Master's Thesis

A Comparison of Customers` Profile and Default Probabilities of Consolidated and Non-consolidated Loans in the Peer-to-peer Lending Market.

by

ISLAM Muhammad Rofiqul

52117612

June 2019

Master's Thesis Presented to

Ritsumeikan Asia Pacific University

In Partial Fulfillment of the Requirements for the Degree of

Master of Business Administration

Table of Contents

List of Tables

List of Figures

Certification Page

I, ISLAM Muhammad Rofiqul (Student ID 52117612) hereby declare that the contents of this Master's Thesis are original and true, and have not been submitted at any other university or educational institution for the award of degree or diploma.

All the information derived from other published or unpublished sources has been cited and acknowledged appropriately.

ISLAM Muhammad Rofiqul

2019/06/29

Acknowledgments

The successful completion of this master thesis may be considered as a remarkable step in my academic career which I believe would not be possible with the active support of the academic supervisor, the university authority, friends and my family members who deserve acknowledgment for their support.

First, I would like to express my sincere gratitude to my academic supervisor Professor OTAKE Toshitsugu who guides me all the way of my thesis form the initial idea generation to final interpretation of the results. I always had easy access to him whenever it is required to discuss issues related to the thesis. I should acknowledge his inspiration that always made my work easy.

Second, The Ritsumeikan Asia Pacific University deserves appreciation for the available resources like the statistical software and the free access to online resources without which this research would be more difficult to be completed.

Finally, I would like to express my gratitude to my parents who believe in myself and keep confidence in my capacity.

Summary

The recent emergence of peer-to-peer lending through online platform shows a significant growth both in terms of its number and geographical expansion. Both the lenders and borrowers are sharing the benefits of no intermediation cost. However, for this added benefits individual or group lenders need to bear the credit risk which is generally taken by conventional financial intermediaries like banks or credit unions. Like formal financial institutions, the lending decision, as well as pricing, depends on the financial, demographic and social characteristics of the borrowers. Lenders in P2P market infer creditworthiness of the borrowers from some structured financial variables of the borrower. This credit assessment process is not significantly different from the traditional banking system. With the growth of the peer-to-peer lending market, it is observed that a significant portion of all the loans is taken to pay off existing loans with formal financial institutions. The growth of the P2P lending market is induced by its ability to offer services with low transaction cost and supported by the competitive market theory. However, the long-term sustainability of the whole P2P lending system largely depends on its capacity to measure and manage the risk associated with lending activities. The major risk develops by the existence of information asymmetry between the lenders and the borrowers transacting in the online P2P lending market. The lending platforms as the market maker of the system employ their efforts to reduce the information asymmetry so that the adverse selection problem is reduced. The risk grading assigned by the lending platforms is the main effort to reduce the information asymmetry which helps the investors in taking a rational decision. Despite their efforts to reduce the information asymmetry problem, the scholars are contributing to their

research to help the stakeholders for taking a rational investment decision.

Our study aims to analyze the consolidated loans in terms of their customers ` profile as well as default probabilities. For this, we formulate our first research question to explore the contrasting features of the customers taking loans for their primary purposes and of those who are taking loans to top-up their existing loan liabilities with other financial institutions. Secondly, we find the determinants of the default risk for the group of the customers to find out any significant differences among them. For our analysis, we have the collected secondary data from the lending club, one of the prominent P2P lending platforms of the U.S.A. To address the first research question, we use descriptive statistic and test statistic for comparing means of variables. We used a binary logistic regression model to measure the default probability. Results of our study show that consolidated borrowers are getting funded with worse credit grade and higher debt burden as compared with the non-consolidated borrowers. The default rate, as well as determinants of the default probability of low-credit grade consolidated borrowers, is higher than that of low-credit grade non-consolidated borrowers. In conclusion, the lending platforms need to develop a separate model to assess the default probabilities of low-credit grade borrowers rather than grouping them with other non-consolidated borrowers under the same credit grade.

Chapter 1 - Introduction

1.1 Introduction

1.1.1 Study background

The emergence of web 2.0[1] technology has induced the rapid establishment of the online markets and the virtual community where an individual can interact virtually to meet their needs. Like the virtual market, in the online peer-to-peer (P2P) lending market the borrowers and the lenders meet virtually through an online platform for processing a lending transaction without a formal conventional financial intermediary. The Lenders and the borrowers in the P2P lending market can share the savings from the traditional intermediation cost. However, the lenders bear the default risk of the borrowers in case of loan default. To deal with the potential default risk of the borrowers, the lenders face the asymmetric information problem where the lenders lack information of borrowers which hampers taking prudent lending decision by the lenders which lead to adverse selection problem on the part of the lenders. Banking theory shows that traditional financial intermediaries like banks, credit unions, etc. can reduce some of the adverse selection problems through hiring expert executives, obtaining guarantees and collaterals, and ensuring post disbursement monitoring (Akerlof, A., 1970). Unlike the traditional financial market, in the online P2P market environment information it is difficult to reduce the effects of the information asymmetry due to the high transaction cost and for the limitations of the virtual environment.

[1] Wikipedia defines the web 2.0 as the website which emphasizes on the user-generated contents, the ease of use, the participatory culture, and the interoperability for end uses.

The lending platforms in the P2P lending market take initiatives to identify trustworthy borrowers to reduce the lending risk associated with information asymmetry. Firstly, platforms use their own screening system to drive out some of the potential borrowers based on some set thresholds. For example, the lending club[2] uses a floor on the FICO score[3] and below that FICO score customers are not able to be listed with the platform. Secondly, to reduce the risk exposures of individual borrowers, platforms set a ceiling of lending limit. Presently the lending club sets the highest limit of USD 35,000 for individual borrowing which makes enables investors to dilute risk among different borrowers. Thirdly, platforms offer portfolio recommendation services for the investors. Platforms with their expertise and scalability can better understand the risk level of borrowers and generate workable recommendation mechanism. Like many other lending platforms, the lending club assigns credit grades and subgrades to its potential borrowers which the investors use as a recommendation considering the risk level of the borrowers. The ultimate investors can reduce the adverse effect of information asymmetry by using the platform assigned credit grades for their lending decision making. In addition to the above, the lending platform provides the management services of the non-performing loans (NPLs). They help investors to engage in collecting agency for recovering the NPLs and to assist in getting legal services for NPL litigation. Platforms, as market makers employ various tools and methods to reduce the problem of information asymmetry for better risk management of the P2P loans nevertheless the credit risks associated with the P2P loans, have not been eliminated

[2] Lending club is one of the largest online lending platforms of the U.S.A. For more information: https://www.lendingclub.com/

[3] FICO score is a credit score representing the credit risk of an individual borrower where FICO stands for Fair Isaac Corporation which is a data analytics company to assign FICO score for consumer credit risk in the USA. For detail information visit: https://www.fico.com/

and there are rooms for further improvement of the decision-making capacity of the investors as well as the market makers.

The pioneer market maker in the online P2P market is ZOPA of the UK which has started its platform business in 2005 followed by the prosper.com of the USA in 2006. Since then the P2P lending market shows a significant growth both in terms of a number of customers and in term of the amount of the transactions. In 2018, one of the leading P2P lending platforms of the USA, the lending club facilitated 6,76,460 loans with a total amount of USD 8.84 b. Borrowers get investment for the purposes ranging from the weddings, the mortgages to the set off the existing loans with other financial institutions. Data from the lending club in 2018 shows that the consolidated loans constitute the highest proportion of the loans with a 57% contribution to the total portfolio of the lending club. Loan purposes are observed as one of the explanatory variables for the default prediction (Serrano-Cinca, C., et. al., 2015). The consolidated loans are expected to have different default risk profile as compared with loans taken for purposes other than loan consolidation. With this hypothesis in mind, this study analyzes the consolidated loans and non-consolidated loans as two separate groups of loans to explore general attributes and to develop two different models for default risk prediction.

1.1.2. Defining the Peer to Peer Lending

The Peer to Peer lending (hereinafter refers as the P2P) can be considered as a type of direct financing mode where the funds flow from a surplus economic unit to a deficit economic unit where lenders and borrowers meet each other through a virtual platform for a loan transaction. The concept of direct finance i.e. private lending is not a new concept in banking literature rather the private one-to-one lending existed well before the origination of the formal banking system. However, the new dimension of this type of private finance uses the internet to interact with borrowers and investors. This marketplace is characterized by the absence of the traditional financial intermediaries which exist between the borrowers and the lenders to facilitate the transaction and to bear the credit risk of the borrowers, however, in the P2P lending, the borrowers using the online P2P lending platform directly meet the lenders and negotiate the loan transactions. The role of P2P lending platforms which may be considered as an intermediary in this market is to facilitate the borrowers and the lenders to meet virtually and to mature the transactions without bearing the credit risk of the borrowers by them. Unlike the traditional lending market where the financial intermediaries bear the credit risk of the borrowers the lending platforms in the P2P marketplace only facilitate transactions and the entire credit risk of the borrowers is taken by the lenders. The P2P lending is a process of establishing a borrower-lender contract through removing the middleman from the process which enables them to save the cost of using a financial intermediary and share this cost-saving benefits by both the lenders and borrowers. The borrowers get access to the credit with a low-interest rate than the traditional credit market and the lenders earn an interest higher than the traditional bank deposits.

The online P2P lending market is considered as the other platform-based businesses where the buyers and sellers interact online to exchange the desired products or services for a price set by a digital auction or at a fixed price or at a price set by the dynamics of the demand and the supply of the market. However, the P2P lending platform business is characterized by a set of distinct features which are absent in the other platform business models like the UBER, the Airbnb, the trivago.com, etc. Firstly, the lending platforms provide financial advice in the form of initial processing of financial information of the borrower. For example, the Lending Club assigns a credit grade to each customer which is considered as financial advice regarding the risk, return, and profitability of the borrower. Secondly, the lending platforms develop a special type of principal-agent relationship under which the lending platforms collect the installment payments from the borrowers and transfer the money to the investors. Thirdly, the lending platforms provide the account management services for the investors enabling them for any subsequent reinvestment or resell of the product to the third party (Davis, K. 2016).

The potential borrowers in the online P2P lending market apply for a loan with an online P2P lending platform like the lendingclub.com by providing the certain required information, Then the platform undertakes initial screening based on some set criteria and only eligible borrowers are got listed with the platform. The platform then assigns a credit grade (some cases with subgrades) considering the risk level of the borrowers. Investors can see the potential borrowers' information along with the credit grades assigned by the lending platform. With this information, investors choose acceptable borrower from the alternatives. Borrowers pay an agreed interest rate on the loan availed and the investors get the agreed interest for the fund they have invested. The lending platforms charge a fee for the services they provide to their customers.

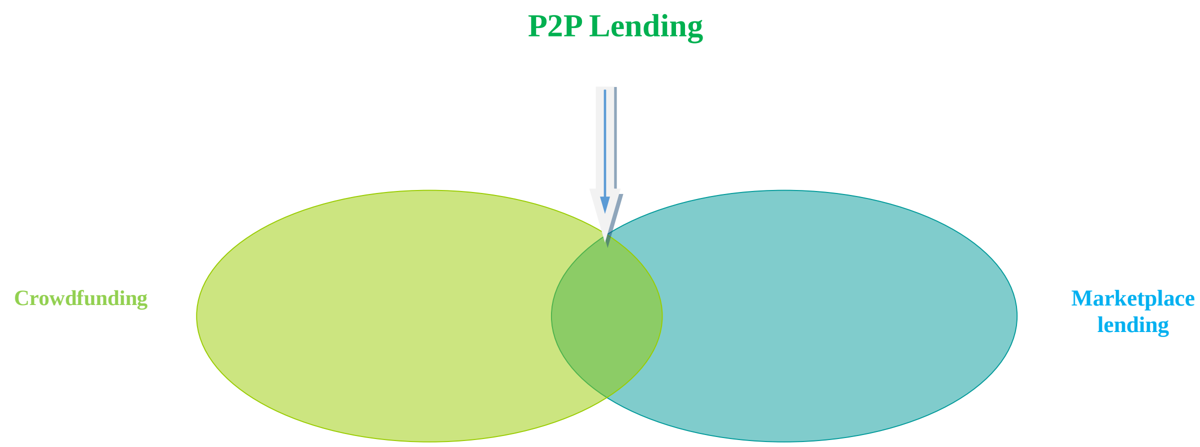

Figure 1 : Comparing the P2P lending with the crowdfunding and the marketplace lending.

The above figure-1 describes the P2P lending showing its similarities with the crowdfunding and the marketplace lending. The P2P are comparable both to the crowdfunding and the marketplace lending however it is different from both the other modes of finance. In crowdfunding, many individuals or group investors general a pool of small capitals for various purposes via internet platforms. On the other hand, in the marketplace lending, the institutional investors extend business credit to other businesses via internet-based platforms. The P2P lending system is characterized by individual investors who lend a small amount of money to other individuals via internet platforms. In this market, loans can also be used for business purposes. The P2P market also allows institutional investors to invest through the online platform. The P2P lending is better described as a hybrid model of finance that bears comparable fears with the crowdfunding and the marketplace lending.

The growth of P2P lending in recent years is the result of its perceived cost and other advantages over traditional lending modes. The fundamental driving force is the exploitation of new technology where the internet facilitates to eliminate the need for the financial intermediary by allowing individual borrowers and lenders to communicate directly to negotiate loan transactions. However, other competitive advantages also contribute to the rapid growth of the P2P market. These competitive advantages can be clubbed under four main categories: (i) offer higher interest rate for the investors than the bank deposits and low price for borrower as compared to the similar loans from conventional banks, (ii) high access to credit as the credit sanctioning criteria in the traditional banking system keep some categories of borrowers out of the formal banking system, (iii) the networking system in the P2P lending develops a positive perception of having good social value, and (iv) players in the P2P lending market experience enhanced service quality swiftly as a result of technical innovation in the sector.

1.2. Aims of the study

The purpose of this study can be viewed from three different aspects. Firstly, the paper explores the features of the high-risk[4] consolidated loans and non-consolidated loans in the P2P lending market to compare these features between them. Secondly, it develops two separate models for the default prediction of these two groups of loans by identifying the factors that can significantly contribute to the default probability prediction. Finally, the similarities and differences in terms of default risk along with its explanatory variables are discussed.

[4] The high-risk loans for the purposes of this reports are those having credit grade “D”, “E”, “F”, and “G” assigned by the lending club.

1.3. Research questions

The first research question of the thesis is related to identifying the features of both the high risk consolidated and the high-risk non-consolidated loans. We work with the question of whether the loan attributes are significantly different between the consolidated and the non-consolidated loans. The second research question is whether two separate models can significantly contribute to the default prediction of these two groups of loans. The last research question for this paper is whether the high risk consolidated loans bear more default risk than the high-risk non-consolidated loans.

1.4. Significance of the study

The previous literate related to the default risk of the loans in the online P2P lending market explores that the interest charge on the high-risk loans is not enough to compensate the risk level of the loans which raises a question mark in the existing risk modeling for high-risk loans. This paper further analyzes the features of these high-risk loans by dividing them into two separate groups depending on the purposes of the loans as high risk consolidated loans and high-risk non-consolidated loans. The result of this study helps investors as well as the lending platforms to make more informed decision where the investors now understand that the high risk consolidated loans bear is more prone to default than the high-risk non-consolidated loans and either choose to invest in the high-risk non-consolidated loans than the high risk consolidated loans considering the other factors as fixed. The lending platforms may also be benefited by understanding the actual risk level of the high risk consolidated loans and design a separate method to assign risk grade for the high risk consolidated loans rather than using the present method of assigning the credit grades of these loan customers.

1.5. Scope

The research work is done on the publicly available data collected from the lending club during the period from January 2018 to December 2018. The variables used in the analysis are those listed as customers' attributes in the lending club website to compare these variables between the high risk consolidated and the high-risk non-consolidated loans and to develop to separate models for predicting the default probabilities of these two group of loans.

1.6. Limitation

This research works for answering the predetermined research questions and employs an adequate level of care in terms of analysis and interpretation however this report bears some limitations which may be considered with the research finding for better interpretation of the results of this research. The followings are the main limitations of this research work: Firstly, the research is done with the publicly available data and for this reason, the research findings are limited by the available data of the website. In addition to this, for some the data the authenticity is not verified which limits the research findings ability to generate a credible result.

1.7. Thesis structure

This thesis paper consists of five chapters describing five broad areas of the report with the subdivisions of ideas as follows:

The chapter-1 titled “Introduction” discusses the general background, aims, research questions, scope, significance and limitation of the study followed by a general structure of the report.

The chapter-2 titled “Literature review” which includes the scholarly contributions related to the funding success and the delinquency of the loans with a discussion on the general research trend of the field.

The chapter-3 titled “Research methodology” discusses the research design, sampling, data collection and preparation along with data analysis techniques.

The chapter-4 titled as “Result and analysis” discusses the classification analysis, default modeling, and financial analysis.

The chapter-5 titled as “Conclusion and recommendations” summarizes the paper findings along with recommendation and further study.

Chapter 2 - Literature review

2.1. Research trend

The emergence of the P2P lending market is a recent phenomenon and hence the scholarly contributions to this area are limited. which can broadly be viewed under three different aspects. One group of scholars have been contributing in researching the emergence of the P2P lending market to explore the reasons for its emergence in addition to the existing financial systems (Hulme, M. & Wright, C., 2006). Other groups of scholars concentrate on finding the factors and mechanisms of how successful transactions take place. Their research area also includes the identification and management of default risk. The third group of researchers contributes to the performance of lending platforms.

2.2. The emergence and growth of the P2P lending

The emergence of a new form of the lending system reflects adapting the financial system with the new social trend to directly respond to the new social trend which generates the demand for a new form of relationship in the financial market under this information age. They also argue that the fundamentals of any social lending scheme are the general understanding among the members in the community, the boarder ethicality, and the transparency. The presence of these factors helps all the stakeholders in the system to form a relationship where they can work in "good faith". However, the traditional financial market relies more on the transactional relationship among the parties involved (Hulme, M. & Wright, C., 2006). Though the research findings of the study by Christensen et. al. (2000) doubts that whether the innovation of the online P2P lending system will become a disruptive technology, the P2P market shows a significant growth since its inception in the year 2005. The rapid growth of P2P lending is explained through two theories- the financial intermediation theory and the market equilibrium theory. The financial intermediation theory states that as the online P2P lending platforms generate credit with more cost-efficient manner then the traditional lending institutions like banks then both the lenders and the borrowers prefer the P2P lending market than the banks and other financial intermediaries. The market equilibrium theory suggests that due credit rationing the financial market a good number of borrowers are kept out of the system even they are ready to pay the high price. On the other hand, in the P2P lending market, the demand and the supply interact to clear the market towards its equilibrium point. The problem of credit rationing is minimum in the P2P lending market which helps the market to grow faster (Serrano-Cinca et al. 2015).

Galloway (2009) shows the role of the lenders, the borrowers, and the platforms in the online P2P lending market. The borrowers get loans without the direct intervention of the formal financial intermediaries which enables them to get a loan with better conditions than the formal banking system. The lenders may consider their funding as the investment models where the credit risk of the funded loans is assumed to be consistent in the credit grade assigned by the platforms. The platforms are considered as the financial service provider where they receive a fee for successful transactions. The limited role of banks in the online P2P lending to facilitate the process of lending is also endorsed by scholars. The activities of lending platforms include market making, credit request processing, and social network building but the system excludes the lending platforms to be engaged in direct decision making regarding the loan sanction. The platforms act as an auxiliary force in the process of decision making (Meyer 2007; & Wang et al. 2009).

Klafft (2008) demonstrates the intermediation role of online platforms in the P2P market. The lending platforms like the lending club summarizes the borrowers' financial characteristics and make available to the lenders so that the lenders can use that information for their decision making regarding the funding decision. The platforms also assign a credit grade to help the lenders to assess the potential credit risk of the borrowers. Some platforms also use their own proprietary software to calculate the credit grade, others take outsourcing services from the credit rating agencies to assess the credit risk of the borrowers. Platforms like prosper.com provide additional financial information such as the utilization of the open credit line and the credit card utilization record.

Both the funding success and the default risk are influenced by the social networking of the borrowers. Strong social connection of the borrowers improves the probability of getting their loans being funded and resulted in lower default rate. Borrowers with good social networking enable them to negotiate the lower interest rate on their loans (Lin, et al., 2013). The involvement of group leader with endorsement increases the funding possibility of a loan with a lower interest rate and the default rates of these loans are lower than the other loans with similar criteria except the involvement of a group leader (Berger SC, & Gleisner F, 2009). Social ties of the borrowers also improve the possibility of getting their loan request to be funded and the interest rate on these types of borrowers is lower as compared with other borrowers (Freedman, S. & Jin, G., 2014).

Davis, K. & Murphy, J. (2016) view the online P2P lending as a classical example of the integration of a series of economic activities where the market operators, the financial service providers, and loan brokers are integrated to the system with their very different roles. The regulatory environment also needs to be integrated to deal with the new emerging market.

2.3. Research on funding outcome

The financial, demographic and social variables are the determinants of the successful lending in the P2P lending market. Some of these variables not only influence the funding probability but also influence the interest on the loan generate through the P2P lending platforms (Alexander. B, et al, 2011). The influence of these variables is significantly different from each other in terms of their contribution towards the successful lending decision. Herzenstein, et al. (2008) showed that the funding success in online P2P market is influenced by borrowers' financial strength, the degree of motivation for listing and publicizing, and demographic variables however the demographic variables like race and gender have very little influence on the funding success as compared to the other variables like financial strength of the borrowers. The financial variables, the personal attributes and the platforms' recommendation in the form of assigning credit grades to the borrowers work as the mediator between the borrowers and the lenders to have a successful lending transaction. He also argues that the lenders in the online P2P market take funding decision more fairly than the formal financial sector in the USA where discrimination is well documented by the scholars. The P2P market has a role to reduce the discrimination practice in the financial sector. However, the study of Pope and Sydnor (2010) shows the presence of discrimination in the bidding process evidenced by the data of prosper.com. The thesis reveals that the black, overweight, and aged applicants are discriminated with higher interest rates as compared with that of the white and young people. People with the military association are favored with better lending terms than the other applicants.

Some scholars argued with the data collected from prosper.com that the funding success is negatively correlated with the credit grade of the borrower with the higher funding success rate for borrowers with low credit grade and vice versa. However, the interest rate for low credit grade borrowers is higher than that of high credit grade borrowers. The credit card limit utilization impacts the success rate of a loan request is being funded. High level of utilization of the bank credit card signals low level of creditworthiness of the potential borrowers and the low to medium level of credit card limit utilization corresponds to high creditworthiness of the borrowers and the funding probability is high for borrowers with low to medium level of credit card utilization (Lin et al., 2012).

Unlike others, Chen (2012) favors the view that the credit grade is insignificant in determining the interest rate and the default risk for high credit grade borrowers is low as evidenced from the analysis done with the data analysis of a Chinese lending platform ppdai.com. He, however, favors the relationship of credit grade and lending success. The study also shows that the higher interest rate and the smaller loan size increase the funding success probabilities and vice versa. So, the borrower can increase the funding success rate either by agreeing to pay higher interest on loans or to reduce the requested loan amount.

The lenders in the online P2P lending market infer borrower's creditworthiness with the help of different banking and financial variables. In the process of selection, there is evidence that the lenders also use non-standard soft information regarding the borrower before taking the lending decision. While dealing with high-risk borrowers the study also suggests that the use of soft information brings better result in evaluating the loan request. This soft information includes the reasons for which the loans are requested and the frequency of friends' endorsements. Lenders decision of funding also negatively influenced by factors like the past default rate, the debt-income ratio, and the most recent loan request frequency. (Iyer et al., 2009). The study of Prystav, F. (2016) also shows similar findings regarding important of soft information in evaluating the loan request by high-risk borrowers. Because the online P2P lending platforms choose which information is to be made available for the investors this research argues that the investors decline loan requests form the high-risk borrowers if they are not convinced with the soft personal information of the borrowers. Michels J. (2012) argues that lenders decision in the P2P market not only affected by the structured verifiable information of the borrowers but also is influenced by the voluntary and unverifiable information disclosures of the borrowers. These types of voluntary and unverifiable disclosures increase the likelihood of their loan proposals are being funded. The study also showed that the borrowers with voluntary unverifiable information disclosures get funded with low-interest rate than the borrowers with no such information disclosures. Duarte, J. et al. (2012) claims that trustworthy appearance matters in financial transactions like funding in the P2P market. They use the photograph of the potential borrowers from online P2P lending platform and found that the borrowers with photographs appearing trustworthy can influence the decision of lender and the funding decision is positively influenced by the trustworthy appearance of the borrowers with higher probabilities of having the loan funded. He also argues that a borrower with trustworthy appearance have a better credit score and hence default less frequently. Scholars in their recent studies show that the borrowers' creditworthiness may not be reflected properly by the credit grades assigned by the online P2P lending platforms and it is suggested that additional information disclosures if used together with the hard information like credit grades may give better result in evaluating borrowers' creditworthiness in the market (Serrano-Cinca, et al. 2015; Tao et al. 2017; Zhu 2018). Chan, D. & Han, C. (2012) studies on the relative importance of the soft and hard information between the USA P2P market and the Chinese P2P market and show that in both the countries both the hard and the soft information play a significant role in lending outcomes. However, the investors in the Chinese P2P market more dependent on the soft information.

The issue of a personal guarantee is studied by Agarwal, S. et al. (2015) with the evidence form the Chinese P2P market. The research reveals that loans with personal guarantees get funded with a higher probability of success with lower processing time and higher bidding activity. The borrowers associated with personal guarantee enjoy a lower interest rate than others. However, the investor's preference for such borrowers is not supported by the fact that the average default rate of those loans is higher than the other loans without a personal guarantee.

2.4. Research on default probabilities

Though the research contribution regarding the P2P lending market is narrow, there are several scholarly contributions related to the credit risk and the default probabilities. Ma and Wang (2016) examined the factors influencing the credit risk in the online P2P lending market viewing the factors from three different perspectives like the borrowers' perspective, the platform`s perspective, and the environmental perspective. Borrowers' moral level and job security are considered as the important variables influencing credit risks related to the borrowers' perspective. The formal control mechanism of the lending platform and the overall policy environment are related to the platform and the environment which affect credit risk of loans in the online P2P lending market. Reddy, S. (2016) observed the relative importance of the variables in explaining the default probabilities of the loans in online P2P market and showed that the credit score is the most important variable which can describe whether the loans would turn into default loans. The immediate next important variable he found is the amount paid as a proportion of the loan amount.

Lin, X., et al. (2017) worked with the data from a large online P2P platform of China and proposed a default prediction model to assess the potential lending risk of the borrowers. They explored that the demographic variables of the borrowers are the determinants of the default probabilities of the borrowers. The gender, marital status, level of education, age, length of service, installment size, loan amount, debt-to-income ratio, and credit history play a significant role in loan default. Serrano-Cinca, C., et al. (2015) also found the credit grades assigned by the platforms are the most important variable which can explain the default probability of the loans. The accuracy of the model can be improved by adding variables like the debt level of the borrowers. In addition to this, they also find other variables like loan purpose, annual income, homeownership, and credit history as significant explanatory variables for default risk prediction. Guo, Y., et al. (2016) proposed an instance-based credit assessment model as an alternative to credit grading-based risk assessment model. The model can evaluate the risk and return of individual loan in the P2P lending market and the performance measures of the model show that the model can efficiently improve investment decision.

Eid, et al., (2016) discusses a relationship between the income rounding tendency and the lending outcomes. The result of his work shows that the borrowers with a tendency to report their income as rounded rather than reporting the accurate income are more likely to default and less likely to make prepayments. The borrowers who round their income get lower interest rate and higher amount of loan than what their actual income might pursue and as a result the investors are exposed to a higher level of risk because of income rounding by the borrowers for which the investors are not compensated.

The study of Kumar, S. (2007) regarding the investor's behavior in setting the risk premium for loans and shows that in general investors behave rationally to set interest rate considering the appropriate risk premium corresponding to the predicted default risk of the loan. They also notice some instances where the investors fail to rationalize the charged risk premium and the variables responsible for loan default. In addition to other findings, Serrano-Cinca, C. et al. (2015) show the relationship of loan default at a point of time-based on the purposes for which the loan is taken. The study shows that the survival rates for different loans based on purposes are different from each other. The loan is taken for one purpose default earlier or later than the other purpose loan. The survival analysis shows that by comparing purposes, the small business is the riskiest among other loans and the wedding loan is the less risky loan. Investors in the P2P lending market can use this information to make a better investment decision in terms of selecting the right borrowers considering the appropriate risk level of the borrowers.

Miller (2011) studies on the causal relationship between the availability of information in the loan application and the final lending outcome. The study shows that when the investors get more information regarding the borrowers this can substantially reduce the default risk of the borrower for high-risk borrowers however fails to establish any relationship of the information availability and the default risk for low-risk borrowers. The higher information availability improves the lending outcomes mainly in two ways. First, it helps the investors to improve their loan screening ability with more information. Secondly, when more information is available the number of participating investors in the bidding process increases. More biddings bring a more efficient result in choosing the right borrower.

Emekter, R. et.al.(2015) undertakes a comprehensive study on the lending club data and explores that the most creditworthy borrowers in terms of their FICO score and higher income level remain out of the P2P lending networks. The P2P lending market fails to attract the most creditworthy borrowers. The good borrowers have a selection bias towards the traditional formal financial market over the online P2P lending market. The study also reveals that the interest charged on the high-risk borrower is not enough to compensate for the risk level of the borrower. The lending platform fails or inefficient to predict adequate risk level of the borrower especially the high-risk borrowers and hence the charged risk premium is less than the actual risk premium which is consistent with the risk level of the risky borrower. With this research findings, the investors in the P2P lending market seem to be more interested to lend money to the high creditworthy borrowers and it is shown that the good borrows have selection bias against the P2P lending market. This dilemma may hamper the growth and sustainability of the P2P lending system. This study focuses on the analysis of high-risk borrowers where existing models fail to assess the actual risk of these customers hence the risk premium for these loans seems lower than the required risk premium corresponding the risk level. The research finding of Serrano-Cinca, C. et al. (2015) shows the difference in risk level depending on the purposes of loans in the P2P market. To analyze the risk level of the high-risk borrowers, this study classifies the high-risk borrowers into two board groups: the one group is the high-risk consolidated loans and the other group is the high-risk non-consolidated loans and compare the variables of both the groups of loans and then propose two separate default prediction model for these groups of loans

Chapter 3 - Research Methodology

3.1. Methodology

This section describes the entire process of how the research work is done for the purposes of achieving research objectives. The section includes topics such as research design, data collection, data processing, and data analysis techniques.

3.1.1. Research design

This research is quantitative in nature which involves quantitative data analysis to answer the research questions. The secondary data is collected from one of the largest online P2P lending platforms of the USA, the lending club. The descriptive statistics analysis is done to explore the similarities and difference in terms of loan attributes between the high-risk consolidated loans and the high-risk non-consolidated loans. The binary logistics analysis is done to develop two separate models for predicting the default probabilities of the high-risk consolidated loans and the high-risk non-consolidated loans.

3.1.2. Sample size and determining target

The lending club successfully facilitated 6,76,460 loans listing during the period from January 2018 to December 2018 with a total disbursement amount of USD 8.84

billion. Out of the total loans listed during the period, 1,31,730 loans are categorized as "high-risk"[1] loan which is the sample size for this thesis.

3.1.2. Data collection and sources

The secondary data for 1,31,730 loan listings are collected from the lending club's website (www.lendingclub.com) which is publicly available. These data are collected in softcopy during the period of March 2019.

3.1.3. Data preparation

The source website contains a total of 6,76,460 data for the period from January 2018 to December 2018. Out of them, we have sorted high-risk loan data for 1,31,730 loans which are related to our target population. Each loan data contains 115 variables and several steps are undertaken to process the raw data for analysis. Duplicate rows and rows with no ID no. or loan no. are excluded from the calculation. This exclusion is insignificant considering their overall proportion in the data set. The variables that do not qualify to be a predictor for measuring default probabilities are removed. Many of the statistical models do not work well with highly correlated variables and to avoid such a situation the highly correlated variables have been removed from the analysis. The data are also cleaned by removing special characters like % sign, $, etc. The categorical variables are recorded according to the order of the categorical variable. The year of experience is recorded as using the scale from 0 to 10 where no experience category is included in 0 groups and the experience of 10 years and more in included in experience class 10. This improves the visualization of the data and helps in data processing smoothly.

3.1.4. Data analysis techniques

The data analysis techniques for this paper are done in two parts. The first part of the analysis is the consist of summarizing the variables parameters for both the high-risk consolidated loans and the high-risk non-consolidated loans by calculating descriptive statistics for the concerned variables using the SPSS software. For comparing the variables of the groups', the 't' test statistics are used. In the second part of the data analysis, we use the binary[5] logistic regression analysis which helps in generating two separate models for predicting the default probabilities of each of the two groups of loans. The target outcome of the models is binary i.e. the occurrence of the event here is "the default" of the concerned loan and the nonoccurrence of the event is the "no default" event. In the proposed default prediction models the dependent variable in the probability of default of a loan and the independent variables are the loan amount, the credit term, the interest rate, the credit grade, the employment length, the homeownership status, the annual income of the borrower, the information verification status, the debt-to-income ratio, and the delinquency record in last 2 years.

The regression analysis is widely used to describe the relationship between the outcome variables with one or more explanatory variables. The linear regression model

[5] For the model derivation and the underlying assumptions, please refer to Hosmer, D. and Lemeshow, S. (2000).

is best fitted in a situation where the outcome variable is continuous in nature. This research explores the relationship between the default probability with the explanatory variables where the outcome variable is binary in nature. To establish a relationship with the binary outcome variable, the binary logistic regression analysis is used (Hosmer and Lemeshow, 2000). Under this model the expected value of the outcome variable can be explained by the below-mentioned equation where E (Y|x) is the conditional expected value of the Y given x:

$$E(Y|x) = \beta_0 + \beta_1 x \qquad (1)$$

The above equation gives the value range for the expected value of the outcome variable from -∞ to + ∞. However, for the binary outcome variable, the expected value of the outcome should be within the range of 0 to 1. To get the desired expected value a mathematical transformation is done as follows which is termed as “the logit transformation” in the binary logistic regression model.

$$d_x = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \qquad (2)$$

Now with this logit transformation (equation 2), the model can explain the binary outcome variable in terms of its explanatory variables.

Chapter 4 - Results and Analysis

The analysis section of this paper deals with the three dimensions of the analysis namely classification analysis, default modeling, and financial analysis. The classification section explores the features of the consolidated loans and non-consolidated loans, the default modeling section identifies the factors influencing the default risk of the loan for each of the two categories of loans, and the financial analysis part tries to rationale the findings of both the sections.

4.1. Classification analysis

The data shown in Table1 are collected from the lending club are first presented in terms of the purposes for which the loan is requested. This data is required disclosures by the borrowers to the lending platforms for their loan request are being listed.

Table 1: Distribution of loans by the purposes

purpose	**N**	**% of Total N**	**Amount**	**% of the Total Sum**
Debt consolidation	387117	57.2%	$2.20 b	60.2%
Credit Card	159427	23.6%	$5.32 b	24.9%
Car	7851	1.2%	$0.06 b	0.7%
Education	423	0.1%	$0.00 b	0.0%

Home improvement	39318	5.8%	$0.49 b	5.5%
House	3035	0.4%	$0.04 b	0.5%
Major purchase	14207	2.1%	$0.14 b	1.6%
Medical	7304	1.1%	$0.06 b	0.6%
Moving	4862	0.7%	$0.03 b	0.4%
Wedding	2346	0.3%	$0.02 b	0.3%
Renewable energy	515	0.1%	$0.00 b	0.1%
Small business	8799	1.3%	$0.13 b	1.4%
vacation	4446	0.7%	$0.03 b	0.3%
Others	36809	5.4%	$0.32 b	3.6%
Total	**676459**	**100.0%**	**$8.85 b**	**100.0%**

*Source:*www.lendingclub.com

Out of 6,76,459 loan distributed in the year 2018, around 60% of the loans are requested for paying off the existing loan liabilities of the potential borrowers. The proportion becomes around 80% of the total disburse loan when credit card loan is added with the other consolidated loans. The credit grades[6] assigned by the platform represent the risk level of the potential borrowers. Presently the lending club assigns six alphabetic grades as 'A', 'B', 'C', 'D', 'E', 'F', & 'G' where the 'A' grade presents the best credit grade with the lowest level of risk and the 'G' grade represent s the worst grade with the highest level of risk. Each credit grade further is subdivided into five sub-groups which makes total thirty-five credit grades.

[6] The lending club assigns a credit grade for each of the borrowing customer based on the customer's profile including the FICO score. There are seven credit grades namely A, B, C, D, E, F. and G where the "A" grade represents the highest credit grades with low level of risk and the "G" grade represents the worst credit grade with the highest level of risk. Then each of these seven credit grads are subdivided into further five subgroups like A1, A2, A3, A4, A5, and so on…

The data show that the high-risk borrowers constitute about 20% of the total loan portfolio with an approximate fifteen percentage of default rate. The dollar value of the loans extended to the high-risk customers is around USD 1.8 billion and the table-2 shows the classification of the high-risk loans according to whether they are consolidated loans or not. This distribution portraits much similarities in their distribution according to the respective credit grades.

Table 2: Comparing the high-risk consolidated and high-risk non-consolidated loans

Credit Grade	**Consolidated**				**Non-consolidated**			
	N	% of Total N	Sum	% of Total Sum	N	% of Total N	Sum	% of Total Sum
D	44003	67.1%	$0.60 b	62.1%	34513	64.7%	$0.39 b	60.7%
E	15841	24.2%	$0.26 b	26.8%	13221	24.8%	$0.17 b	26.3%
F	4805	7.3%	$0.09 b	9.2%	4622	8.7%	$0.06 b	10.2%
G	886	1.4%	$0.02 b	1.9%	1003	1.9%	$0.02 b	2.8%
Total	**65535**	**100.0%**	**$0.97 b**	**100.0%**	**53359**	**100.0%**	**$0.64**	**100.0%**

Source: www.lendingclub.com

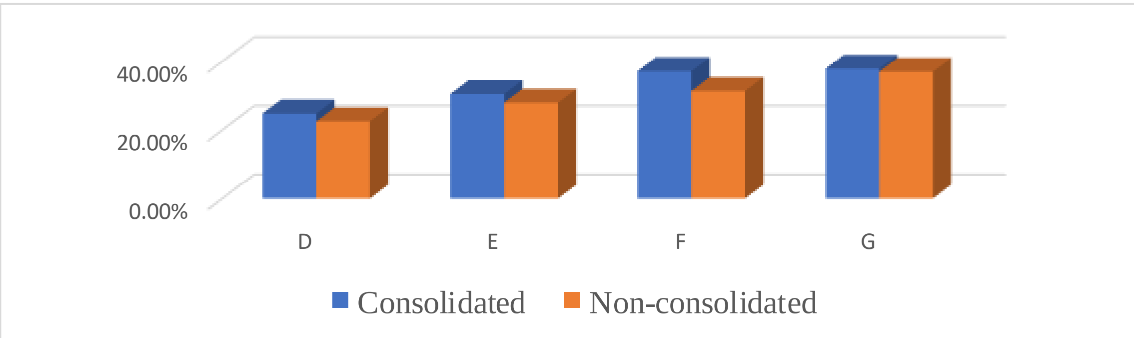

Figure 2: The comparison of default proportion in high-risk consolidated and non-consolidated loans.

The figure-2 shows a comparison between consolidated and non-consolidated loans with different credit grades. The vertical axis represents the percentage of default loans whereas the horizontal axis represents the credit grades in which the yellow bars are non-consolidated loans and the blue bars are consolidated loans. With almost similar distribution these loans show a significant difference in terms of their default risk. The data demonstrate that for each of the credit grades the default risk is higher in consolidated loans than that of non-consolidated loans.

The descriptive statistical analysis shows that the mean value, as well as the nature of the distribution of the variables, are significantly different for the high-risk consolidated loans. The test statistics for equality of means also demonstrate the significant difference in test statistics between these two groups of loan. The t-test results show that the high-risk consolidated loans and the high-risk non-consolidated loans are significant in terms of variables like Loan amount, Interest rate, Installment, Employment length, Annual income, and Debt-to-income ratio at 99% significance level (Appendix-7).

Table 3: The non-parametric test for differences in important variables

	Variables	Consolidated Loans	Non-consolidated Loans
1	Default Rate	Higher	Lower
2	Credit Grade	Higher	Lower
3	Interest Rate	Lower	Higher
4	Loan amount	Higher	Lower
5	Debt-to-income ratio	Higher	Lower

The table-3 shows the non-parametric test results for different variables between the consolidated and non-consolidated loans. The non-parametric test also reveals the difference between the high-risk consolidated loans and the high-risk non-consolidated loans. The Mann-Whitney test results show that in terms of the variables such as the credit terms, the credit grades, the verification status, and the loan status, the high-risk consolidated loans are different from the high-risk non-consolidated loans at 99% confidence level. However, the homeownership status does not show any significant difference between these two groups of loans (Table-3).

4.2. Default probability analysis

The credit risk is the major concern for the lenders in the online P2P lending like any lending environment. From the lenders' perspective, the understanding of default probability along with the influencing factors makes more sense than simply to classify them in different groups. The lenders are benefited if they can understand whether the loan is going to default or not. A default probability model can help the lender to analyze the factors of borrowers to predict the default probability. With this knowledge, the lender can improve their ability to choose among different borrowers waiting to be funded. Borrowers' features with the help of a default model help lenders to make a better decision in funding a loan. Evidence from literature shows that the importance of the proper default prediction even high for the high-risk borrowers and the interest charges on the high-risk borrower fails to compensate for the higher probability of default. From the data of the lending club for the year 2018 of the risky borrowers

(borrowers with credit grade ‘D’, ‘E’, and ‘F’, and ‘G’), out of 6,76,460 loans listed 1,00774 loans are not paid (Fully or partially) as per agreed terms which constitutes a 14.90% of the funded loans are defaulted or at late in their agreed payment.

The non-parametric statistical test in the previous section explores the significant difference in term of borrowers’ features between consolidated and non-consolidated loans. This section develops two separate models to predict default probabilities of these two different group of loans and explores how the different factors contribute differently in default prediction for these groups. The binary logistic modeling is used to develop the logistic equation to measure the probability of the loans in terms of their potential binary status ‘Defaulted’ or ‘Not Defaulted /Regular’. For this modeling, the ‘Defaulted’ loans include loans with status ‘Late’, ‘Charged Off’, and ‘Default’ whereas the ‘Not Defaulted /Regular’ loans mean the loans having status ‘Current’, ‘Fully Paid’, and ‘In Grace Period’.

We assume that y_iC and y_iNC are the continuous numbers representing the default probabilities of the consolidated loans and the non-consolidated loans respectively. The higher value of y indicative of the higher probability of default and vice versa. As the outcome of the loans is best represented by the binary outcomes like ‘Defaulted’ or ‘Not Defaulted /Regular’, the dependent variable in binary logistic regression is the probability of an event being occurred which in this case is the default. So, the dependent variable can only take a value between ‘0’ and ‘1’ where the ‘0’ represent the probability of not occurring the event and ‘1’ represent the probability of occurring the event. To convert the open-ended continuous numbers y_iC and y_iNC to the numbers between the '0' and the '1', the following binary logit transformation is used,

$$d\_iC = \frac{1}{1+e^{-y\_iC}} \quad \text{for consolidate loans} \qquad (3)$$

$$d\_iNC = \frac{1}{1+e^{-y\_iNC}} \quad \text{for non-consolidated loans} \qquad (4)$$

Here the *d_iC* represents the probability of default for consolidated loans and *d_iNC* represents the probability of default for non-consolidated loans. We further assume that the explained variables *y_iC* and *y_iNC* are represented by n independent explanatory [7]variables in the modeled binary logistic functions which can be written as,

y_iC = b_0 + b_1C CG_1Ci + b_2C CG_2Ci + b_3C CG_3Ci + b_4C T_Ci
+ b_5C r_Ci + b_6C EL_Ci + b_7C HO_Ci + b_8C VS_Ci + b_9C DIR_Ci
+ b_10C LA_Ci + b_11C IS_Ci + b_12C DR_Ci + epsilon_Ci (5)

and

y_iNC = b_0 + b_1NC CG_1NCi + b_2NC CG_2NCi + b_3NC CG_3NCi
+ b_4NC T_NCi + b_5NC r_NCi + b_6NC EL_NCi + b_7NC HO_NCi
+ b_8NC VS_NCi + b_9NC DIR_NCi + b_10NC LA_NCi + b_11NC IS_NCi
+ b_12NC DR_NCi + epsilon_NCi (6)

Here CG_1Ci, CG_2Ci, CG_3Ci, T_Ci, r_Ci, EL_Ci, HO_Ci, VS_Ci, DIR_Ci, LA_Ci, IS_Ci, DR_Ci represent Credit grade-1("D"), Credit grade-2("E"), Credit grade-3("F"),

[7] The explanatory variables considered for the binary logistic regression are the loan amount, the interest rate, the installment size, the credit terms, the credit grade, the home ownership status, the information verification status, the debt-to-income ratio, and the delinquency records of the borrowers.

Term, Interest rate, Employee length in years, Home ownership status, Information verification status, Debt-to-income ratio, Loan amount, Installment size, and Delinquency Record respectively for the consolidated loans. The terms with NCi represent corresponding variables for non-consolidated loans.

Table 4: Binary logistic regression results of the high-risk consolidated loans

Consolidated Loans	β	SE	Wald test statistics	Significance	Exp(β)
Credit grade			77.734	0.000	
Credit grade 1	-0.256***	0.031	66.097	0.000	0.774
Credit grade 2	-0.077***	0.024	10.222	0.001	0.926
Credit grade 3	0.134***	0.031	18.545	0.000	1.143
Term 1	-0.082***	0.012	47.683	0.000	0.921
Interest rate	1.883***	0.620	9.224	0.002	6.570
Employee length	-0.010***	0.002	15.874	0.000	0.991
Home ownership	-0.117***	0.014	72.114	0.000	0.889
Verification status	0.141***	0.025	33.067	0.000	1.152
Debt-to-income ratio	0.012***	0.001	117.520	0.000	1.012
Constant	-1.149***	0.137	70.010	0.000	0.317
Hosmer and Lemeshow's Test: Chi-square = 10.7					

The above table-4 displays the binary logistic analysis results for consolidated loans at 1% significance level where the extreme left column shows the statistically significant variables and the other columns show the other measures of the analysis.

The results[8] of the binary logistic regressions show that the default probabilities of the consolidated and non-consolidated loans are explained differently in terms of covariates as well as the coefficient of the covariates. The binary logistic regressions are performed using the SPSS software. At first, the model parameters are estimated by following the forward stepwise likelihood method and then the process is repeated with

[8] The base value of the variable credit term is the 36-month duration loans. The base value for credit grade is the lowest credit grade "G". Therefore, the models' constants can determine the default risk for the credit grade "G". *** represents a 1% significance level and ** represents a 5% significance level.

backward likelihood method and the results are similar in using both the methods.

Table 5: Binary logistic regression results of the high-risk non-consolidated loans

Non-consolidated Loans	β	SE	Wald test statistics	Significance	Exp(β)
Credit grade				0.000	
Credit grade 1	-0.381***	0.033	134.868	0.000	0.683
Credit grade 2	-0.078***	0.025	9.983	0.002	0.925
Credit grade 3	0.114***	0.032	12.483	0.000	1.121
Interest rate	-1.495**	0.634	5.563	0.018	0.224
Employee length	-0.012***	0.003	19.872	0.000	0.988
Home ownership	-0.102***	0.015	44.127	0.000	0.903
Verification status	0.177***	0.026	47.030	0.000	1.194
Debt-to-income ratio	0.012***	0.001	105.515	0.000	1.012
Constant	-0.769***	0.139	30.704	0.000	0.463
Hosmer and Lemeshow's Test: Chi-square = 13.7					

The table-5 displays the binary logistic analysis results for non-consolidated loans both at 1% and 5% significance level where the extreme left column shows the statistically significant variables and the other columns show the others measures of the analysis.

For the logistic regression analysis of the high-risk consolidated loans, the nine variables are used and out of these nine variables, seven variables can significantly explain the default probability of the loans. The analysis is done considering both 1% and 5% confidence level and the result shows that all these seven variables are significant at 1% significant level. The model can predict the expected outcome with the correctness of 72.7% and the Chi-square value of 10.769 advocates for the acceptable goodness to fit of the model. The smaller SE's values for all the coefficients can be explained with the low level of the multicollinearity and the model's R^2 value in the final step of the model is 2.4%.

By incorporating the covariates and their calculated coefficients in the binary logistic regression model for high-risk consolidated loans as shown in equation (3), the predicted default probability of a typical high-risk consolidated loan can be determined by using the following equation,

$$y_{iC} = b_0 + b_{1C}\, CG_{1Ci} + b_{2C}\, CG_{2Ci} + b_{3C}\, CG_{3Ci} + b_{4C}\, T_{Ci} + b_{5C}\, r_{Ci} + b_{6C}\, EL_{Ci} + b_{7C}\, HO_{Ci} + b_{8C}\, VS_{Ci} + b_{9C}\, DIR_{Ci} + \epsilon_{Ci} \quad (7)$$

To explain the model, for a high-risk consolidated loan with interest rate 17% p.a., credit grade of 1 (the formal credit grade ‘D’), with verified information (numeric value 1), lives in a rented home (numeric value 2), with 5 years of employment and having 50% debt-to-income ratio, the default probability is 21.35%[9]. By using the model for borrowers with credit grade ‘E’, ‘F’, and ‘G’ are predicted assuming all other variables in the above equation (7) remain same except for interest rate[3] as 20%, 23%, 24% respectively. The predicted default probabilities are 25.57%, 30.98% and 28.58% respectively.

Another similar model is developed using the same method with the same variables for the high-risk non-consolidated loans and the result shows that out of nine variables similar six variables are statistically significant. The loan term is statistically

[9] $y_{iC} = -1.149 + 1.883 \times 0.17 - 0.082 - 0.256 - 0.141 - 0.117 \times 2 - 0.010 \times 5 - 0.012 \times 0.50 = -1.30389$, $d_{iC} = \frac{1}{1+e^{-(-1.30389)}} = 0.2135$

significant for consolidated loans but not significant in this case. The variables are statistically significant at 1% significance level except for the interest rate which is statistically significant at 5% level. The default prediction model for high-risk non-consolidated loans looks as follows,

$$y\_iNC = b\_0 + b\_1NC\ CG\_1NCi + b\_2NC\ CG\_2NCi + b\_3NC\ CG\_3NCi + b\_5NC\ r\_NCi + b\_6NC\ EL\_NCi + b\_7NC\ HO\_NCi + b\_8NC\ VS\_NCi + b\_9NC\ DIR\_NCi + epsilon\_NCi \tag{8}$$

The predicted default probabilities for the high-risk non-consolidated loans by using the model shown in equation (8) and using the same variable parameters as used for the high-risk consolidated loan for the credit grade 'D', 'E', 'F', and 'G' are 18.47%, 22.67%, 25.35%, and 22.99% respectively. The predicted default probabilities support our claim that the risk level of the consolidated loans and the non-consolidated loans for high-risk borrowers are different and the consolidated loans attribute high default rate for each credit grades from the 'D' grade and worse.

Table 6: Calculated default rates for each of the credit grades

Credit Grade →	D	E	F	G
Credit Group ↓				
Non-Consolidated	18.47 %	22.67 %	25.35 %	22.99 %
Consolidated	21.35 %	25.57 %	30.98 %	28.58 %

The above table-6 shows the calculated default risk using the developed models both for the consolidated loans and the non-consolidated loans. The different columns of the table represent the default risk at different credit grades. The predicted default rate using the models show that the high-risk consolidated loans are more likely to be defaulted as compared with that of non-consolidated loans. This result is consistent with the results of the non-parametric test to compare the means of these two groups of loans where the t-test shows that the mean default rate for high-risk consolidated loans is more than that of the high-risk non-consolidated loans.

In addition to default predictability, the models further explore new insights both in terms of the significance of explanatory variables and in terms of a coefficient of a variable. Earlier studies on default probabilities claim that along with other explanatory variables the credit history of the borrowers can significantly contribute towards default prediction and the default models developed by other scholars include the delinquency records of the customer as an explanatory variable to calculate the value of dependent variables. However, in contrast with the earlier studies both of our models fail to consider the credit history of the borrowers (variable 'delinquency in last 2 years') in the model as the variable is not even significant at 5% significance level. For high-risk loans, in the online P2P credit market, the credit history cannot help in predicting the default probabilities of the loans. Another aspect of the models is the coefficient of the interest rate where the interest rate shows an inverse relationship means the high-interest rate predicts higher default probability for high-risk consolidated loan customers whereas for non-consolidated loans the relationship is reverse that predict low default rate for high interest paying customers.

4.3. Financial analysis

The binary logistic analysis in the previous section develops two separate models for predicting default of the loans in their respective group. The regression analysis shows a contrasting behavior of one of the explanatory variables, the interest rate on the loan. For consolidated loans, the higher interest rate corresponds to the higher default probability on the other hand, for non-consolidated loans the higher interest rate corresponds to the low default probabilities. The financial analysis explains the contrasting interest rate behavior using economics theory. The demand-supply dynamics to achieve market equilibrium in financial market behave differently than the traditional goods and services market where the demand and the supply interact to reach an equilibrium point where each firm can earn a normal profit. The shape of the supply curve in the loanable fund market is shown in figure 3 where the supply curve is concave shape rather than the usual upward sloping supply curve. The concave shaped supply curve in loanable fund market is due to the presence of adverse selection problem.

Borrowers and lenders virtually interact to make a transaction with the help of the platforms where high interest on the loan does always brings the desired benefits for the investors.

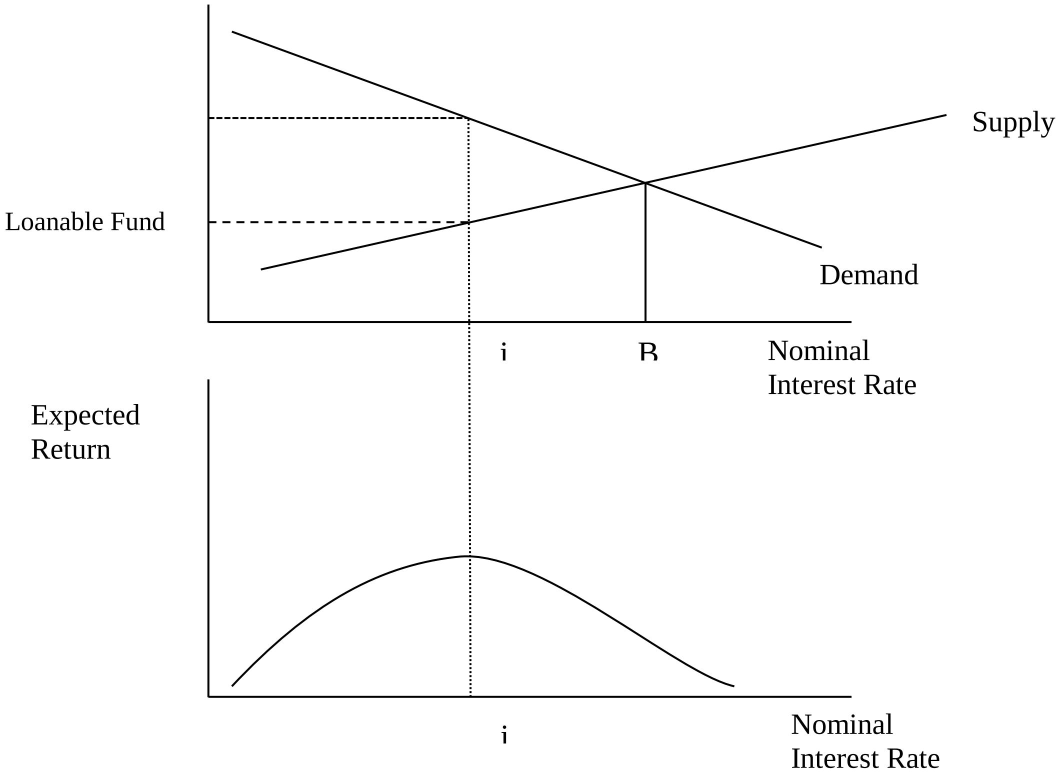

Figure 3: **Loanable fund market equilibrium point**

The vertical line i on figure 2 shows a position where investors get the highest benefits with a low level of default rate. If the market stands at a point left side of the line, increase of interest rate on loan brings more benefits and if the market stands at a point right side of the line any increase in interest rate will enable the bad borrowers to avail loans from the market which is considered because of adverse selection in financial transaction. The Banks and the other formal financial intermediaries address this issue through credit rationing (Stiglitz, J., & Weiss, A., 1981).

Chapter 5 - Conclusion and Recommendations

This chapter concludes this research study to demonstrate the achievement of the research objectives though answering the research questions and make some recommendations for the stakeholders of the P2P online market.

5.1 Conclusion

Though the growth of the online P2P market is supported by its ability to provide efficient, swift services with the lower transaction cost as compared to the traditional financial intermediaries, the sustainability of the system depend on its ability to address the credit risk issue and assign appropriate risk premium in the pricing. The study of Emekter, R. et.al.(2015) reveals a concern for the sustainability of the online P2P lending system showing that the interest rate on the loans of high-risk borrower is not enough to compensate the risk corresponding to the loans and hence suggest to book low-risk borrowers for sustainability of the innovative lending system. Our study focuses on the high-risk borrowers for measuring the appropriate credit risk which subsequently helps in assigning required interest rate considering the risk. This thesis analyzes the high-risk loans of the lending club segregating them into two separate groups depending on the purposes for which the loans are availed—the high-risk consolidated loans and the high-risk non-consolidated loans. Initial classification data shows that the proportion of consolidated loans is much higher than all other loans together both in numbers and amount of the loans. The borrower's profile, as well as the lending outcome, is different between the groups. The borrowers who borrow for consolidation purposes avail funding form lending club with an average higher amount, better credit grade, and they have on average higher default rate as compared with those of non-consolidated loan customers. The high-risk consolidated customers are favored by the investors with a higher amount of loans and lower interest rate as compared with the high-risk non-consolidated borrowers. The credit grade assigned by the lending club also shows upward bias as it shows that with better credit grade the high-risk consolidated borrowers have a higher probability of default. Although the consolidated loans have higher default records the lending platform fail to recognize that this higher level of risk needs to be adjusted in the form of the risk premium and the analysis shows that the consolidated loan customers get funded with same interest rate. The grouping of loans in terms of their credit grades shows that for all the high-risk credit grade the consolidated loans have a higher default rate than the non-consolidated loans.

The two separate models are developed using binary logistic regression analysis for the high-risk consolidated and the non-consolidated borrowers to predict the default risk associated with each of the two groups of loans. The analysis shows that the interest rate and the credit grades are the strongest explanatory variables in predicting the default probabilities of both the high-risk consolidated and non-consolidated loans. The models also predict the default probabilities for all the high-risk credit grades which are consistent with the actual calculated default risk of the loans. Most of the scholars who study delinquency in the P2P market finds a positive relationship between the credit history of the borrowers and the default rate. However, our models reveal that there is no significant relationship between the credit history of the high-risk borrowers irrespective of their purposes of the loans.

5.2 Recommendations

Within the scope of this thesis, the research findings may be useful for the investors, the borrowers, and the online P2P lending platforms. The investors are the primary beneficiary of the credit risk assessment in the online P2P lending system as the investors bear the default risk. With these research findings, investors are able to identify the risky customers even within a credit grade which generally symbolizes the same level of risk. For high-risk borrowers, investors will prefer non-consolidated borrowers, ceteris paribus, over consolidated borrowers or charge more interest on the consolidated borrowers considering the higher level of risk. Of the high-risk borrowers, the non-consolidated borrowers can bid for lower interest rate, ceteris paribus, than the consolidated borrowers. The lending platform should develop a new model to assign the credit grade so that the grades can accommodate the risk differences between the high-risk consolidated and the high-risk non-consolidated borrowers.

References

Agarwal, S., Y. Li, C. Liu, & J. Zhang (2015). Personal Guarantee and Peer-to-Peer Lending: Evidence from China. *Social Science Electronic Publishing,* SSRN Electronic Journal, DOI: 10.2139/ssrn.2638546.

Akerlof, G. A., (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism, *The Quarterly Journal of Economics*, Volume 84, Issue 3, August 1970, Pages 488–500, https://doi.org/10.2307/1879431

Alexander. B, et. al, (2011). Online Peer-to-Peer Lending – A Literature Review. *Journal of Internet Banking and Commerce,* Vol. 16 (2). Retrieved from https://www.researchgate.net/publication/288764128_Online_Peer-to-Peer_Lending_-_A_Literature_Review/download

Berger, SC, & Gleisner, F. (2009). Emergence of Financial Intermediaries in Electronic Markets: The Case of Online P2P Lending. *BuR - Business Research, Official Open Access Journal of VHB* 2: 39-65.

Chan, D. & Han, C. (2012). A Comparative Study of online P2P Lending in the USA and China. *Journal of Internet Banking and Commerce*, Vol. 17, No. 2, pp.1-15.

Chen, DY., (2012). Is online peer-to-peer lending market effective? A study on herding behaviour in China, Working Paper (School of Management, Fuzhou University).

Christensen, Clayton M., & Overdorf, M., (2000). Meeting the Challenge of Disruptive Change, *Harvard Business Review* 78, no. 2 (March–April 2000): 66–76.

Davis, K., & Murphy, J. (2016). Peer to Peer Lending: Structures, Risks and Regula tion, *JASSA: The Finsia Journal of Applied Finance*, 2016:3, 37-44. Available at SSRN: https://ssrn.com/abstract=2862252

Duarte, J., Siegel, S. & Young, L. (2012). Trust and Credit: The Role of Appearance in Peer-to-Peer Lending. *The Review of Financial Studies*, Vol. 25, No. 8. pp. 2455-2482.

Eid, N., Maltby, J & Talavera, O., (2016). Income Rounding and Loan Performance in the Peer-to-Peer Market. *Mpra Paper*:1 –38

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, DOI: 10.1080/00036846.2014.962222

Freedman, S. & Jin, G. (2014). The signaling value of online social networks: lessons from peer-to-peer lending, NBER Working Paper No. 19820. Available at http://www.nber.org/papers/w19820 (accessed 30 September 2018).

Galloway, I. (2009). Peer-to-Peer Lending and Community Development Finance. Community Development Investment Center Working Paper. San Francisco: Federal Reserve Bank of San Francisco. Retrieved from http://ideas.repec.org/p/fip/fedfcw/2009-06.html.

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. doi:10.1016/j.ejor.2015.05.050.

Hulme, M. & Wright, C. (2006). Internet based social lending: past, present and future, Working Paper, Social Futures Observatory, UK.

Herzenstein, M., Andrews, R., & Dholakia, U. et al. (2008). The democratization of personal consumer loans? Determinants of success in online peer-to-peer lending communities, Working Paper. Available at SSRN www.prosper.com (accessed 30 September 2014).

Hosmer, D. & Lemeshow, S. (2000). Applied Logistic Regression, 2nd edition, John Wiley, New York.

Iyer, R., Khwaja, AI., Luttmer, EF., & Shue, K. (2009). Screening in new credit markets: Can individual lenders infer borrower creditworthiness in peer-to-peer lending? In AFA 2011 Denver meetings paper.

Klafft, M. (2008). Peer to Peer Lending: Auctioning Microcredits over the Internet. Proceedings of the 2008 Int’l Conference on Information Systems, Technology and Management (pp. 1-8). Dubai: IMT. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1352383

Kumar, S., (2007). Bank of One: Empirical Analysis of Peer-to-Peer Financial Marketplaces, AMCIS 2007 Proceedings. 305. http://aisel.aisnet.org/amcis2007/305

Lin, MF, Prabhala, NR., & Viswanathan, S. (2012). Judging borrowers by the company they keep: Social networks and adverse selection in online peer-to peer lending, Western Finance Association 2009 Annual Meeting Paper. Available at SSRN: https://ssrn.com/abstract=1355679 or http://dx.doi.org/10.2139/ssrn.1355679

Lin, X., Li, X. & Zheng, Z. (2017). Evaluating Borrowers’ Default Risk in Peer-to-Peer Lending: Evidence from a Lending Platform in China. *Applied Economics,* Vol.49 No.35, pp.3538-3545.

Ma, H., & Wang, X., (2016). Influencing factor analysis of credit risk in P2P lending based on interpretative structural modeling, *Journal of Discrete Mathematical Sciences and Cryptography,* 19:3, 777-786, DOI: 10.1080/09720529.2016.1178935

Michels, J. (2012). Do Unverifiable Disclosures Matter? Evidence from Peer-to-Peer Lending. *The Accounting Review,* Vol. 47, No. 4, pp. 1385-1413.

Miller, S. (2011). Information and Default in Consumer Credit Markets: Evidence from a Natural Experiment (June 27, 2011). Available at SSRN: http://dx.doi.org/10.2139/ssrn.1873232

Meyer, T. (2007). Online P2P lending nibbles at banks' loan business. Deutsche Bank Research.

Pope, D. G., & Sydnor, J. (2010). What's in a picture? Evidence of discrimination from prosper.com, *Journal of Human Resources*, 46(1), 53-92.

Prystav, F. (2016). Personal information in peer-to-peer loan applications: Is less more? *Journal of Behavioral and Experimental Finance*, 9, 6–19.

Reddy, S. & Gopalaraman, K. (2016). Peer To Peer Lending, Default Prediction-Evidence from Lending Club, *Journal of Internet Banking and Commerce,* Vol. 21, No. 3, pp.1-19.

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

Stiglitz, J., & Weiss, A. (1981). Credit Rationing in Markets with Imperfect Information. *The American Economic Review, 71*(3), 393-410. Retrieved from http://www.jstor.org/stable/1802787

Tao, Q., Dong, Y., & Lin, Z. (2017). Who Can Get Money? Evidence from the Chinese Peer-to-Peer Lending Platform. *Information Systems Frontiers*, *19*(3), 425-441. https://doi.org/10.1007/s10796-017-9751-5

Wang, H., Greiner, M., & Aronson, J. E. (2009). People-to-people lending: The emerging e-commerce transformation of a financial market. *Value creation in E-business management* (pp. 182–195). Berlin and Heidelberg: Springer.

Zhu, Z. (2018). Safety promise, moral hazard and financial supervision: Evidence from peer-to-peer lending. *Finance Research Letters*. https://doi.org/10.1016/j.frl.2018.07.002

Appendices

Appendix 1 – Description of variables

Item	Description
Loan Amount	The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
Installment	The monthly payment owed by the borrower if the loan originates.
Term	The number of payments on the loan. Values are in months and can be either 36 or 60.
Interest rate	Interest Rate on the loan
Credit Grade	Lending club assigned loan grade
Sub-grade	Lending club assigned loan sub-grade
Homeownership	The homeownership status provided by the borrower during registration or obtained from the credit report. Our values are: RENT, OWN, MORTGAGE, OTHER
Verification status	Indicates if income was verified by Lending Club, not verified, or if the income source was verified
Debt-to-income ratio	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.
Delinquency in 2 years	The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years
Loan purpose	A category provided by the borrower for the loan request.
Employment Length	Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
Annual Income	The self-reported annual income provided by the borrower during registration.
Issue date	The month which the loan was funded
Loan status	Current status of the loan

Appendix 2: Distribution of loans by their credit grades assigned by the lending club

Credit Grade	N	% of Total N	Amount	% of Total
A	144228	21.3%	$2.00 b	22.7%
B	220255	32.6%	$2.78 b	31.4%
C	180248	26.6%	$2.27 b	25.6%
D	87158	12.9%	$1.11 b	12.5%
E	32154	4.8%	$0.48 b	5.4%
F	10352	1.5%	$0.17 b	1.9%
G	2064	0.3%	$0.04 b	0.5%
Total	**676459**	**100.0%**	**$8.85** b	**100.0%**

Source: Author's compilation for classification purposes.

Appendix 3: Distribution of loans by their status

Credit Grade	N	% of Total N	Amount	% of Total Sum
Current	11482	1.7%	$0.15	1.7%
Fully Paid	563764	83.3%	$7.36	83.2%
In grace period	440	0.1%	$0.01	0.1%
late 16-30 days	148	0.0%	$0.00	0.0%
late 31-120 days	786	0.1%	$0.01	0.1%
Charged off	99837	14.8%	$1.32	14.9%
Default	2	0.0%	$0.00	0.0%
Total	**676459**	**100.0%**	**$8.85**	**100.0%**

Source: Author's compilation for classification purposes.

Appendix 4: Comparing high-risk consolidated and high-risk non-consolidated loans by their status

Credit Grade	D	E	F	G
Consolidated	24.80%	30.51%	37.14%	37.89%
Non-consolidated	22.49%	27.95%	31.37%	36.99%

Source: Author's compilation for classification purposes.

Appendix 5: Descriptive statistics for high-risk consolidated loans

Variables	N	Min	Max	Mean	Std. Deviation	Skewness	Kurtosis
						Statistic	Statistic
Loan amount	65535	700	35000	14846.38	9167.428	.712	-.462
term	65535	0	1	.23	.424	1.250	-.437
Interest rare	65535	.0600	.2849	.184977	.0244747	.829	.574
installment	65535	23	1445	488.77	294.470	.892	.201
Credit Grade	65535	3	6	3.43	.688	1.557	1.860
Employment Length	65535	0	10	5.54	3.808	-.104	-1.532
Home ownership	65535	0	3	1.62	.657	.590	-.651
Annual income	65535	0	8900060	65682.74	55865.284	63.157	9586.269
Verification status	65535	0	1	.80	.398	-1.523	.320

Loan status	65535	0	1	.27	.446	1.017	-.965
Debt-to-income	65534	.00	672.52	19.8575	8.77518	6.412	466.267
Delinquency 2yrs	65535	.00	20.00	.3607	.92876	4.935	42.105
Valid N (listwise)	65534						

Appendix 6: Descriptive statistics for high-risk non-consolidated loans

Variables	N	Min	Max	Mean	Std. Deviation	Skewness	Kurtosis
						Statistic	Statistic
Loan amount	53359	500	35000	11905	9094	1.073	.265
term	53359	0	1	.18	.385	1.661	.757
Interest rare	53359	.0600	.2899	.1865	.0258	.752	.434
installment	53359	5	1445	398	296.274	1.204	.924
Credit Grade	53359	3	6	3.48	.732	1.461	1.459
Employment Length	53359	0	10	5.24	3.797	.017	-1.527
Home ownership	53359	0	3	1.62	.680	.616	-.693
Annual income	53359	1770	7500000	67964	62108	35.372	3872.759
Verification status	53359	0	1	.76	.428	-1.213	-.530
Loan status	53359	0	1	.25	.432	1.162	-.650
Debt-to-income	53359	0	999	17.63	9.829	18.828	1861.329
Delinquency 2yrs	53356	0	18	.37	.934	4.762	37.365
Valid N (listwise)	53356						

Appendix 7: The comparison of means between consolidated and non-consolidated loans.

	Independent Samples Test t-test						
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	99% Confidence Interval of the Difference	
						Lower	Upper
Loan amount ($)	-57.4	115156	.000	-2941.0	51.199	-3072.9	-2809.2
Interest rate	11.3	110502	.000	.001605	.000141	.001240	.001970
Installment	-54.6	114125	.000	-90.677	1.659	-94.950	-86.404
Employment length	-13.9	114835	.000	-.297	.021	-.352	-.242
Annual income	7.0	104248	.000	2330.3	331.691	1475.91	3184.710
Debt-to-income ratio	-41.9	105318	.000	-2.21	.05274	-2.3498	-2.07815
Delinquency	1.06	114368	.287	.00557	.00524	-.00792	.01906

The above table shows the t-test statistics for the variables to compare the means of the consolidated and non-consolidated loans. For the variables Loan amount, Interest rate, Installment, Employee length, Annual income, and Debt-to-income ratio, the means are different between the groups. This result is statistically significant at 1% level. This test statistics fail to reject the claim that the means are the same for the variable "Delinquency".