

Chen, Xiao; Huang, Bihong; Ye, Dezhu

Working Paper

The role of punctuation in P2P lending: Evidence from the People's Republic of China

ADB Working Paper, No. 787

Provided in Cooperation with:

Asian Development Bank Institute (ADBI), Tokyo

Suggested Citation: Chen, Xiao; Huang, Bihong; Ye, Dezhu (2017) : The role of punctuation in P2P lending: Evidence from the People's Republic of China, ADB Working Paper, No. 787, Asian Development Bank Institute (ADBI), Tokyo

This Version is available at:

<https://hdl.handle.net/10419/190208>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-nd/3.0/igo/>



ADB Working Paper Series

THE ROLE OF PUNCTUATION IN P2P LENDING: EVIDENCE FROM THE PEOPLE'S REPUBLIC OF CHINA

Xiao Chen, Bihong Huang,
and Dezhu Ye

No. 787
October 2017

Asian Development Bank Institute

Xiao Chen is a PhD student at the Department of Finance, Jinan University, in Guangzhou, People's Republic of China. Bihong Huang is research fellow at the Asian Development Bank Institute in Tokyo, Japan. Dezhu Ye is professor at the Department of Finance, Research Institute of Finance, Jinan University, in Guangzhou, People's Republic of China.

The views expressed in this paper are the views of the author and do not necessarily reflect the views or policies of ADBI, ADB, its Board of Directors, or the governments they represent. ADBI does not guarantee the accuracy of the data included in this paper and accepts no responsibility for any consequences of their use. Terminology used may not necessarily be consistent with ADB official terms.

Working papers are subject to formal revision and correction before they are finalized and considered published.

The Working Paper series is a continuation of the formerly named Discussion Paper series; the numbering of the papers continued without interruption or change. ADBI's working papers reflect initial ideas on a topic and are posted online for discussion. Some working papers may develop into other forms of publication.

The Asian Development Bank recognizes "China" as the People's Republic of China.

In this report, "\$" refers to US dollars.

Suggested citation:

Chen, X., B. Huang, and D. Ye. 2017. The Role of Punctuation in P2P Lending: Evidence from the People's Republic of China. ADBI Working Paper 787. Tokyo: Asian Development Bank Institute. Available: <https://www.adb.org/publications/role-punctuation-p2p-lending-evidence-prc>

Please contact the authors for information about this paper.

Email: chenxiao427@163.com, bihuang@adbi.org, gzydz@126.com

Asian Development Bank Institute
Kasumigaseki Building, 8th Floor
3-2-5 Kasumigaseki, Chiyoda-ku
Tokyo 100-6008, Japan

Tel: +81-3-3593-5500
Fax: +81-3-3593-5571
URL: www.adbi.org
E-mail: info@adbi.org

© 2017 Asian Development Bank Institute

Abstract

This paper investigates the role of punctuation in the peer-to-peer (P2P) lending market. Using data from Renrendai, one of the largest P2P lending platforms in the People's Republic of China, we investigate how the amount of punctuation used in loan descriptions influences the funding probability, borrowing rate, and default. The empirical evidence shows that the amount of punctuations is negatively associated with the funding probability and borrowing rate. We propose that the usage of punctuation affects the readability of a loan description and reflects borrowers' self-control and cognitive ability. Within a given number of words, excessive usage of punctuation makes loan description informal and reduces the readability of the text, thereby impairing investors' trust in borrowers. Moreover, borrowers that overuse punctuation may have lower ability of self-control, and tend to underestimate the risk of borrowing and offer lower borrowing rate due to overconfidence.

Keywords: P2P lending, information asymmetry, word, punctuation

JEL Classification: G10, G11, G14, G20, G23, G29

Contents

1.	INTRODUCTION	1
2.	HYPOTHESIS DEVELOPMENT	3
3.	DATA AND EMPIRICAL METHODOLOGY.....	4
3.1	Data Source.....	4
3.2	Key Variables and Summary Statistics.....	5
3.3	Empirical Methodology	8
4.	EMPIRICAL RESULT	9
4.1	Punctuation and Funding Probability.....	9
4.2	Punctuation and Borrowing Rate.....	12
4.3	Punctuation and Loan Default.....	12
4.4	Robustness Checks	14
5.	CONCLUSION.....	17
	REFERENCES	18

1. INTRODUCTION

With the advance of digital technology, online peer-to-peer (P2P) lending has emerged as an alternative to traditional lending institutions around the world (Greiner and Wang 2010). Bypassing banks, online P2P lending is a special type of credit market in which individual lenders make microloans to individual borrowers without collateral or intermediation from financial institutions (Lin, Prabhala, and Viswanathan 2013). Compared with the traditional credit market, online P2P lending is easy to access (Mild, Waitz, and Wock 2015), provides a new investment channel, and improves the utilization efficiency of social funds (Duarte, Siegel, and Young 2012). However, information asymmetry remains a critical issue in this emerging market, and is likely to be more exaggerated than in the traditional credit market (Luo and Lin 2013) where a financial intermediary is the main organization judging borrowers' loan requests and deciding whether to lend the money. In contrast, in the P2P lending market, there are no financial intermediaries to investigate the creditworthiness of the borrowers. Moreover, both lenders and borrowers are anonymous and don't have opportunities to meet each other. The lenders make their decisions mainly based on the information provided by the borrowers (Ashta, Assadi, and Johnson 2010).

Existing researches have investigated a wide range of mechanisms that might mitigate information asymmetry in the P2P lending market (Burtch, Ghose, and Wattal 2014). For example, using data from Prosper.com, the leading P2P lending platform in the United States (US), Duarte, Siegel, and Young (2012) show that borrowers appearing to be more trustworthy are more likely to have their borrowing requests granted. Gonzalez and Loureiro (2014) find that attractive applicants have a higher probability of obtaining loans and paying lower rates, but have similar default rates to average-looking borrowers. Pope and Sydnor (2011) present evidence of significant racial disparities. Loan listings by blacks are less likely to receive funding than those by whites with similar credit profiles while the interest rates paid by blacks are higher than those paid by comparable whites. Moreover, in the P2P lending platform, lenders tend to make use of social capital, particularly information regarding friends and groups in the P2P lending market, to judge the credibility of borrowers (Everett 2014; Lee and Lee 2012; Lin, Prabhala, and Viswanathan 2013; Michels 2012).

In addition to gender, race, appearance, and social capital, the text written by borrowers in an effort to describe the purpose of borrowing and convince potential investors of their trustworthiness is the most common form of information disclosure across all P2P lending platforms (Michels 2012). Despite its importance, research on the role of text in alleviating information asymmetry in the P2P lending process is still limited. Using data from Prosper.com, Gao and Lin (2015) show that investors believe that textual descriptions and text features can explain and predict loan default. Analyzing the description texts on Prosper for a short period of time, Iyer et al. (2016) reveal the predictive power of self-reported loan purpose and text characteristics on the default probability. Comparing the transaction and loan applications on the two German P2P platforms, Dorfleitner et al. (2016) find that spelling errors, text length, and keywords evoking positive emotion predict funding probability. Using data from the peer-to-peer lending website Prosper.com, Herzenstein, Sonenshein, and Dholakia (2011) find that as the number of identity claims in narratives increases, loan funding probability increases whereas loan performance deteriorates.

In this paper, we study the usage of punctuation in loan descriptions and its role in attenuating information asymmetry in this nascent but fast-growing online P2P lending market. Punctuation is used to connect the organization of the text (Ferreiro and Pontecorvo 1999; Baron 2001) and balance the rhythm and syntactic structure of sentences (Bruthiaux 1993). It shows a writer's personal characteristics and has a significant impact on a reader's understanding of written language (Liu, Wang, and Jin 2010).¹ Similarly, in P2P lending, the usage of punctuation reflects the borrower's personality and has a significant impact on lenders' judgement of borrowers' trustworthiness. Moreover, although language itself changes slowly, the Internet has speeded up the process of change (Steffes and Burgee 2009). Being an indispensable part of online P2P lending, loan descriptions are inevitably influenced by "Internet language."²

Using the data obtained from Renrendai, one of the largest P2P lending platforms in the PRC, we investigate the role of punctuation in bridging the information gap between borrowers and lenders. The PRC has developed the biggest and fastest-growing market for P2P lending. While as yet there are no verifiable data on the volume of P2P lending in the PRC, it is estimated to have been as high as \$40 billion by the end of 2015 (ACCA 2015). Compared with high-income countries, the social credit system is underdeveloped in the PRC. In high-income countries where the credit system is well established, many platforms, like Smava in Germany, only allow loan applications with a minimum credit score (Dorfleitner et al. 2016). On such platforms, investors rely heavily on hard factors such as credit scores or suggested interest rates to screen borrowers while the effect of soft factors on the funding success and default rate is minimal. In the PRC, most people never have credit scores. As of 2014, the People's Bank of China maintained credit histories for around 350 million citizens, less than one-third of the adult population, while in America 89% of adults have credit scores (The Economist 2016). Under such conditions, lenders in the P2P lending market have to rely heavily on soft information to judge the creditworthiness of borrowers and the issue of information asymmetry is likely to be more exaggerated. Exploring the various mechanisms through which the information asymmetry between borrowers and lenders could be moderated is of critical importance.

Our empirical results show that punctuation can predict the funding probability and borrowing rate in the P2P lending market. All else being equal, an additional ten punctuation marks is associated with a 9.6% reduction in the funding possibility and 0.56% reduction in the borrowing rate. However, we don't find any significant relations between the amount of punctuation and the default probability. Various tests confirm the robustness of our findings. Our explanations for these empirical findings are as follows. First, a large amount of punctuation accompanying a lengthy loan description text results in information overload, making it difficult for investors to judge the quality of borrowing requests. Second, a loan description with an excessive amount of punctuation usually bears the imprint of Internet language that is full of slang. It makes the loan description informal, reduces the readability of the text, and impairs the perception of trustworthiness of the applicants. Third, the usage of punctuation reflects the borrowers' self-control and cognitive ability. Borrowers who overuse punctuation may have lower ability of self-control and tend to be overconfident in their loan listings.

¹ For example, 弟弟找不到爸爸妈妈, 很着急. 弟弟找不到爸爸, 妈妈很着急.

In the Chinese context, the meaning of the first sentence is: the younger brother is very anxious because he cannot find his parents. Additionally, the second sentence means: the mother is very anxious because her son cannot find his father.

² For instance, I have a physical store only to increase the credit grade and I need support from you guys....

They are more likely to underestimate the risk of borrowing and offer a lower funding rate.

To the best of our knowledge, this is the first research to study punctuation and its impact on P2P lending. We contribute to the existing literature in several ways. First, our study deepens and enriches the understanding of loan description text in P2P lending. With the development of P2P lending, a growing number of researches have been carried out on this market. However, no attention has been given to the role of punctuation. Our paper fills this gap in the literature by showing that the amount of punctuation used in loan descriptions can predict the funding probability and borrowing rate. Second, this research fits into the existing literature on information asymmetry that is the main feature of financial markets. Identifying the mechanisms that mitigate information asymmetry is of critical importance. The empirical findings of this paper indicate that the amount of punctuation used in loan descriptions significantly influences lenders' evaluation of borrowers' trustworthiness even when hard facts like credit scores are not available. Third, our research is the first to suggest that the usage of punctuation can reflect important characteristics of borrowers including self-control, cognitive capability, and confidence in the online credit market. With given length of loan description, excessive number of punctuations may indicate borrowers' low ability of self-control and overconfidence.

The remainder of this paper is organized as follows. Section II reviews previous research and develops testable hypotheses; Section III describes our data set and empirical methodology; Section IV reports the main results; and Section V concludes the paper.

2. HYPOTHESIS DEVELOPMENT

In this section, we review the relevant literature to derive the testable hypothesis on the effect of punctuation on P2P lending. Existing researches have shown that information embedded in the loan description can help lenders to screen borrowers (Gao and Lin 2015). Punctuation, as a component of written language, expresses voice and rhythm (Baldwin 1978), activates words, conveys semantic information (Schriefers, Meyer, and Levelt 1990), and has an indispensable role in affecting text comprehension, word recognition, and reading time (Cohen, Douaire, and Elsabbagh 2001). Fuchs and Krivokapic (2016) provide evidence of significant differences in reading pause durations between words at commas and at full stops. Pynte and Kennedy (2007) show that punctuation marks carry word class information. Several researches have shown that the readability of financial reports significantly affects investors' willingness to invest (Rennekamp 2012; Loughran and McDonald 2014; Tan, Wang, and Zhou 2014). According to Reber and Schwarz (1999), people like information that is easy to process and perceptual fluency affects judgments of truth. In contrast, less readable filings of financial reports are associated with a reduction in small investors' trading activity (Miller 2010).

In most cases, the amount of punctuation used in the text increases with the amount of content (Roux 2008; Whissell 2013). In P2P lending, the amount of punctuation also grows with the amount of information that a borrower wishes to express. However, faced with a large volume of information, people tend to screen and read it selectively because attention is a scarce cognitive resource (Kahneman and Tversky 1973). Peng and Xiong (2006) have shown that limited investor attention leads to category-learning behavior. In the P2P lending market, a large amount of punctuation accompanying a lengthy loan description text may result in *information overload*, making it difficult for

investors to make decisions. Larrimore et al. (2011) and Dorfleitner et al. (2016) have demonstrated the negative impact of long loan descriptions.

As all P2P borrowing requests are posted online, the expression of loan descriptions is inevitably influenced by “Internet language” where informal writing and excessive use of punctuation are common. The imprint of Internet language and excessive use of punctuation make loan requests *informal*. (For instance, this loan is only for an increased credit limit, I need support from you guys...) Undoubtedly, using informal expression in a formal lending market will reduce the lender’s trust in a borrower enormously.

Hypothesis 1: All else being equal, the more punctuation is used in loan descriptions, the less likely it is that loan applications will be granted.

Within a given amount of words, a loan description containing too many punctuation marks could be interpreted as an indication of borrowers’ limited *self-control* and *cognitive capacity*. Self-control, which refers to a person’s control of his current behavior, affects his consumption and savings (Ye, Ng, and Lian 2015) and predicts his behavior in the future (Thaler and Shefrin 1977). In addition, self-control is a form of *cognitive bias* (Hirshleifer 2001). Being an important channel for lenders to understand borrowers, informal expression of loan descriptions affects lenders’ judgment of borrowers’ creditworthiness. Despite this reality, some borrowers continue to overuse punctuation because they lack self-control or are overconfident about their loan requests. Such borrowers will usually overreact to private information, overestimate their judgment, and underestimate the risks (Gervais, Heaton, and Odean 2011). They tend to believe that the interest rate they set is correct, whereas in fact it deviates from the real risk.

Hypothesis 2: All else being equal, the more punctuation borrowers use in their loan descriptions, the more likely it is that borrowers will set a lower interest rate.

Hypothesis 3: All else being equal, the more punctuation borrowers use in their loan descriptions, the more likely it is that borrowers will default.

3. DATA AND EMPIRICAL METHODOLOGY

3.1 Data Source

The data used in this study are obtained from Renrendai, one of the largest peer-to-peer lending platforms in the PRC. Founded in 2010, it now has over 1 million members located in more than 2,000 cities and counties across the country. Moreover, the reputation of Renrendai has been well recognized in the PRC. In 2014 and 2015, it was awarded the level of an AAA (the highest level) online lending platform by the Internet Society of China and China Academy of Social Science. It ranked no. 53 in a list of the PRC’s top 100 Internet companies in 2015 released by the Internet Society of China and the Ministry of Industry and Information.

The transactions taking place at Renrendai are typical examples of P2P lending. On Renrendai, borrowers can post loan requests with the required information of the loan title, amount of borrowing, interest rate, description of loan usage, and monthly installment. Renrendai provides verification services with national identification cards, credit reports, and borrowers’ addresses. It assigns a credit score to each borrower according to his or her borrowing/lending history and the amount of verified information. As in the case of Prosper.com, Renrendai’s profit mainly comes from borrowers’ closing fee and lenders’ servicing fee. To increase the probability of having their

loan requests granted, borrowers may provide personal information such as gender, education, income, marriage status, etc. Since the verification and credit rating provided by Renrendai are limited, it is of critical importance for the lenders to identify the trustworthiness of the borrowers from the observable information disclosed on the platform. In particular, when creating the loan listings, borrowers are encouraged to disclose additional information regarding the purpose of the loan and other personal information in a freeform text field called the “loan description.” Once a loan listing is posted online, lenders may place bids by stating the amount they want to fund. With a minimum bid amount of RMB 50, a listing typically requires dozens of bids to become fully funded. A listing that achieves 100% funding is a “successful” listing; otherwise, the borrower receives no funding.

The transaction module of Renrendai is comparable to that of Prosper, the largest lending platform in the US. The existing research (Greiner and Wang 2010; Herzenstein, Sonenshein, and Dholakia 2011; Pope and Sydnor 2011; Michels 2012; Yum, Lee, and Chae 2012; Zhang and Liu 2012; Lin, Prabhala, and Viswanathan 2013; Luo and Lin 2013) mainly uses data obtained from Prosper. On Prosper, borrowers post personal loan requests while investors (individual or institutional) can fund any amount ranging from \$2,000 to \$35,000 per loan request. In addition to credit scores, ratings, and histories, investors can use borrowers’ personal loan descriptions, endorsements from friends, and community affiliations to make investment decisions. Prosper handles the servicing of the loan and collects and distributes borrower payments and interest back to the loan investors. Prosper verifies borrowers’ identities and selects personal data before funding loans and manages all stages of loan servicing.

This study uses all loan listings created on Renrendai between 1 March 2012 and 1 December 2014. We eliminate the data earlier and later than this period to avoid the initial launch period and truncation of loan repayments, respectively. To prevent estimation biases, we drop loan requests with incomplete information, those from borrowers younger than 22 years old, and those with institutional guarantee and field certification. As a result, our sample includes 170,817 loan listings, of which 11,136 were successfully funded while the remaining 159,681 were not funded. Among all listings that were successfully funded, there were 809 defaults. It is worth mentioning that except for the 809 on the default list, all other loans were repaid on time.

3.2 Key Variables and Summary Statistics

To analyze the role of punctuation in P2P lending, we collect three categories of data. The first category is the information related to loan listings, including the term, interest rate, and borrowing amount, etc. The amount of punctuations used in the loan description is our key variable of interest and the number of words is included to control the length of the loan description. The second category is the credit of borrowers, including their credit score, mortgage, and car loans. The third is borrowers’ personal characteristics such as age, education, income, working hours, and marriage status, etc. Moreover, we also control the year effect and change in the design of the platform. The definition of each variable is shown in Table 1.

Table 1: Variables and Definitions

Variable	Name	Definition
Probability of Funding	Success	1 if a loan listing is fully funded and 0 otherwise
Probability of Default	Default	1 if the funded loan has been defaulted and 0 otherwise
Number of words	Word	The number of Chinese characters used by the borrower in a loan description
Number of Punctuation	Punctuation	The number of punctuation marks used in a loan description
Interest Rate (in %)	Rate	The rate that the borrower pays on the loan
Loan Amount (in RMB)	Amount	Loan amount requested by the borrower
Loan Term (in months)	Time	Loan term requested by the borrower
Credit Grade in platform	Credit	Credit grade of the borrower at the time the listing was created. Credit grade takes on values between 1 (high risk) and 7 (AA)
Home Loans	House_I	1 if borrower has a mortgage and 0 otherwise
Car Loans	Car_I	1 if borrower has a car loan and 0 otherwise
Gender of borrower	Gender	Dummy variable taking value of 1 if borrower is a man and 0 otherwise
Age	Age	Age of the borrower in years
Marital Status	Marry	1 if borrower is married and 0 otherwise
Income Level (in RMB)	Income	Income level of the borrower, 1=Less than 1,000, 2=1,001–2,000, 3=2,001–5,000, 4=5,001–10,000, 5=10,001–20,000, 6=20,001–50,000, 7=More than 50,000, per month
Education level	Education	Education level of borrower, 1=middle/high school, 2=3-year college, 3=4-year college, 4=graduate school
Homeowner indicator	House	1 if borrower is a homeowner and 0 otherwise
Car owner indicator	Car	1 if borrower is a car owner and 0 otherwise
Years of work experience	Worktime	Borrower's working experience, 1=Less than 1 year, 2=1–3 years, 3=3–5 years, 4=more than 5 years
Redesign	Redesign	Renrendai platform made some changes in October 2013; the indicator equals 1 if the loan listings were created after this time
Year	Year	Year dummies, 2012–2014

Table 2 provides summary statistics of all variables used in this study. As indicated by Table 2, 6.5% of loan listings are funded. A loan description on average contains 53 words and nine punctuation marks. The kurtosis and skewness of the amount of punctuation are 18.034 and 3.560, respectively, implying a “sharp peak and heavy tail” distribution. The kurtosis and skewness of the number of words are 10.416 and 2.391, respectively, indicating a “platykurtic and dissymmetric” distribution. We winsorize the loan lists whose amount of punctuation marks and words are at the top 1% level to eliminate the extreme values. After winsorization, the minimum amount of punctuation does not change, but its maximum value decreases by 576% from 480 to 71 while the maximum number of words decreases by 113% from 234 to 504. At the same time, the kurtosis and skewness of punctuation and words decline substantially, making their distributions become normal. The interest rate offered by borrowers is 15% on average with a maximum of 24.4%. The average amount of loan is approximately RMB 68,600, indicating that small loans are a major form of credit offered by P2P lending platforms. The credit grades of borrowers are universally low, with 96% rated as high risk (HR). Among all borrowers, 13.7% have a mortgage and 5.4% have car loans.

Table 2: Summary Statistics

Variable	Mean	Sd	Min	Max	N
Success	0.065	0.247	0	1	170,817
Default	0.072	0.259	0	1	11,136
Word	52.495	37.484	18	236	170,817
Punctuation	8.597	10.819	0	71	170,817
Rate	14.98	3.489	9.5	24.4	170,817
Amount	68,599	109,700	3,000	1,000,000	170,817
Time	15.59	10.11	1	36	170,817
Credit	1.063	0.375	1	7	170,817
Credit=HR	0.962	0.192	0	1	164,264
Credit=E	0.023	0.153	0	1	4,091
Credit=D	0.008	0.089	0	1	1,391
Credit=C	0.003	0.058	0	1	589
Credit=B	0.002	0.044	0	1	342
Credit=A	0.0003	0.018	0	1	58
Credit=AA	0.0005	0.021	0	1	82
House_l	0.137	0.344	0	1	170,817
Car_l	0.053	0.226	0	1	170,817
Gender	0.862	0.344	0	1	170,817
Age	30.56	6.714	22	73	170,817
Marry	0.508	0.5	0	1	170,817
Income	3.985	1.213	1	7	170,817
Income<=1000	0.003	0.058	0	1	594
Income=1~2000	0.021	0.146	0	1	3,720
Income=2~5000	0.409	0.492	0	1	69,955
Income=5~10000	0.306	0.461	0	1	52,284
Income=1~20000	0.126	0.332	0	1	21,483
Income=2~50000	0.076	0.266	0	1	13,046
Income>50000	0.057	0.232	0	1	9,735
Education	1.848	0.791	1	4	170,817
Edu<=HighSchool	0.385	0.487	0	1	65,818
Edu=JuniorCollege	0.395	0.489	0	1	67,535
Edu=Bachelor	0.205	0.404	0	1	34,968
Edu>=Postgraduate	0.014	0.12	0	1	2,496
House	0.444	0.497	0	1	170,817
Car	0.249	0.432	0	1	170,817
Worktime	2.431	1.003	1	4	170,817
Worktime<=1year	0.166	0.372	0	1	28,421
Worktime=1~3year	0.451	0.498	0	1	77,054
Worktime=3~5year	0.168	0.374	0	1	28,680
Worktime>=5year	0.215	0.411	0	1	36,662
Redesign	0.699	0.459	0	1	170,817
Year	2,013.5	0.712	2,012	2,014	170,817
Year=2012	0.129	0.335	0	1	22,042
Year=2013	0.276	0.447	0	1	47,155
Year=2014	0.595	0.491	0	1	101,620

Table 2 also shows that men are more active than women in the P2P lending market, accounting for 86.2% of all borrowers. Some 50.8% of borrowers are married with an average age of 32. In addition, 71.5% of borrowers are in an income range of RMB 2001–10000. A proportion of 62.5% of borrowers attain an educational level of college or above. According to the data garnered from the sixth population census, only 8.93% of PRC people (except for those from Hong Kong, China; Macau, China; and Taipei, China) have an educational level of colleges or above, suggesting that the education level of participants in the P2P lending market is substantially higher than the national level. Furthermore, 44% of borrowers have houses, 25% have cars, and 83% have been working for more than one year. In brief, most borrowers on the P2P lending platform are young and well-educated people with a medium income and certain working experience.

3.3 Empirical Methodology

In this paper, we employ the logit model to estimate the impact of punctuation on the Chinese P2P lending market. The empirical equation is as follows:

$$\Pr(Y_i = 1) = \beta_0 + \beta_1 Punctuation_i + \sum_{i=1}^{i=n} \beta_i X_i + \varepsilon_i \quad (1)$$

where the dependent variable Y_i is a binary variable equal to 1 if the borrowers successfully have their loan request granted (or default after receiving funding) and 0 otherwise. The main explanatory variable of interest is $Punctuation_i$, the amount of punctuation used in a loan description. X_i is a vector of control variables, including the number of words in a loan description, amount of borrowing, interest rate, and term as well as borrowers' characteristics such as credit score, age, education, income, marriage status, working experience, etc. ε is the random disturbance term.

Equation (1) is estimated using logistic regression, which is widely employed to predict the binary dependent variable. We assume that y is the outcome of a Bernoulli trial coded as “1” and “0” and use π and $1-\pi$ to describe the corresponding probabilities of these two outcomes. We further assume that π is affected by a series of variables, and their correlation can be expressed by a linear function, $\pi = \beta X$, where β is the coefficient vector. However, given that the linear regression may make nonsensical predictions, we convert the binary variable into a continuous one by logistic transformation. We first take the odds of the event happening as:

$$\Omega_i = \frac{\pi_i}{1 - \pi_i} \quad (2)$$

In other words, it is the ratio of the probability of $Y=1$, π , and the probability of $Y=0$, $(1-\pi_i)$. Without doubt, those odds are continuous but cannot be negative.

Second, we take the logarithm to calculate the logit or log odds

$$\log it(\pi_i) = \ln(\Omega_i) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right) \quad (3)$$

Consequently, the logistic regression is likely to widen the value range of the probability π from (0, 1) to the whole real number axes via the logit transformation. The transformation function is defined as follows:

$$L : \ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \Pr(Y_i = 1) = \beta_0 + \sum_{i=1}^{i=n} \beta_i X_i + \varepsilon \quad (4)$$

where X is the vector of explanatory variables, β is the coefficient vector, and ε is random disturbance.

4. EMPIRICAL RESULT

4.1 Punctuation and Funding Probability

In this subsection, we test Hypothesis 1 by examining the relationship between the number of punctuation marks used in loan descriptions and the possibility of obtaining loans, and report the estimation results in Table 3-1 and Table 3-2.

In Table 3-1, Column (1) summarizes the regression result on main variables that have been used in the previous literature to explain the probability of funding success. In line with existing researches, loan requests with lower interest rates, lower requested amounts, and longer listing durations are more likely to be funded (Liu et al. 2015; Mild, Waitz, and Wockl 2015; Dorfleitner et al. 2016; Iyer et al. 2016). In Column (2), we report the result for an extended model including the amount of punctuation as the key explanatory variable and number of words as an additional control variable. As expected, the coefficient on punctuation is negatively significant, implying that more amount of punctuation is associated with a lower probability of having a loan request granted. The marginal effect reported in Column (3) indicates that ten additional punctuation marks is associated with a reduction in the possibility of funding success of 9.6 percentage points ($0.000629 \times 10 / 0.065$). Column (4) shows the regression result including the square of the number of punctuation marks (Punctuation²) as an additional explanatory variable and Column (5) reports the corresponding marginal effects. The negative coefficient on the Punctuation² implies that the funding success rate is linearly related to the amount of punctuation as the coefficient on it is always positive. The estimated coefficient on "Word" is positive, implying that a longer loan description increases the funding success rate by disclosing more information to the investors. To understand the relationship between punctuation and words, we include the interaction term between the amount of punctuation and the number of words and present the estimation results in Column (6). The corresponding marginal effect is shown in Column (7). The positively significant coefficient on the interaction term suggests that the length of the loan description can help to moderate the negative effects of punctuation on the funding success rate. With a larger number of words, more detailed contents of the loan description will be disclosed. This helps to attenuate information asymmetry between lenders and borrowers.

Table 3-1: Logit Regression Results on Funding Success**Dependent variable: probability of getting loan funded**

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Punctuation		-0.0134*** (-9.43)	-0.000629*** (-9.424)	-0.0027 (-0.79)	-0.000127 (-0.795)	-0.0182*** (-7.92)	-0.000856*** (-7.915)
Punctuation2				-0.0002*** (-3.41)	-8.00e-06*** (-3.412)		
Punctuation_Word						0.0000*** (2.72)	1.97e-06*** (2.721)
Word		0.0065*** (16.85)	0.000307*** (16.80)	0.0061*** (14.83)	0.000286*** (14.80)	0.0058*** (11.87)	0.000271*** (11.85)
Rate	-0.1763*** (-40.02)	-0.1800*** (-40.69)	-0.00846*** (-39.85)	-0.1806*** (-40.77)	-0.00849*** (-39.93)	-0.1798*** (-40.62)	-0.00845*** (-39.79)
lnAmount	-0.7468*** (-63.93)	-0.7635*** (-64.84)	-0.0359*** (-62.26)	-0.7637*** (-64.84)	-0.0359*** (-62.26)	-0.7641*** (-64.87)	-0.0359*** (-62.28)
Time	0.0241*** (17.19)	0.0244*** (17.37)	0.00115*** (17.31)	0.0243*** (17.34)	0.00114*** (17.29)	0.0244*** (17.41)	0.00115*** (17.35)
Credit	1.2470*** (57.57)	1.2437*** (57.54)	0.0585*** (58.51)	1.2429*** (57.47)	0.0584*** (58.44)	1.2451*** (57.57)	0.0585*** (58.54)
House_I	0.1794*** (5.59)	0.1859*** (5.78)	0.00874*** (5.780)	0.1852*** (5.76)	0.00870*** (5.758)	0.1859*** (5.78)	0.00873*** (5.779)
Car_I	-0.0138 (-0.30)	-0.0127 (-0.27)	-0.000598 (-0.274)	-0.0128 (-0.28)	-0.000603 (-0.276)	-0.0107 (-0.23)	-0.000504 (-0.231)
Gender	0.0220 (0.67)	0.0269 (0.81)	0.00126 (0.812)	0.0249 (0.75)	0.00117 (0.751)	0.0261 (0.79)	0.00123 (0.787)
Age	0.0331*** (17.89)	0.0320*** (17.25)	0.00151*** (17.19)	0.0321*** (17.30)	0.00151*** (17.24)	0.0319*** (17.19)	0.00150*** (17.13)
Marry	0.2247*** (8.57)	0.2267*** (8.63)	0.0107*** (8.622)	0.2263*** (8.62)	0.0106*** (8.608)	0.2269*** (8.64)	0.0107*** (8.631)
Income	0.4090*** (40.15)	0.3992*** (38.98)	0.0188*** (38.51)	0.3994*** (38.98)	0.0188*** (38.51)	0.3989*** (38.95)	0.0187*** (38.47)
Education	0.2491*** (18.08)	0.2475*** (17.93)	0.0116*** (17.88)	0.2467*** (17.86)	0.0116*** (17.81)	0.2479*** (17.95)	0.0117*** (17.90)
House	-0.1120*** (-4.09)	-0.1113*** (-4.06)	-0.00523*** (-4.060)	-0.1114*** (-4.06)	-0.00524*** (-4.063)	-0.1105*** (-4.03)	-0.00520*** (-4.033)
Car	0.3119*** (11.06)	0.3127*** (11.07)	0.0147*** (11.06)	0.3126*** (11.07)	0.0147*** (11.06)	0.3128*** (11.07)	0.0147*** (11.06)
Worktime	0.3588*** (29.20)	0.3623*** (29.40)	0.0170*** (29.11)	0.3620*** (29.37)	0.0170*** (29.08)	0.3623*** (29.39)	0.0170*** (29.10)
_cons	1.6219*** (13.30)	1.5781*** (12.87)		1.5480*** (12.59)		1.6474*** (13.15)	
Redesign	Yes	Yes		Yes		Yes	
Year	Yes	Yes		Yes		Yes	
N	170,817	170,817	170,817	170,817	170,817	170,817	170,817
r2_p	0.2566	0.2600		0.2601		0.2601	

Note: *p<0.1, **p<0.05, ***p<0.001; in parentheses are the Z statistical values.

Table 3-2: Logit Regression Results on Funding Success**Dependent variable: probability of getting loan funded**

Variable	(1)	(2)	(3)	(4)
Ten_Punctuation	-0.1803*** (-5.50)			
Five_Punctuation		-0.4177*** (-8.07)		
Ninety_Punctuation			-0.3767*** (-8.28)	
Ninetyfive_Punctuation				-0.4742*** (-8.28)
Word	0.0037*** (12.40)	0.0037*** (12.96)	0.0059*** (16.16)	0.0055*** (16.31)
Rate	-0.1797*** (-40.67)	-0.1801*** (-40.73)	-0.1795*** (-40.62)	-0.1802*** (-40.73)
lnAmount	-0.7589*** (-64.58)	-0.7601*** (-64.63)	-0.7612*** (-64.72)	-0.7627*** (-64.80)
Time	0.0244*** (17.41)	0.0244*** (17.39)	0.0244*** (17.36)	0.0244*** (17.40)
Credit	1.2356*** (57.27)	1.2358*** (57.31)	1.2444*** (57.57)	1.2420*** (57.48)
House_l	0.1841*** (5.73)	0.1846*** (5.75)	0.1853*** (5.76)	0.1869*** (5.81)
Car_l	-0.0106 (-0.23)	-0.0097 (-0.21)	-0.0107 (-0.23)	-0.0135 (-0.29)
Gender	0.0247 (0.74)	0.0250 (0.75)	0.0274 (0.83)	0.0258 (0.78)
Age	0.0329*** (17.74)	0.0330*** (17.78)	0.0323*** (17.39)	0.0323*** (17.40)
Marry	0.2283*** (8.70)	0.2279*** (8.68)	0.2269*** (8.64)	0.2272*** (8.65)
Income	0.4013*** (39.24)	0.4018*** (39.27)	0.4000*** (39.10)	0.4009*** (39.17)
Education	0.2454*** (17.78)	0.2454*** (17.78)	0.2463*** (17.85)	0.2465*** (17.85)
House	-0.1167*** (-4.26)	-0.1180*** (-4.31)	-0.1134*** (-4.14)	-0.1131*** (-4.13)
Car	0.3100*** (10.98)	0.3106*** (11.00)	0.3124*** (11.06)	0.3119*** (11.05)
Worktime	0.3623*** (29.42)	0.3623*** (29.42)	0.3630*** (29.46)	0.3625*** (29.42)
_cons	1.5623*** (12.74)	1.5737*** (12.84)	1.4857*** (12.14)	1.5142*** (12.38)
Redesign	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
N	170,817	170,817	170,817	170,817
r2_p	0.2592	0.2597	0.2597	0.2597

Note: *p<0.1, **p<0.05, ***p<0.001; in parentheses are the Z statistical values.

To understand the impact of both an extremely small and a large number of punctuation marks on the funding success rate, we do more tests by creating several new variables: (1) *Five_Punctuation* is equal to one if the number of punctuation marks used in the loan description is under the 5th percentile, and zero otherwise; (2) *Ten_Punctuation* is equal to one if the number of punctuation marks used in the loan description is under the 10th percentile, and zero otherwise; (3) *Ninety_Punctuation* is equal to one if the number of punctuation marks used in the loan description is among the top 10%, and zero otherwise; and (4) *Ninetyfive_Punctuation* is equal to one if the number of punctuation marks used in the loan description is among the top 5%, and zero otherwise. Table 3-2 reports the estimation results. Columns (1) and (2) indicate that an extremely small amount of punctuation is negatively related to funding probability. Within a given number of words for a loan description, too few punctuation marks would impair the readability of the text and hence lower the probability of having a loan request granted. Columns (3) and (4) suggest that an excessive amount of punctuation also reduces the funding success rate. All these results demonstrate that the appropriate usage of punctuation does indeed predict the funding probability in the P2P lending market.

4.2 Punctuation and Borrowing Rate

As discussed in the previous section on hypothesis development, borrowers whose loan description contains a large number of punctuation marks tend to offer a lower interest rate to the investors because excessive use of punctuation may reflect the borrowers' cognitive bias and overconfidence in his loan request. Table 4 presents the estimation results on the relationship between the number of punctuation marks and the borrowing rate. Comparing the two columns, we find that including the number of punctuation marks enhances the R square and hence the explanatory power of the model. The sign on the Punctuation coefficient is -0.0084 and is statistically significant. This suggests that all else being equal, the interest rate will fall by approximately 0.56% ($0.0084 \times 10/14.98$) when the number of punctuation marks in the loan description increases by ten.

In formal Chinese writing, the number of punctuation marks generally increases with the number of words. However, the loan descriptions for P2P lending have been considerably affected by Internet language where the number of punctuation marks may not be constrained by the length of the text.³ Overuse of punctuation will not only make loan descriptions informal, but also reflect the cognitive bias of borrowers. Borrowers using punctuation excessively in their loan description are more likely to be overconfident and offer borrowing rates lower than the market rate. Our empirical results strongly support this argument and prove that Hypothesis 2 is correct.

4.3 Punctuation and Loan Default

The empirical evidence presented in the previous two sections confirms that the number of punctuation marks can predict the probability of funding and the borrowing rate in the P2P lending market. This subsection tests Hypothesis 3 that the more punctuation borrowers use in their loan descriptions, the more likely it is that borrowers will default.

³ For instance, I have a physical store only to increase the sales volume and I need support from you guys.....

Table 4: OLS Regression Results on Loan Interest Rate**Dependent variable: borrowing rate**

Variable	(1)	(2)	Variable	(1)	(2)
Punctuation		−0.0084*** (−8.50)	Marry	−0.1766*** (−9.77)	−0.1719*** (−9.53)
Word		0.0071*** (23.79)	Income	0.1218*** (15.59)	0.1089*** (13.94)
InAmount	−0.0158** (−2.00)	−0.0285*** (−3.63)	Education	−0.2612*** (−25.64)	−0.2658*** (−26.14)
Time	0.0321*** (34.50)	0.0321*** (34.56)	House	−0.0628*** (−3.26)	−0.0668*** (−3.48)
Credit	−0.8597*** (−39.39)	−0.8669*** (−39.78)	Car	−0.4932*** (−22.26)	−0.4985*** (−22.54)
House_l	−0.2736*** (−10.64)	−0.2651*** (−10.32)	Worktime	0.0149* (1.69)	0.0185** (2.10)
Car_l	−0.0253 (−0.66)	−0.0199 (−0.52)	_cons	16.9084*** (208.65)	16.7080*** (205.00)
Gender	0.2488*** (10.83)	0.2561*** (11.16)	Redesign	Yes	Yes
Age	−0.0006 (−0.44)	−0.0016 (−1.12)	Year	Yes	Yes
			N	170,817	170,817
			r2_a	0.1284	0.1317

Note: *p<0.1, **p<0.05, ***p<0.001; in parentheses are the T statistical values.

Table 5: Logit Regression Results on Loan Performance**Dependent variable: probability of default**

Variable	(1)	(2)	(3)	Variable	(1)	(2)	(3)
Punctuation		−0.0020 (−0.35)	−0.0020 (−0.351)	Marry	−0.0935 (−0.98)	−0.0928 (−0.97)	−0.0928 (−0.973)
Word		−0.0005 (−0.31)	−0.0004 (−0.314)	Income	0.1123*** (2.99)	0.1141*** (3.04)	0.1140*** (3.036)
Rate	0.1669*** (10.67)	0.1676*** (10.68)	0.1680*** (10.68)	Education	−0.4678*** (−8.83)	−0.4673*** (−8.82)	−0.4670*** (−8.818)
InAmount	0.4030*** (7.09)	0.4054*** (7.08)	0.4050*** (7.085)	House	0.2434*** (2.65)	0.2439*** (2.65)	0.2440*** (2.653)
Time	−0.0180*** (−3.26)	−0.0181*** (−3.27)	−0.0181*** (−3.273)	Car	−0.3643*** (−3.57)	−0.3640*** (−3.56)	−0.3640*** (−3.562)
Credit	−4.7347*** (−10.49)	−4.7320*** (−10.48)	−4.7320*** (−10.48)	Worktime	−0.0666 (−1.45)	−0.0675 (−1.47)	−0.0675 (−1.471)
House_l	−0.6454*** (−5.13)	−0.6468*** (−5.14)	−0.6470*** (−5.136)	_cons	−3.3184*** (−4.56)	−3.2949*** (−4.52)	
Car_l	0.2147 (1.33)	0.2127 (1.32)	0.2130 (1.317)	Redesign	Yes	Yes	Yes
Gender	0.4913*** (3.64)	0.4918*** (3.64)	0.4920*** (3.637)	Year	Yes	Yes	Yes
Age	0.0159** (2.38)	0.0160** (2.38)	0.0160** (2.384)	N	11,136	11,136	11,136
				r2_p	0.2200	0.2202	

Note: *p<0.1, **p<0.05, ***p<0.001; in parentheses are the Z statistical values.

In Table 5, Column (1) summarizes the estimation result on the probability of default using the main explanatory variables employed by previous literature. Column (3) reports the estimated marginal effects of Column (2). The coefficient on Punctuation is −0.0027, and the marginal effect of the Punctuation is −0.000015. However, both are

statistically insignificant. In addition, Pseudo R^2 in Column (4) is only 0.03% higher than that in Column (1), indicating that the number of punctuation marks and words failed to explain repayment performance.

4.4 Robustness Checks

This subsection performs several robustness checks to verify the validity of the estimation results reported in the previous three subsections.

Table 6: Probit Regression Results on Funding Success and Default

Variable	Funding Success		Default	
	(1)	(2)	(3)	(4)
Word	0.0020*** (13.36)	0.0033*** (16.39)	-0.0004 (-0.67)	-0.0001 (-0.19)
Punctuation		-0.0069*** (-9.45)		-0.0011 (-0.37)
Rate	-0.0902*** (-42.22)	-0.0908*** (-42.38)	0.0927*** (10.64)	0.0926*** (10.62)
lnAmount	-0.4100*** (-68.09)	-0.4127*** (-68.35)	0.1993*** (6.51)	0.1985*** (6.47)
Time	0.0122*** (17.66)	0.0122*** (17.62)	-0.0110*** (-3.91)	-0.0110*** (-3.90)
Credit	0.6676*** (62.84)	0.6705*** (63.03)	-2.0661*** (-13.75)	-2.0664*** (-13.75)
House_l	0.1049*** (6.40)	0.1048*** (6.38)	-0.3435*** (-5.38)	-0.3440*** (-5.39)
Car_l	-0.0060 (-0.25)	-0.0064 (-0.26)	0.1071 (1.25)	0.1077 (1.26)
Gender	0.0146 (0.89)	0.0158 (0.96)	0.2639*** (3.83)	0.2646*** (3.83)
Age	0.0168*** (17.63)	0.0165*** (17.26)	0.0085** (2.37)	0.0085** (2.36)
Marry	0.1072*** (8.27)	0.1066*** (8.21)	-0.0420 (-0.85)	-0.0420 (-0.85)
Income	0.2109*** (40.09)	0.2100*** (39.84)	0.0684*** (3.42)	0.0684*** (3.42)
Education	0.1355*** (19.48)	0.1359*** (19.51)	-0.2500*** (-9.09)	-0.2496*** (-9.07)
House	-0.0582*** (-4.27)	-0.0566*** (-4.14)	0.1273*** (2.62)	0.1274*** (2.62)
Car	0.1616*** (11.20)	0.1622*** (11.23)	-0.1994*** (-3.69)	-0.1999*** (-3.70)
Worktime	0.1809*** (29.38)	0.1811*** (29.37)	-0.0293 (-1.22)	-0.0292 (-1.22)
_cons	0.7738*** (12.23)	0.8093*** (12.75)	-2.2348*** (-6.70)	-2.2271*** (-6.66)
Redesign	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
N	170,817	170,817	170,817	170,817
r2_p	0.2676	0.2687	0.2196	0.2196

Note: *p<0.1, **p<0.05, ***p<0.001; in parentheses are the Z statistical values.

First, to eliminate possible errors in the selection of measurement methods, we use the Probit Model to re-estimate the association of the number of punctuation marks with the probability of funding success and default. The results are shown in Table 6. The coefficient on the number of words is positively significant while the coefficient on the number of punctuation marks is negatively significant. This result again verifies that Hypothesis 1 is true.

Second, we use robust and bootstrap standard errors (50 times) rather than ordinary standard errors to do the OLS regression on the funding rate. In Table 7, Columns (1) and (2) report the robust standard errors while Columns (3) and (4) list the bootstrap standard errors. All the test results indicate that the number of punctuation marks is negatively associated with the funding rate. This again proves that Hypothesis 2 is true.

Table 7: Robust and Bootstrap Standard Errors
Dependent variable: borrowing rate

Variable	(1)	(2)	(3)	(4)
Word	0.0055*** (22.59)	0.0071*** (22.27)	0.0055*** (24.84)	0.0071*** (22.96)
Punctuation		-0.0084*** (-7.82)		-0.0084*** (-8.23)
lnAmount	-0.0263*** (-3.13)	-0.0285*** (-3.39)	-0.0263*** (-3.66)	-0.0285*** (-3.96)
Time	0.0322*** (36.87)	0.0321*** (36.76)	0.0322*** (41.80)	0.0321*** (41.77)
Credit	-0.8700*** (-53.33)	-0.8669*** (-53.16)	-0.8700*** (-66.94)	-0.8669*** (-66.89)
House_l	-0.2659*** (-10.87)	-0.2651*** (-10.83)	-0.2659*** (-12.50)	-0.2651*** (-12.43)
Car_l	-0.0190 (-0.54)	-0.0199 (-0.56)	-0.0190 (-0.78)	-0.0199 (-0.82)
Gender	0.2522*** (11.24)	0.2561*** (11.42)	0.2522*** (12.50)	0.2561*** (12.73)
Age	-0.0012 (-0.83)	-0.0016 (-1.12)	-0.0012 (-0.97)	-0.0016 (-1.31)
Marry	-0.1712*** (-9.45)	-0.1719*** (-9.49)	-0.1712*** (-9.33)	-0.1719*** (-9.41)
Income	0.1102*** (13.88)	0.1089*** (13.72)	0.1102*** (14.45)	0.1089*** (14.26)
Education	-0.2653*** (-26.21)	-0.2658*** (-26.26)	-0.2653*** (-28.14)	-0.2658*** (-28.15)
House	-0.0685*** (-3.55)	-0.0668*** (-3.46)	-0.0685*** (-3.47)	-0.0668*** (-3.39)
Car	-0.4990*** (-23.51)	-0.4985*** (-23.49)	-0.4990*** (-23.35)	-0.4985*** (-23.42)
Worktime	0.0182** (2.07)	0.0185** (2.11)	0.0182** (2.24)	0.0185** (2.28)
_cons	16.6815*** (194.44)	16.7080*** (194.68)	16.6815*** (227.62)	16.7080*** (225.56)
Redesign	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
N	170,817	170,817	170,817	170,817
r2_a	0.1313	0.1317	0.1313	0.1317

Note: *p<0.1, **p<0.05, ***p<0.001; in parentheses are the Z/T statistical values.

Third, we exclude default loan samples, rather than full sample data, to re-estimate the logit regression on the probability of funding success and the OLS regression on the borrowing rate. The results shown in Table 8 indicate that punctuation is significantly and negatively related to loan funding success and the borrowing rate. This also further proves that Hypotheses 1 and 2 are robust.

Table 8: Excluding the Default Sample

Variable	Funding Success		Default	
	(1)	(2)	(3)	(4)
Word	0.0040*** (13.50)	0.0066*** (16.22)	0.0055*** (23.90)	0.0071*** (23.89)
Punctuation		-0.0132*** (-8.96)		-0.0085*** (-8.58)
Rate	-0.1876*** (-39.82)	-0.1887*** (-39.97)		
InAmount	-0.7568*** (-61.85)	-0.7622*** (-62.12)	-0.0292*** (-3.71)	-0.0315*** (-3.99)
Time	0.0252*** (17.29)	0.0252*** (17.26)	0.0322*** (34.56)	0.0321*** (34.44)
Credit	1.3237*** (59.53)	1.3304*** (59.71)	-0.8769*** (-40.17)	-0.8738*** (-40.03)
House_l	0.2258*** (6.80)	0.2265*** (6.81)	-0.2709*** (-10.52)	-0.2701*** (-10.48)
Car_l	-0.0172 (-0.36)	-0.0198 (-0.41)	-0.0248 (-0.64)	-0.0257 (-0.67)
Gender	-0.0035 (-0.10)	-0.0022 (-0.06)	0.2528*** (10.99)	0.2567*** (11.16)
Age	0.0302*** (15.53)	0.0296*** (15.16)	-0.0008 (-0.54)	-0.0012 (-0.84)
Marry	0.2312*** (8.45)	0.2290*** (8.37)	-0.1703*** (-9.41)	-0.1710*** (-9.45)
Income	0.3755*** (35.21)	0.3738*** (34.98)	0.1162*** (14.78)	0.1149*** (14.62)
Education	0.2794*** (19.50)	0.2797*** (19.50)	-0.2678*** (-26.26)	-0.2683*** (-26.31)
House	-0.1295*** (-4.52)	-0.1254*** (-4.37)	-0.0695*** (-3.60)	-0.0678*** (-3.51)
Car	0.3211*** (10.93)	0.3234*** (11.00)	-0.4959*** (-22.34)	-0.4954*** (-22.32)
Worktime	0.3639*** (28.39)	0.3636*** (28.33)	0.0196** (2.22)	0.0199** (2.26)
_cons	1.5102*** (11.85)	1.5833*** (12.38)	16.6887*** (204.32)	16.7157*** (204.54)
Redesign	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
N	170,008	170,008	170,008	170,008
r2_a			0.1319	0.1322
r2_p	0.2698	0.2709		

Note: *p<0.1, **p<0.05, ***p<0.001; in parentheses are the Z/T statistical values.

5. CONCLUSION

The information asymmetry in the online P2P lending market motivates us to explore the role of loan descriptions in bridging the information gap between borrowers and investors. In particular, using the data from Renrendai, one of the largest peer-to-peer lending platforms in the PRC, we investigate the impact of punctuation on the funding probability, borrowing rate, and default. We find that the amount of punctuation is negatively associated with the funding probability and borrowing rate. Within a given amount of words, more punctuation makes loan descriptions informal and reduces the readability of the text, thereby impairing investors' trust in borrowers. Moreover, borrowers who use punctuation excessively may have lower ability of self-control and tend to be over confident. When setting interest rates, they tend to underestimate the risk and offer a lower borrowing rate. In conclusion, investors are able to identify creditworthy borrowers with the help of the punctuation used in the loan description even when hard facts like credit scores are not available. Our study has important implications for both the platform operators and participants. To reduce the noise caused by the excessive use of punctuation, platforms should provide guidance for borrowers on how to write their loan description in a formal and standard format.

REFERENCES

- ACCA, 2015. The rise of peer-to-peer lending in China: an overview and survey case study.
- Ashta, A., Assadi, D., Johnson, S., 2010. Online or offline?: the rise of "peer-to-peer" lending in microfinance. *Journal of Electronic Commerce in Organizations* 8, 26–37.
- Baldwin, R.S., 1978. Psycholinguistic approaches to a theory of punctuation. *Journal of Literacy Research* 10, 363–375.
- Baron, N.S., 2001. Commas and canaries: the role of punctuation in speech and writing. *Lang Sci* 23, 15–67.
- Bruthiaux, P., 1993. Knowing when to stop: investigating the nature of punctuation. *Lang Commun* 13, 27–43.
- Burtch, G., Ghose, A., Wattal, S., 2014. Cultural differences and geography as determinants of online prosocial lending. *Mis Quart* 38, 773–794.
- Cohen, H., Douaire, J., Elsabbagh, M., 2001. The role of prosody in discourse processing. *Brain Cognition* 46, 73–82.
- Dorflleitner, 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. *J Bank Financ* 64, 169–187.
- Duarte, J., Siegel, S., Young, L., 2012. Trust and credit: the role of appearance in peer-to-peer lending. *Rev Financ Stud* 25, 2455–2483.
- Everett, C.R., 2014. Origins and development of credit-based crowdfunding. Working Paper.
- Ferreiro, E., Pontecorvo, C., 1999. Managing the written text: the beginning of punctuation in children's writing. *Learning & Instruction* 9, 543–564.
- Fuchs, S., Krivokapic, J., 2016. Prosodic boundaries in writing: evidence from a keystroke analysis. *Front Psychol* 7, 1678.
- Gao, Q., Lin, M., 2015. Lemon or cherry? The value of texts in debt crowdfunding. Working Paper.
- Gervais, S., Heaton, J.B., Odean, T., 2011. Overconfidence, compensation contracts, and capital budgeting. *J Financ* 66, 1735–1777.
- Gonzalez, L., Loureiro, Y.K., 2014. When can a photo increase credit? The impact of lender and borrower profiles on online peer-to-peer loans. *Journal of Behavioral and Experimental Finance* 2, 44–58.
- Greiner, M.E., Wang, H., 2010. Building consumer-to-consumer trust in e-finance marketplaces: an empirical analysis. *Int J Electron Comm* 15, 105–136.
- 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. *J Marketing Res* 48, S138–S149.
- Hirshleifer, D., 2001. Investor psychology and asset pricing. *J Financ* 56, 1533–1597.
- Iyer, R., Khwaja, A.I., Luttmer, E.F.P., Shue, K., 2016. Screening peers softly: inferring the quality of small borrowers. *Manage Sci* 62, 1554–1577.

- Kahneman, D., Tversky, A., 1973. On the psychology of prediction. *Psychol Rev* 80, 237–251.
- Larrimore, L., Jiang, L., Larrimore, J., Markowitz, D., Gorski, S., 2011. Peer-to-peer lending: the relationship between language features, trustworthiness, and persuasion success. *J Appl Commun Res* 39, 19–37.
- Lee, E., Lee, B., 2012. Herding behavior in online P2P lending: an empirical investigation. *Electron Commer R A* 11, 495–503.
- Lin, M.F., Prabhala, N.R., Viswanathan, S., 2013. Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending. *Manage Sci* 59, 17–35.
- Liu, B.L., Wang, Z.N., Jin, Z., 2010. The effects of punctuations in Chinese sentence comprehension: an ERP study. *J Neurolinguist* 23, 66–80.
- Liu, D., Brass, D.J., Lu, Y., Chen, D.Y., 2015. Friendships in online peer-to-peer lending: pipes, prisms, and relational herding. *Mis Quart* 39, 729–742.
- Loughran, T., McDonald, B., 2014. Measuring readability in financial disclosures. *J Financ* 69, 1643–1671.
- Luo, B.J., Lin, Z.X., 2013. A decision tree model for herd behavior and empirical evidence from the online P2P lending market. *Inf Syst E-Bus Manag* 11, 141–160.
- Michels, J., 2012. Do unverifiable disclosures matter? evidence from peer-to-peer lending. *Account Rev* 87, 1385–1413.
- Mild, A., Waitz, M., Wockl, J., 2015. How low can you go? Overcoming the inability of lenders to set proper interest rates on unsecured peer-to-peer lending markets. *J Bus Res* 68, 1291–1305.
- Miller, B.P., 2010. The effects of reporting complexity on small and large investor trading. *Account Rev* 85, 2107–2143.
- Peng, L., Xiong, W., 2006. Investor attention, overconfidence and category learning. *J Financ Econ* 2006 80, 563–602.
- Pope, D.G., Sydnor, J.R., 2011. What's in a picture? Evidence of discrimination from prosper.com. *J Hum Resour* 46, 53–92.
- Pynte, J., Kennedy, A., 2007. The influence of punctuation and word class on distributed processing in normal reading. *Vision Res* 47, 1215–1227.
- Reber, R., Schwarz, N., 1999. Effects of perceptual fluency on judgments of truth. *Consciousness & Cognition* 8, 338–42.
- Rennekamp, K., 2012. Processing fluency and investors' reactions to disclosure readability. *J Account Res* 50, 1319–1354.
- Roux, J.P., 2008. An interlocutory analysis as a methodological approach in studying cognitivo-linguistic mediations: interest, difficulties, and limits. *Eur J Dev Psychol* 5, 609–622.
- Schriefers, H., Meyer, A.S., Levelt, W.J., 1990. Exploring the time course of lexical access in language production: picture–word interference studies. *Journal of Memory & Language* 29, 86–102.
- Steffes, E.M., Burgee, L.E., 2009. Social ties and online word of mouth. *Internet Res* 19, 42–59.

- Tan, H.T., Wang, E.Y., Zhou, B., 2014. When the use of positive language backfires: the joint effect of tone, readability, and investor sophistication on earnings judgments. *J Account Res* 52, 273–302.
- Thaler, R.H., Shefrin, H.M., 1977. An economic theory of self-control. *J Polit Econ* 89, 392–406.
- The Economist, 2016. Credit in China: just spend. 19–25 November.
- Whissell, C., 2013. Titles in highly ranked multidisciplinary psychology journals 1966–2011: more words and punctuation marks allow for the communication of more information. *Psychol Rep* 113, 969–986.
- Ye, D.Z., Ng, Y.K., Lian, Y.J., 2015. Culture and happiness. *Soc Indic Res* 123, 519–547.