

Analysis and Prediction of P2P Online Lending Platform—Based on Binary Logistic Regression Model

Yuanyuan Zhang⁺, Yuelin Shen²

¹School of Management, Shanghai University, Shanghai200444, China

²College of Business Administration, Zhejiang University of Finance and Economics, Hangzhou 310018, China

Abstract: This paper uses 17 indicators to analyze the critical factors that affect the P2P online lending platforms based on a binary logistic regression model. It shows that the problematic P2P online lending platforms are closely related to four indicators: the average interest rate, the top ten borrowers in terms of the repayment amount, the operating duration, and the top ten investors to receive the repayments. The accuracy of the regression model is up to 73% to predict whether the P2P network lending platform will be in a problem.

Keywords: binary logistic regression model, problematic platform, online loan, peer-to-peer

1. Introduction

P2P (Peer-to-Peer) online lending is a kind of personal lending mode on the Internet. P2P online lending begins in England in 2005, and becomes popular in China in 2007. It has developed more than 4300 network lending platforms. However, with this rapid development, P2P (Peer-to-Peer) online lending also has exposed many problems, such as herd behavior of investors, malignant run-away incidents of the lending platforms, cash difficulties of the platform, and other undesirable phenomena. Since December 2015, due to the “e-rent treasure”, P2P online lending becomes under the pressure of public, and investors for the P2P network lending platform selection is first established in the advertising and brand effect, e-rent treasure event reduced investor’s confidence to invest P2P online lending platform. In the P2P online lending system, the most critical node is the P2P network lending platform, as the center of the system, if there is a problem, there is no transaction infrastructure, and unable to achieve lending behavior. Many problems in the P2P online lending system, the most serious problem is the platform has problems.

This paper is the new research on the data of a third party rating agencies, Online Lending House. At present, Online Lending House has been collecting the information of the existing platforms which meet the criterion of its rating, helping investors understand the quality of the existing P2P online lending platforms. However, there are some weaknesses with their rating system. Firstly, the number of P2P network lending platforms is increasing every day, and the new P2P platforms which cannot meet its criterion for more than three months are not allowed to participate in the rating; hence, the rating system is not timely. Secondly, investors cannot find a platform beyond the 100 platforms in their rating system; therefore, the investors who desire to explore the opportunities in those new platforms are not able to find some guidances.

This paper uses the method of binary logistic regression model and builds discriminant indexes to find the problems with platforms. Such a method helps investors to find problematic platform, reduce property

⁺ Corresponding author. Tel.: + 18800203778; fax: +0515-85486601.
E-mail address: 1505539455@qq.com.

loss, reduce the problematic platform speed, promote the normalization of P2P network lending industry, which makes the entire Internet financial environment more healthy and stable.

2. Review

Foreign research on P2P network loans is more detailed, which is mainly based on the quantitative analysis and chooses the participants of network loans as research objects. Wei and Lin [1] research the participants of the whole P2P market, transaction results and the impact of social welfare in the two forms of auction price and published price on the platform, and point out that although the published price on the platform is conducive to finishing the transaction quickly, to the whole market, auction price mechanism is more conducive to the benign development of the market. Lin, Prabhala and Viswanathan [2] find that the friendship among money lenders can be regarded as a signal of credit quality, which increases the probability of successful lending, reduces the interest rate of the loan, and reduces the probability of default after the loan. The friendship economy based on the role and identity of the friend shows an amazing level. The paper discusses the role of their findings in disintermediation of the financial market and decentralized electronic market design. Zhang and Liu [3] point out that there are herding behaviors in P2P network loans.

The research on P2P online lending in China focuses on the qualitative research on its model and regulation. In 2012, Xie and Zou [4] put forward the concept of "Internet financial model" for the first time, interpreted the operation mode of the P2P network loans platform of "peer-to-peer lender", and deeply interpreted the prerequisite for the borrowing on P2P network loans platforms. Based on the potential problems in the development process of P2P network loans model in China, Qinke Ye [5] puts forward the suggestions of preventing the risks of P2P network loan platforms. Also, a small number of scholars research P2P network loans from the perspective of quantitative research. Chen, Ding [6] and others point out that binary logic regression can be used to judge whether users are out of date from four dimensions, including credit, individual, contract object and past borrowing.

For current research, there are few researches on the influencing factors of the problems existing on P2P network loans platforms from the quantitative point of view and whether problems occur on the existing P2P network loan platforms.

3. Analysis of the Causes of P2P Online Lending Platform

P2P online lending model is a double-edged sword, while improving the efficiency of investment and financing, but also gave birth to a variety of risks. For investors, the problems encountered platform, they may be loss money. These events are not conducive to the healthy development of P2P network lending industry. P2P network platform problems can be roughly divided into the following three cases.

(1) Net loan platform does not have the ability to operate financial

Because the P2P network loan in a three-no state before December 2015, there are no barriers, no industry standards, no regulators, which resulting in some enterprises do not have the financial background into the P2P network loan industry. Platform like this is generally not suitable for investors long-term investment, its development is very unstable, there are serious defects in risk control, easy to cause the platform to leverage funding imbalance, loss of the number of investment, the platform cumulative amount to be also unable to detect qualified borrowers but don't accumulate [7], a snowball growth form, finally the general will appear to cash difficulties phenomenon.

(2) Net loan platform founder treat a platform as a channel for rapid financing

Net loan industry become more and more popular, there is a part of the interested motives of the founder of online loan platform, started with the "POND'S scam" preparations and plans [8]. Platform to attract investor A's funds into the platform, then pay back to investor B. When the platform accumulated to a certain asset, it will eventually run away, this behavior does harm to the P2P net loan industry, investors will bear the pain of loss of funds. These platforms are similar in volume, registered capital, operating hours. For example MAYIDAI, YOUYIWANG.

(3) P2P online lending platform profit model has not yet formed

China's online lending platform is much higher than the interest rate of bank loans, but borrowers have lower qualifications. In order to maintain the platform operation, part of the platform through self built pool of funds, frequent demolition standard, self financing, so on, strive to expand the scale of business. However, these methods of operation itself is a huge risk, and does not meet regulatory requirements, there is a problem sooner or later [9]. In particular, the development of China's online credit industry lags behind, many platforms with the help of online audit, collection of borrower information, high operating costs, has deviated from the principle of low-cost Internet banking business [10]. In addition, China's network loan platform credit evaluation technology is still immature, far behind the u.s.

4. Methods and Data

In order to discover the problematic platforms in the P2P online lending platforms in time, reduce the loss of investors, and promote the P2P net loan market stability, on the basis of previous research, including the proposed average interest rate, the top ten borrowers also amount to be accounted for, operation time, the top ten investors accounted for 17 indicators ,using the method of binary logistic regression model, through the collected data, to establish a forecasting model of P2P net loan problematic platform.

4.1. Data

Data collection from March 2016 to August 2016, the volume, the average interest rate, the average loan period, the cumulative amount to be also, the per capita amount of loan number, the number of borrowers, the top ten borrowers also amount to be accounted for, time weighted volume, operation time, registered capital, the next 60 days to, financial leverage, investment, the number of full scale use, per capita the amount of investment, the top ten investors accounted for 17 of the amount to be received input indexes, using the method of binary logistic regression model analysis of the data. Monthly comparison to the P2P platform for the network platform may be a problem to predict, guidance for investors to invest in the P2P platform, but also to the supervisory function of the existing platform, standardize the network loan order, maintain good credit atmosphere.

Criteria for the evaluation of the problematic platform: the month of P2P network lending platform does not appear in the data collection within the scope of Online Lending House next month is regarded as a problematic platform.

4.2. Variable

In this paper, the explanatory variables is that the P2P network lending platform my be a problematic platform, if the platform is the problematic platform, marked 1, if not, marked 0. Through the above analysis of the problems of the P2P platform and the related literature, this paper selects 17 explanatory indicators of Table 1.

Table 1: Indicators and their interpretation

Index	Index Interpretation
Volume (x1)	Platform in a certain period of time, the total amount of investment absorbed by investors, also known as the transaction size, volume of transactions, transaction size, etc.
Average interest rate (x2)	The weighted average interest rate of all the subjects on a monthly basis.
Average loan period (x3)	Weighted average value of all the loan items.
Cumulative amount to be paid (x4)	P2P net loan platform since the establishment of the total repayment has not yet been reflected in the payment pressure P2P platform.
Loan amount per capital (x5)	The ratio of the total turnover of the platform to the total number of borrowers.

Number of borrowings (x6)	P2P net loan platform issued all the subject matter of the mortgage, credit subject, the subject of security and the number of seconds.
Number of borrowers (x7)	The number of new loans.
The top ten borrowers to be accounted for (x8)	All borrowers in the top 10 of the borrower's outstanding repayment amount accounted for the proportion of total outstanding repayment of all borrowers.
Time weighted volume (x9)	The weighted average value of the total investment of the investor in a given period of time.
Operation time (x10)	P2P net loan platform from the date of establishment of the month.
Registered capital (x11)	The amount of the property owned by the state for the operation and management of an enterprise as a legal person.
The amount to be paid in the next 60 days (x12)	P2P online lending platform needs to pay to investors in the next month.
Leverage ratio (x13)	Registered capital of P2P net loan platform and the proportion of the platform investors to receive money.
The number of investors (x14)	New investors.
The time needed to complete the loan transaction (x15)	P2P online lending platform to release the average value of all the subject full time, less full time, indicating that the higher the degree of enthusiasm for the platform investors, the financing ability of the platform is stronger.
Average investment amount per person (x16)	The ratio of the total turnover of the platform to the total investment.
The proportion of the top ten investors to be received (x17)	All investors in the top 10 investors have not yet recovered the total amount of investment accounted for the proportion of total investment of all investors have not yet recovered.

4.3. Model selection

The binary logistic regression model method is suitable for the regression analysis of the dependent variable. This paper discusses the factors of P2P network lending platform run into problems, the dependent variable is "P2P network lending platform whether there is a problem", because only "yes" and "no", binary logistic regression model belong to the typical model, so the binary logistic regression model of the P2P network lending platform is studied.

The general expression of the model is:

$$\theta(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

P , which means that the probability of the problem of the platform, $1 - P$ means the platform does not have the probability of the problem, x_i means the i th factors, β_i means the i th indicator of the regression coefficient, β_0 is a constant, ε is the error term.

In this paper, the Clementine software is used to collect the data for logistic regression, so as to get the value of β_i , on this basis, forecast the existing platform is likely to become a platform.

4.4. Descriptive statistical analysis

This is a collection of regional data from March 2016 to August 2016, a total of 912 samples, the number of platform there will be some changes every month, it illustrates that some P2P network loan industry

platforms are eliminated, these fluctuations of the platform exit the market, because of their own ability is not enough to exit the market, or because of the beginning with the purpose of fraud into the market, waiting for an opportunity to escape.

5. Empirical Analysis

Using the method of binary logistic regression model to carry on the regression, to the target screening, retained the important value to be bigger than 0.95 targets, on this foundation uses the forward method to carry on the further screening to the target, carries on the logical regression to the data. The final retention indexes are shown in table 2.

After screening, leaving 4 important indicators, there are average interest rate (x2), the top ten borrowers to be accounted for (x8), operation time (x10) and the proportion of the top ten investors to be received (x17).

Operating time (x10) can be used as a reference to investors the most basic factors. General business platform for a long time to ensure a lasting business status.

Table 2: The final forecast variables

	B	S.E.	Wald	df	Sig.	Exp(B)
X2	-13.848	3.197	18.761	1	0.000	0.000
X8	-1.930	0.326	35.151	1	0.000	0.145
X10	-0.035	0.006	29.411	1	0.000	0.965
X17	-3.125	0.640	23.838	1	0.000	0.044
Constant	3.706	0.491	56.879	1	0.000	40.694

Table 3: Prediction results of training set data

	Observed value	Predicted value		Accuracy rate	
		0	1		
1st	y	0	157	125	55.7
		1	40	241	85.8
	Average prediction accuracy				70.7
2nd	y	0	189	93	67.0
		1	65	216	76.9
	Average prediction accuracy				71.9
3rd	y	0	205	77	72.7
		1	87	194	69.0
	Average prediction accuracy				70.9
4th	y	0	207	75	73.4
		1	77	204	72.6
	Average prediction accuracy				73.0

In table 2, observe the Wald value can be found in the top ten borrowers also amount to be accounted for is the most important index, the regression coefficient is negative, that the top ten borrowers also amount to be accounted for more of the lower platform occurrence probability. The top ten investors accounted for more than the amount to be received, the lower the probability of occurrence of the platform, and is the most important indicator of the top ten borrowers also amount to be accounted for is consistent, which shows that a healthy development of the P2P network lending platform, the borrower is not also amount to be as low as possible, it needs more traffic continuous circulation. Operating time regression coefficient is negative, it is not difficult to understand, the longer the operation time of the P2P network lending platform, the lower the likelihood of problems. First of all, it can continue operations, in the P2P network loan industry operation mode of business is not clear, has its own management system a more sophisticated management, P2P online lending platform this is the main control system of wind company, for example, the first domestic P2P network loan company which has a pat on the loan, the mature mirror system, for the rating of borrowers, help investors rational investment. The average interest rate is relative important value of the lowest index, the average interest rate is P2P platform for the main means to attract customers, set P2P online lending platform can not exceed the interest rate of bank interest rates 24%, at present stage, after the introduction of

a series of regulatory policies, interest rates gradually to the 10%.General set of high interest rates platform, with a higher risk, more than 24% of the bank interest rate platform, mainly want to attract more customers in a short time, there are more than POND'S scam phenomenon.

Run the logistic regression, after 4 steps to remove the model contribute less variables and Wald test, the average rate (x2)and eventually left the top ten borrowers also amount to be accounted (x8), operation time(x10), the top ten investors accounted for of the amount to be received(x17), as the final model variables. As can be seen from table 3, the accuracy of the model reaches 73% at the time of the fourth step. In order to clear the final model, the model was tested by Hosmer-Lemeshow.

Table 4: Test results of Hosmer-Lemeshow

Chi square	df	Sig.
28.069	8	0.000

From table 4, we can find that, after the test, the P value of the model is less than 0.05, and the model fitting effect is significant, and the results show that the model has a significant effect on. P indicates that the probability of the problem of the platform, the regression model can be expressed as:

$$P = \frac{EXP(3.706 - 13.848X_2 - 1.930X_8 - 0.035X_{10} - 3.125X_{17})}{1 + EXP(3.706 - 13.848X_2 - 1.930X_8 - 0.035X_{10} - 3.125X_{17})}$$

The model has been subjected to an internal inspection using a training sample, and an external inspection is performed with the previously reserved sample. Bring the July data into the model. The regression value P of each sample was calculated, and the results were classified according to the principle of taking 1 greater than 0.5 and less than 0.5. As shown in table 5.

Table 5: Results of test model

Observed value		Predicted value		Accuracy rate
		0	1	
y	0	263	55	82.7
	1	10	21	67.7
Average prediction accuracy				81.4

6. Conclusion

The basic data collection based on third party rating agencies , Online Lending House, using the method of binary logistic regression model, to find out the important index, which are the average interest rate (x2), the top ten borrowers also amount to be accounted for (x8), operation time the top ten investors(x10), the amount to be received accounted for (x17).Through the establishment of regression model, the prediction accuracy of the training sample 73% and the prediction accuracy of the test sample of 81.4% were obtained.The results show that the model established in this paper can be used to judge whether the P2P network lending platform run into a problem. The research results show that investors choose to invest in P2P network lending platform analysis platform to the attention of the top ten are of interest rate, the borrower still amount to be accounted for, the operation time and the top ten investors accounted for the amount to be received in the 4 important factors, to eliminate the interference caused by the secondary property on the platform.

In this paper, the selected indicators only selected the quantitative indicators, such as the platform's trading volume, platform operation, platform dispersion and so on. Without consideration other effects,so the regression model could not reflect the influence of other aspects of the platform, secondly, whether the P2P network loan problem will happen needs from the two aspects of quantitative and qualitative considerations, such as area and other factors, can further the qualitative index in modeling.

7. Acknowledgement

This project is in part supported by National Science Foundation of China (Grant Number: 71371114).

8. References

- [1] Zaiyan Wei, Mingfeng Lin. Market mechanisms in online peer-to-peer lending. *Management Science*, 2016.
- [2] Mingfeng Lin, Nagpurnan and R. Prabhala, Siva Viswanathan. Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending. *Management Science*. 2016, 59(1):17-35
- [3] Juanjuan Zhang, Peng Liu. Rational Herding in Microloan Markets. *Management Science*. 2016, 58(5):892-912.
- [4] Ping Xie, Chuanwei Zou. Research on Internet Financial Model. *Financial Research*, 2012, 12:11-22.
- [5] Qinke Ye. Research on the problems and countermeasures of P2P network lending platform. *Market Modernization*, 2015, 11:156-157.
- [6] Xiao Chen, Xiaoyu Ding, Beifen Wang. A study of the overdue behaviors in private borrowing. *Finance Forum*, 2013, 11:65-72.
- [7] Yumei Zhang. Research on P2P micro-finance model. *Productivity Research*, 2010, (12):162-165.
- [8] Li Yi. Discussion on the drawbacks and management of P2P network lending platform in China. *Commercial Age*, 2014, (14):65-67.
- [9] Jie Ding, Zhu Ma. P2P network loan model alienation and risk control in China. *New Finance*, 2015, (09):55-59.
- [10] Songqing Wang, Yun Tian, Xia Shen. An analysis on credit risk for P2P online lending mode from credit reference perspective. *Credit Investigation*, 2014, (12):49-52.