Determinants of Loan Performance in P2P Lending

Author: Nilas Möllenkamp
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT
This research paper investigates the influential factors of loan performance in the online P2P lending industry. Unlike in traditional banking, lenders in P2P lending are mostly private and small investors which do not have the expertise to evaluate credit risks appropriately. They suffer from information asymmetry and are generally in disadvantage compared to the borrower. In order to solve information asymmetry, P2P lending platforms provide information regarding borrower characteristics and loan characteristics and assign a credit grade that should predict the default risks of loans. This study analysed 143,654 P2P loans that were funded on the P2P lending platform Lending Club between 2012 and 2013 with binary logistic regressions and finds evidence that the assigned credit grade is the most influential factor on loan performance. Further predictors of loan performance are the loan amount, annual income of the borrower, debt-to-income ratio and the number of inquires in the last 6 months. The actual influence of each determinant on loan performance is observed to be changing between different credit grades.

Supervisors:
Dr. X. Huang
Dr. S.A.G. Essa

Keywords
P2P lending, loan performance, P2P loan default, P2P risk, P2P credit scoring, online lending

9th IBA Bachelor Thesis Conference, July 5th, 2017, Enschede, The Netherlands.

1. INTRODUCTION

In the past years, the market of peer-to-peer 'P2P' lending rapidly grew and captured international interest of borrowers, small lenders and even institutional lending companies. Market analysts predict the loan origination volume to be around $90.0 billion in 2020, according to the U.S. Treasury (2016). In the view of many lenders, the industry is a modern and alternative way of investing money and getting attractive returns. However, default risk of loans is omnipresent and must be investigated further in order to state whether P2P lending is not only an alternative and modern, but also an attractive opportunity for investors.

In the U.S. and Europe, P2P platforms are facing an increasing user base and higher amounts of loans are funded every year (Milne & Parboteeah, 2016) whereas in China, the peak of this hype was accompanied with many discrepancies and fraud so that new regulations and governmental intervention were necessary (Huang, 2017). However, P2P lending is still a popular investing opportunity that attracts a lot of attention.

P2P lending is based on the idea of crowd funding. On P2P lending platforms like Prosper.com and Lending Club, private borrowers apply for funding their microloans that are supposed to be used for different purposes. Investors can select these unsecured loans on the basis of different information like borrower' characteristics and loan details and lend a portion of the overall loan amount. The borrowers are paying back their debt in monthly payment rates or repay the full amount earlier. While borrowers have to pay an interest rate on borrowings, lenders gain profit. The interest rate for borrowers includes the investment profit of lenders as well as transaction fees that have to be paid for using the P2P platform. P2P lending fills the gap for microloans starting from $1,000 with comparably low interest rates. Usually, traditional credit institutes do not target this segment.

There are several advantages of P2P lending for both parties, lenders and borrowers, compared to traditional credits. P2P loans are easy accessible for borrowers over the Internet and funding is straightforward and accomplished in less than 14 days. Interest rates are lower than with bank credits and therefore more attractive for borrowers. Lenders are said to benefit from lending since P2P loans are expected to deliver appropriate risk-return relationships.

P2P lending is also accompanied with some drawbacks that are mostly on investor's side. Besides possible regulatory issues and platform default, it is generally more risky than traditional lending of banks. Typically, collateral is absent, therefore the loans are not secured. Although P2P platforms take legal actions in order to collect funds, the default risk is still high.

The most discussed issue regarding the default risk and loan performance is information asymmetry. While investors always seek to find high performing loans that will not default, they are in a disadvantage towards the borrower. Lenders naturally have less available information and data about the credibility of borrowers than borrowers have about themselves. Loan selection, therefore, can cause moral hazard and adverse selection (Boot & Thakor, 1994; Edelberg, 2004). Credit institutes and banks usually use precise monitoring and get access to in-depth information about the borrowers, their credit scores and other information and data from the credit file. However, private P2P lenders just have very limited capabilities to monitor borrowers.

Banks are using credit grading that is accurate and done by professional credit institutes, while P2P platforms have no access to these expertise assessments. Instead, P2P platforms calculate a credit rating that is a replacement of the missing professional credit grading. However, credit grades from P2P platforms are not expected to be as accurate as credit ratings from credit institutes since platforms put less effort in gathering all relevant information and have less monitoring capabilities. The accuracy is questionable. The credit grade is generally determined by the default probability that is estimated by the platform in a forward-looking analysis of the borrower. On the basis of the credit grading, an interest rate is charged for each individual loan. At Lending Club, the interest rates are actually between 6,03% and 26,06%. Interest rate and credit grade also represent the estimated default probability of a loan that is predicted by the P2P platform even before the loan is funded in an ex-ante evaluation.

When a loan matures, there are two possible outcomes: the borrower paid back the whole funded amount plus an interest rate or the borrower did stop her payments at a certain time and so, after 121 days, the loan is incurred on lenders' side as 'charged off' or 'defaulted'. Loan performance, therefore, can be described as the ex post default rate or success rate. For investors it is essential to know the determinants of loan performance in order to focus on the most relevant influential factors when estimating whether a particular loan is worth an investment or not.

This paper aims to analyse the factors that explain the loan success and default of P2P loans at Lending Club and explores past data on the loan performance of different credit grades. The P2P platform publishes a variety of different information concerning the loan itself like loan amount and loan purpose as well as many borrower characteristics like the annual income, the number of total credit lines and the debt-to-income ratio. Lending Club rates each loan on a scale of seven different credit grades from A (lowest risk) to G (highest risk). With this mechanism, the P2P platform tries to display the default risk of a particular loan to the lender as an information basis that supports individual loan selection. To estimate the actual relationship between credit grade and loan performance in practice is goal of this study. It also tests the correlation of the variables loan amount, annual income, debt-to-income ratio, the number of inquires in the last 6 months, the number of open and total credit lines as well as the revolving credit balance with loan performance. The effects of each variable on the loan success rate are tested for all loans in one sample as well as sub-samples, divided by credit grade classes. This paper, therefore, strongly follows the research question:

What are the determinants of loan performance in P2P lending?

Investors are always seeking to minimise their investment risk. With knowing the most important determinants of loan performance and their weighting is it possible to decrease the overall default risk through only investing in the most attractive loans. The issue of information asymmetry can actively be lowered and the investment profit potentially increased.

P2P lending is accompanied with a lot of literature; however, the evaluation of credit risk and loan performance as well as the question, which factors are actually the best predictors for loan performance are not approached in its entirety. Four prior studies already analysed the influences of different variables on the loan performance of P2P loans. However, the results differ and loan data might not fit actual situations due to unusual economic conditions like the financial crisis at the time periods of study. This paper is the first one that provides evidence for influential factors on loan performance from matured loans in a time period that is better comparable with more recent situations. It contributes to the small number of analysis of P2P loans and delivers a new insight into the determinants the loan performance. This study is also the first to deliver predictors of loan performance for single credit grades and shows differences in the effects of influential factors.

The study finds a positive relationship between the credit grade and the loan performance. With a higher credit grade, the risk of loan default decreases. Furthermore, based on the gathered data, there is evidence for four more determinants of loan performance that successfully predict the probability of loan success. Loan amount, annual income, debt-to-income ratio as well as inquires in the last 6 months are significant influential factors in a loan sample of all credit grades. The number of open and total credit lines as well as the revolving credit balance are found out to be only predictive in some credit grades.

This paper is organised as follows: In 2.1, a literature review is provided on the basis of the general trend of P2P lending (2.1.1), the risks in P2P lending for investors (2.1.2) and the determinants of default (2.1.3). Paragraph 2.2 delivers the development of hypothesis for this study. In 3.1, the empirical methodology and regression model are explained whereas paragraph 3.2 gives an overview about the regression variables. The data section in 4.1 includes the data selection and paragraph 4.2 gives a general overview using the descriptive statistics and correlation matrix. Results of the binary logistic regression are presented in 5.1. Paragraph 6 contains the final conclusion of this study and 6.1 gives directions for further studies. Limitations are described in 8 and the references as well as appendices can be found in 9.

2. LITERATURE REVIEW

2.1.1 P2P lending as an emerging trend

Everett (2015) does not see P2P lending in a completely new context of banking. There are strong similarities between traditional financial intermediaries and P2P lending, hence P2P lending is a new and modern alternative in the banking sector. Käfer (2016) follows a similar approach, but categorises P2P lending in the topic of 'shadow banking' since it is generally more risky for the borrower than traditional banking. While established for private investors, P2P lending is already relevant for institutional investors. Mateescu (2015) provides a general overview about how P2P lending is used by professional agencies and summarises the idea of P2P lending networks. She sees efficiency, financial inclusion through improvements in underwriting as well as more transparency in the process of loan providing as the main advantages of this technology.

Borrowers at P2P lending platforms are described by literature as "debt-laden, middle-to-high income, individuals who are consolidating credit cards and other debt." (Morse, 2015, p. 4). Emekter, Jirasakuldech and Lu (2015) state that, however, high-income borrowers do generally not participate in P2P lending markets. P2P lending is, concerning to a discussion paper from De Roure, Pelizzon and Tasca (2016), attracting customer that were already rejected by traditional credit institutes and are now accepting higher interest rates and apply for P2P loans. This is especially true for the market segment of loans with high risk and small size. Banks and traditional financial intermediaries are often not willing to operate in this sector for various reasons.

Another important stream of research was conducted in order to measure the determinants for funding success of P2P loans. As mentioned by Lee and Lee (2012), herding behaviour plays an important role for whether a loan is funded or not. This does not help to overcome information asymmetry, however, is regarded as an emotional aspect that must be considered when talking about the successfulness of funding specific loans. Earlier, Herzenstein, Dholakia and Andrews (2011) described the herding behaviour in P2P loans as 'strategic herding'. An increasing number of bids attract other bidders so that partially funded P2P loan auctions (in case of Prosper.com) are becoming more and more popular until they are fully funded. Also on the basis of Prosper.com data, Lin, Prabhala and Viswanathan (2013) found out that members of relational friendship networks generally get faster funding of loan applications and additional suffer less default.

2.1.2 Risks in P2P lending

In P2P lending, two types of default risk are evident that have differences in meanings in various contexts: ex ante and ex post risk. Both generally refer to the same, namely the estimation of the probability of loan default, but the occurrence, so the timing, is different. Thus, also some determinants of default risk can be unique. With the ex ante risk, so the calculated default risk before the loan is made, the interest rates for loans are determined dependent on the estimated default risk. Some papers in literature are focusing on the risk before the loan is made and therefore bring adverse selection as the most relevant information asymmetry in (Iyer, Khwaja, Luttmer & Shue, 2009) However, adverse selection and ex ante risk are less relevant for this study.

This paper analyses the ex post default risk, so the loan performance after the loan is matured and searches for the probabilities of default and the determinants of loan performances. Research in this field is still very limited. There are some unique factors and risks in this type of loans that differentiate from traditional loans and are usually not considered there. For example, P2P loans suffer more information asymmetry in the case of moral hazard. Gorton and Winton (2003) studied on the roles of banks and state that one important issue for financial intermediaries is to produce information and monitor borrowers. Both roles are just partly satisfied by P2P platforms. While in traditional lending, information advantages on borrowers' side and moral hazard effects are mitigated in a form that borrowers are permanently monitored as part of the asset services to ensure they behave according to contractual obligations (Diamond & Dybvig, 1986), this does not happen in P2P lending. In P2P lending, monitoring is usually not possible and P2P platforms refuse to collect that information since it can be very costly. Borrowers can easily chose another purpose to spend the money, for example. Dependent on the actual loan purpose, this can increase the loan default probability and therefore lower the lenders' security. Anonymity and inexperience of private lenders to handle information asymmetry raise the issue of information asymmetry and make the default risk less predictable. Monitoring is usually priced in the interest rates of banks that are higher than in P2P lending (Diamond, 1984). Low interest rates are combined with lesser service, such as less monitoring capabilities.

A chance for better monitoring in P2P lending is personal relationships. Lin et al. (2013) explored a decrease in the likelihood of default when personal relationships in P2P lending are noticeable. Nevertheless, it is not clear whether the better loan performance is a direct consequence of more monitoring efforts. However, Everett (2015) indeed found relationship banking in P2P lending helps to mitigate moral hazard issues. Lin (2009) recommends for the extension of the P2P lending business model to also provide capabilities for lenders to monitor borrowers or outsource these tasks in order to decrease information asymmetry.

P2P lenders should also recognise loan default risks that do not stem from the borrower or the loan itself, but from the platform. In China, for example, some P2P platforms committed fraudulent activities that led to bankruptcies and government interventions (Yu, 2017). Mostly, platform fraud also results in a default on lenders' side. It is therefore another kind of default risk that has to be noted. As Kirby and Worner (2014) state, the risk of fraud in P2P lending is unpredictable and must be considered. This applies to all crowdfunding and crowdlending platforms and activities. Wei et al. (2015) also sees fraud as one of the major risks that investors are facing when lending money on P2P platforms. The capitalisation of companies that are active in the Lending Club industry is mostly low (Wei, 2015), especially in China where an outstanding number of platform defaults already happened. Therefore, it is questionable whether they will offer their service over the expected time period of a loan. Platform failure also could affect the execution of transactions and therefore lead to loan default.

Legal systems are not in every country well prepared for the new challenges that could come with the modern technology of P2P lending. Regulatory and legal issues are in fact very much relevant when it comes to the successfulness of loans. In order to establish a legal framework for P2P lending, bank involvement is necessary (Galloway, 2009). However, not every country has strict rules concerning this technology and the managing of borrower-lender relationships. In 2012, Verstein made clear that at least in the U.S. regulation for P2P businesses was poor and the SEC did not understand what P2P lending really was about. Misregulation made P2P lending systems costly and decreased security for lenders. When a P2P platform breaks down or is forbidden after investors funded loans, they are facing another kind of ex-post loan default risk, other than typical charge offs. The risk of platform default due to fraudulent, economical and legal issues is much less in the traditional banking industry, making P2P loans more risky.

The risk of cyber-attacks is also mentioned in the working paper of Kirby and Worner (2014). Since P2P platforms are relatively new and some might not have the capabilities and financial opportunities to entirely exclude the potential risk of cyber-attacks, cybercrime could also harm the platforms users. The effects of cyber-attacks on the borrowers and lenders are not appraisable since private bank details are provided on P2P lending platforms.

2.1.3 Determinants of default

P2P lending is risky for investors, because loans are not secured. A chance to get a more profitable outcome out of every investment, therefore, is to decrease the information asymmetry between borrowers and investors and use loan specific information and borrower characteristics as provided by the P2P lending platform. Lending Club, for example, delivers a wide range of data concerning the borrowers and their credit history. Furthermore, every borrower is categorised into a credit grade ranging from A to G with the expected lowest risk in credit grade A. Lenders use this information to estimate the default probability of a specific loan and decide on whether the loan is attractive to fund or not.

Several studies concerning these determinants of default risk were examined in the past. Iyer et al. (2009) found credit score, number of current delinquencies, total delinquencies, debt-to-income ratio and loan amount as significant determinants. Furthermore, some non-standard variables like membership in a group were explored as influential predictors for loan default. Everett (2015) stated credit score, borrower age, home ownership, endorsements as well as loan amount as significant for predicting the default risk of P2P loans. Guo, Zhou, Luo, Liu and Xiong (2016) used kernel regression for finding the risk determinants of FICO score, number of inquires in the last six months, loan amount, homeownership and debt-to-income ratio. However, Iyer et al. (2009) and Everett (2015) used data from P2P lending platform Prosper.com while Guo et al. (2016) focused on a relatively small sample of only 2,016 Lending Club loans and 4,128 loans from Prosper.com. Since Prosper.com uses a Dutch auction mechanism in order to fund loans, the studies, samples and determinants of loan success seem inappropriate to compare with the data gathered from Lending Club.

Studies that focus on Lending Club data and analysed the determinants of loan default are Emekter et al. (2015), Serrano-Cinca, Gutiérrez-Nieto, and López-Palacios (2015), Carmichael (2014) and Li, Yao, Wen and Yang (2016). They all used similar approaches and logistic regressions, however, results differ and

Table 1: Summary of Default Determinants

Study	Data	Methodology	Determinants of loan default
Emekter et al. (2015)	Loans from May 2007 to June 2012; 36 & 60 months maturity	Binary logistic regression	Credit grade (Lending Club), debt-to-income ratio, FICO score, revolving credit line utilisation
Li et al. (2016)	Loans from Q3 2007 to Q3 2015; 36 months maturity	Multinomial logistic regression	Bankruptcy fillings, debt-to-income ratio, federal funds, FICO score, GDP growth rate, home ownership (mortgage), interest rate, loan to annual income ratio, open credit lines, payment due to balance, payment due to income, public record, total credit lines
Carmichael (2014)	Loans from June 2007 to November 2013; 36 & 60 months maturity	Dynamic logistic regression	Annual income, credit grade (Lending Club), credit history length, FICO score, inquires in the last six months, loan amount, loan description, loan purpose, months since last delinquency, revolving credit utilisation, unemployment level, subgrade (Lending Club)
Serrano-Cinca et al. (2015)	Loans from January 2008 to December 2011; 36 months maturity	Survival analysis (Cox regressions) and logistic regression	Annual income, credit grade (Lending Club), credit history length, debt-to-income ratio, delinquency in past two years, homeownership, inquires in the last six months, loan purpose, open credit lines, revolving credit line utilisation

the determinants of default vary. In all papers, except of Li et al. (2016), credit grade is the most predictive determinant. Furthermore, debt-to-income ratio, FICO score and revolving credit line utilisation are mentioned in three studies.

The discrepancy between all four studies can be explained on the basis of three factors. Not all studies used the same variables and control variables as coefficients for the regression models. Small differences in the used research methods also bear reasons for various outcomes. Serrano-Cinca et al. (2015) used a survival analysis with 33 Cox regressions, followed by a logistic regression whereas Emekter et al. (2015) only focuses on a binary logistic regression. The third aspect why determinants of loan default differ is data itself. Different loan maturities, time frames and loan status are used which can lead to diverse outcomes. Information on the different studies and the predictors of default are summarized in Table 1.

2.2 Derived hypothesis

It is to assess the relevance of the information and data that the P2P lending platform provides for lenders in order to decrease the information asymmetry and make prudent investment decisions. Therefore, influential factors that determine the loan performance of P2P loans have to be analysed. One of the most important determinants is expected to be the credit grade that is assigned by Lending Club on the basis of the expected credit risk of a particular loan. It usually relies on information like the individual FICO score of a lender and data that was gathered from the credit file. The credit grade and shows the credibility and therefore an estimate about the loan default risk of a borrower. At Lending Club, seven credit grades from A (lowest risk) to G (highest risk) are assigned. On the basis of the credit grade, the interest rate for a loan is determined. Lenders and investors at P2P platforms expect a negative relationship between credit grades and default probability. With a higher credit grade and therefore more creditworthiness, the default risk decreases. Therefore, this study firstly sets the hypothesis:

H1: The higher the credit grading, the less likely is P2P loan default.

With H1 fulfilled, Lending Club assigns credit grades successfully and gives the lender an appropriate overview about the direction of loan performance she can expect from a particular loan in that loan category. The P2P lending platform delivers an opportunity to easily decrease the information asymmetry and gives the investor a better overview about the risk that is combined with loans.

Hypothesis 2 is designed in order to explore more drivers of loan performance. Many investors in P2P lending do not only rely on the credit grade but also take a closer look at other data and information offered by Lending Club in order to the loan performance. They actively use determinants of loan default like the debt-to-income ratio and annual income in order to predict whether an investment will deliver a success (loan is fully paid) or not (loan defaults). This study intends to find out which factors indeed determine the loan performance in past data of Lending Club. Therefore, the study also follows the hypothesis:

H2: The borrower and loan characteristics loan amount, annual income, debt-to-income ratio, inquires in the last 6 months, the number of open credit lines, revolving credit balance and the number of total credit lines are significant predictors of loan success in all risk classes.

The fulfilment of H2 indicates that loan performance is determined on borrower and loan characteristics other than the assigned Lending Club credit grade and not only the credit grade must be taken into account when evaluating credit risk.

3. METHODOLOGY

3.1 Regression Model

This paper mainly follows the approaches from Emekter et al. (2015), Serrano-Cinca et al. (2015), Carmichael (2014) and Li et al. (2016). Although there are many different statistical models in order to predict the default probability, most researchers are using logistic regressions in order to assess the determinants of loan performance for P2P loans and calculate the probabilities of default. The advantages of logistic regressions are good capabilities for predictions and a high accuracy (Thomas, 2010). This study uses the binary logistic regression in order to estimate the default determinants of P2P loans and tests the hypothesis whether a high credit grade indeed bears a significant lower P2P loan risk as well as the presumption that borrower characteristics are relevant factors for predicting the default probability of P2P loans.

With a logistic regression, the likelihood of an event depending on different variables is estimated (Hosmer & Lemeshow, 2013). Independent variables can be categorical or metric; dependent variables are dichotomous or multinomial. In the case of estimating the P2P loan performance, only two outcomes are possible – ‘ default’ and ‘fully paid’ (respectively ‘0’ or ‘1’). Therefore, the dependent variable is dichotomous and a binary logistic regression is run since linear regressions cannot deal with categorical dependent variables.

For the binary model, y and x are defined as:

$$P\,(y = 1 \mid x) = P\,(y = 1 \mid x_1, x_2, \dots, x_n) \qquad (1.1)$$

In this study, y = 1 is the response and therefore the dependent variable that shows the successful repayment of a loan. y = 0 is, respectively, defined as loan default. x is set as the full set of explanatory variables that are described in detail in Appendix 1 and contain, for instance, credit grade, loan amount and the annual income of the borrower. All explanatory variables and the related probabilities are summarized in variable d:

$$d_i = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon_i \qquad (1.2)$$

β_0 is constant, β_k are the regression coefficients. In order to convert the result of Equation 1.2 to a calculation of the probability of default that is between 0 and 1, the transformation function (T) in Equation 1.3 is needed:

$$\mathrm{P}\,(y = 1 \mid x) = T(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) = T(x\beta) \quad (1.3)$$

Therefore, the logistic function with the transformation included can be expressed as:

$$p_i = T(d) = \frac{1}{1 + e^d}\,, d = x\beta \qquad (1.4)$$

To estimate the linearity in the non-linear logistic model, the Maximum Likelihood Estimation (MLE) is used.

$$\frac{\partial \mathrm{T}(x\beta)}{\partial x_i} = \frac{dT\,(x\beta)}{d\,(x\beta)}\,\beta_i \qquad (1.5)$$

Equation 1.5 gives the effect of every variable x on the probability of default, namely the log Odds.

There are some assumptions that have to be fulfilled in order to ensure that the binary logistic regression is applicable in this case. Since there are only two possible outcomes (‘default’ and ‘fully paid’), the assumption of a dichotomous dependent variable is met. Both groups are mutually exclusive and exhaustive. Furthermore, all observations are independent from another. Whilst logistic regression does not assume linearity between independent and dependent variable, it assumes linearity between the independent variables and the log odds. This is satisfied since all variables were transformed to an ordinal level. Lastly, the Maximum Likelihood Method needs at a high amount of cases for each parameter. The independent variables generally satisfy this assumption, however, a small amount of only 37 cases of the variable *credit grade* for grade 'G' should be considered when only focusing on credit grade G loans.

Coefficients in a binary logistic regression cannot be interpreted as direct and linear influential factors on probabilities. The algebraic sign gives a first guess about the direction in which a coefficient pushes the calculated probability. In the model of this study, a negative coefficient shows a decrease in the probability of loan success and a positive coefficient an increase in the probability of loan success. The relationship between independent and dependent variable can further be described using the Odds ratio. With the Odds ratio, the probability for an event (in this case 'loan success') is set in relation to the probability that this event does not occur (in this case 'loan default'). The Odds ratio is expressed with the $Exp(\beta)$. For explanation purposes, this value is inversed by the logarithm function $\ln \beta$. It answers the question of how much the probability of full loan payment changes by adding one more unit of the corresponding coefficient, given all other variables in the model remain constant.

3.2 Variables

3.2.1 Dependent variable

Only matured loans are chosen and therefore the loan performance can only be 'default' or 'fully paid'. Since just two outcomes are possible, the logistic regression is dichotomous. 'Default' is set as y = 0, 'fully paid' is y = 1, as already explained. Default is defined as a status in which no more repayments of the credit from the borrower are incurred. At Lending Club, the platform of study, the status 'defaulted loan' is given when a loan is 121 days past due. It happens if a borrower is not able or willing to repay her credit any longer. The invested money cannot be regained. Therefore, it must be written off on lenders' side. Loan success occurs when matured loan was fully paid.

3.2.2 Explanatory variables

There are more than 120 variables that are provided in the loan statistics of Lending Club. However, not all are of interest for this study.

The variables of interest for this study are selected on the basis of the expected relevance and the results of past studies on that topic. A comprehensive summary of the 8 most relevant determinants found out in all four studies and expected to have an influence on the default probability is used as the explanatory variables. All are allocated to the categories of 'borrower characteristics' and 'loan characteristics'. Detailed descriptions of the explanatory variables and all control variables can be found in Appendix 1.

In order to test hypothesis H1, the *credit grade* assigned by Lending Club is the first variable that is used. The Lending Club credit grade is a general classification measure that bases on borrower characteristics and is calculated with a non-published algorithm by the platform itself. The credit grade is the most important determinant for the set interest rate. The higher the credit grade, the less risky a loan of that category should be.

The additional explanatory variables that are considered are two self-reported variables that are entered by the borrowers themselves when registering at Lending Club or applying for funding, *annual income* and the funded *loan amount*. Besides, also *debt-to-income ratio*, *inquires in the last 6 months*, the number of *open credit lines*, the number of *total credit lines* and the *revolving credit balance* are considered as variables of interest. Detailed descriptions on these variables can be found in Appendix 1. The borrower can usually influence the self-reported borrower characteristics; data from the credit report file that is gathered from credit bureaus is fixed. All chosen variables are expected to have an effect on the loan performance. Variables that are provided by Lending Club but not included in this study or selected as control variables are of minor interest since they are mostly only sub-categorical variables of the ones chosen. The selected variables provide an appropriate overview about the financial status quo and position of a borrower and allow a comprehensive sight onto the risk status and therefore the likely performance of loan and borrower.

In prior literature about the determinants of default probability in P2P loans, past studies already found out some influential variables. However, the results are not distinct and vary widely. An explicit overview about the determinants of loan performance provided by the four studies of Emekter et al. (2015), Li et al (2016), Carmichael (2014) and Serrano-Cinca et al. (2015) on which this paper relies is provided in Table 1. Differences in the results of the logistic regressions in all four cases can be explained on the basis of three main reasons as stated in part 2.1.3 about the determinants of default. The selected variables of this study are in accordance with the tendency that is provided by past studies concerning the variables and determinants of loan performance.

FICO score, also seen as an important determinant of default probability, however, is not included in this analysis. To the 1st of June 2017, Lending Club excluded the FICO score from the available loan statistics data. Therefore, an analysis whether the FICO score is significant for predicting default risk, cannot be made.

4. DATA

4.1 Data selection

This study uses data that is delivered online by Lending Club, the largest P2P platform in the United States. Lending Club was launched in 2006 and has issued loans of around $21 billion until May 2017.

From 2012 to 2013, 143,654 with a maturity of 36 months and an overall loan amount of over $1,7 billion were funded. At Lending Club, loans with a maturity of three or five years are available. For this study, only already matured loans will be analysed in order to get an insight into the actual loan performance. Thus, it is secured that every loan has a fixed loan status of 'default' or 'fully paid'. Loans are further comparable with each other's so that a basis for assessment is provided. All used data for this study is publicly available at www.lendingclub.com.

From the selected 143,654, 18,249 loans (12,7%) defaulted, 125,405 loans (87,3%) were fully paid (Table 2).

Borrowers at Lending Club can apply for loan funding for specific purposes. Appendix 2 shows the loan purpose distribution at Lending Club and indicates that most of the loans are supposed to be used for debt consolidation (56.9%) and credit card purposes (24.5%). The least popular reason for borrowing money at Lending Club is to spend it for renewable energy (0.1% of all loans). The highest default rate is observable in loans for small businesses (22.0%) whereas the highest loan success rate can be found in loan purposes regarding cars (90.0%), closely followed by major purchases (89,9%).

Table 2 shows how the loans are distributed among the different credit grades and gives the distributions of loan amounts. 19,3% of the loans belongs to credit grade A and therefore are expected to bear the lowest credit risk. Only 5.6% of all loans in category A defaulted.

Most issued loans fit to credit grade B (39,7%). The loans are also combined with the highest loan amount per grade, around $700 million (39.58%). The percentage of defaulted loans relative to total credit grade B loans is higher than in category A and equals to 10.60%. For credit grade C, D and E, this value increases to 15.96%, 20.61% and 22.99% respectively. The peak is reached in credit category F with 25,64% and decreases to the highest risk grade G 18,92%.

Out of the 18,249 defaulted loans in total, most (33,13%) stem from credit grade B loans. Only 1,554 loans from credit grade A defaulted. The highest amount of fully paid loans is also found in credit grade B with 51,004 loans out of the total of 57,050 loans were fully paid when matured.

Table 2: Frequency Distribution by Credit Grade

Grade	Frequency (%)	Amount (%)	Defaulted Loans (%)	Amount defaulted loans (%)	Fully paid loans (%)	Amount fully paid loans (%)	Ratio defaulted to all loans (%)
A (lowest risk)	27,767 (19.3)	$377,978,925 (21.27)	1,554 (8.52)	$19,910,225 (9.12)	26,213 (20.90)	$358,068,700 (22.97)	5.60
B	57,050 (39.7)	$703,508,625 (39.58)	6,046 (33.13)	$72,764,000 (33.34)	51,004 (40.67)	$630,744,625 (40.46)	10.60
C	34,531 (24.0)	$416,220,100 (23.42)	5,510 (30.19)	$64,254,575 (29.44)	29,021 (23.14)	$351,965,525 (22.57)	15.96
D	19,552 (13.6)	$222,941,150 (12.54)	4,029 (22.08)	$47,303,400 (21.67)	15,523 (12.39)	$175,637,750 (11.27)	20.61
E	4,011 (2.8)	$48,212,575 (2.71)	922 (5.05)	$11,768,725 (5.39)	3,089 (2.46)	$36,443,850 (2.33)	22.99
F	706 (0.5)	$7,589,300 (0.43)	181 (0.99)	$2,084,800 (0.96)	525 (0.42)	$5,504,500 (0.35)	25.64
G (highest risk)	37 (0.0)	$924,625 (0.05)	7 (0.04)	$174,525 (0.08)	30 (0.02)	$750,100 (0.05)	18.92
Total	143,654 (100.0)	$1,777,375,300 (100.0)	18,249 (100.0)	$218,260,250 (100.0)	125,405 (100.0)	$1,559,115,050 (100.0)	12.70

There is a special situation for credit grade G that is accompanied with the highest credit risk and therefore the worst loan performance. 18,92% of the loans defaulted which is less than in category D, E and F. From 2012 to 2013, only 37 loans with a maturity of 36 months were counted and so less than 0,1% of the overall amount of loans. The data have to be analysed with caution since bias is likely due to the small amount of analysed loans.

4.2 Descriptive statistics

Table 3 shows the descriptive statistics of the sample regarding the most relevant variables that are delivered by the Lending Club statistics. On the basis of the sample of 143,654 loans with a maturity of 36 months, the average loan amount is $12,372.61 and the average interest rate charged is 13.14%. Borrowers earn around $67,011.63 per year and have a debt-to-income ratio of 0.1672, 20 16.72%. 0.24 delinquencies in the last 2 years as well as 0.78 inquires in the past half year are the means. On average, the last delinquency per borrower was 35.26 months ago with a standard deviation of 21.56. The mean number of open accounts is 10.81; the number of total accounts is equal to 23.88. As the descriptive statistics show, the average revolving credit balance is $14,005.23 and the revolving credit line utilisation is 0.57738.

Appendix 3 compares the descriptive statistics of this study to prior literature on loan performance. Emekter et al. (2015) included loans with a maturity of 36 months and 60 months into their study, whereas Serrano-Cinca et al. (2015) and Li et al. (2016) only rely on loans with a term of 36 months, just as this study. In the cases of Emekter et al. (2015) and Serrano-Cinca et al. (2015), the sample sizes are much lower than the one of this study, however, both include a wider timeframe (May 2007 until June 2012 and January 2008 until December 2011, respectively). Li et al. (2016) used data from Q3 2007 until Q3 2015.

In the cases of Emekter et al. (2015) and Serrano-Cinca et al. (2015), the sample sizes are much lower than the one of this study, however, both include a wider timeframe (May 2007 until June 2012 and January 2008 until December 2011, respectively). Li et al. (2016) used data from Q3 2007 until Q3 2015.

In order to handle possible outliers and extreme values, the sample data of this study was winsorised for some variables before use. With winsorising, extreme values are limited in order to decrease the effect of outliers on the output. A 90% winsorisation was used. This means that the 5th percentile replaces all data below the 5th percentile and the 95th percentile replaces data above the 95th percentile. Especially differences between the minimum and maximum values of this study and Emekter et al. (2015) can be explained by the cleaned data. On the basis of very high values (e.g. $6,000,000 annual income) it is likely that Emekter et al. (2015) did not remove all outliers and trimmed the data. The study of Carmichael (2014) that also provides a view on default risk at Lending Club cannot be easily compared to this research since an exact number of loans and data to many other variables are not given.

Means and standard deviations of the variables of this study and prior studies are in most cases very much comparable with each other. However, major differences in the delinquencies in the last 3 years are observable. Different explanations are thinkable, e.g. a shift of the whole user base of Lending Club towards borrowers that were more delinquent than other studies suggest since the timeframe of this sample is more recent. It is also possible that lenders reacted more averse to borrowers with high delinquencies and did not fund these loan applications like in the past.

By comparing the descriptive statistics of variables of all four papers, similar observations can be made. Appendix 4 gives an overview about the loan distribution per grade of all four papers. Besides the small amount of values in the lowest credit grade (G), this study is expected to deliver similar results since data does vary little from prior researches data.

risk. However, credit grade F is an outlier in that case. The β value is little lower than the one of credit grade E, which indicates that default is less likely. On the other hand, Table 2 shows that the ratio of defaulted loans to all loans in the respective credit grade category is higher in grade F than in grade E. Different reasons for this unexpected outcome are thinkable. For example, the actual values for every variable in credit grade class F must be on average apparently higher than in grade class E in order to compensate for the lower standard risk. Since none of the explanatory variables in credit grade F remains significant, this guess must be tested further with more data, however this is beyond the scope of this study.

Table 3: Descriptive Statistics

Variable	Number of observations	Minimum	Maximum	Mean	Standard Deviation
Loan amount	143,654	1,000	35,000	12,372.61	7,286.83
Interest rate	143,654	0.0600	0.2589	0.1314	0.0391
Annual income	143,654	26,445	143,000	67,011.63	31,556.71
Debt-to-income ratio	143,654	0.00	0.3499	0.1672	0.0761
Inquires in the last 6 months	143,654	0	29	0.24	0.705
Months since last delinquency	61,372	0	8	0.78	1.013
Open credit lines	143,654	0	152	35.26	21.560
Revolving credit balance	143,654	1	62	10.81	4.571
Revolving credit utilisation	143,654	2,400	37,070	14,005.23	9,569.639
Total credit lines	143,654	0.156	0.920	0.57738	0.220622

Appendix 5 shows the correlation matrix of all explanatory variables that are non-categorical. It is based on the entirety of 143,654 observations for all credit grades. The highest correlation of 0.667 can be found between the *number of total credit lines* and the number of *open credit lines*. *Loan amount* and *annual income* bear the second largest correlation with 0.479. The lowest correlations are, on average, with the variable *inquires in the last 6 months*.

5. EMPIRICAL RESULTS

In order to determine the loan performance of P2P loans of different credit grades and to find the significant explanatory variables that can be regarded as 'determinants of loan performance', binary logistic regressions for all credit grades were run. Estimates are made on the basis of an iterative maximum likelihood method. This section describes the results of the regressions that show the effect of the variables on the likelihood of loan success and default for each credit grade.

Table 5 presents the results of the binary logistic regressions, including all described explanatory variables as well as some control variables[1] in order to assess the determinants for loan performance. In credit grade class 'A-G', the results for the full sample of loans and credit grades can be found. Out of the eight chosen variables, five variables (*credit grade*, *loan amount*, *annual income*, *debt-to-income ratio* and *inquires in the last 2 years*) are significant at the 1% level in most of the credit grade classes. *Number of open credit lines*, *revolving credit balance* and the *number of total credit lines* did not deliver clear results since all are significant in the overall regression for all credit grade classes, but in less than half of the single credit grade classes.

As expected, the results of the binary logistic regressions give evidence that a higher credit grade is accompanied with a lower

The binary logistic regression was run again for every credit grade class, respectively. By doing so, influential factors and the impact of those can be compared between different credit grades and the overall class of all grades. The results can be found in Table 5. Due to the very limited data on credit grade G and more parameters than cases, the binary logistic regression did not deliver information on other determinants than the intercept, so, the influence of the credit grade itself. There are also concerns about the reliability of the results for credit grade F since none of the expected influential factors is significant. Both grades, G and F, should be tested further using larger amounts of data in order to find reliable and significant coefficients and influential factors of default. In this study, both grades will be left out since results are not comparable with others.

As expected, the major influence on the loan performance stems from the credit grade that is provided by Lending Club. The intercept of 2.724 for the credit grade class 'A-G' refers to the highest credit grade 'A' that is associated with the lowest risk. Included in Equation 1.4, the base default probability of credit grade A is set at 6.16%. While using the coefficient of -.831 for credit grade C in the same class, a base default rate of 13.09% is calculated. When taking single credit grade-classes into account, the probabilities differ (7.56% in grade 'A', 11.44% for class 'C', for example). As expected, the results suggest evidence that the higher the credit grade, the less risky is the loan. The logarithm function indicates that a decrease in credibility from credit grade A to B in class 'A-G' lowers the loan success chance by 47,97% and accordingly increases the loan default probability.

Binary logistic regression results show a descending order from loan grade A to E concerning the likelihood of loan success. For

[1] As control variables, the following are used: delinquencies in the last 2 years, revolving utilisation line, total current balance of all accounts, total revolving credit limit, total credit limit, total credit balance excluding mortgage, total bankcard credit limit, average current balance of all accounts

these credit grades, a higher credit class is connected with a lower default probability. Hypothesis 1 is therefore supported. However, results for credit grades F and G are diverse since the default risk of credit grade F is lower than in E and, regarding the single grade classes, even lower than in D and C.

In all credit grade categories is the loan amount coefficient negative. This indicates decreasing probabilities for loan success with higher amounts of money lend. The coefficient seems to be quite low and therefore is associated with low influence, however, has to be interpreted on a financial basis. With one more US-Dollar added to the loan amount, the probability of loan success decreases in credit grade class 'A-G' by 0.0026%, given all other coefficients remain the same. The small contribution per US-Dollar to the probability is due to typically high values of loan amounts. Usually, loans differ from one another in terms of hundreds and thousands of Dollars; the impact of one US-Dollar is marginal. The loan amount, however, has a larger influence on

Table 5: Binary Logistic Regression Results

Variable	Credit grade class							
	A-G	A	B	C	D	E	F	G
	β ($\ln\beta$)							
Intercept	2.724*** (2.724)	2.503*** (2.503)	2.095*** (2.095)	2,047*** (2.047)	1.752*** (1.752)	1.144*** (1.144)	2.167** (2.167)	1.099 (1.009)
Grade (B)	-.480*** (-.480)							
Grade (C)	-.831*** (-.830)							
Grade (D)	-1.115*** (-1.115)							
Grade (E)	-1.275*** (-1.277)							
Grade (F)	-1.180*** (-1.181)							
Grade (G)	-1.356 (-1.355)							
Loan Amount	-.000026*** (-.000026)	-.000034*** (-.000034)	-.000019*** (-.000019)	-.000021*** (-.000021)	-.000036*** (-.000036)	-.000030*** (-.000030)	-.000018 (-.000018)	
Annual Income	.000008*** (.000008)	.000006** (.000006)	.000008*** (.000008)	.000008*** (.000008)	.000008*** (.000008)	.000006** (.000006)	.000008 (.000008)	
Debt-to-Income	-.014*** (-.015)	-.026** (-.025)	-.017*** (-.016)	-.012*** (-.012)	-.017*** (-.017)	-.002 (-.002)	-.010 (-.010)	
Inquires in the Last 6 Months	-.086*** (-.086)	-.045 (-.045)	-.071*** (-.071)	-.114*** (-.114)	-.062** (-.062)	-.103** (-.102)	-.141 (-.140)	
Open Credit Lines	-.027*** (-.027)	-.023 (-.023)	-.009 (.009)	-.034*** (-.034)	-.039*** (-.040)	-.027 (-.027)	-.062 (-.061)	
Revolving Credit Balance	.000009** (.000009)	.000005 (.000005)	.000008 (.000008)	.000014 (.000010)	.000008 (.000008)	-.000006 (-.000006)	.000043 (.000043)	
Total Credit Lines	.007*** (.007)	-.003 (-.003)	.005 (.005)	.008*** (.008)	.012*** (.012)	.007 (.007)	-.008 (-.008)	
N	143,654	27,767	57,050	34,531	19,552	4,011	706	37

Notes: The highest credit grade (grade A) is used as the base value. The probability of default for loans with credit grade A therefore is determined by the Intercept of 2.2687. 'Grade (B)' shows the determinant for credit grade B. * shows significance at the 1% level and ** indicates significance at the 5% level. Due to the small sample, no data could be gathered for the binary logistic regression of credit grade class 'G'.**

the success probability when values for this coefficient are high. Taking a closer look at the loan amount coefficients indicates the highest contribution per US-Dollar on the loan performance in credit category D with an $\ln(\beta)$ of -0.000036, so a decrease of loan success likelihood of 0.0036% with one added US-Dollar. The coefficients of annual income are to interpret on the same basis, however values of annual income are expected to be higher than loan amount since the mean of annual income is about 5.5 times higher than the mean of the loan amount (Table 3). All annual income coefficients in every credit grade class are positive. This indicates a positive influence on the probability of loan success. A higher annual income, therefore, decreases the loan risk. In all credit classes, the annual income coefficient is either 0.00008 or 0.00006, so the influence on loan success is relatively equal, independently from credit grade.

A higher *debt-to-income ratio* decreases loan success probability in all credit grades, except of grade E. The regressions nevertheless draw a coherent picture. One more percentage point in *debt-to-income ratio* decreases the loan success probability by 2.53% in credit grade A, which is the highest effect *debt-to-income ratio* has among all categories. In all other credit grade categories, the effect is inferior. A similar outcome can be observed for *inquires in the last 6 months*. The loan success probability decreases with an increase of inquires. All coefficients with the exception of credit grade A are significant. Differences from -.114 (grade C) to -.062 (grade D) indicate that the number of inquires is not equally influential among the different credit grades.

Open credit lines and *total credit lines* are only significant predictors for loan success when taking all loans (‘A-G’) or the single grades C and D into account. In all other classes, coefficients remain insignificant. The magnitude coefficients for the number of *open credit lines* are generally higher than for the *total number of credit lines*, possibly because the average is also more than doubled going from *open credit lines* to *total credit lines*. Therefore, one more *open account* is more influential on the loan success probability than one more *credit line in total*. However, both variables are changing the success probability in different directions. While an increasing number in *open credit lines* decreases the loan success likelihood, the *total number of accounts* is positively related to loan success. One more *open credit line* increases loan default probability in credit grade E by 3.98%. In the same grade class, with one more *credit line*, the loan success probability increases by 1.19%. The more *open credit lines* a borrower has, the less likely is her paying off the loan. However, the more *credit lines* a borrower has in total, the more likely she is paying off the loan. Actual *credit lines* therefore have a negative influence while the number of *total credit lines* is positively related to loan success.

Revolving credit balance is only a significant predictor for the loan grade class ‘A-G’. One more US-Dollar increases the loan success probability by 0.0009%, given all other variables remain constant. Since the binary regression results for single credit grades are insignificant, the effects must be handled with caution.

The coefficients are screwed over the different credit grades and patterns for the distribution of highest and lowest coefficient values are not observable. However, it can be stated that all coefficients provide a clear tendency that is given by the algebraic sign and does not change over the various credit grades. All tested variables are significant for predicting loan success when taking into account the whole sample of loans (credit grade class ‘A-G’). *Loan amount*, *annual income*, *debt-to-income ratio* and *inquires in the last 6 months* are significant influential factors for loan success in (nearly) all credit grades. For these variables, hypothesis 2 is supported. Since the *number of open* and *total credit lines* and *revolving credit balance* are not significant in many of the credit grades, they cannot be regarded as explanatory variables with an influence on loan success. For these three borrower characteristics, hypothesis 2 is not confirmed.

6. CONCLUSION

Credit risk is the most important concern when it is about to assess the loan performance of P2P loans. This study analysed 143,654 matured P2P loans from the P2P platform Lending Club that were funded between 2012 and 2013 and finds the determinants of loan performance that have an influence on the probability of loan success.

The study suggests that loan performance can be explained by five different influential factors, namely the borrower’s *credit grade* that is assigned by Lending Club, the *loan amount, annual income* of the borrower, *debt-to-income ratio* and *inquires in the last 6 months*. The *numbers of open and total credit lines* as well as the *revolving credit balance* are, contrary to the hypothesis, not significant for every single credit grade.

The most relevant influential factor for loan performance with the highest effect is the *credit grade*. With a lower credit grade, the probability of successful loan payment decreases; therefore the default risk is higher. The relationship between *loan amount* and successful loan repayment is negative: the higher the loan, the less likely is repayment. The exact opposite describes the relationship between *annual income* and loan performance. More annual income increases the chances for successful repayment in every credit grade category. Increases in *debt-to-income ratio* and *inquires in the last 6 months* have negative impact on the loan repayment probability.

Prior already identified some determinants of loan performance. However, none of these papers investigated the effects of the determinants in different loan credit grade classes. This analysis gives a deeper understanding of the predictors of loan performance and is generally in-line with prior studies, however, did not find evidence for some influential factors that were previously declared as significant predictors. The study further suggests that borrower characteristics and credit grade can be used in order to determine the loan performance. However, especially for the lower credit grades, further studies have to be undertaken to also find significant determinants in these cases.

7. REFERENCES

Boot, A. W., & Thakor, A. V. (1994). Moral hazard and secured lending in an infinitely repeated credit market game. *International Economic Review*, 899-920.

De Roure, C., Pelizzon, L., & Tasca, P. (2016). How does P2P lending fit into the consumer credit market?.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *The review of economic studies*, *51*(3), 393-414.

Diamond, D. W., & Dybvig, P. H. (1986). Banking theory, deposit insurance, and bank regulation. *The Journal of Business*, *59*(1), 55-68.

Edelberg, W. (2004). Testing for adverse selection and moral hazard in consumer loan markets.

Emekter, R., Tu, Y., Jirasakuldech, B., & Lu, M. (2015). Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. *Applied Economics*, *47*(1), 54-70.

Everett, C. R. (2015). Group membership, relationship banking and loan default risk: the case of online social lending.

Galloway, I. (2009). Peer-to-peer lending and community development finance. *Community development investment center working paper*, (39), 19-23.

Gorton, G., & Winton, A. (2003). Financial intermediation. *Handbook of the Economics of Finance*, *1*, 431-552.

Guo, Y., Zhou, W., Luo, C., Liu, C., & Xiong, H. (2016). Instance-based credit risk assessment for investment decisions in P2P lending. *European Journal of Operational Research*, *249*(2), 417-426.

Iyer, R., Khwaja, A. I., Luttmer, E. F., & Shue, K. (2009). Screening in new credit markets: Can individual lenders infer borrower creditworthiness in peer-to-peer lending?.

Herzenstein, M., Dholakia, U. M., & Andrews, R. L. (2011). Strategic herding behavior in peer-to-peer loan auctions. *Journal of Interactive Marketing*, *25*(1), 27-36.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Huang, R. H. (2017). Online P2P Lending and Regulatory Responses in China: Opportunities and Challenges

Käfer, B. (2016). *Peer to peer lending: A (financial stability) risk perspective* (No. 22-2016). Joint Discussion Paper Series in Economics.

Kirby, E., & Worner, S. (2014). Crowd-funding: An infant industry growing fast. *IOSCO, Madrid*.

Lee, E., & Lee, B. (2012). Herding behavior in online P2P lending: An empirical investigation. *Electronic Commerce Research and Applications*, *11*(5), 495-503.

Li, Z., Yao, X., Wen, Q., & Yang, W. (2016). Prepayment and Default of Consumer Loans in Online Lending.

Lin, M. (2009). Peer-to-peer lending: An empirical study. *AMCIS 2009 Doctoral Consortium*, 17.

Lin, M., Prabhala, N. R., & Viswanathan, S. (2013). Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. *Management Science*, *59*(1), 17-35.

Mateescu, A. (2015). Peer-to-Peer Lending. *Data & Society Research Institute*, 19-25.

Milne, A., & Parboteeah, P. (2016). The Business Models and Economics of Peer-to-Peer Lending.

Morse, A. (2015). Peer-to-peer crowdfunding: Information and the potential for disruption in consumer lending. *Annual Review of Financial Economics*, *7*, 463-482.

Serrano-Cinca, C., Gutiérrez-Nieto, B., & López-Palacios, L. (2015). Determinants of default in P2P lending. *PloS one*, *10*(10), e0139427.

Thomas, L. C. (2010). Consumer finance: Challenges for operational research. *Journal of the Operational Research Society*, *61*(1), 41-52.

U.S. Treasury, (2016). Opportunities and Challenges in Online Marketplace Lending. *Washington, DC, http://bit. ly/27bRLDC*.

Verstein, A. (2011). The misregulation of person-to-person lending.

Wei, S. (2015). Internet lending in China: Status quo, potential risks and regulatory options. *Computer Law & Security Review*, *31*(6), 793-809.

Yu, F., FY. (2017, January 08). Chinese P2P Lending Bubble Quietly Bursts. *The Epoch Times*, Retrieved from http://www.theepochtimes.com/n3/2208086-chinese-p2p-lending-bubble-quietly-bursts/

8. APPENDIX

Appendix 1: Descriptions of Variables

Variable	*Description of variable*
Annual Income (self-reported)	Every borrower provides information about her annual income during registration at Lending Club.
Home Ownership (self-reported)	Lending Club offers three statuses that are self-reported by the borrower during their registration: own, mortgage, rent.
Loan Amount (self-reported)	The amount of money that the borrower is applying for to get funded.
Loan Purpose (self-reported)	When applying for a loan, borrowers can choose between 13 category purposes for which the loan is used: debt consolidation, credit card, home improvement, car, house, major purchases, medical, moving, renewable energy, small business, vacation, wedding, other.
Credit Grade	The credit grade is set in respect of the default risk of a borrower and loan that is calculated by Lending Club and ranges between A (lowest risk) and the highest credit grade G (highest risk). For explaining credit risk in the credit grade, Lending Club uses several variables and the credit grade is calculated with a non-public algorithm. Basis of the credit grade is a borrowers' individual FICO score. Further adjustments are taken on the basis of the number of accounts currently open, credit utilisation, number of recent credit inquires, length of credit history, ratio of loan requested to established guidance limits and term of the loan. The interest rate of a specific loan is set on the basis of the credit grade.
Debt-to-Income Ratio	Monthly debt payments, excluding mortgage and Lending Club loan, are divided by the monthly income of the borrower. The maximum debt-to-income ratio that allows borrowers to apply for loan funding is 0.3499.
Delinquencies in the Last 2 Years	The number of delinquencies in the last 2 years is part of the credit file of each borrower and shows the number of incidences in which a borrower was 30 days or more past due paying a loan off.
Inquires in the Last 6 Months	The variable shows how many inquires were counted by officials in the past half year, excluding auto and mortgage inquiries. The variable is part of the borrower's credit file.
Months since Last Delinquency	Months since the last delinquency of a borrower was recorded in the credit file.
Open Account	The number of total credit lines that is reported in the credit file.
Total Account	The number of current credit lines that is reported in the credit file.
Revolving Credit Balance	The total credit revolving balance that is part of the credit file of each borrower. It describes the amount of money that is not paid at the end of a billing cycle, so is the credit that offers a credit institute on a credit card.
Revolving Credit Utilisation	The revolving credit line utilisation is described as the ratio of the amount of credit a borrower is using in respect to the total amount that is available as a revolving credit.

Appendix 2: Loan Purposes and Frequencies

Loan purpose	*Frequency (%)*	*Defaulted (%)*	*Fully paid (%)*
Debt consolidation	81,806 (56.9)	10,700 (13.1)	71,106 (86.9)
Credit card	35,188 (24.5)	3,817 (10.8)	31,371 (89.2)
Home improvement	7,716 (5.4)	858 (11.1)	6,858 (88.9)
Car	1,620 (1.1)	162 (10.0)	1,458 (90.0)
House	796 (.6)	106 (13.3)	690 (86.7)
Major purchase	2,952 (2.1)	299 (10.1)	2.653 (89.9)
Medical	1,348 (.9)	219 (16.2)	1,129 (83.8)
Moving	944 (.7)	147 (15.6)	797 (84.4)
Renewable energy	107 (.1)	19 (17.8)	88 (82.2)
Small business	2,099 (1.5)	462 (22.0)	1,637 (78.0)
Vacation	850 (.6)	131 (15.4)	719 (84.6)
Wedding	1,162 (.8)	149 (12.8)	1,013 (87.2)
Other	7,066 (4.9)	1,180 (16.7)	5,886 (83.3)
Total	143,654 (100.0)	18,249 (12.7)	125,405 (87.3)

Appendix 3: Descriptive Statistics in Comparison to Prior Studies

Variable	Number of observations	Minimum	Maximum	Mean	Standard Deviation
Loan amount	143,654	1,000	35,000	12,372.61	7,286.83
Emekter et al. (2015)	*61,451*	*500*	*35,000*	*11,604.20*	*7,575.7465*
Serrano-Cinca (2015)	*24,449*	*///*	*///*	*9,449*	*6,253*
Li et al. (2016)	*///*	*///*	*///*	*///*	*///*
Interest rate	143,654	0.0600	0.2589	0.1314	0.0391
Emekter et al. (2015)	*61,451*	*0.0542*	*0.25*	*0.1240*	*0.0393*
Serrano-Cinca (2015)	*24,449*	*///*	*///*	*0.110*	*0.032*
Li et al. (2016)	*160,956*	*///*	*///*	*0.1272*	*0.0386*
Annual income	143,654	26,445	143,000	67,011.63	31,556.71
Emekter et al. (2015)	*61,451*	*0.00*	*6,000,000.00*	*69,563.4*	*5,212.4929 (monthly)*
Serrano-Cinca (2015)	*24,449*	*///*	*///*	*67,432*	*66,843*
Li et al. (2016)	*///*	*///*	*///*	*///*	*///*
Debt-to-income ratio	143,654	0.00	0.3499	0.1672	0.0761
Emekter et al. (2015)	*61,451*	*0.00*	*0.35*	*0.1381*	*0.0677*
Serrano-Cinca (2015)	*24,449*	*///*	*///*	*0.1286*	*0.0668*
Li et al. (2016)	*143,719*	*///*	*///*	*0.1603*	*0.0768*
Delinquencies 2 years	143,654	0	29	0.24	0.705
Emekter et al. (2015)	*61,422*	*0*	*13*	*0.1469*	*0.5107*
Serrano-Cinca (2015)	*24,449*	*///*	*///*	*0.15*	*0.49*
Li et al. (2016)	*///*	*///*	*///*	*///*	*///*
Inquiries in last 6 months	143,654	0	8	0.78	1.013
Emekter et al. (2015	*61,442*	*0*	*33*	*0.9914*	*1.3923*
Serrano-Cinca (2015)	*24,449*	*///*	*///*	*0.85*	*1.06*
Li et al. (2016)	*///*	*///*	*///*	*///*	*///*
Months since last delinquency	61,372	0	152	35.26	21.560
Emekter et al. (2015)	*21,749*	*0*	*120*	*36.0016*	*22.1773*
Serrano-Cinca (2015)	*24,449*	*///*	*///*	*33.64*	*22.40*
Li et al. (2016)	*///*	*///*	*///*	*///*	*///*
Open credit lines	143,654	1	62	10.81	4.571
Emekter et al. (2015)	*61,442*	*1*	*49*	*9.5593*	*4.45*
Serrano-Cinca (2015)	*24,449*	*///*	*///*	*9.13*	*4.40*
Li et al. (2016)	*160,956*	*///*	*///*	*10.61*	*4.75*
Revolving credit balance	143,654	2,400	37,070	14,005.23	9,569.639
Emekter et al. (2015)	*61,422*	*0*	*1,207,359*	*14,315.60*	*19,741.3993*
Serrano-Cinca (2015)	*///*	*///*	*///*	*///*	*///*
Li et al. (2016)	*///*	*///*	*///*	*///*	*///*
Revolving credit utilisation	143,654	0.156	0.920	0.57738	0.220622
Emekter et al. (2015)	*61,338*	*0*	*1.19*	*0.5156*	*0.2778*
Serrano-Cinca (2015)	*24,449*	*///*	*///*	*0.46*	*0.28*
Li et al. (2016)	*160,956*	*///*	*///*	*0.54*	*0.25*
Total credit lines	143,654	2	105	23.88	11.025
Emekter et al. (2015)	*61,422*	*1*	*90*	*22.2256*	*11.3375*
Serrano-Cinca (2015)	*///*	*///*	*///*	*///*	*///*
Li et al. (2016)	*160,956*	*///*	*///*	*24.05*	*11.58*

Appendix 4: Loans and Credit Grade Distribution

	A	*B*	*C*	*D*	*E*	*F*	*G*	*Total*
This sample	27,799 (19.3%)	57,067 (39.7%)	34,537 (24.0%)	19,557 (13.6%)	4,013 (2.8%)	709 (0.5%)	37 (<0.1%)	143,719
Emekter et al. (2015)	15,015 (24.4%)	18,707 (30.4%)	12,545 (20.4%)	8,317 (13.5%)	4,513 (7.3%)	1,746 (2,8%)	608 (1.0%)	61,451
Serrano-Cinca et al. (2015)	7,901 (32.3%)	7,757 (31.7%)	4,927 (20.2%)	2,826 (11.6%)	785 (3.2%)	198 (0.8%)	55 (0.2%)	24,449
Li et al. (2016)	34,347 (21.3%)	57,077 (35,4%)	39,912 (24.8%)	22,007 (13.7%)	5,890 (3.7%)	1,411 (0.9%)	312 (0.2%)	160,956

Appendix 5: Correlation Matrix

	Annual income	*Loan amount*	*Debt-to-income ratio*	*Inquires in the last 6 months*	*Open credit lines*	*Revolving credit balance*	*Total credit lines*
Annual income	1	.479***	-.234***	.098***	.224***	.431***	.332***
Loan amount	.479***	1	.008***	-.002	.189***	.470***	.222***
Debt-to-income ratio	-.234***	.008***	1	.013***	.297***	.223***	.226***
Inquires in the last 6 months	.098***	-.002	.013***	1	.126***	-.015***	.152***
Open credit lines	.224***	.189***	.297***	.126***	1	.327***	.667***
Revolving credit balance	-431***	.470***	.223***	-.015***	.327***	1	.307***
Total credit lines	.332***	.222***	.226***	.152***	.667***	.307***	1