50000	35000		
>50000	50000		

Notes: We assigned the median value to each group as proxy variables of income and work experience. Monthly income is measured in Chinese yuan (CNY).

Notes.

[1]Another strand of literature studies the impact of cultural differences between lenders and borrowers on loan spreads and interest rates (Chui et al., 2016 and the references therein).

[2] See Figure 1 and Table A2 describing the number of dialects in each city.

[3]They are the first who calculated and used the concept of dialect diversity.

[4] More details are given in Table A3 in the Appendix.

[5] By "borrower's location", we mean their permanent residence.

[6] 210,841 loans were issued during the period 2013–2018. Someone may have applied several times during the observed period.

[7] Following the literature (Chen et al., 2020; Liu et al., 2020), two special administrative regions (i.e. Hong Kong and Macao) are not included due to their special status and restricted P2P loan markets for the residents there. We have also not considered Tibet and Taiwan since the data is missing for these two regions. There are 23 provinces, 4 municipalities, 5 autonomous regions and 2 special administrative regions in China. Following the literature (Chen et al., 2020; Liu et al., 2020), we denote them as provinces for short.

[8] According to Renrendai, investors can choose to bid and lend money to borrowers; the bid provides the information about loans and the borrowers.

[9] This ID number only includes the first and the last three digits for privacy reasons. The first two digits determine the province where a borrower is from. We cannot determine the borrower's city as that requires four digits.

[10] We have also replaced GDP with the one-year lag of GDP. The empirical results remain unchanged. We wish to thank an anonymous reviewer for mentioning this robustness test.

[11] With Mandarin being the official language used in the classroom.