Evaluating the risk of the Chinese borrowers' default in peer-to-peer lending: using the Logistic regression model for ordinal response variables

Yusi Lia and Xiaolong Lib*

aDana and David Dornsife College of Letters, Arts and Sciences, University of Southern California, Los Angeles, USA; bSchool of Economics, Peking University, Beijing, China

*Xiaolong Li: tell714@gmail.com, Yusi Li: yusili@usc.com

Evaluating the risk of the Chinese borrowers' default in peer-to-peer lending: using the Logistic regression model for ordinal response variables

The interpersonal lending has appeared since the historical record. The key difference between the new P2P and previous interpersonal lending is that the borrows and lenders no longer need to meet each other before the transition completed. Using data from a P2P lending platform in China, this article explores the P2P load characteristics, evaluate the risk of the Chinese borrowers' default. The article finds that the ordinal categories of the number of late repayment days and the ordinal logistic regression is much more precise than the binary variable and the binary logistic regression when solving this problem. By using the ordinal logistic regression, the article finds that older male and divorced or widowed borrowers who have children and have ever repaid late, with lower education level, lower monthly earning and larger amount of load money are more likely to repay late. The P2P lending platform in China must find ways to attract younger, married borrowers with higher education level, higher monthly earning and great borrowing history.

Keywords: Peer-to-Peer lending; default risk; ordinal logistic regression; ordinal categories

Subject classification codes: G10

I. Introduction

The interpersonal lending has appeared since the historical record. With the emergence of the first online P2P lending platform "Zopa" the new lending model raised attention for the first time in the year 2006 (Hulme & Wright 2006). However, it was Prosper.com, who caused a wave of scientific contributions by making the entire platform's data public in 2007. The key difference between the new P2P and previous interpersonal lending is that the borrows and lenders no longer need to meet each other before the transition completed. (P2P Research Group 2017) Within these platforms borrowers generally describe the purpose of their loan request and provide information about their current financial situation, like income or open credit lines. Lenders then have opportunity to offer a loan with an interest rate derived upon this information. For borrowers, online P2P lending is a way to receive a loan without a financial institution involved in the decision process and might also be a possibility to receive better conditions than in the traditional banking system. For lenders, it can be seen as an investment model where the investment risk is coupled to the credit rating of the funded loans. The platforms themselves often benefit by raising fees for successful realized transactions (Galloway 2009).

Information asymmetry is the fundamental problem in online P2P lending. The challenge is to overcome the principal-agent problem (Jensen & Meckling 1976). While the lender wants to get as much valid information about the borrower as possible, the borrower might be interested in hiding some of his characteristics in order to get an interest rate as low as possible. In order to allow lenders to make an informed decision based on valid information, P2P lending platforms force their borrowers to provide financial information that have been validated by external agencies. Additionally, many platforms demand users to supply demographic information, like gender, race or age. Borrowers are also often given the opportunity to provide social information, which cannot be validated, like hobbies, the family background or a photo. We call these characteristics determinants of P2P lending, since they have major influence on the successful funding of a borrower's loan-listing and the demanded interest-rate. (Alexander Bachmann 2011) Some platforms like prosper.com provide additional financial information about their borrowers like current open credit lines or bankcard utilization (Klafft 2008).

As for the determinants in P2P lending, Alexander Bachmann divide them into financial characteristics, also called hard-factors, and soft-factors like demographic characteristics and group intermediation. We conclude with ideas for future research in online P2P lending. (Alexander Bachmann 2011; Freedman & Jin 2008) reveal in their study that the average funding rate on ***prosper.com*** rose from 8.51% in the time-period 11/2005 to 03/2007 to 10.14% between 06/2006 and 07/2008. They assume that the higher funding rate is a result of the improved information that ***prosper.com*** provides to its lenders (on February 12, 2007, prosper.com added more detailed financial information about the borrower and the possibility that borrowers report their current income, employment-status and occupation). Research shows that discrimination based on demographic characteristics other than race have only little impact on the likelihood of funding and interest rates (Herzenstein et al. 2008; Pope & Sydnor 2008; Ravina 2007). The borrower's race can be an important determinant in P2P lending. Pope & Sydnor (2008) show that the chances of African American loan listings to get fully funded are 25 to 34 percent smaller than those of whites with similar credit ratings. These findings are confirmed by Herzenstein et al. (2008) Pope & Sydnor (2008) analyze the effects of the borrowers' age on funding success. Compared to a base group of 35-60 year olds, there is a 40 to 90 basis points higher chance of getting funded for those who appear younger than 35. Those who appear to be 60 years and older are between 1.1 and 2.3 percentage points less likely to succeed in acquiring a loan. Barasinska (2009) investigates the question if the lenders' gender is relevant for return and risk characteristics of the loan. To her surprise, she finds that female lenders are less risk-averse than male lenders.

At present, two of the biggest P2P platforms are ***prosper*** and ***the lending Club***. Although the Chinese P2P lending platform started later than those of western developed countries, the gross trading amount of China has been the global number one since 2014. The gross trading amount of the Chinese P2P lending platform has been $53,760,000,000 in the 2014, while that of America was $8,800,000,000, that of England was $1,900,000,000. The emerging of the P2P lending platform offered a low-cost and convenient platform to the small amount lenders and borrower. One of the technology requirements for the P2P lending platforms to develop healthily and fast is that they should acquire the related information to evaluate the credit risk of the borrowers in order to manage the credit risk of borrowers. Compared with the risk management system of Chinese P2P lending platform, the western mature credit risk rating system and the FICO score standard provide the references for the development of the risk management system of Chinese P2P lending platform.

Chinese seemed rigorous laws and large number of the national stated banks may not be a good place to lend to the financial creativity. However, these characteristics are exactly why Chinese Peer-to-peer leading market increased rapidly. Chinese entrepreneurs made the Chinese Peer-to-peer leading market have the first history breaking increase by using the grey area of the law. (P2P Research Group 2017) In 2015, the Chinese largest Peer-to-peer lending platforms are ***Shanghai Lujiazui International Financial Asset Exchange***, ***Hongling Capital***, ***Wealth Evolution***. Until August, 2016, "The management interim measures of network lending information intermediary activity" was published, the peer-to-peer lending platform began to develop standardly. Then in the next year of 2016, Chinese largest Peer-to-peer lending platforms are ***Shanghai Lujiazui International Financial Asset Exchange***, ***Hongling Capital*** and ***CreditEast***. The peer-to-peer leading platform in China began to shuffle. The development status of the Chinese Peer-to-peer lending platforms is shown in Table 1.

In generally, the default of the Peer-to-peer leading platform has two kinds: First is the voluntary default and the second is the negative default. The voluntary default is due to the repaying capability and repaying willingness of individuals. The negative default is due to the macroeconomic environment, including the unemployment caused by economic depression, the systematic risk of peer-to-peer lending platform, the imperfect of the domestic credit system, the volatility of financial markets. Another significant factor which can lead to default is the "free-rider" mentality of the lenders. Because the load amount is always small, once a borrower repays late, some of the lenders would rely on those lenders with large loss. Generally, if a borrower repays late, the platform would phone to the borrower. In the severe cases, the platform would open the borrower's personal information. To the peer-to-peer lending platform which want to develop healthily, it must have the risk pricing ability and make the estimation and the prediction of the probability of the default borrower crowds.

In America, the Peer-to-peer lending platforms always use the credit rating and FICO score to estimate the probability of the default. Among them, ***Proster*** is the simple intermediary model, both the borrowers and lenders have the completely autonomous choice. The ***lending club*** connect the Facebook to use the effective information to recognize the suitable borrowers and constraint the borrowers except using the credit rating system. ***Zopa*** in England uses the intermediary and regulation system and make the rigorous credit system to the borrowers. However, no matter in the peer- to-peer lending market or in the traditional banks, securities market, China doesn't have the perfect financial credit rating system. At present, most of the Chinese peer-to-peer lending platforms use the personal information and the default history of the borrowers to assess the risk of default. This article use the ordinal logistic regression to help the Chinese peer-to-peer lending platforms to find the better borrowers much more precisely. In the meantime, some emerging peer-to-peer platforms in China have began to use the FICO score system to estimate the probability of default. In the future, using the big data to establish the risk management model and decision making engine is the development trend of the risk estimation system of the peer-to-peer lending platforms.

When it comes to the factors that influence the possibility of loan default. In the past, most study choose the binary variable and the binary logistic regression. However, ordinal categories are common in research situations where the assignment of numbers representing successive categories of an attribute, construct, or behavior coincides with meaningful directional differences. Knapp (1999) used ordinal ratings to assess severity of illness with scale categories such as mild (1), moderate (2), and severe (3). The primary characteristic of ordinal data is that the numbers assigned to successive categories of the variable being measured represent differences in magnitude, or a "greater than" or "less than" quality (Stevens 1946). Some examples of ordinal data include rubrics for scaling open-ended writing responses or essays and the solutions to arithmetic problems for which responses are scored based on improving levels of quality (e.g., 0 = poor, 1 = acceptable, 2 = excellent). In contrast, nominal-level data occur when the numeric values used to measure a variable simply identify distinct qualitative differences between categories (i.e., gender as 1 = male or 2 = female; geographic description of school attended as 1 = rural, 2 = urban, 3 = suburban, etc.); nominal data do not possess the directional characteristics of ordinal data. On the other hand, variables measured on an interval- level or ratio-level scale do use scale values to indicate the "greater than" or "less than" quality of ordinal variables but in addition maintain a property of equal-distance or equal-interval length between adjacent values across the scale.

The choice of numbers used to represent the progressively more severe categories conveniently preserves the "greater than" or "less than" quality of the underlying attribute defining the categories themselves. If the value of 3 represents a state that is more critical than the state represented by the value 2, and the value 2 represents a state more critical than the condition represented by the value 1, then the property of transitivity implies that the condition represented by the value of 3 is also more critical than the condition represented by the value of 1 (Cliff & Keats 2003; Krantz, Luce, Suppes, & Tversky 1971). When the possible responses for an outcome variable consist of more than two categories and are ordinal in nature, the notion of "success" can be conceived of in many different ways. Regression models for ordinal response variables are designed for just this situation and are extensions of the logistic regression model for dichotomous data. (Ann A. O'Connell 2006)

The remainder of the article is organized as follows. In the next section, the article describes our data and summarizes the descriptive statistics of online P2P in China. In Section III, we present the descriptions of methodologies and empirical results for evaluating and predict the risk of a borrower's default by using the personal information and borrow information in history. Section IV discusses how each variable influences the risk of a borrower's default. The last section makes the conclusion and provide the suggestion to the peer-to-peer lending club in China.

II. Data

In this section, the article describes and summarizes the descriptive statistics of the data used in our study, including the personal information of the borrowers and the borrow information in history. This article uses 52017 borrowers from 1 January 2015 to 31 May 2015 obtained from a P2P lending platform in China. Through data preprocessing and data screening, we excluded the cases of P2P borrowers which lack the vital information and eventually we get 49,449 valid cases of P2P borrowers.

According to the available data, the explained variable is the risk of a borrower's default and the explanatory variables of each cases include Gender, Geography status, Marital Status, Children status, Educational level, Monthly income, Company scale, Working years, Age, Lending periods, Load amount and Default history. In particular, we define the explained variables (the risk of a borrower's default) and some of the explanatory variables (Gender, Geography status, Marital Status, Children status, Educational level and Default history) as ordinal categories. The detailed definition of variables is as Table 2.

Based on the sample of the 49,449 borrowers, more than 75% the borrowers repay on time; late payment within 7 days, late payment from 7 to 32 days and late payment from 32 to 62 days cases all together are more than 9%; late payment over 62 days cases are less than 1%. The male borrowers are more than female borrowers. More than 99% of the borrowers are live in the western China, which means this explanatory variable is unrepresentative. More than 25% of the borrowers are married, more than 45% of the borrowers are single and more than 1% of the borrowers are divorced or widowed. More than 75% of the borrowers don't have the children. More than 9% of the borrowers have secondary school degree; more than 25% of the borrowers have high school degree; less than 25% of the borrowers have technical secondary school degree; more than 15% of the borrowers have junior college degree; more than 4% of the borrowers have the bachelor degree; less than 1% of the borrowers have elementary school degree, master degree and doctor degree. More than 95% of the borrowers have the monthly earning 7000RMB or less than 7000RMB. The largest amount of the borrowers is 200,000RMB. More than 90% of the borrowers are in the companies which have less than 200 employees; more than 5% of the borrowers are in the companies which have 1000 employees. More than 95% of the borrowers have worked 6 or less than 6 years. Half of the borrowers are 24 years old or younger than 24 years old. More than 10% of the borrowers' load period is 6 months and 24 months separately. Other load periods include 9, 12, 15, 18 and 21 months. More than 99% of the load amount is less than 5888RMB. Less than 10% of the borrowers have ever repaid late. The descriptive statistics of the loan data used in this study is shown in Table 3, including the general characteristics of borrowers and loans.

III. Empirical results

In this section, the article explores the factors that influence the possibility of loan default. In order to evaluate and predict the risk of a borrower's default by using the personal information and borrow information in history, we choose the logistic regression model for ordinal response variables. Some of statistics in our study, for which the explained variable (on time payment, late payment within 7 days, late payment from 7 to 32 days, late payment from 32 to 62 days, late payment over 62 days) and some of the explanatory variables (such as Marital Status and Educational level), although not a Likert-type scale1, is never the less ordinal. We could estimate a linear regression model or classical logistic regression. There are, however, some obvious problems. First and foremost, classical linear regression

1 The typical Likert-type scale has five categories (e.g., strongly disagree, disagree, undecided, agree, strongly agree) to gauge one's response to a question, though it may have anywhere between three and seven or more response categories.

assumes a continuous dependent variable with equally spaced, ordered response categories. A Likert-type scale, or any other ordinal scale, is, albeit ordered, not necessarily equally spaced between categories. Second, and perhaps more important, such a scale would not give the normal distribution that the classical linear regression or classical logistic regression assumes the data to display. As the Geography status is not representative and the load period is not a prominent variable, the explanatory variables included in the Logistic regression model for ordinal response variables are Gender, Marital Status, Children status, Educational level, Monthly income, Company scale, Working years, Age, Load amount and Default history.

（1）Logistic regression model for ordinal response variables

Assume that we have an ordinal scales variable constituted by the J Category Y (Y=1, …, J),

$$L_j(X) = \text{logit}\left[F_j(X)\right], (j = 1, \dots J - 1) \tag{1}$$

$$= \log\left[\frac{P(Y \le j|X)}{P(Y > j|X)}\right] \tag{2}$$

$$= \log\left\{\frac{P(Y \le j|X)}{[1 - P(Y > j|X)]}\right\} \tag{3}$$

Among which, $F_j(X) = P(Y \le j|X)$ is the cumulative probability function (c.d.f.) of the J Category.

If Y is independent from X, then

$$L_j(X) = \alpha_j \tag{4}$$

If not, then

$$L_j(X) = \alpha_j - \beta X \tag{5}$$

With different X, for instance X_1 and X_2, then

$$L_j(X_1) - L_j(X_2) = \beta(X_2 - X_1) \tag{6}$$

To certain explained variable ($\leq$ j), the possibility of explaining variable X_1 to the possibility of explaining variable X_1 is $\exp^{[\beta(X_2 - X_1)]}$.

In our study, $Y_i = 1$ (on time payment), if $-\infty < Y_j^* \leq \alpha_1$; $Y_i = 2$ (late payment within 7 days), if $\alpha_1 < Y_j^* \leq \alpha_2$; $Y_i = 3$ (late payment from 7 to 32 days), if $\alpha_2 < Y_j^* \leq \alpha_3$; $Y_i = 4$ (late payment from 32 to 62 days), if $\alpha_3 < Y_j^* \leq \alpha_4$; $Y_i = 5$ (late payment over 62 days), if $\alpha_4 < Y_j^* \leq \alpha_5$; Here, $\alpha_1 < \alpha_2 < \alpha_3 < \alpha_4 < \alpha_5$ is the cut off points.

$$Y^* = \beta X + \alpha \tag{7}$$

$$P(Y = j|X) = F(\alpha_1 - \beta X) - F(\alpha_{j-1} - \beta X) \tag{8}$$

Using the OLM Model in STATA, the result is as Table 4.

The estimated value of y: when $y^* \leq 2.192$(cut1), y=1 (on time payment); When $2.192 < y^* \leq 2.718$(cut2), y = 2 (late payment within 7 days); When $2.718 < y^* \leq 3.884$(cut3), y = 3 (late payment from 7 to 32 days); When $3.884 < y^* \leq 5.298$ (cut4), y = 4 (late payment from 32 to 62 days); When $y^* > 5.298$ (cut4), y = 5 (late payment over 62 days).

Add other explanatory variables into the model: Gender (X1), Marital Status (X3), Children status (X4), Educational level (X5), Monthly income (X6), Company scale (X7), Working years (X8), Age (X9), Load amount (X11), Default history (X12). Using the OLM Model in STATA, the result is as Table 5.

IV. Discussion

(1) Men have the more possibility than women to repay late. If other variables are constant and the borrower is woman rather than man, compared with the possibility of late payment over 62 days, the possibility of on time payment, late payment within 7 days, late payment from 7 to 32 days and late payment from 32 to 62 days all increase by 121%.

(2) Divorced or widowed borrowers are more likely to repay late than single borrowers or married borrowers. The single borrowers are more likely to repay late than married borrowers. If other variables are constant and the borrower is single rather than married, compared with the possibility of on time payment, the possibility of late payment within 7 days, late payment from 7 to 32 days, late payment from 32 to 62 days and the possibility of late payment over 62 days all decrease by 19%.

(3) The borrowers who have children have the more possibility to repay late than those who don't have children. If other variables are constant and the borrower have children rather than don't have children, compared with the possibility of on time payment, the possibility of late payment within 7 days, late payment from 7 to 32 days, late payment from 32 to 62 days and the possibility of late payment over 62 days all decrease by 15%.

(4) When the education level of the borrowers increases, the possibility of late payment decreases. If other variables are constant and the education level of the borrowers increases by one level, compared with the possibility of late payment over 62 days, the possibility of on time payment, late payment within 7 days, late payment from 7 to 32 days and late payment from 32 to 62 days all increase by 17%.

(5) When the working year of the borrowers increases, the possibility of late payment decreases. If other variables are constant and the working year of the borrowers increases by one year, compared with the possibility of late payment over 62 days, the possibility of on time

payment, late payment within 7 days, late payment from 7 to 32 days and late payment from 32 to 62 days all increase by 3%. The probability predictions graph of working years is as Figure 1.

(6) When age of the borrowers increases, the possibility of late payment increases. If other variables are constant and the age of the borrowers increases by one year, compared with the possibility of on time payment, the possibility of late payment within 7 days, late payment from 7 to 32 days, late payment from 32 to 62 days and the possibility of late payment over 62 days all decrease by 4%. The probability predictions graph of age is as Figure 2.

(7) When load amount of the borrowers increases, the possibility of late payment increases. If other variables are constant and the load amount of the borrowers increases by 100 RMB, compared with the possibility of on time payment, the possibility of late payment within 7 days, late payment from 7 to 32 days, late payment from 32 to 62 days and the possibility of late payment over 62 days all decrease by 3%. The probability predictions graph of load amount is as Figure 3.

(8) The borrowers who have ever repaid late have the more possibility than the borrowers who have never repaid late. If other variables are constant and the borrower have ever repaid late rather than have never repaid late, compared with the possibility of late payment over 62 days, the possibility of on time payment, late payment within 7 days, late payment from 7 to 32 days and late payment from 32 to 62 days all increase by 3607%.

(9) The estimated value of y: when $y^* \leq -9.21(\text{cut1})$, y=1 (on time payment); When -9.21 $< y^* \leq$ -8.39(cut2), y = 2 (late payment within 7 days); When -8.39 $< y^* \leq -6.83(\text{cut3})$, y = 3 (late payment from 7 to 32 days); When -6.83 $< y^* \leq$ -5.30 (cut4), y = 4 (late payment from 32 to 62 days); When $y^* > -5.30$ (cut4), y = 5 (late payment over 62 days)

V. Conclusion

Online P2P lending has gained scientific relevance over the past years. The availability of data about markets and transactions allows researchers from different disciplines to investigate the various determinants that play a role in the process of funding. (Alexander Bachmann 2011)

As for the risk management of P2P platforms in China, P2P platforms have extended different amount of advancement in 2015. The amount of advancement provided by P2P platform has been stable from Q4 2014 to Q1 2015. Some P2P platforms have set up risk reserve funds. The risk reserve fund balance of four was less than 4%of their total outstanding loans, and one platform's reserve fund balance reached approximately 10%. No matter advancements or risk reserve funds, they indicate P2P platforms are sharing credit risks and are functioning as banks rather than as information intermediaries. Since 2014 there has been a gradual increase in the amount of overdue loans. The absolute amount and growth rate of overdue loans have increased rapidly. Because of different business models, mature structures for overdue loans on different platforms are also difficult. For example, the majority of platform 10's overdue loans have terms of approximately 90-180 says; however, the majority of overdue loans of platform 12 and 16 have term lengths of 30 days or less. Overall, the vast majority of overdue loans on the four platforms had a maturity of less than 180 days. (P2P Research Group 2017)

As an information intermediary between borrowers and lenders, whether or not a P2P platform has access to and distributes client funds is seen as an important sign to determine if the platform's operations are in compliance with rules and regulations. The article employs the data to evaluate the default risk of the peer-to-peer lending platforms by using the ordinal logistic regression and ordinal categories. The study finds that gender, marital status, children status, educational level, monthly income, company scale, working years, age, load amount and the default history have the significant influence on the risk of a borrower's default. Older male and divorced or widowed borrowers who have children and have ever repaid late, with lower education level, lower monthly earning and larger amount of load money. While the findings indicated that characteristics of borrowers with low default risk are female gender, young adults, long working time, stable marital status, high educational level, working in large company, low monthly payment, low loan amount, low debt to income ratio and no default history. (Xuchen Lin, Xiaolong Li & Zhong Zheng 2017) The peer-to-peer lending platforms in China must find ways to attract younger, married borrowers with higher education level, higher monthly earning, great borrowing history.

On the basis of the survey, more than 60% of P2P platforms in China are using fund custody services. The kind of institutions that can provide custody services for the platforms include third party payment institution (57%), banks (36%) and asset management companies (7%). Over half of the platforms use a third party payment institution as a custodian. Furthermore, over half of P2P platforms have introduced fund guarantor. There is a variety of guarantor institutions, including guarantee companies, micro loan companies, investment management and consulting companies, as well as pawnbrokers, factoring companies, and financial leasing companies. There is no major difference in the interest rate level between platforms that have not introduced a guarantor and those that have. (P2P Research Group 2017) Compared with the peer-to-peer lending platforms in America and UK such as Prosper, Lending Club and Zapo, which have the perfect credit rating system, Chinese peer-to-peer lending platform should gradually establish a more mature credit rating system. borrowers' credit grade is an indicator of financial strength and ability to repay the loan, which is an important input into the decisions of Prosper lenders. Similarly, other financial indicators, demographic characteristics, and effort measures may sway lenders to bid on one auction and not the other even when the auction decision variables are equal. (MICHAL HERZENSTEIN 2008) To establish a much more scientific and mature risk management system has been the significant factor of the Chinese peer-to-peer lending platform to develop.

VI. Appendices

Table 1. Development status of the Chinese Peer-to-peer lending platforms

Year	Number of platform	Year-on-year growth （%）	Amount of transition （Unit: 100,000,000RMB）	Yea-on-year growth rate （%）
2007	1	-	-	-
2008	1	0	<0.1	-
2009	5	400	1.5	-
2010	15	200	13.7	813.3
2011	50	233	84.2	514.6
2012	148	196	228.6	171.5
2013	523	253	897.1	292.4
2014	1,942	271	2528	181.8
2015	2,769	94	6381	152.4

(Statistics from www.wdzj.com and www.ce.com)

Table 2. Definition of variables

Dimension of variables	Name of variables	Definition and assignment of variables
explained variable	The risk of a borrower's default (Y)	Use the number of late payment days to evaluate the risk of a borrower's default: on time payment is 1; late payment within 7 days is 2; late payment from 7 to 32 days is 3; late payment from 32 to 62 days is 4; late payment over 62 days is 5.
explanatory variables	Gender (X1)	Male is 1; female is 2.

	Geography status (X2)	Eastern China is 1; central China is 2; western China is 3.
	Marital Status (X3)	Married is 1; single is 2; divorced or widowed is 3.
	Children status (X4)	Don't have child is 1; have child is 2.
	Educational level (X5)	Elementary school is 1; secondary school is 2; high school is 3; technical secondary school is 4; junior college is 5; bachelor is 6; master is 7; doctor is 8.
	Monthly income (X6)	The earning of a borrower monthly. (Unit: RMB)
	Company scale (X7)	The number of the employees in a borrower's company.
	Working years (X8)	The number of the years a borrower has worked.
	Age (X9)	The age of a borrower.
	Load periods (X10)	The number of the lending months.
	Load amount (X11)	The amount of lending money.
	Default history (X12)	Ever late payment over 7 days in the past is 1; never late payment over 7 days in the past is 2.

Note: Eastern China includes Beijing, Shanghai, Tianjin, Fujian, Guangdong, Hainan, Hebei, Jiangsu, Liaoning, Shandong, Zhejiang; central China includes Anhui, Henan, Heilongjiang, Hubei, Hunan, Jilin, Jiangxi, Shanxi; western China includes Gansu, Guangxi, Guangzhou, Ningxia, Shanxi, Sichuan, Xizang, Xinjiang, Yunnan, Chongqing.

Table 3. Descriptive statistics

Name of variables	N	Min	Max	Mean	SD	Variance	Skewness	Kurtosis

The risk of a borrower's default (Y)	49,448	1	5	1.188	0.622	0.388	3.646	16.512
Gender (X1)	49,448	1	2	1.367	0.482	0.232	0.550	1.302
Geography status (X2)	49,448	1	3	2.997	0.078	0.006	-24.528	613.051
Marital Status (X3)	49,448	1	3	1.762	0.521	0.272	-0.216	2.760
Children status (X4)	49,448	1	2	1.242	0.429	0.184	1.202	2.445
Educational level (X5)	49,448	1	8	3.604	1.280	1.638	0.220	1.975
Monthly income (X6)	49,448	0	200,000	3,706.632	2,791.742	7,993,825	19.426	819.125
Company scale (X7)	49,448	5	1,000	113.106	235.244	55,339.53	3.044	11.300
Working years (X8)	49,448	0	83	1.842	3.270	10.695	7.618	112.308
Age (X9)	49,448	16	55	25.026	6.077	36.930	1.579	5.890
Load periods (X10)	49,448	6	24	12.830	5.186	26.900	0.867	3.233
Load amount (X11)	49,448	800	10,000	3,471.173	1,050.903	1,104,397	0.110	3.085
Default history (X12)	49,448	1	2	1.858	0.349	0.121	-2.057	5.231

Table 4. Ordinal logistic regression results of the explained variable

y	Coef.	Std. Err.	[95% Conf. In > terval]
Cut 1	2.192	0.015	2.162703 0.221341
Cut 2	2.718	0.019	2.681017 0.754146
Cut 3	3.884	0.032	3.820849 0.946266

Cut 4	5.298	0.064	5.173083 0.423633

Table 5. Ordinal logistic regression results

Y	β	Std. Err.	[95% Conf. In > terval]	exp$^{(-\beta)}$
X1	-0.7917***	0.0416	-0.8736 -0.7099	2.2071
X3	0.2140***	0.0626	0.0913 0.3366	0.8073
X4	0.1655*	0.0884	-0.0078 0.3387	0.8475
X5	-0.1571***	0.0147	-0.1859 -0.1284	1.1701
X6	-0.0006***	0.0000199	-0.0006 -0.0006	1.0006
X7	-0.000821	0.0000761	-0.0002 -0.0000672	1.0008
X8	-0.0284***	0.0069	-0.0419 -0.0150	1.0288
X9	0.0375***	0.0036	0.0305 0.0446	0.9632
X11	0.0003***	0.0000176	0.0002 0.0003	0.9997
X12	-3.6127***	0.0380	-3.6873 -3.5382	37.0660
Cut1	-5.0346	0.2319	-5.4891 -4.5800	/
Cut2	-4.2168	0.2310	-4.6696 -3.7639	/
Cut3	-2.6581	0.2314	-3.1117 -2.2046	/
Cut4	-1.1305	0.2378	-1.5967 -0.6644	/

Note: * represents significance at the 10% level, and *** represents significance at the 1% level.

Figure 1. The probability predictions graph of working years

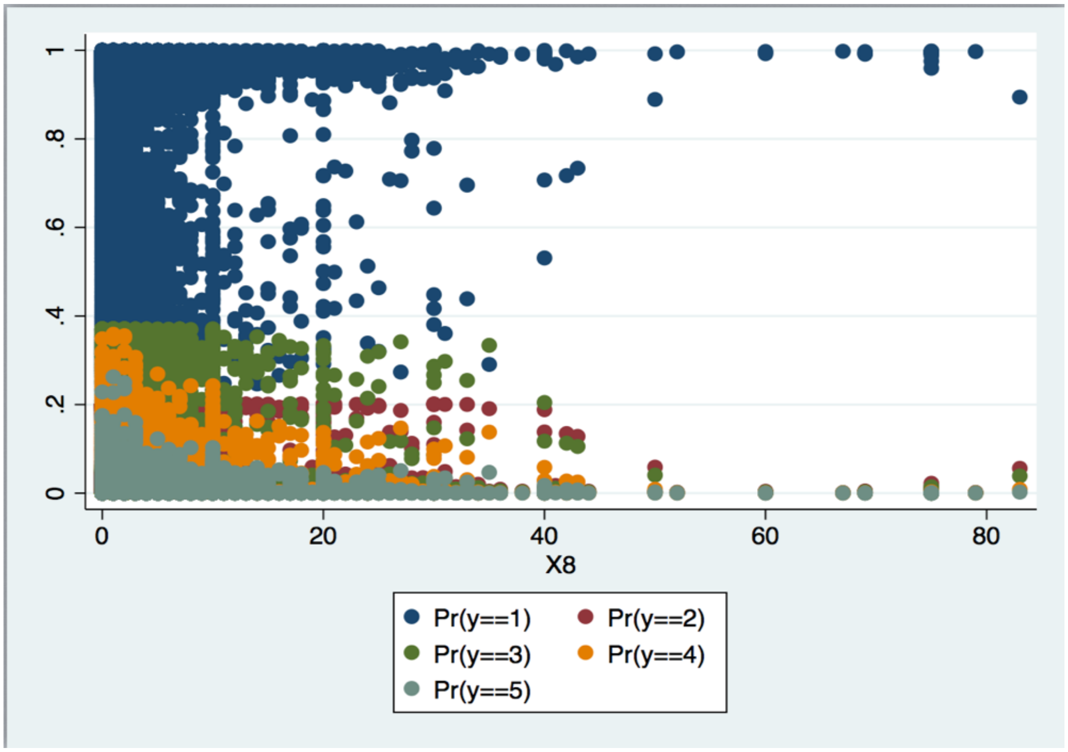


Figure 2. The probability predictions graph of age

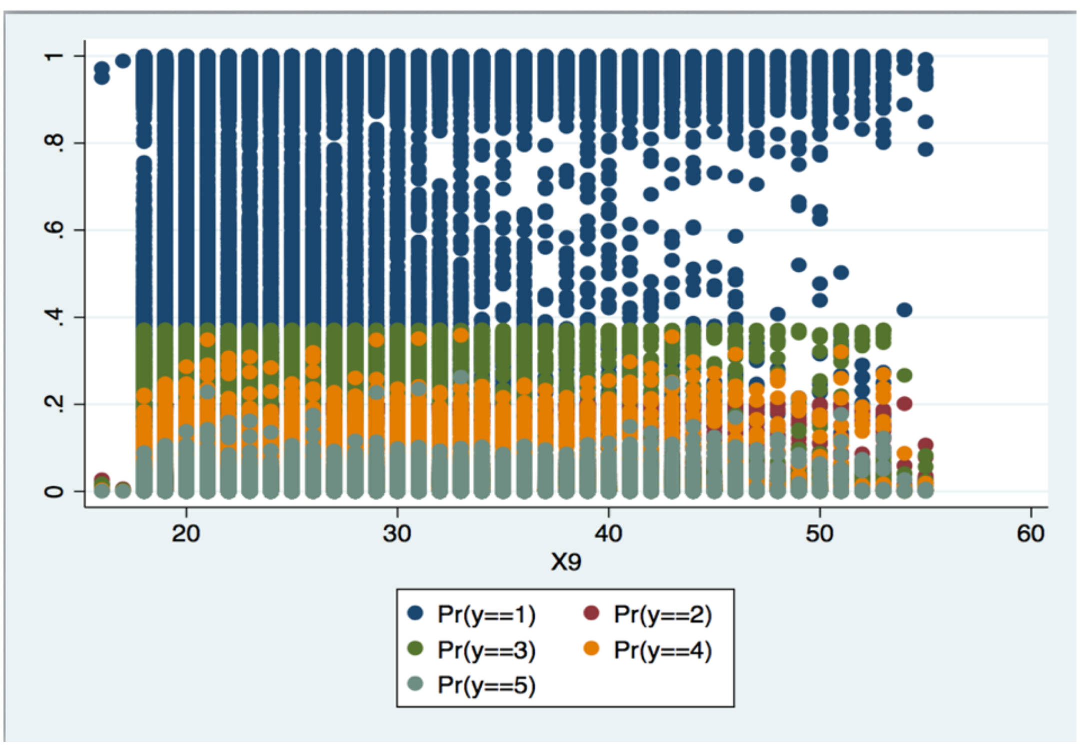
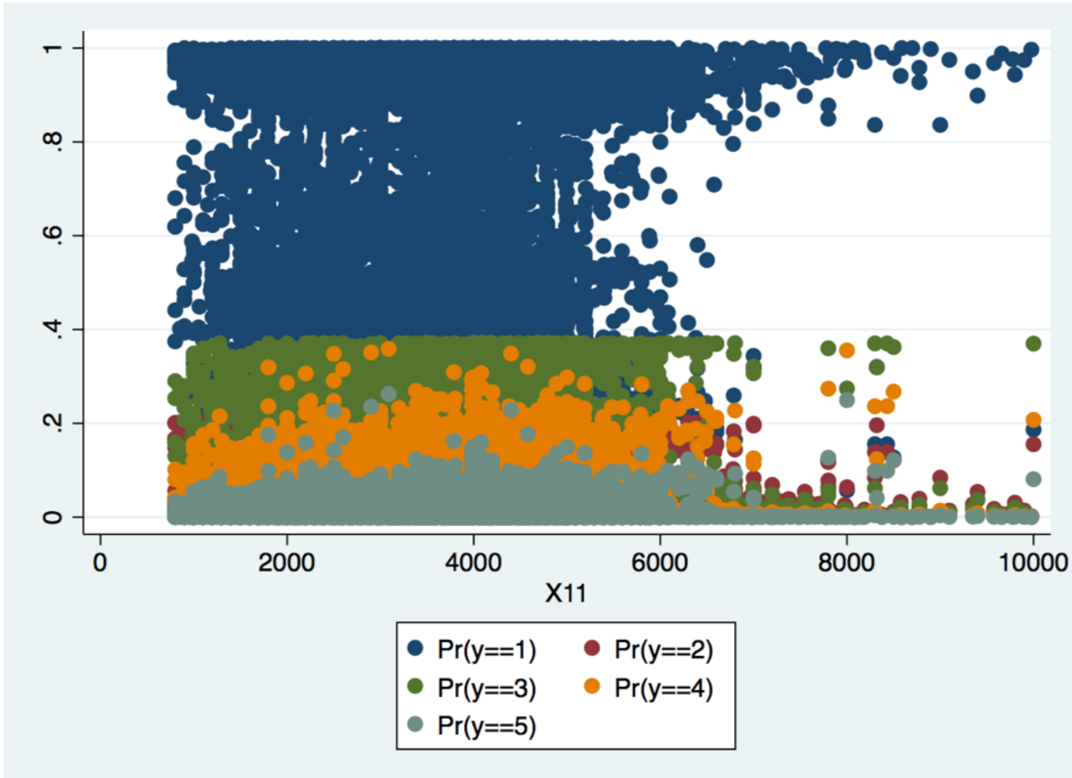


Figure 3. The probability predictions graph of load amount



Acknowledgements: We would like to express my gratitude to all those who helped us during the writing of this thesis. We gratefully acknowledge the help of Zhong Zheng, who provided the data. We also owe a special debt of gratitude to Xuchen Lin, who cleaned the data.

References

O'CONNELL, A. *LOGISTIC REGRESSION MODELS FOR ORDINAL RESPONSE VARIABLES* (1 ed., Vol. 121, pp. 1-121). London, United Kingdom: International Educational and Professional Publisher, 2008.

P. (2006). *PEER_TO_PEER LENDING EITH CHINESE CHARACTERISTICS* (1 ed., Vol. 121, pp. 1-99). Shanghai, China: Shanghai Finance Institute, 2006.

Xuchen Lin, Xiaolong Li & Zhong Zheng X. “Evaluating borrower's default risk in peer-to-peer lending: evidence from a lending platform in China.” *Applied Economics* 8 (2006): 1-8.

Jacobs, Jr., M. “An empirical study of the return on defaulted debt.” *The Journal of Financial Markets* 30 (2010): 1-30.

Riza Emekter, Yanbin Tu, Benjamas Jirasakuldech & Min Lu. “Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending.” *Applied Economics* 17 (2014): 54-70.

Dimitris Gavalasa & Theodore Syriopoulos. “Maritime Policy & Management: The flagship journal of international shipping and port research.” *Maritime Policy & Management* 22 (2014): 1-22.

Yanhong Guo, Wenjun Zhou, Chunyu Luo, Chuanren Liu, Hui Xiong. “Instance-Based Credit Risk Assessment for Investment Decisions in P2P Lending.” *European Journal of Operational Research* 33 (2015): 1-33.

Lili Wang a, b, Wei Lu a, Naresh K. Malhotra. “Demographics, attitude, personality and credit card features correlate with credit card debt: A view from China.” *Journal of Economic Psychology* 15 (2010):179-193.

Shun Cai, Xi Lin, Di Xu, Xin Fu. “Judging online peer-to-peer lending behavior: A comparison of first-time and repeated borrowing requests.” *Information & Management* (2016): 858-867.

Mingfeng Lin, Nagpurnanand R. Prabhala, Siva Viswanathan. “Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending.” *MANAGEMENT SCIENCE* 19 (2013): 17-35.

QINGYAO WAN, DONGYU CHEN, WEIHUA SHI. "ONLINE PEER-TO-PEER LENDING DECISION MAKING: MODEL DEVELOPMENT AND TESTING". *SOCIAL BEHAVIOR AND PERSONALITY* 44, 1 (2016): 117-130.

Lauri Puro, Jeffrey E. Teich, Hannele Wallenius, Jyrki Wallenius. "Borrower Decision Aid for people-to-people lending." *Decision Support Systems* 49, 52-60 (2010): 53-60.

Hulme, M. K., & Wright, C. "Internet Based Social Lending: Past, Present and Future." *Social Futures Observatory*, (2006).

Freedman, S., & Jin, G.Z. "Do Social Networks Solve Information Problems for Peer-to-Peer Lending? Evidence from Prosper. com." Working Papers. College Park, MD: *NET Institute* (2008)

Galloway, I. "Peer-to-Peer Lending and Community Development Finance." Community Development Investment Center Working Paper. San Francisco: Federal Reserve Bank of San Francisco (2009).

Jensen, M. C., & Meckling, W. H. "Theory of the Firm : Managerial Behavior , Agency Costs and Ownership Structure." *Journal of Financial Economics* 3, 4 (2009): 305- 360.

Klafft, M. "Peer to Peer Lending: Auctioning Microcredits over the Internet." Proceedings of the 2008 Int'l Conference on Information Systems, Technology and Management. Dubai: IMT (2008): 1-8

Herzenstein, M., Andrews, R. L., Dholakia, U. M., & Lyandres, E. "The Democratization Of Personal Consumer Loans? Determinants Of Success In Online Peer-To-Peer Lending Communities." Online. Newark, DE ; Houston, TX. Retrieved from http://www.prosper.com/downloads/research/democratization- consumer-loans.pdf (2008).

Pope, D. G., & Sydnor, J. R. "What's in a Picture? Evidence of Discrimination from Prosper. com." *Journal of Human Resources*. Philadelphia, PA (2008)

Barasinska, N. "The Role of Gender in Lending Business: Evidence from an Online Market for Peer-to-Peer Lending." *The New York Times*. Berlin (2009).