

Article

P2P Network Lending, Loss Given Default and Credit Risks

Guangyou Zhou ¹ , Yijia Zhang ¹ and Sumei Luo ^{2,*}

¹ School of Economics, Fudan University, Shanghai 200433, China; zgy@fudan.edu.cn (G.Z.); 16210680104@fudan.edu.cn (Y.Z.)

² School of Finance, Shanghai University of Finance and Economics, Shanghai 200433, China

* Correspondence: luosumei@shufe.edu.cn; Tel.: +86-21-6590-8396

Received: 1 February 2018; Accepted: 25 March 2018; Published: 29 March 2018



Abstract: Peer-to-peer (P2P) network lending is a new mode of internet finance that still holds credit risk as its main risk. According to the internal rating method of the New Basel Accord, in addition to the probability of default, loss given default is also one of the important indicators of evaluation credit risks. Proceeding from the perspective of loss given default (LGD), this paper conducts an empirical study on the probability distribution of LGDs of P2P as well as its influencing factors with the transaction data of Lending Club. The results show that: (1) the LGDs of P2P loans presents an obvious unimodal distribution, the peak value is relatively high and tends to concentrate with the decrease of the borrower's credit rating, indicating that the distribution of LGDs of P2P lending is similar to that of unsecured bonds; (2) The total asset of the borrower has no significant impact on LGD, the credit rating and the debt-to-income ratio exert a significant negative impact, while the term and amount of the loan produce a relatively strong positive impact. Therefore, when evaluating the borrower's repayment ability, it is required to pay more attention to its assets structure rather than the size of its total assets. When carrying out risk control for the P2P platform, it is necessary to give priority to the control of default rate.

Keywords: P2P network lending; LGD; credit risks; Lending Club

1. Introduction

Since 2011, with the rapid development of internet finance in China, peer-to-peer (P2P) network lending has also sprung up suddenly, giving rise to Alibaba, Suning Micro-finance, and a large number of P2P lending platforms. P2P is the main form of network lending, and, according to the “Guiding Opinions on Promoting the Healthy Development of Internet Finance” released by The People's Bank of China and other ten ministries and commissions on 18 July 2015, network lending includes individual network lending (i.e., P2P network lending) and online network microfinance. Individual network lending refers to the direct lending between individuals through the internet platform. Network microfinance refers to the petty loans provided by internet enterprises to customers by making use of the internet through the microfinance companies under their control. As intermediary organs, service agencies of network lending provide information release, risk assessment, credit consulting, transaction management, and customer services for lending activities, and offer services to the borrower and the lender so as to obtain service charges. P2P gives play to the advantages of the internet and directly connects to the borrower and the lender without the media of commercial banks, which greatly reduces transaction costs and meets the needs of China's current economic development to a large extent.

P2P lending, since its emergence, has kept a sound momentum of development. In 2005, Zopa, the first network lending website, was born in Britain. In 2007, Prosper, America's first P2P lending website, was founded. Since then, P2P platforms have sprung up like mushrooms around the world.

Due to the imperfect financial system, China is facing a more serious phenomenon of credit rationing, there is a strong demand for financing and lending in society, and traditional commercial banks find it hard to meet this demand. Under this historical background, China's P2P platforms have ushered in an explosive growth. According to the data of "Annual Report of China's Network Lending Industry in 2017" released by WDZJ, by the end of December 2017, the number of normal operating platforms of the network lending industry has reached up to 1931, a decrease of 517 compared with the end of 2016, and the number of normal operating platforms throughout the whole year had been declining unilaterally. Since the process of platform rectification has not yet been completed, it is expected that the number of operating platforms of the network lending industry will continue to decline further in 2018, and the specific rate of decline depends on the filing and compliance. From the current information, it is estimated that the number of operating platforms of the network lending industry may fall to about 800 by the end of 2018. In the meantime, the rectification process of the network lending industry in 2017 has come to the ending stage, the number of platforms exiting the industry has dropped to a substantial extent compared with that in 2016, the number of platforms that suspend business and have problems in is 645 in 2017 while the number of platforms, in that case, is 1713 in 2016. The ratio of the number of problematic platforms continues to decrease, the number of problematic platforms in 2017 only takes up 33.49%, and 66.51% of the platforms choice to benignly withdraw. All the above data indicate that the regulation of China's network lending industry is highly effective and fruitful, and the industry development environment will be increasingly healthy in the future. Thus, it can be seen that China's online lending industry has moved from the stage of "savage development" to the new stage of "standardized development".

Due to the lack of regulation, domestic P2P faces enormous legal risks and platform risks, and at the same time, as the credit business, P2P is also confronted with the credit risks due to the default of the borrower, but this cannot obliterate the positive significance P2P brings to the society. In view of the current situation, domestic and foreign scholars have conducted research on P2P lending in various aspects so as to help the platform carry out risk control and promote the healthy development of the industry. In the meantime, as China's credit system and credit rating system are not sound enough, the study on theories related to P2P network lending can also offer certain theoretical guidance for formulating interest rates of China's network lending and conducting credit rating.

Different from the existing research, this paper has made some progress in the following two aspects: first of all, this paper deeply analyzes the influencing factors of P2P network credit from the perspective of loss given default (LGD). Secondly, it selects the transaction data of Lending Club to depict the probability distribution characteristics and influencing factors of LGDs in P2P network lending. It has been discovered the LGDs of P2P loans present an obvious unimodal distribution, the peak value is relatively high and tends to concentrate with the decrease of the borrower's credit rating. The total assets of the borrower have no significant impact on LGD, the credit rating and the debt-to-income ratio exert a significant negative impact, while the term and amount of the loan produce a relatively strong positive impact.

The next research of this paper mainly includes the following parts: the second part is the literature review, the third part presents the theory analysis and research hypothesis, the fourth part displays the empirical analysis, and the fifth part includes the conclusions and suggestions.

2. Literature Review

A large number of scholars have been investigating P2P lending, a new lending mode. The related studies mainly concentrate on the factors influencing P2P default and P2P credit risks, etc.

In respect of empirical research, the data concerning the empirical research of network lending theories mainly come from the transaction data of Prosper and Lending Club. The initial research mainly focuses on the influencing factors of interest rate, success rate and default rate of network lending. Ravina (2007) [1] conducted a more comprehensive empirical analysis on network lending, and he carried out an empirical research on network lending from the perspectives of lending success

rate, interest rate and default rate with the transaction data of Prosper. It is found that not only the borrower's hard financial information and the loan treaty itself will directly affect the success rate of the loan, other factors (such as the loan amount, term and interest rate) will have a certain impact on the success rate; besides, good hard financial information, credit line of limit, race and other factors can significantly affect interest rates as well. Based on the empirical research of Prosper, Klafft (2008) [2] found that the "hard information" related to the default rate, such as the transaction status of the borrower's bank authentication account and the borrower's credit rating, produces a significant positive impact on the transaction rate of P2P. Marques, Garcia and Sanchez (2013) [3], based on the expert scoring, establish the regression model to measure the borrower's risk situation by regarding financial indexes as the explaining variables. Serrano-Cinca (2015) [4] used single factor mean test and survival to analyze 24,449 loan sample data of Lending Club platform from 2008 to 2014, and explained that the default factors were loan purpose, annual income, current housing status, credit records, and liabilities. Malekipirbazari (2015) [5] compared different machine learning methods in order to identify high-quality P2P borrowing customers, the results show that random forecasting (RFS) is significantly superior to FICO scoring and LC grade identification in identifying the best borrowers.

In terms of credit risks, in order to maximize the capacity of batch processing with the internet and big data technology, many transactions on the P2P platform adopt an unsecured mode of credit loans, and the biggest risk comes from the borrower's credit risks. Under this constraint, in order to reduce the information asymmetry between the borrower and the lender, the P2P platform has also introduced a great number of advanced information control methods. For instance, when rating the loan credit, Prosper referred to the group mode of Grameen Bank, that is, all groups develop strict admission standards to help members get low borrowing rates in the future, so the influence of social information is also one of the hot spots in the study. Though research on the platform of Prosper, Herzenstein (2008) [6] found that one of the essential conditions for the financing success of the borrower is the ideal credit scoring. Rating is the further deepening of scoring. Herrero (2009) [7] social group and other information in network lending cannot merely enhance the availability of loans, but fill the disadvantages of personal high borrowing rate, low credit rating, etc. Lin (2009) [8] also found that social network capital can raise the possibility of network lending, reduce the interest rate and lower the default rate simultaneously. Leow et al. (2014) [9], taking retail loan data in the UK as the research object, found that macroeconomic factors have a certain impact on LGD, for mortgage loans, macroeconomic factors make the estimation effect of LGD better; but for personal loans, macroeconomic factors cannot bring about the improvement of prediction accuracy. Robert et al., (2015) [10] without considering the correlation between PD and LGD, conducted modeling for LGDs of small and medium-sized loans with complex collateral to get the expression form of LGD containing collateral and risk exposure. Without the quantized data similar to bank lending, investors cannot convert available information into appropriate market behavior, thus threatening the sustainable development of P2P lending (Mild et al., 2015) [11]. Traditionally, Emekter et al. (2015) [12] analyzed the platform data of Lending Club from May 2007 to June 2012, constructed a logistic regression (LR) model to predict the default probability of the borrower, and made use of the debt-to-income ratio, FICO scoring and circulating credit amount to conduct credit rating for the borrower, and the empirical evidence suggests that credit rating plays a crucial role in reducing the default of the lender.

In China, with the rapid development of P2P network lending, related studies have also emerged in endlessly. Li Yuelei, et al. (2013) [13] found that the basic properties of loan orders, the basic information of borrowers and the social capital of borrowers exert significant effects on the lending success rates in the Chinese market. At the same time, it is found that investors in China's P2P microfinance market exhibit obvious herding behavior characteristics, and these herding behaviors have an important influence on the success rates of borrowing. Liao Li, et al. (2014) [14] used the transaction data of Renrendai to empirically study the risk identification effect of interest rates in P2P lending. The result turns out that in the case of asymmetric information, the non-complete

market-oriented interest rates generated by Renrendai reflect the borrower's default risks, but still a high proportion of default risks is not reflected in the interest rate, and some basic public information of the borrower contributes to predicting this part of risks to some extent. The empirical results of Liao Li, et al. (2015) [15] demonstrate that the repayment probability of borrowers with high academic qualifications as agreed is higher, and the length of higher education has enhanced the borrower's self-discipline ability. However, investors do not prefer borrowers with high academic qualifications and there exist biases in their behaviors of identifying credit risks through educational level. The research results of He Qizhi, et al. (2016) [16] suggest that the fluctuations of interest rates of network lending have the effects of agglomeration and risk accumulation instead of leverage effect and have generally consistent response to bull and bear information, which means that the risk of the network lending market is strong while the risk awareness of market participants is not strong. The research of Xuchen Lin, et al. (2017) [17] Empirical results reveal that gender, age, marital status, educational level, working years, company size, monthly payment, loan amount, debt to income ratio and delinquency history play a significant role in loan defaults. Empirical results reveal that gender, age, marital status, educational level, working years, company size, monthly payment, loan amount, debt to income ratio and delinquency history play a significant role in loan defaults.

To sum up, though P2P lending is different from loans from commercial banks in terms of lending style, they are both the loan relations generated based on credit essentially, and the biggest risk is still the borrower's credit risks. According to the internal rating theory introduced by the New Basel Capital Accord, the analysis of credit risk valuation can be carried out from two aspects: namely, the default rate reflecting the possibility of default and LGD reflecting the severity of the loss after default. Scholars both at home and abroad have achieved fruitful research findings of the default rate. These findings have also substantiated the negative influence of information increase and hard finance conditions of the borrower on the default rate. However, research into the loss given default (LGD) is still insufficient. The default rate and the LGD are two expression factors of equal importance to describe losses caused by loan defaults. They are two indispensable aspects to fully describe default losses. Therefore, the author considers that in addition to the indicator of probability of default, the study on LGD of P2P lending should also be an important issue. Toward this end, this paper analyzes the factors of influencing LGD of P2P network credit, and on this basis, tentatively selects relevant data of Lending Club, factors influencing LGD of P2P lending. Theoretically, this research enriches research into risks of loan default and provides a theoretical basis for a comprehensive description of default losses. Meanwhile, research findings of this paper about LGD can guide P2P platforms to more accurately evaluate credit risks and monitor platform operation risks. In this way, the P2P lending industry can seek more reasonable development.

3. Theoretical Analysis and Research Hypothesis

3.1. Calculation of LGD (LGD)

Default loss refers to the loss of the debtor in the event of default, while LGD is the ratio of the risk exposure loss when the default occurs, and Recovery Rate (RR) and it meet the following relational formula: $LGD = 1 - RR$. Currently, there are three main methods for LGD measurement: Market LGD, Workout LGD and Implied Market LGD.

Lending Club is the most influential P2P lending website in the US. It was founded in 2007, and it is the first P2P network lending platform registering bills as securities through U.S. Securities and Exchange Commission. It provides customers with convenient, transparent and friendly services, and has achieved great success in the United States. By the end of March 2018, it has contributed to a total amount of 33 billion US dollars. On Lending Club, any individual with a US social security account and over 18 years old can release a loan and then deliver it to Lending Club for review. Lending Club will grade the loan applied by the customer based on the borrower's FICO credit score and other performances, and finally, the interest rate of the loan will be decided by credit rating. The customer

passing the review can get the loan easily and choose to pay the money ahead of time without paying extra fees. Since the loans in Lending Club are not publicly traded, it is difficult to use Market LGD; besides, Lending Club delivers the default loans to a third party collection company but does not announce its collection and clearing costs, so Workout LGD is not feasible; Implied Market LGD needs to obtain the risk information of the credit spread in the recent non-default bonds, and then deduce the LGD, and because of the irrationality and complexity the market, it is also difficult to adopt this method in this paper. In order to simplify the calculation of LGD, in the calculation of LGD, this paper does not consider the time cost of currency or the recovery cost and management cost of Lending Club arising from the future recovery, but directly uses the most basic book value for calculation, and defines the LGD required in this paper as follows:

$$\text{LGD} = 1 - \frac{\text{total_payment}}{\text{total_amount}} \quad (1)$$

Thereinto, the numerator total payment indicates the total amount of the loan that has been recovered, the denominator total amount means the total amount of principal and interest of the loan that should have been recovered, based on the basic data published by Lending Club, we can further obtain the two data.

3.2. Analysis of the Factors of Influencing LGD

3.2.1. The Relationship between the Default Rate and LGD

In the supervision of credit risks, the ultimate direct monitoring object is default loss, and the default rate and LGD are only two performance factors of default loss. If they are mutually independent of each other, then the following relational expression is established: $EL = PD \times LGD$, so the relationship between LGD and PD is also a factor of investigation, and some scholars have also done relatively in-depth research in this regard. Frye measured the default rate and LGD with a single system factor, and its results show that the system factor has a negative effect on default rate and LGD, thus leading to a positive correlation between default rate and LGD (Schuermann, 2004) [18]. Altman (1996) [19] used corporate bonds in the United States from 1982 to 2001 for measurement and found that the default rate has a significant impact on LGD, and the correlation is positive. Jarrow (2001) also drew similar conclusions [20]. But Gordy (2000) [21] found that there was no significant correlation between the two and consider that they are two independent variables. While Hu and Perraudin (2006) [22] made use of Moody's historical data and found that the correlation coefficient between the recovery rate and the default rate between 1983 and 2000 is positive, which is 0.22. Although there is a negative relationship between RR and PD from the intuitive sense, the correlation still needs to be confirmed from the research results of scholars.

3.2.2. Influencing Factors of LGD

Mature studies on the LGD of bonds have been made in foreign countries, and relatively perfect default databases have been set up as well. By studying the distribution of recovery rates of all bonds of Moody's Corp from 1970 to 2003, Til Schuermann (2004) [18] found that bonds of different guarantee types have different LGD distribution characteristics, the LGDs of senior secured bonds present a fairly uniform distribution between 20–70%, while the LGDs of senior unsecured and subordinated bonds show a unimodal distribution, with peak values concentrated at around 85%. However, after overlaying all the bonds, LGDs present a double peak distribution, with the main peak near 80% and the side peak near 20%. On the basis of foreign research, domestic scholars conducted a detailed study on the LGD of Chinese loans combined with China's national conditions. Chen Muzi found that the LGD distribution of Chinese loans takes on a U-shaped double-peak with peaking at around 0 and 1 [15]. According to the results of existing literature, LGD is mainly affected by four major factors: project factors, macroeconomic cycle factors, industry factors and enterprise factors.

(1) Project factors

Project factors include the priority of bonds, debt type, etc., the better the quality of collateral and the more the quantity of collateral, the lower the LGD will be.

(2) Macroeconomic cycle

Macroeconomic cycle can affect LGD as well, and the LGD is higher in the recession period than the booming period, especially for low-rated loans, the economic cycle is more influential. Economic cycle is also an important factor of LGD, and more research data show that in the recession period and booming period of macro economy, the overall LGD levels will be different to some extent. For instance, as shown in Figure 1, through the study of recovery rates of default loans in Moody's database, Til Schuermann (2004) found that during the period of economic prosperity, the overall distribution of loan recovery rates is higher and gentler than that in the recession period, indicating that the performance of recovery rates in the booming period is superior to that in the recession period [18]. With the same data, Frye (2000) [23] also found that the overall recovery rate during the recession period is nearly one third lower than that during the booming period. Figure 1 lists the distribution of recovery rates under different economic cycles.

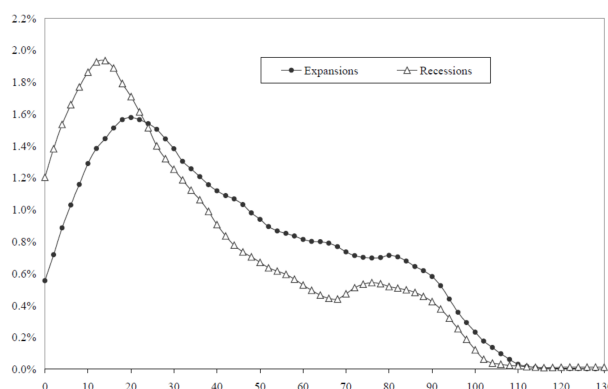


Figure 1. Distribution of recovery rates under different economic cycles (Moody's 1970–2003).

(3) Industry factor

The different industries in which debtors are engaged in also have an impact on LGD. First of all, similar to the impact of macroeconomic cycles, the cycle of the industry itself can produce an impact on default loss. In the next place, the capital structure characteristics of the industry will influence LGD as well. Studies have found that the LGD of real-asset-intensive industries such as utilities and service industries will be higher than that of other industries. The influence and utility of industries need further discussion, but in the study of LGD, it is required to pay due attention to industries. In the first place, industry characteristics indeed will lead to differences in the asset structures and debt ratios of enterprises, which could lead to enterprises' liquidation of debts and thus affect the LGD level of enterprise loans. In addition, the industry itself is also cyclical. When the industry as a whole is in a recession, the bank will strictly review the industry and public investors will also be cautious in investing in the industry, and as a result, corporate financing of the industry will be more difficult and thus affect the enterprise's repayment of debts.

(4) Enterprise factors

Enterprise factors include the capital structure, operating condition and scale of the enterprise, and the enterprise capital structure reflects the financing leverage ratio of the enterprise. The empirical evidence shows that the higher the financing lever, the worse the operating condition and the greater the LGD; but the impact of enterprise scale on LGD is not significant.

Combined with the previous research results, this paper carries out a study by analogy, and comprehensively analyzes the impact of four factors on the LGD of P2P loans, namely project,

macroeconomic cycle, industry and enterprise. But compared with traditional lending, P2P also has its particularity. First of all, P2P was born late and it was just over a decade ago, Lending Club was not established until 2007, and the time span of data generated is not large, so the impact of economic cycle cannot be studied. Secondly, P2P loans mainly include personal consumption loans and small micro-owners operating loans, and there are no obvious industry characteristics, so it is also difficult to measure the impact of industry factors. In addition, P2Ps are mostly credit loans, while previous studies on LGD are mainly concentrated on mortgage loans. Due to the complexity of liquidation of mortgage loans, the studies on LGD are more complex, so it is necessary to consider the legal cost and management cost of collateral market discount and bankruptcy liquidation, etc. Unlike mortgage loans, credit loans are hard to be recovered after the default, but the measurement of default loss is relatively simple, which also provides some convenience for the study of this paper. Taking into account of the above reasons, this paper will start with project factors and enterprise factors.

Project factor refers to the loan contract itself, and the research of traditional lending has found that the collateral quality, priority and type of debts will have a significant impact, but for P2P credit loans, the role of collateral cannot be distinguished, only contract attributes of the loan itself can be considered, including loan rating, interest rate, amount and duration and other factors. Enterprise factor refers to the borrower factor, traditional research mainly focuses on the enterprise capital structure and enterprise scale, this paper maps the enterprise capital structure to the borrower's capital condition, and maps enterprise scale to the borrower's total assets scale. In addition, P2Ps all involve personal loans, so personal credit records are particularly important, and personal credit records undergo great changes, so this paper adds the attributes of personal recent credit situation in the borrower factor.

3.3. Research Hypotheses

Based on the above analysis, this paper makes the following hypotheses concerning factors influencing the LGD of P2P lending.

3.3.1. Influencing Factors of Loans

Hypothesis 1 (H1). *The credit rating of loans has a negative influence on the LGD. The higher the credit rating is, the safer the loans can be and the lower the LGD is. Previous research has already proved that debts with a higher degree of priority and safety are lower in the LGD.*

Hypothesis 2 (H2). *The lending amount has a positive influence on the LGD. The higher the lending amount is, the higher the defaulted earnings of the borrower can be and the higher the moral risks are. Under the condition, the LGD is likely to increase.*

Hypothesis 3 (H3). *The lending term has a positive influence on the LGD. The longer the lending term is, the higher the uncertainties are facing the debt investor and the higher the LGD can be caused.*

Hypothesis 4 (H4). *The interest rate has a positive influence on the LGD. Similar to the lending amount, the higher the interest rate is, the heavier the debt burden is imposed on the borrower and the more likely the default will be to happen. As a result, more default losses will be triggered.*

3.3.2. Influencing Factors of the Borrower

Hypothesis 5 (H5). *The total assets of the borrower have a negative influence on the LGD. The more assets the borrower has, the stronger the borrower's loan repayment capacity can be and the lower the risks of default losses are.*

Hypothesis 6 (H6). *The income of the borrower has a negative influence on the LGD. Income increase can enhance the willingness of borrowing. As the amount of loan principal increases, the effect is similar to that brought by the increasing income.*

Hypothesis 7 (H7). *The financial status of the borrower has a negative influence on the LGD. Previous research suggests that the lower the financing leverage of a company is, the lower the LGD will be. Likewise, in P2P2 lending, the better the financial status of the borrower is, the lower the probability of default will be and the lower the LGD is.*

Hypothesis 8 (H8). *The recent default records of the borrower have a positive influence on the LGD. In general, the poorer the credit of a person is, the higher the default losses can be caused by him.*

4. Empirical Analysis

4.1. Data and Variables

This paper adopts the transaction data of Lending Club as the research object. The transaction data of Lending Club from June 2007 to December 2017 are selected to form a dataset. The original dataset contains 1,341,582 borrowing records. The loans already charged off are set to be defaulted ones. The data with the income not yet verified by Lending Club as well as the data providing incomplete information are screened out. Finally, 41,717 data constitute samples for empirical analysis of this paper. Taking advantage of these existing data, this paper selects appropriate loans and borrower attributes as explaining variables and makes a detailed empirical analysis of the LGD of loans on Lending Club.

In view of the loan factors, this paper selects the amount of each loan, loan term, interest rate and credit rating and other variables to conduct studies. It should be noted that the credit rating of loans in Lending Club is obtained by converting the borrower's FICO credit score, loan amount and term. Lending Club first conducts a preliminary classification of credit rating of loans based on the borrower's FICO score, and then considers the amount and term of the loan, adds a risk modification, and finally obtains the specific credit rating of the loan. That is to say, the credit rating of the loan is a mixed variable of the borrower's credit rating, loan amount and loan term. Lending Club's interest rate is derived from the credit rating of the loan plus risk premium rather than determined by negotiation between the borrower and the investor, resulting in a strong correlation between the interest rate and the credit rating of the loan. Therefore, this paper cannot examine the influencing factors of interest rate alone and can only regard them as an alternative variable of the credit rating of the loan.

The borrower factors mainly include his financial position, total assets, and recent credit record. Since it is unable to obtain direct data of these attributes, this paper uses some variables published by Lending Club for alternative analysis. The debt-to-income ratio can well reflect the repaying capability of the borrower, so it is used to measure the borrower's financial position. Income and housing can reflect the current capital condition of the borrower very well while working years reflect the historical accumulation of the borrower, so this paper measures the total asset of the borrower by employing information such as working years, housing ownership and income. Lending Club directly announced the number of credit defaults of the borrower within two years, which can reflect the recent credit conditions well. The detailed description of all variables is shown in Table 1.

All the selected explaining variables are statistically analyzed, and the results are shown in Table 2. Then the variables are statistically analyzed according to the credit rating, and Table 3 lists the statistical information of variables under different credit ratings. Among the default samples, the highest interest rate is 30.9% while the lowest is 5.23%, which differs to a large extent; then from the perspective of loan amount, compared with the average loan amount of 18,166 US dollars, the maximum loan amount of 40,000 cannot be counted as high, indicating that the amounts of default loans in Lending Club are relatively concentrated; by observing the annual income, it is found that the highest annual income among the samples reaches 7,500,000 US dollars, compared with an average annual salary of 73,673 US dollars, the standard deviation of income is 58024, indicating that the distribution of the defaulter's annual incomes is relatively scattered.

Table 1. Definitions of variables.

Variables		Definitions of Variables	
Explained Variables	LGD	LGD of the Loan	
Explaining variables	Loan information	amount	Loan amount (one thousand US dollars)
		term	Loan term (3 years/5 years), expressed with the virtual values 0 and 1 in the data
		int_rate	Loan interest rate
		grade	The credit rating of the loan
	The borrower's information	emp_length	The borrower's working years. It is recorded as 0.5 if it is below one year and 10 is it is above 10, and the rest will be recorded according to the actual number of years
		home_ownership	Housing ownership, which is divided into four categories, namely "no house", "leased house", "house loan" and "own house", respectively expressed with the virtual values 0, 1, 2, 3, and the larger the number, the better the housing ownership
		annual_inc	The borrower's annual income (one thousand US dollars), which has been verified by Lending Club
		dti	The borrower's debt-to-income ratio, monthly repayment amount/monthly income. Housing loans and loans of Lending Club that have been applied Lending Club are eliminated here.
		deliq_2yrs	Number of poor credit records within two years

Table 2. Statistical information of all variables.

Variable Value	Sample Number	The Maximum Value	The Minimum Value	Average Value	Median	Standard Deviation
Amount (one thousand US dollars)	41,717	40.00	1.00	18.17	17.10	9.2494
term (month)	41,717	60.00	36.00	47.32	36.00	11.9809
int_rate	41,717	30.99%	5.32%	17.37%	16.99%	0.0480
grade	41,717	35.00	1.00	17.04	16.00	6.7890
emp_length	41,717	10.00	0.50	6.24	7.00	3.5502
home_ownership	41,717	3.00	0.00	1.65	2.00	0.6341
annual_inc	41,717	7500	6.00	73.67	64.16	58.0241
dti	41,717	149.46%	0.00%	20.60%	20.64%	8.7202
deliq_2yrs	41,717	18.00	0.00	0.33	0.00	0.8996

Table 3. Average values of variables in various ratings.

Credit Rating	Principal Amount	Loan Term	Interest Rate	Working Years	Annual Salary	The Debt-to-Income Ratio	Number of Defaults within Two Years	Sample Number	Proportion
A	16,276	37.44	7.80%	6.08	86,005	17.00%	0.19	1235	2.96%
B	15,961	40.66	11.54%	6.16	74,077	19.31%	0.30	5971	14.31%
C	16,858	45.20	14.70%	6.31	72,383	20.51%	0.33	11,964	28.68%
D	18,097	46.71	18.09%	6.21	72,157	20.83%	0.35	10,078	24.16%
E	20,274	52.55	21.49%	6.24	75,078	21.37%	0.32	7259	17.40%
F	21,117	55.82	24.68%	6.25	73,439	21.58%	0.35	3895	9.34%
G	22,002	56.92	27.06%	6.29	76,536	21.84%	0.36	1315	3.15%

Table 3 shows the performance of loans of all credit ratings. Through analysis, it is found that the proportion of default loans with the credit rating of B–E is the largest, reaching 84.55%, while the A level only accounts for 2.96%, and F and G take up 9.34% and 3.15% respectively. Compared with the total loans, the A level accounts for 16.73%, while the F and G levels only take up 2.13% and 0.68% respectively, thus the default rate increases significantly with the decline of the rating. Figure 2 lists the proportions of loans of all credit ratings in total loans and default loans. In addition, what is interesting is that credit ratings among the default samples have a negative relationship with working years, from an average of 6.08 to A G level of 6.29 years, the longer the working age, the higher the credit score.

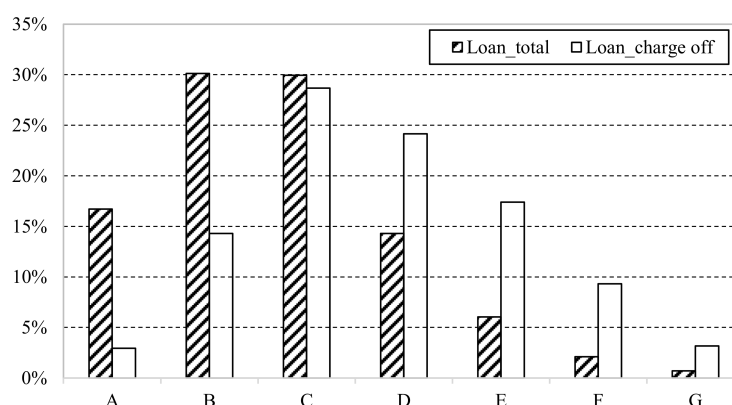


Figure 2. Proportions of various credit ratings in loans.

Table 4 displays the correlation coefficient matrix of 9 dependent variables, which demonstrates that loan interest rate is strongly correlated with credit rating in Lending Club. This is because Lending Club set interest rates on the loans according to the credit rating, with 5.05% as the benchmark rate, and once credit rating declines by a small level, the risk interest rate of 0.5–1% will be floating upward accordingly, from 5.32% at the highest A1 level to 30.99% at the worst G5 level. The lending amount is found to have a significantly positive correlation with the lending term and annual income, respectively. The significantly positive position is also observed in the lending term with the interest rate and the credit rating, respectively. However, the positive correlation is very weak among the working years, poor credit records in two years and credit rating.

Table 4. Correlation coefficients of all variables.

	Amount	Term	Interest Rate	Credit Rating	Working Years
Amount	1				
Term	0.3551	1			
Interest rate	0.1873	0.4033	1		
Credit rating	0.1987	0.4222	0.9641	1	
Working years	0.0903	0.0741	0.0058	0.0047	1
Housing situation	0.1168	0.0602	−0.0361	−0.0404	0.1282
Annual income	0.3614	0.0608	−0.0192	−0.0058	0.0523
The debt-to-income ratio	−0.0460	−0.0148	0.1084	0.0965	0.0229
Number of defaults within two years	0.0172	−0.0295	0.0135	0.0191	0.0135
	Housing situation	Annual income	the debt-to-income ratio	Number of defaults within two years	
Housing situation	1				
Annual income	0.0942	1			
the debt-to-income ratio	−0.0002	−0.1643	1		
Number of defaults within two years	0.0400	0.0677	−0.0248	1	

4.2. Distribution Characteristics of LGD

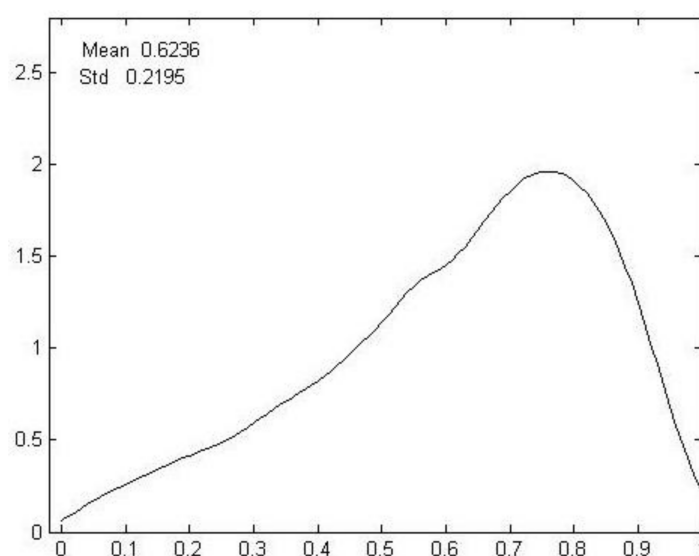
According to the definitions given in Formula (1), this paper analyzes the default data released by Lending Club, records the loans that have been charged off and delayed for more than 120 days as a breach of contract, and calculates the LGDs of all default loans with Formula (1), and the statistical results are shown in Table 5:

Table 5. The statistical information of loss given default (LGD).

Variable Value	Sample Number	Maximum Value	Minimum Value	Average Value	Median	Standard Deviation
LGD	41,717	1.0000	−1.4111	0.6236	0.6650	0.2195

As shown in the statistical results of LGDs, the average value is 0.6236 and the standard deviation is 0.2195, indicating that the total default loss is relatively large and distributed in a relatively concentrated manner. In addition, it is noted that the minimum value is negative, this is because in the calculation of this paper, only the book value of the debt is taken into consideration, although some loans have defaulted, plus the late fees and total penalty interest charged later, it may exceed the total amount of principal and interest, so the loss rate may appear negative.

In order to display the probability distribution characteristics of LGD more intuitively, this paper uses Kernel method to estimate the probability distribution of the LGD of the default loans of Lending Club and obtains the distribution result shown in Figure 3. It can be found that the LGDs of the loans of Lending Club display a distinct unimodal distribution with peak values concentrated around 75%. This shows that LGDs of loans are generally high, and though only a small part of loans default, it causes only minor losses. Analysis results of Lending Club's defaulted bonds are close to statistical results of Til Schuermann about subordinate bonds and limited uncovered bonds. Both demonstrate a unimodal distribution and the peak value of the two is close to each other. This suggests that the priority for credit loans generated by Lending Club is relatively low and the LGD is comparatively larger.

**Figure 3.** Distribution of probability density of LGDs from 2007 to 2013 (Lending Club).

Then, the same probability distribution analysis of LGDs of loans of different credit ratings is analyzed respectively, and the results shown in Figure 4 are obtained. In Figure 4, the LGDs of A-level loans are distributed relatively uniformly, and there is a relatively wide and gentle distribution between 0.1 and 0.9. But with the decline of the credit rating of the loan, the LGD curve gradually steepens and moves to the right and finally converges to the single peak of 0.85. The G level loan exhibits most obviously, and its LGD probability density has a high peak value near 0.85. The information shown in Figure 4 is similar to that of Til Schuermann, and as the credit rating of the loan declines, the mean value of LGDs increases gradually and the distribution tends to be concentrated.

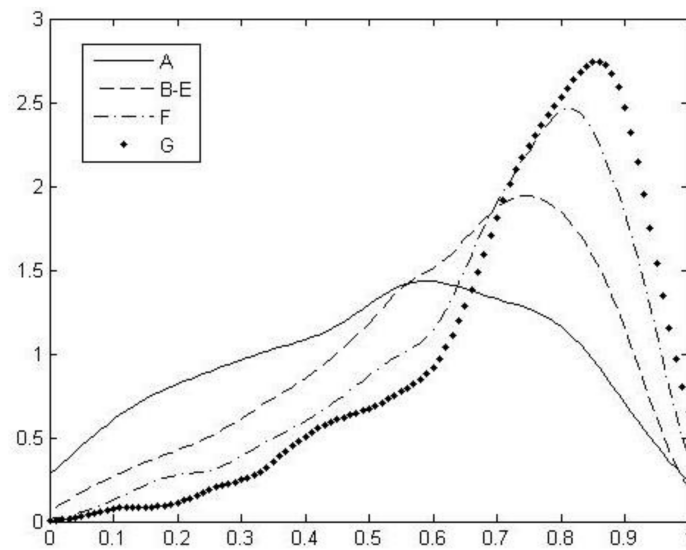


Figure 4. Distribution diagram of probability densities of LGDs of loans of different ratings.

4.3. Result Analysis

The previous text lists probability distribution characteristics of LGDs of Lending Club, but what factors affect LGD? This section will conduct an empirical study of the impact of loans and borrower characteristics on LGDs, and the empirical model adopts multivariate linear regression model, defined as follows:

$$LGD_i = \alpha + \beta X_i + \gamma W_i + \mu_i$$

In the above formula, α is a constant term, X is the feature information vector of the loan, including the credit evaluation, principal, term and interest rate of the loan, while W represents the borrower's feature information vector, including the lender's working years, annual income, debt-to-income ratio and number of defaults within two years (see Table 1 for specific information). Table 6 lists the results of multiple linear regression analysis of this model.

Table 6. Multiple linear regression results of LGD.

Loan Information	(1)	(2)	(3)	(4)
amount (\$1000)	−0.0002 (0.1431)	−0.0002 (0.1988)	−0.0002 (0.1118)	−0.0002 (0.1113)
term	0.0285 *** (0.0000)	0.0297 *** (0.0000)	0.0263 *** (0.0000)	0.0275 *** (0.0000)
int_rate	0.8661 *** (0.0000)		0.8771 *** (0.0000)	
The borrower's information				
grade		0.0056 *** (0.0000)		0.0057 *** (0.0000)
emp_length	−0.0003 *** (0.0000)	−0.0032 *** (0.0000)		
home_ownership	−0.0053 *** (0.0015)	−0.0053 *** (0.0017)		
annual_inc (\$1000)	5.56×10^{-5} *** (0.0047)	4.59×10^{-5} ** (0.0200)		
dti	0.0016 *** (0.0000)	0.0017 *** (0.0000)	0.0015 *** (0.0000)	0.0016 *** (0.0000)

Table 6. Cont.

Loan Information	(1)	(2)	(3)	(4)
delinq_2years	0.0028 ** (0.0170)	0.0027 ** (0.0225)	0.0026 ** (0.0236)	0.0025 ** (0.0337)
Constant term	0.4536 *** (0.0000)	0.5058 *** (0.0000)	0.4300 *** (0.0000)	0.4822 *** (0.0000)
R	0.0584	0.0534	0.0551	0.0503
Observation sample number	41,717	41,717	41,717	41,717

Description: the data in parentheses are *p* values of the corresponding variables, **, *** represent the significant levels of 5% and 1% respectively. The first column removes the variable of credit rating, and the second column removes the variable of loan interest rate. The third, fourth columns remove the variable of the borrower's total assets.

Because of the strong correlation between the loan interest rate and the borrower's credit rating in the sample data, in order to remove the correlation of independent variables, decompose this model into two steps to seek regressive solution respectively. In the first step, remove the variable of credit rating in the regression model to obtain the first column of data in Table 5. In the second step, remove the variable of loan interest rate in the regression model to obtain the second column of regression results.

Analysis of the data presented in Table 6 shows that LGDN and credit rating are significantly negatively correlated. The higher the grade is, the poorer the credit rating is. On the contrary, the LGD is significantly positively correlated with the loan interest rate. This is because Lending Club has set a strict ratio between the grade and the interest rate. The credit rating is to decide the interest rate of a loan. Therefore, H1 is substantiated, but H4 cannot be verified. Besides, the relationship between the LGD and the lending amount is not significant. This suggests that H2 cannot be verified. The LGD and the lending term are significantly positively correlated with each other. This provides solid evidence for H3. Among loan factors, apart from hypotheses about the interest rate and the lending amount, the other two hypotheses can both be substantiated.

Then with respect to the borrower factor, firstly, Regression results indicate that the LGD is significantly negatively correlated with the borrower's working years and housing ownership. The latter two factors can both reflect the borrower's financial status. This means the more the total assets the borrower has, the lower the risks of debt default losses will be. In contrast, the LGD and the borrower's income are significantly positively correlated. This suggests that increase in the income cannot enhance safety of debts. Rather, with the increase in the income, the borrower's loaning willingness strengthens. As the loan principal increases, the LGD is on an upward trend, Hypothesis 5 and Hypothesis 6 are established. In the next place, LGD has a significantly positive correlation with the borrower's debt-to-income ratio and number of defaults within two years, indicating that the borrower's financial position exerts a significant negative impact on LGD, while the recent poor credit standing has a significant positive impact on LGD, so Hypothesis 7 and Hypothesis 8 are established. The third and fourth columns are the regression analysis removing the borrower's total assets in the model, and the results are the same as those of the previous analysis.

5. Conclusions and Inspirations

Through the empirical analysis of default loans of the Lending Club, this paper describes probability distribution characteristics and influencing factors of LGDs in P2P network lending. It is found that the probability density of the P2P lending's LGD is generally in a unimodal distribution with the value peaking at around 0.75. The characteristic is similar to the statistical results obtained by previous research about the subordinated bonds and limited uncovered bonds. This means that the P2P lending has a lower degree of priority, so its LGD is higher. This further demonstrates the importance of the LGD in deciding the final default losses. Besides, after solving the probability density distribution of the LGD of different credit ratings in samples, this paper observes that, as the loan credit rating declines, the LGD keeps on rising and the LGD probability density distribution

tends to concentrate. Based on that, the negative correlation between the credit rating the LGD can be verified. To sum up, P2P lending has a lower degree of priority. Once default happens, the losses of P2P lending are often serious. This necessitates control of the LGD by P2P lending platforms so as to avoid negative influence brought by occurrence of default. Taken as a whole, the poorer the loan credit rating is, the longer the lending term will be, and the higher the LGD will be. Among the borrower's factors, the borrower's total assets are significantly negatively correlated with the LGD, while the borrower's income is proved to have a significantly positive correlation with the LGD. The borrower's financial status is significantly negatively correlated with the LGD, while the borrower's recent poor credit records are significantly positively correlated with the LGD. On the whole, from the perspective of the borrower, the higher the total assets the borrower has, the better the borrower's financial status can be and the lower the LGD will be. When the borrower's income is studied, the extra loaning willingness brought by the borrower's consumption needs should also be taken into account, which can lead to an increase in the LGD.

In the meanwhile, this paper puts forward the following inspirations:

- (1) LGDs of network lending are generally high, so when carrying out risk control for P2P platform, it is necessary to give priority to the control of default rate and take preventive measures in order to ensure the safe operation of the platform.
- (2) The credit rating of the borrower has a strong negative impact on LGD of network lending, and it is required to be more cautious when the P2P platform is reviewing loans of a low credit rating. While the credit rating of loans is mainly formulated based on the credit points of the borrower, and it reveals the significance of the credit reporting system in the risk control of network lending. In view of the imperfect credit reporting system in China at present and in order to seek better development of the P2P industry in the future, all countries should strongly support the construction of credit reporting system, all platforms should take the initiative to shoulder their own responsibilities and strengthen the cooperation with the credit investigation of Central Bank, and cooperate with the government to establish and improve the credit reporting system so as to achieve a win-win situation.
- (3) When the network lending platform carries out loan review, it is supposed to bear in mind that the borrower's repayment will always be more important than his repayment ability, and especially, it is due to pay more attention to the borrower's credit performance in the short term. Only in this way can we reduce the loan loss, reduce the loan risk, and thus maintain the interests of investors.
- (4) When the network lending platform reviews the borrower, be careful not to be deceived by the size of the total asset of the borrower but attach more importance to the borrower's financial situation at that time, clearly investigate other debt burdens of the borrower, and explore such issues as how the debt-to-income ratio is and reducing the adverse selection.

Acknowledgments: The research for this paper was supported by the National Natural Science Foundation of China (No. 71573050; No. 71573170) and the Shanghai Social Science Fund (No. 2015BJB003; No. 2014BJB003).

Author Contributions: Guangyou Zhou performed the theory analysis and contributed to drafting this manuscript. Sumei Luo conceived and empirical analysis. Yijia Zhang analyzed the data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ravina, E. *Beauty, Personal Characteristics, and Trust in Credit Markets*; Social Science Electronic Publishing: Rochester, NY, USA, 2008.
2. Klafft, M. *Peer to Peer Lending: Auctioning Microcredits over the Internet*; Social Science Electronic Publishing: Rochester, NY, USA, 2008; pp. 1–8.
3. Marques, A.I.; Garcia, V.; Sanchez, J.S. A literature review on the application of evolutionary computing to credit scoring. *J. Oper. Res. Soc.* **2013**, *9*, 1384–1399. [[CrossRef](#)]

4. Carlos, S.C.; Begoña, G.N.; Luz, L.P. Determinants of Default in P2P Lending. *PLoS ONE* **2015**, *10*, e0139427.
5. Malekipirbazari, M.; Aksakalli, V. Risk assessment in social lending via random forests. *Expert Syst. Appl.* **2015**, *10*, 4621–4631. [[CrossRef](#)]
6. Herzenstein, M.; Andrews, R.L.; Dholakia, U.M. The Democratization of Personal Consumer Loans? Determinants of Success in Online Peer-to-Peer Lending Communities. *Bull. Univ. Del.* **2008**, *1*, 274–277.
7. Herrero-Lopez, S. Social Interactions in P2P Lending. In Proceedings of the Workshop on Social Network Mining & Analysis, Paris, France, 28 June 2009; Volume 33, pp. 1–8.
8. Lin, M.; Prabhala, N.; Viswanathan, S. Social Networks as Signaling Mechanisms: Evidence from Online Peer-to-Peer Lending. 2009. Available online: <https://pdfs.semanticscholar.org/01c5/07b28ec8e2f8a1382c4da4b75f31257715b0.pdf> (accessed on 26 March 2018).
9. Leow, M.; Christophe, M.; Lyn, T. The Economy and Loss Given Default: Evidence from Two UK Retail Lending Data Sets. *J. Oper. Res. Soc.* **2014**, *3*, 363–375. [[CrossRef](#)]
10. Frontczak, R.; Rostek, S. Modeling loss given default with stochastic collateral. *Econ. Model.* **2015**, *44*, 162–170. [[CrossRef](#)]
11. Mild, A.; Waitz, M.; Wöckl, J. 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.* **2015**, *6*, 1291–1305. [[CrossRef](#)]
12. Emekter, R.; Tu, Y.B.; Jirasakuldech, B.; Lu, M. Evaluating credit risk and loan performance in online Peer-to-Peer lending. *Appl. Econ.* **2015**, *1*, 54–70. [[CrossRef](#)]
13. Li, Y.L.; Guo, Y.; Zhang, W. Analysis of Factors Influencing the Borrowing and Lending Success Rate on China's P2P Microfinance Market. *J. Financ. Res.* **2013**, *7*, 126–138.
14. Liao, L.; Li, M.R.; Wang, Z.W. The Intelligent Investor: Not-Fully-Marketized Interest Rate and Risk Identity. *Econ. Res. J.* **2014**, *7*, 125–137.
15. Liao, L.; Ji, L.; Zhang, W.Q. Education and Credit: Evidence from P2P Lending Platform. *J. Financ. Res.* **2015**, *3*, 146–159.
16. He, Q.Z.; Peng, M.S. Research on the Characteristics of P2P Lending Interest Rate Based on Internet Finance. *J. Financ. Res.* **2016**, *10*, 95–110.
17. Lin, X.C.; Li, X.L.; Zheng, Z. Evaluating borrower's default risk in peer-to-peer lending: Evidence from a lending platform in China. *Appl. Econ.* **2017**, *35*, 3538–3545. [[CrossRef](#)]
18. Schuermann, T. What do We Know about Loss Given Default? In *Credit Risk Models and Management*, 2nd ed.; Shimko, D., Ed.; Risk Books: London, UK, 2004; pp. 249–274.
19. Altman, E.; Kishore, V. Almost every-thing you wanted to know about recoveries on defaulted bonds. **1996**, *6*, 57–64. [[CrossRef](#)]
20. Jarrow, R.A. Default parameter estimation using market prices. *Financ. Anal. J.* **2001**, *5*, 75–92. [[CrossRef](#)]
21. Gordy, M. Credit VAR Models and Risk-Bucket Capital Rules: A Reconciliation. In Proceedings of the 36th Annual Conference on Bank Structure and Competition of the Federal Reserve Bank of Chicago, Chicago, IL, USA, 27 April 2000; pp. 406–415.
22. Perraudin, W.R.M.; Hu, Y.T. *The Dependence of Recovery Rates and Defaults*; Social Science Electronic Publishing: Rochester, NY, USA, 2011.
23. Frye, J. Depressing Recoveries. *Risk* **2000**, *13*, 108–111.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).