6-2015

Peer-to-Peer Lending Industry and Risk Control Measures

Ran Wang
Union College - Schenectady, NY

Peer-to-Peer Lending Industry and Risk Control Measures

By

Ran Wang

* * * * * * * * *

Submitted in partial fulfillment
of the requirements for
Honors in the Department of Economics

UNION COLLEGE
June, 2015

ABSTRACT

WANG, RAN. Peer-to-Peer Lending Industry and Risk Control Measures in the Big-Data Age. Department of Economics, June 2015.

ADVISOR: Professor Eshragh Motahar and Professor Roger Hoerl

With the rise of the Internet, a new form of financing, peer-to-peer lending (P2PL), has embraced its opportunities in the 21st Century. After Zopa, the world's first financial company that offers P2P loans, was founded in the UK, the U.S. also seized the trend and witnessed the launch of Prosper in 2006, followed by Lending Club. The IPO of Lending Club in 2014 created a faster momentum for the development of similar companies in the industry and cleared some concerns regarding SEC regulations. However, given the business model that P2PL companies adopt and the economic characteristics of P2P loans borrowers, the industry is still facing controversies over default risk controls. Therefore, it is important to understand the industry and its risk control measures. This paper presents an overview of the historical background of the P2PL industry and discusses its advantages and disadvantages that lead to the important role of risk control. By adopting the linear probability model and the logistic regression model, this paper proposes a method of measuring the default risk of P2P loans using the 2007-2011 Lending Club loan dataset. It finds that 8 variables in particular, employment length, inquiries by creditors in the last 6 months, installment, interest rate, annual income, public record, revolving line utilization, and term. While only annual income and public record have a

negative relation with default risk, all the remaining 6 variables will contribute to a higher default risk.

TABLE OF CONTENTS

CHAPTER ONE
PEER-TO-PEER LENDING INDUSTRY

1.1 Peer-to-Peer Lending and Its Development in the 21st Century

Peer-to-peer lending (P2PL), the idea of lending money to unacquainted individuals via online platforms or off-line networks with a focus on "disintermediation", has been one of the essential protagonists in recent developments of the so-called Internet Finance. From the launch of Zopa, the world's first P2PL company based in the U.K., to the initial public offering of Lending Club in 2014, the largest P2PL company in the U.S., P2PL has rapidly grown into a promising industry within short ten years. It keeps drawing attention not only from finance professionals, but also from ordinary people who are in search of alternative ways to obtain credit, and from investors who need better returns during the time of low interest rates. Although the general practice of lending and borrowing among peers in a society was already adopted long time ago, the popularization of P2PL occurred not until recently, as a result of the spread of the Internet, technological advances, and social changes in a much more interconnected world today. However, the general practice of lending and borrowing among social peers was already adopted a long time ago.

1.2 The Early History of Peer-to-Peer Lending

Many people believe that P2PL only came to existence in the 21st century as a result of the rise of Internet Finance. However, concepts relating to microloans are not new.

They can be traced back to as early as 4th century, when the first private credit union, called "Lun Hui" was founded in China (Zhang and Liu, 2012). This is considered to be the earliest origin of the P2PL concept. In the West, the idea was first realized in the 18th century when Jonathan Swift created the Irish Loan Funds (Zhang and Liu, 2012). Hollis and Sweetman (1996) claim that the funds were lending to around 20% of Irish households at that time and the system was remarkably successful at transferring capital to the poor on a large scale over time. They find out that the funds' structure allowed sufficient flexibility for the institution to even survive the Great Famine. The idea then took off and was injected new values when the Internet brought people together.

1.3 Current Development: the U.K, China, and the U.S. Markets

The world's P2PL industry is now mainly dominated by three markets: the U.K., where the modern-day P2PL companies were first founded, China, which is an emerging economy with ample credit supply, and the U.S., which has witnessed the first-ever initial public offering of an online P2PL company.

Chart 1 shows how the U.K. market has developed over time. It is based on the survey data conducted by altfi DATA, a U.K. data company that specializes in providing information about the P2PL industry of the U.K. financial sector. Nothing significant happened during the 2005-2009 period. Starting from 2010, the market has then skyrocketed to around £3.4 billion in cumulative lending volume up until mid-2015.

Chart 2 shows the increasing trend of P2PL investors. The aggregated number of investors increased from less than 20,000 in 2009 to more than 140,000 at the end of the first quarter of 2015. Investors are becoming more interested in seeking profits within the alternative finance sector in the U.K.

The development of the P2PL industry has also shown a great potential in the Chinese market. Because the financial sector is still tightly controlled by the government, the limited sources of investing drives the profit-seeking and well-funded Chinese investors to the relatively new area of finance. Chart 3 shows the increasing trend of P2P loans issued from 2010 to 2014. In short four years, total P2P loans in China skyrocketed from merely 100 million Yuan to around 50 billion Yuan. As shown in Chart 4, the number of P2PL companies increased from only a few to around 1200 as of June 2014. Chart 5 shows the 2015 market condition. China now has 1728 P2PL platforms, outnumbered the U.K. market. The flow of investors and borrowers each month is very dynamic. Monthly number of investors now reach 0.92 million, much more than the accumulated investors in the U.K. market as a whole.

The rapid development of the Chinese P2PL market has brought many competitions for companies. In order to separate themselves from peers, Chinese companies are eager to innovate their business models. Some companies are now doing both online and offline operations. A typical example is Credit Ease, the largest P2PL company in China which

first started off doing businesses offline in its early years when the use of the Internet was not prevalent. With the increasing popularity of smart phones and laptops in China, the company decisively grabbed the timing and founded online lending platforms to complement its offline chain. Some companies were only doing P2PL businesses at the beginning, but gradually incorporated themselves into wealth management firms that not only sell P2PL-related products, but also other financial products as well. Nearly every P2PL company in China claims to have some sort of third-party partners to make insurance on the loans issued. In this case, the key concept "disintermediation" in P2PL seems not that important in the Chinese market, which subsequently separates China from other markets that mainly operate online. Chapter 2 of this paper will come back to this point when discussing risk control measures.

The success IPO of Lending Club has attracted investors' attention to the U.S. P2PL market, as Lending Club is now the only publicly traded P2PL company in the U.S. However, Prosper was the first to operate in the U.S. It pioneered the U.S. market until Lending Club launched in May 2007 as a Facebook application and soon took over the market. But both companies now are still competing with each other and together they dominate the P2PL industry. Therefore, to understand P2PL in the U.S., one often looks at the two companies at the same time. Lending Club alone issued more than $7 billion loans as of 2014, compared with negligible amount in 2007, as shown in Chart 6. Chart 7

shows accumulative loans data included loans issued by Prosper, the second largest P2PL company in the U.S. The two giants issued around $10 billion loans in total as of 2014.

The extremely fast development of the three markets strongly shows that the expansion of the P2PL industry is a large-scale phenomenon and the market will continue to attract more lenders and borrowers in the future. Therefore, the increasing importance of this new alternative financing should not be ignored and is worthy of discussion.

1.4 How does a P2PL Platform Work

In order to understand the industry, one has to be familiar of how a P2PL platform works. The following four steps are generally shared by every online platform:

A. A borrower first submits a loan application to the platform he chooses to get services from. He cannot post any loan information online for now.

B. The platform then issues a credit report based on the information provided in the application and assigns a risk grade to the loan. It also settles a certain interest rate.

C. If the information is considered to be able to reveal the borrower's personal and financial conditions comprehensively, and the borrower's credit score is within a good range, the platform will then approve the loan request. Now the borrower is allowed to post loan information online for investors to pick. However, this does not mean the loan is successful.

D. The loan is successfully launched only if there are enough lenders to fund, as a single loan is oftentimes divided into many portions, each with a much small value for investors to diversify their loan portfolios.

A platform usually profits from operating fees on the loans charged on investors as well as the origination fees charged on borrowers. Some platforms also profits from service fees via special services, such as helping investors retrieve back their funds if a loan has a high probability of default.

Chart 8 shows a loan-listing page from the Lending Club platform. Investors can invest in a loan for as low as $25. If an investor wants to build a portfolio with $2,500, he can invest $25 in 100 different loans, which is a great way to spread the risk. Chart 9 shows the information of a single loan disclosed by Lending Club for investors to refer to. Besides the often-required loan data such as credit scores and income, Lending Club also offers information on recent public records, revolving credit, and so on.

1.5 Microfinance and Peer-to-Peer Lending

Microfinance, also known as microcredit, is often discussed along with P2PL. The two terms are not mutually exclusive, as the general idea of how they function is the same. However, one can separate them based on the methods and goals that they have. Microfinance is more concerned with low-income families or individuals who lack credit

access in the developing world that usually has incomplete financial systems (Lending Club). The concept is pioneered by Muhammad Yunus, economist and Nobel Peace Prize winner for his efforts in promoting economic and social developments from below, with his founding of the Grameen Bank in Bangladesh. Then microfinance platforms such as Kiva and Zidisha have made lending across boarder a worldwide mission. Since Kiva was founded in 2005, it has secured more than 600 million dollars in loans for 1 million lenders and worked with field partners in 86 different countries (Kiva). Zidisha also helped fund more than 6000 entrepreneurial projects in 9 developing countries since 2009 (Zidisha). Microfinance offers an innovative way to solve development problems using old idea.

P2PL is generally referred to as for-profit financial transactions occurring directly between individuals without the intermediation of a traditional financial institution. It is considered more as a profit-oriented lending method. However, in China, P2PL companies often add some social components to the services. Some founders of P2PL companies also market their companies as a mixed entity of a microfinance institution and a for-profit financial services company. A typical example is Credit Ease. The founder of Credit Ease, Tang Ning, met Yunus himself and spent years in Bangladesh exploring village economy. When Yunus won Nobel Prize in 2006, Tang Ning founded Credit Ease and introduced the concepts of microfinance and P2PL to the Chinese market.

He described his company as a leading wealth management company specializing in P2PL services and fulfilling its social responsibility by helping small businesses and rural families achieve better life. Some of the products that the company now is offering, such as Yinongdai that targets low-income rural peasants, show distinguished features of microfinance. But at the same time, the firm also provides wealthy investors in large cities with better asset management strategies such as buying P2PL products. The company profits from management fees, which shows the for-profit side of P2PL. In China, P2PL companies are more diversified in operations. P2PL can absorb good and innovative practices from other areas of the financial industry, which usually involves using social causes as eye-catchers to attract customers who pay attention to charities.

But no matter whether an organization is a microfinance platform or a P2PL company, it cannot ignore risk control measures. Even if microfinance lenders care less about returns, they might also want to lend money to those who are really capable of making projects work. Therefore, the risk control measures discussed in this paper are relevant to both microfinance and P2PL.

CHAPTER TWO
RISK CONTROL MEASURES

Risk control measures are very important in the successful operation of a P2PL platform. They also help investors reduce make wiser decisions to reduce losses. However, before discussing the measures, one needs to understand what makes risk control so crucial.

2.1 The Advantages of Peer-to-Peer Lending

Many platform users find that the attractive advantages of P2PL are what make them choose not to borrow money from a traditional commercial bank or to invest in the products. Borrowers mainly treasure two advantages. First, the efficient assessing strategies adopted by P2PL platforms greatly reduce processing time, as everything is operated online. This is very beneficial for those who are in an urgent need of credits. Borrowers also find that the interest rates charged on P2P loans are often lower on average compared to credit card APRs or the interest rates of installment loans. Chart 10 shows the APRs of various credit cards. On average, they all exceed 15% per year. However, according to Lending Club data in Chart 11, the annual average interest rate of all loans issued over time is hardly beyond 15%.

Investors mainly value three advantages of P2PL platforms. First, the risk-adjusted annual returns of P2P loans are much higher than those of bonds available in the

marketplace, which make P2P loans extremely attractive when the interest rates are still kept low after the 2008 financial crisis. Chart 12 shows the average annual rate of P2P loans in the U.K. marketplace since inception. The rate is maintained at around 5.5% and is even capable of securing a good level of 4.5% during the 2008 financial crisis. As shown in Chart 5, the average annual return of P2P loans in the Chinese market exceeds 15% in 2015, even higher than the returns in the stock markets. In the U.S., the average annual returns of P2P loans issued by Lending Club and Prosper all exceed 5% for all grade levels during the post-2008 financial crisis period. As mentioned earlier, low investment limits allow those who do not have many savings at hand to invest as well. Like buying stocks, investors of P2P loans are also granted much freedom to make decisions according to their own judgment. The autonomy of investments especially brings satisfaction to the financially savvy.

The advantages of P2PL have made it a revolutionary force in consumer finance. In P2PL, lenders and borrowers are the major decision-makers. The “disintermediation” of P2PL subsequently allows P2PL companies to offer attractive rates to both sides without the expensive branches that typically make high street banks so cautious (Shadbolt, 2015). Besides the freedoms embedded in transactions, many P2PL companies also provide enough information transparency to their customers, which is nowhere to be found in large banks. Prosper and Lending Club give the public the access to key data associated

with each successful or default loan. The disclosure of major business data offered by P2PL companies encourages a democratic way of participating financial activities for investors and provides transformational insights that are shaping the conventional consumer finance industry dominated by megabanks. It seems that after all these years, the financial industry has a trend to return to its old way of channeling capital, just like "Lun Hui" and the Irish Loan Funds, but with the extra help of the Internet.

2.2 The Disadvantages of Peer-to-Peer Lending

If P2PL is such an attractive market with many advantages, it should be well known to investors and borrowers after a decade of development. However, this is generally not true, as its disadvantages have raised many concerns for both government authorities and consumers.

The democratic nature of P2PL also brings controversies. Although many platforms set credit score standards to limit subprime borrowers from entering the market, the motives of qualified borrowers still tend to convey high default risks in P2P loans. In most cases, borrowers seek P2PL when they face urgent financial conditions and cannot get loans from commercial banks in a short time, or when they cannot provide all the official documents needed to secure loans from banks. Chart 13 shows the grade mix of P2P loans issued by Lending Club. A means the best loan quality with the least default rate. G means the worst case. Although the grade mix has improved a little over time, the

loans pool is still dominated by those with a grade lower than B. Chart 14 shows the reported use of the loans. The two major loan purposes are refinancing and credit card payoff, which implies incurring new debts to pay off old debts, a risky financial behavior. Since the financial drives of P2P borrowers reveal higher risks for P2P lenders, some investors become hesitated when it comes to whether or not to enter the market.

From lenders' perspective, because P2PL platforms only serve a minimal role in transactions, it is largely up to lenders themselves to determine what loans to make. Therefore, lenders rely heavily on the information provided by borrowers on the platforms and on the level of financial sophistication they already grasp. Those who do not have good financial knowledge can be lured to make bad loans based on incomplete or distorted information. On a regulatory perspective, P2P loans are not subject to deposit insurance and have no lender-of-last-resort protection from a central bank (Frezza, 2013). In the U.S., although the SEC began to regulate P2PL companies since 2008, additional regulatory oversight by Consumer Financial Protection Bureau is still vacant. Therefore, given current regulatory environment of P2PL, borrowers of P2P loans in the U.S. have to count on their own judgments to implement risk controls.

Other disadvantages of P2PL reveal the risky character of this industry. First, as mentioned earlier, limited diversification of borrower types makes P2P loans easily subject to macroeconomic fluctuations. Even though the return of P2P loans in the U.K.

market was still strong during the financial crisis, the performance of loans issued by the two giants in the U.S. market was unfavorable. Chart 15 shows the annual returns of Lending Club and Prosper loans in recent years. During 2006-2008, the return of P2P loans became negative. The return of Lending Club loans in 2007 also became negative, and the return was merely above 0% in 2008. Due to the financial characteristics of the majority of P2PL borrowers, P2PL loans are more easily influenced by macroeconomic shocks. In addition, the short history of P2PL platforms may suggest that investors are unable to see systematic risks, if they exist.

2.3 Current Risk Control Measures

Since the P2PL industry is full of advantages, opportunities, and potentials, how we could improve risk management strategies for investors and even platforms in assessing personal loans becomes extremely important to the healthy development of the market.

In China, two distinctive business models adopted by P2PL companies put risk control responsibilities onto their own shoulders. Due to special cultural and financial background, P2PL companies in China play a more important role in transactions. Most of them have features of an asset management company, taking in the funds offered by investors and lending out the money to others on behalf of investors. To expand market and compete with each other, P2PL companies in China usually claim to promise a certain return to lenders. In order to do that, they set up various operational branches as

well as a key risk control department to select good loans from bad loans. Although investors still face default risk, this type of risk usually comes from the failing of P2PL companies rather than the default of individual loans.

Another type of business model resembles the one currently adopted by major P2PL platforms in the U.S., operating mainly online and granting lenders the freedom to make investment choices, but it has something new added. Many P2P online platforms in China that adopt the second model usually claim that their loans are protected and secured by third-party insurance companies with which they cooperate. In addition to that, loans listed online will first be scrutinized by the platforms themselves and then posted to investors. Even if a lender defaults on his loan at the end, platforms will first help investors retrieve the loan and then offer full compensation from third-party insurance companies if the loss cannot be recovered. In this case, platforms themselves rather than government institutions invent various safety nets as a way of marketing their business. It all seems too good to be true for some investors, as P2P loans look like the perfect investment with high return and almost zero risk, but they might ignore the risk mentioned earlier, the failing of P2PL companies. In 2014, users reported problems with 275 P2PL platforms out of a total of 1,575 in China. 60 were reported to have owners who ran away and 71 were labeled as “scams” (Wildau, 2015). Therefore, for Chinese investors, the major risk associated with P2P loans comes from platforms. The more

important question that matters to investors is which P2PL company to choose, not which specific borrower to lend money to.

In the U.K. and the U.S. markets, investors are mainly responsible for risk control. However, during the four steps of the operation process of an online P2PL platform mentioned in Chapter One, platforms do actually engage in risk control at certain level. For example, when a platform assigns a risk grade to a loan, it is actually telling investors to pay attention to the potential risk. In addition, the rule that a loan is only successfully launched if there are enough investors set by platforms offer investors some sort of "collective intelligence." Investors could draw on what others think about a loan according to what percentage of the loan amount has already been funded by others. However, this could be problematic, as Shen et el. (2010) claim that there were significant herding effects when lenders made their investment decisions on loan listings, and these effects did not lead investors to make better investment decisions. Lenders did not make rational investment based on risk and returns, but followed the herd instead. Therefore, to make the best decisions, P2P loans investors may need to rely heavily on their own financial knowledge. Thanks to the openness of P2PL data provided by Lending Club and Prosper, communities of investors have formed to share their experiences in risk control using the data. In addition, educational websites about P2PL, such as Lend Academy and LendingStats.com, are often seen by new investors as a good

source to gather information from.

2.4 Literature Review on Risk Control Measures

It is crucial to carefully analyze what variables are the important factors that affect the default risk of P2P loans. By building a probit regression model using Prosper data, Michels (2012) claims that the size of the loan strongly predicts future default, as does having a prior loan with poor performance. Among all the controlled variables that he uses, only having a prior loan with current payments is significantly associated with a lower risk of future default. However, besides Michels research, there is not much existing literature discussing the risk control measures of the P2PL industry. Although some literature explores what factors lead to the successful funding of a loan, scholars have not started to pay much attention to the relatively new branch of consumer finance. Therefore, this paper will provide some original thoughts about helping investors reduce risk of P2P loan investments. Chapter Four will discuss the risk model in a much greater detail.

CHAPTER THREE
DATA DESCRIPTION

3.1 Lending Club Dataset

This paper uses the 2007-2011 Lending Club Dataset that includes all loans issued through the time period. The reason for choosing this particular dataset while leaving out the more recent ones is that because the term of all the Lending Club loans is either 36 months or 60 months, only loans issued from 2007-2011 could have the final outcome of either "default" or "fully paid". In order to build the model, the final outcome of loans is necessary. Therefore, the more recent data will not satisfy the requirement.

There are 42538 observations in total, but some loans of 60-month term are still in payment period, which means that the final outcome of these loans cannot be determined so far. Therefore, loans with a status "Current" are deleted. The remaining 36117 observations with definite final outcomes will be used for analysis.

Each loan has 52 variables. However, not all variables can imply information about the future default rate of a loan. Hence, variables such as "member id" and "url" are deleted. Other variables, such as "next payment date" and "collection recovery fee", are only created after a loan is successfully issued, therefore, they can not help investors choose a loan. These variables are also deleted. Finally, for similar variables such as "number of open accounts " and "number of total accounts", the more exclusive ones are

kept. The following rearrangements of the dataset are also made for the purpose of building statistical models:

A. For the variable “verification status of income”, “verified” is replaced by 1 and “not verified” is replaced by 0.

B. For the variable “grade”, A is represented by 1, B is represented by 2, etc.

C. For the descriptions that the borrowers provide regarding the loans, the number of characters in the description is calculated to replace the original texts.

D. For the variable “employment length”, “n/a” is considered as 0, “<1” is considered as 1 year, and “10+” is treated as 10 years.

E. For the variable “loan status”, only “fully paid” is given a value of 0, which means “no default”, while all the other conditions, such as “charge-off” and “late payment”, are all treated as “default” with a value of 1. Even “late payment” and “in grace period” could still signal that a loan could be recovered, still they are considered to be the same as default, as payments have passed the solid deadline on the contract.

Table 1 shows a list of the 19 variables left for discussion and their definitions.

3.2 Summary Statistics

The summary statistics is shown in Table 2. Around 15% of the borrowers default on the loans. The median annual income of Lending Club borrowers is about $58,000. The

average DTI ratio is 13.21%, which is at a reasonable level. The average employment length of borrowers is 4.77 years. Average interest rate is 12%. Mean loan amount is around $11,000. 44% of the borrowers have mortgage, 48% rent home, while only 6% of the borrowers own their own homes. Most of the loans are short 3-year term. A little more than half of the borrowers have verified income source. Among the 19 variables, LOAN_STATUS is the dependent variable and the rest of them are potential factors that could influence the default rate of a loan.

CHAPTER 4
MODELS AND EMPIRICAL RESULTS

This chapter outlines the two statistical models used in the analysis as well as the empirical results generated by the various specifications.

4.1 Estimating the Default Risk of Peer-to-Peer Loans: Linear Probability Model

This section discusses the linear probability model and its regression results.

4.1.1 Model Set-Up

To examine the default risk of P2P loans, LOAN_STATUS with value 0 (fully paid) and 1 (default) is the dependent variable in the analysis. Since it has a binary value, linear probability model is a suitable methodology. It is used when the probability of observing a 0 or 1 is dependent on explanatory variables. This relation is a straightforward one, as it can be fitted using simple linear regression.

Among the 18 explanatory variables, some may have strong correlations with the others, which could affect the accuracy of the model. Table 3 shows the strong correlations among several variables. Because GRADE is strongly correlated with three variables, which has correlations over 40%, it will not be included in the model. A possible explanation is that grade largely determines the interest rate of a loan and when Lending Club calculates grades, it treats TERM and REVOL_UTIL as important indicators. Because loan amount directly impacts loans installment, they have strong a

correlation. Therefore, only INSTALLMENT is used in the model. ANNUAL_INC is replaced by its log term to reduce skewedness.

The linear probability model is as follows:

$$\Pr(LOAN\_STATUS = 1) = \beta_0 + \beta_1 DELINQ\_2YRS^{+} + \beta_2 DESC\_CHAR^{+} + \beta_3 DTI^{+} + \beta_4 EMP\_LENGTH^{-} + \beta_5 HOME\_OWNER^{-} + \beta_6 INQ\_6MTHS^{+} + \beta_7 INSTALLMENT^{+} + \beta_8 INT\_RATE^{+} + \beta_9 LOG\_INC^{+} + \beta_{10} MORTAGE^{+} + \beta_{11} PUB\_REC^{+} + \beta_{12} REVOL\_UTIL^{+} + \beta_{13} TERM^{+} + \beta_{14} TOTAL\_ACC^{+} + \beta_{15} VERIFICATION^{-} + \varepsilon,$$

where ε is the error term and $\Pr(\mathrm{LOAN\_STATUS} = 1)$ is the probability of default. The estimated effects of the explanatory variables are denoted as positive (+) or negative (-) in their upper right.

DELINQ_2YRS is expected to increase the default risk of P2P loans, as recent delinquencies may suggest that a borrower would have the tendency to fall behind payments again. It is also a sign showing the borrower's financial ability is less trustworthy. DESC_CHAR is expected to increase the probability of default, as a less honest borrower in urgent need of credits or a borrower whose credit history is very questionable is highly likely to make up a good story to compensate for their disadvantages. DTI is expected to have positive affect. The higher the ratio, the more likely a loan will default in the future. EMP_LENGTH is expected to reduce default rate, since the longer length of employment history oftentimes suggests financial stability. Home ownership is always included in the standard banking credit check. Those who

own a house are expected to be in a better financial condition while those who rent or have mortgages are considered to face more financial pressure. Therefore, HOME_OWNER has a positive sign, whereas MORTGAGE has a negative sign. RENT is not included to avoid multicolinearity problem. INQ_6MTHS, like DELINQ_2YRS, might also suggest problems in the borrower's credit history, and therefore, it has a positive sign. Higher INSTALLMENT, INT_RATE, and REVOL_UTIL values as well as more TOTAL-ACC could put more financial pressure on a borrower, so they all have positive signs. LOG_INC has a negative sign, as a higher income usually means a better ability to make payments. PUB_REC is positive, as derogatory records could put a borrower's credibility in question. TERM is positive, since longer terms signal more financial uncertainty. VERIFICATION is negative, because a verified income source has more credibility than unverified, and therefore, may reduc2e default risk.

4.1.2 Regression Results

The regression results of the basic run of the model are reported in Table 4. LNQ_LAST-6MTHS, INSTALLMENT, INT_RATE, PUB_REC, REVOL_UTIL, and TERM are strongly significant. As expected, INSTALLMENT, INT_RATE, PUB_REC, REVOL_UTIL, and TERM all positively affect default risk. Increasing them by one unit is expected to add 0.27%, 1.26% (treat 1% as one unit for INT_RATE and REVOL-UTIL), 4.8%, 0.05%, and 0.7% on average, respectively, ceteris paribus.

Increase annual income by 1% is estimated to reduce default risk by 0.06% on average, holding other variables constant. Increase INQ_LAST_6MTHS by 1 unit is estimated to increase the default risk by 1.8% on average, ceteris paribus. Therefore, if the number of inquires by creditors in the last 6 months is large, this could be a strong signal telling that the borrower may undergo serious financial pressure, which leads to suspicious patterns in the debt repayments of several credit lines. However, the variable DELINQ_2YRS, which is a similar indicator, is not statistically significant. A possible explanation is that different time line may cause different outcomes. If a borrower had delinquencies 2 years ago that subsequently affected his chance to successfully get credits from large commercial banks, the borrower may see the opportunity given by Lending Club as very special. He may then try to keep up with the payments to compensate for the bad credit history in the past.

However, contradicting to the expectation, there is no strong relation between types of home ownership and default risk. The number of character in a borrower's description is also not an important factor. Surprisingly, DTI, a crucial indicator that shows a borrower's indebtedness, is not statistically significant. A possible explanation is that, as shown in Table 2, even the maximum DTI does not exceed 30% and the median is around 13%. This means that most Lending Club borrowers are not heavily in debt, and hence, other debts they hold do not seriously impact their ability to repay Lending Club

loans. TOTAL_ACC fails to be statistically significant as well, which could mean that the total number is not necessarily equal to the number of active accounts. Some borrowers may open many accounts, but only use a very few. VERIFICATION is not at all statically significant, which means that most borrowers are quite honest about their annual income when applying for P2P loans, and verification does not play a key role in determining the final outcome of a loan.

The model is then modified by deleting all the very insignificant independent variables, DELINQ_2YRS, DESC_CHAR, DTI, HOME_OWNER, MORTGAGE, TOTAL_ACC, and VERIFICATION, in the basic run. Table 5 reports the new regression results. All the significant variables in the basic run remain significant. In addition, EMP_LENGTH, which is not significant at the 99% confidence level in the basic run, is now statistically significant at the level. However, it is estimated to have a positive effect instead. Increasing EMP_LENGTH by one year will lead to a rise in default risk by 0.1% on average, ceteris paribus. This rejects the previous conjecture. Borrowers with longer years of employment have higher default risk. A possible explanation is that EMP_LENGTH to some extent also implies age. People with longer employment history may be in their middle age when mortgages and the expenses of raising children put much financial burden on them. Lending Club does not provide information on age, so it is hard to separate the effects of employment length and age.

Another major change is that the coefficient of PUB_REC changes from positive to negative. This may due to the fact that the deletion of insignificant variables reveals a more accurate relation between PUB_REC and default risk in the new model. However, a positive sign contradicts the original expectation, as now more public records lead to lower default risk. A possible explanation is one that is very similar to the insignificance of DELINQ_2YRS, people with public records may be rejected loans from commercial banks, and therefore, they treasure the opportunity given by P2PL by repaying installments on time. Another possible reason is that, because the negative coefficient is too small, although it is significant, it has much weaker effects than other variables.

Now the model has 8 independent variables that are all significant at the 99% confidence level. The entire model explains around 9.1% of the changes in default risk. The Durbin-Watson statistics is close to 2, which means that there is no serial correlation in the sample. The linear probability model seems to be reasonable. However, some inherited drawbacks of the linear probability model may cause problems. For example, the predicted probabilities can be greater than 1or less than 0, which does not make sense in denoting the default risk. Therefore, the second model, logistic regression, is introduced in the next section.

4.2 Estimating the Default Risk of Peer-to-Peer Loans: Logistic Regression Model

This section discusses the logistic regression model and its regression results.

4.2.1 Model Set-Up

Logistic regression, also known as logit model, is a probability model that deals with dependent variable with binary values, such as the default risk of P2P loans with 1 as default and 0 as fully paid. It is able to constrain the predicted values within the binary range and solve the heteroskedasticity problem of the error terms in linear probability models. The model is constructed as follows:

$$In[\frac{\Pr(LOAN\_STATUS=1)}{(1-\Pr(LOAN\_STATUS=1))} = \beta_0 + \beta_1 EMP\_LENGTH + \beta_2 INQ\_LAST\_6MTHS + \beta_3 INSTALLMENT + \beta_4 INT\_RATE + \beta_5 LOG\_INC + \beta_6 PUB\_REC + \beta_7 REVOL\_UTIL + \beta_8 TERM + \varepsilon,$$

where ε is the error term, and $\Pr(\mathrm{LOAN\_STATUS} = 1)$ is the probability of default. $\frac{\Pr(\mathrm{LOAN\_STATUS}=1)}{(1-\Pr(\mathrm{LOAN\_STATUS}=1))}$ is called the "odds ratio" and $\mathrm{In}[\frac{\Pr(\mathrm{LOAN\_STATUS}=1)}{(1-\Pr(\mathrm{LOAN\_STATUS}=1))}$ is the log odds ratio, or simply "logit".

4.2.2 Regression Results

The regression results of the logistic model are presented in Table 6. The variables are all statistically significant. In particular, INQ_LAST_6MTHS, INT_RATE, LOG_INC, REVOL_UTIL, and TERM are all significant at the 99% confidence level, while EMP_LENGTH, INSTALLMENT, and PUC_REC are significant at the 95%

confidence level, which shows that they are less dominated factors in this model than the previous ones.

The coefficients in the logistic model have different meanings from those in the linear probability model. For example, the coefficient of EMP_LENGTH means that increase the employment history by one year will increase the log odds ratio by 0.6%, on average, ceteris paribus. In order to understand what the actual probability of default is, a transformation of the log adds ratio is necessary. The final model is as follows:

$\Pr(LOAN\_STATUS = 1) = \frac{1}{[1+exp(M)]}$,

where $M = -(0.510 + 0.006EMP\_LENGTH + 0.086INQ\_LAST\_6MTHS + 0.0001INSTALLMENT + 6.058INT\_RATE - 0.327LOG\_INC - 0.001PUB\_REC + 0.263REVOL\_UTIL + 0.024TERM$.

Therefore, according to the two models in the paper, when making P2P loan investments, lenders need to pay particular attention to the 8 variables. They provide some rationale for investors to objectively evaluate the risk of P2P loans.

CHAPTER FIVE
CONCLUSION

5.1 Summary of Findings

This paper provides a historical review of the P2PL industry. It evaluates the current development of the industry in three major markets: the U.K, China, and the U.S., and discusses the similarities as well as distinctions among them. While P2PL companies in the U.K. and the U.S. largely operate through online platforms, the competitive market in China forces some Chinese P2PL companies to invent new business models to deal with competition. But overall, all the three major markets have seen dramatic development in less than a decade.

The P2PL industry has many advantages that conventional commercial banks do not have, and they allow P2PL companies to continue to grow in the future. However, the drawbacks, typically, the investment risks for both consumers and platforms that do risk control on their own have generated some doubts about the market.

Very few existing literature discusses risk control measures of P2P loans. Therefore, by using the 2007-2011 Lending Club dataset, this paper creates two models to assess the default risk of P2P loans. It finds evidence that 8 variables in particular, employment length, inquiries by creditors in the last 6 months, installment, interest rate, annual income, public record, revolving line utilization, and term, have large impacts on the

default risk of P2P loans.

5.2 Suggestions for Future Research

This paper only uses two methods to build the risk model. During the era of big data, big data analytics, such as support vector machines and artificial neural networks, are also good ways to analyze large datasets like P2P loan information. Future research could use these methods to detect if there are more underlying relations among the variables.

As the P2PL industry continues to grow, its significance in consumer finance is increasing as well. Financial information of P2P loan borrowers may suggest a general trend of the financial soundness of this particular consumer group as a whole, and may subsequently reflect the condition of the current macroeconomic environment. Since the 2007-2011 Lending Club dataset include time-series data, scholars could conduct research to explore how financial crisis has influenced the P2PL industry, and whether the information provided by borrowers in 2007 and after the crisis is able to offer some economic implications.

BIBLIOGRAPHY

altfiDATA. www.altfi.com. Accessed May 1, 2015.

Frezza, Bill. "Caveat Emptor Banking: Peer-To-Peer Lending Challenges Too-Big-To-Fail Status Quo." Forbes. August 13, 2013. Accessed March 28, 2015. http://www.forbes.com/sites/billfrezza/2013/08/13/caveat-emptor-banking-peer-to-peer-lending-challenges-too-big-to-fail-status-quo/.

Herzenstein, Michal, Scott Sonenshein, and Utpal M. Dholakia. "Tell me a good story and I may lend you money: the role of narratives in peer-to-peer lending decisions." *Journal of Marketing Research* 48, no. SPL (2011): S138-S149.

Herzenstein, Michal, Rick L. Andrews, Uptal Dholakia, and Evgeny Lyandres. "The democratization of personal consumer loans? Determinants of success in online peer-to-peer lending communities." *Boston University School of Management Research Paper* 2009-14 (2008).

Hollis, Aidan, and Arthur Sweetman. *The Evolution of A Microcredit Institution: The Irish Loan Funds, 1720-1920*. University of Toronto, Department of Economics, 1996.

Kiva. www.kiva.org. Accessed March 2, 2015.

Lend Academy. www.lendacademy.com.

Lending Club. www.lendingclub.com. Accessed March 20, 2015.

Lendstats. www.lendstats.com. Accessed March 3, 2015.

Michels, Jeremy. "Do unverifiable disclosures matter? Evidence from peer-to-peer lending." *The Accounting Review* 87, no. 4 (2012): 1385-1413.

Peer-2-Peer Finance Association. P2pfa.info. Accessed March 16, 2015.

Prosper.www.prosper.com. Accessed March 5, 2015.

Shen, Dawei, Coco Krumme, and Andrew Lippman. "Follow the Profit or the Herd? Exploring Social Effects in Peer-to-Peer Lending." In*SocialCom/PASSAT*, pp. 137-144. 2010.

Shadbolt, Peter. "No Bank Required: How Peer-to-peer Is Changing the Lending Landscape" CNN. January 5, 2015. Accessed March 5, 2015.

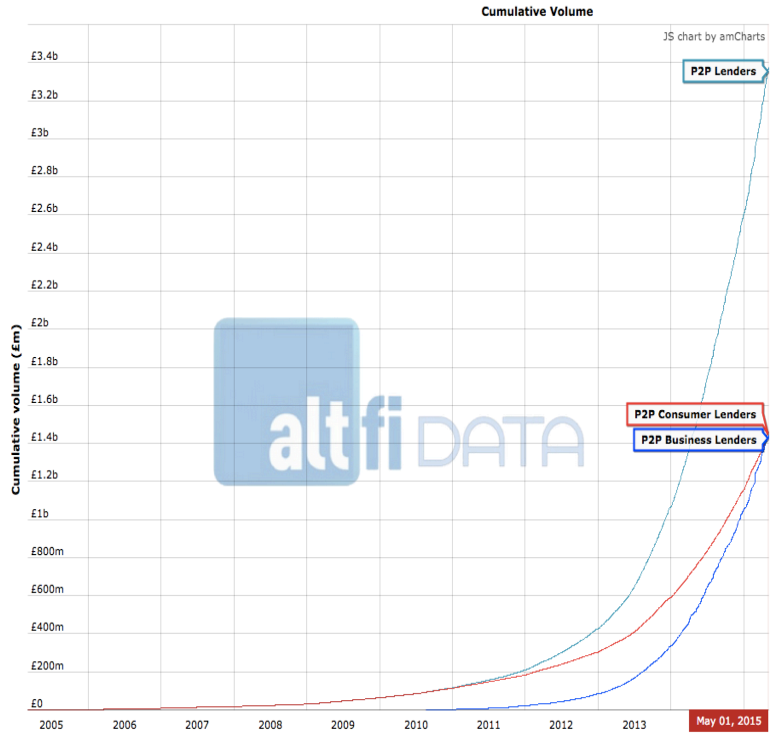
Wildau, Gabriel. "Democratising Finance: China's P2P Industry Attracts Scammers" Financial Times. January 28, 2015. Accessed February 10, 2015.

Zhang, Juanjuan, and Peng Liu. "Rational herding in microloan markets."Management science 58, no. 5 (2012): 892-912.

Zidisha.org. www.zidisha.org. Accessed March 17, 2015.

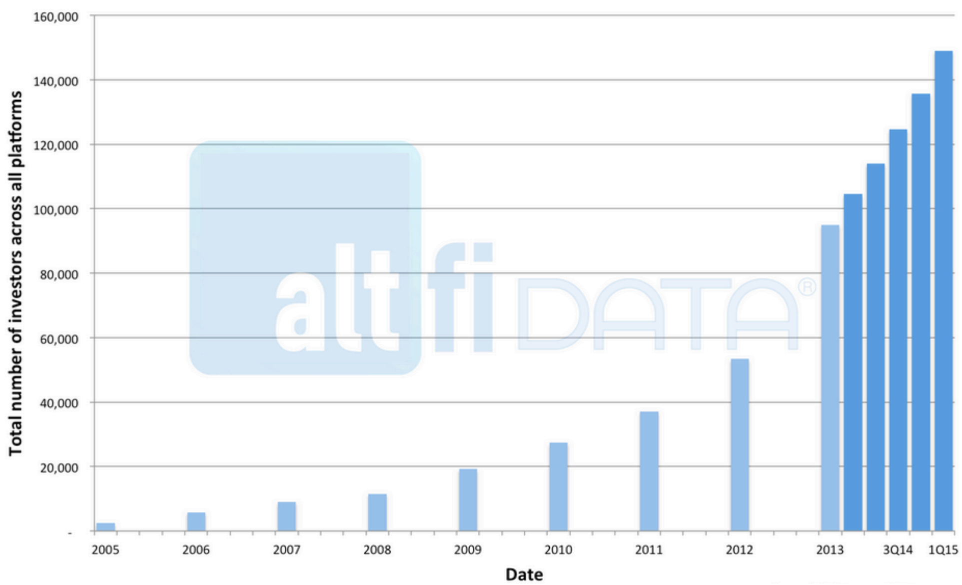
CHART

Chart 1: The Development of the U.K. Market



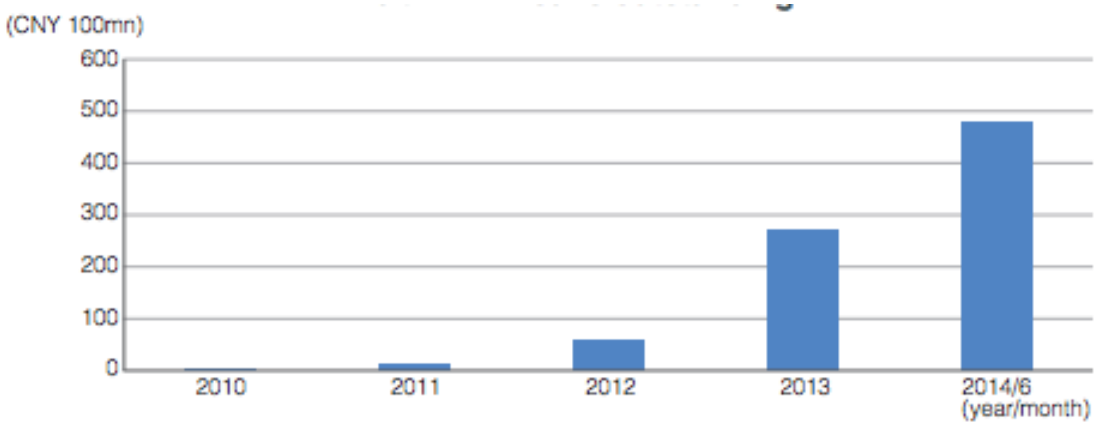
Source: altfi DATA. www.altfi.com

Chart 2 The Trend of P2PL Investors in the U.K. Market



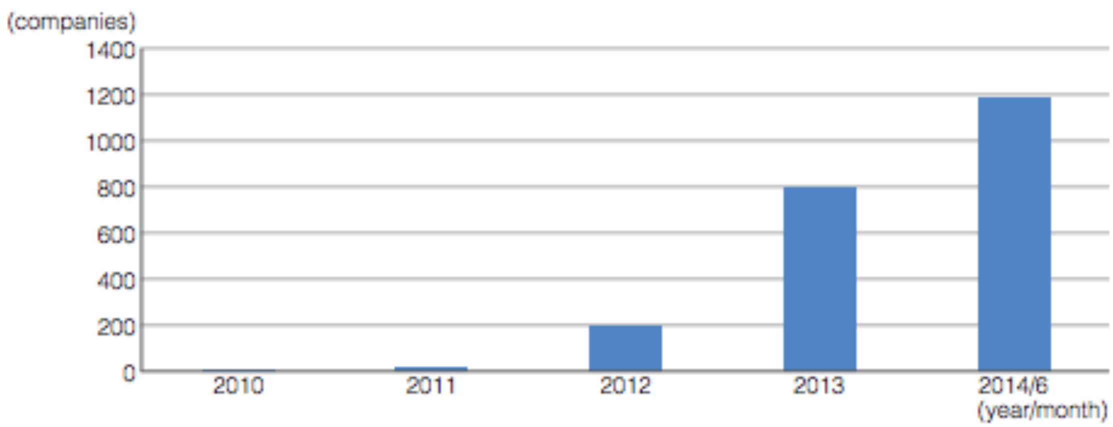
Source: altfi DATA. www.altfi.com

Chart 3: P2P Loans Value in China, 2010-2014/6



Source: Nomura Research Institute, Ltd., based on Wangdaizhijia data.

Chart 4: Number of P2PL Companies in China, 2010-2014/6



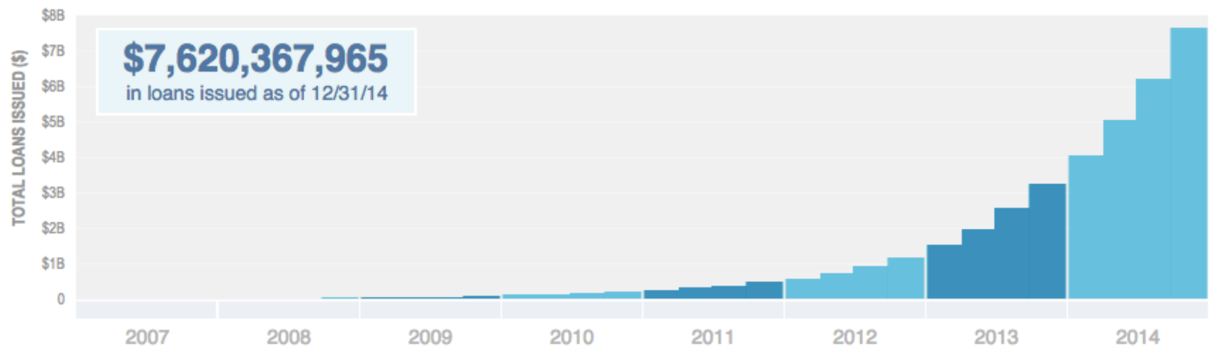
Source: Nomura Research Institute, Ltd., based on Wangdaizhijia data.

Chart 5: P2P Loans in China as of 2015

	Successful Deal	Number of Platform	Average Annual Return	Average Loan Period	Number of Investors/ Month	Number of Borrowers/ Month
National Level	49.26 billion Yuan	1728	15.02%	6.71 months	0.92 million	0.18 million

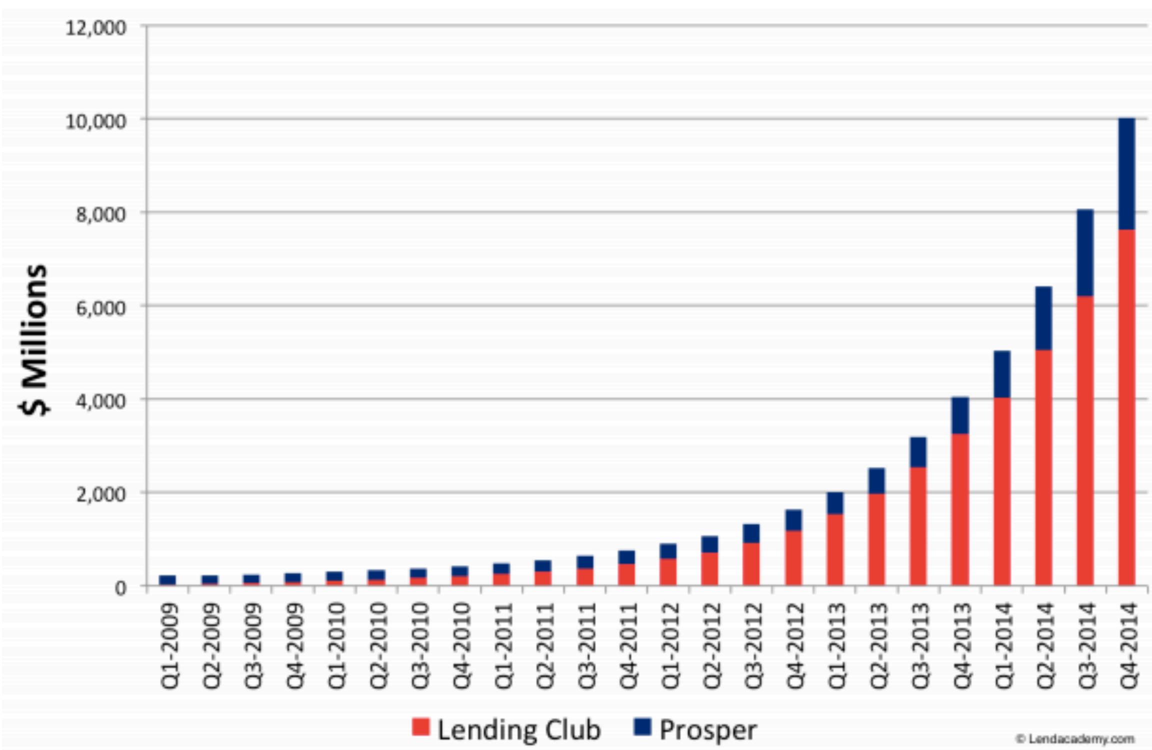
Source: Wangdaizhijia data.

Chart 6: Total Loan Issuance by Lending Club, 2007-2014



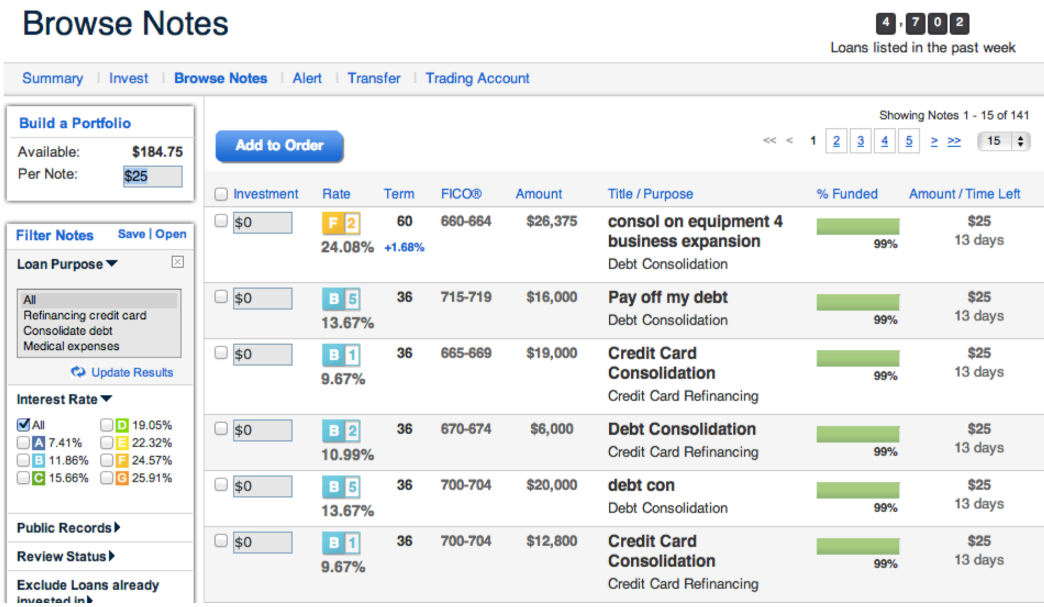
Source: Lending Club.

Chart 7: Lending Club and Prosper Cumulative Loan Totals 2009-2014



Source: Lendacademy.com

Chart 8: Lending Club Loan Listings



Source: Lending Club.

Chart 9: A Lending Club Loan Listing

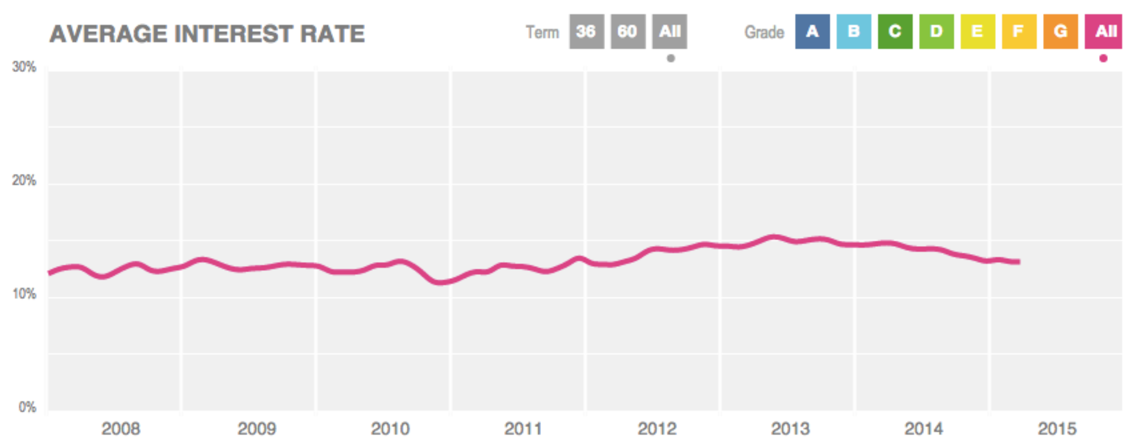
Source: Lending Club.

Chart 10: Credit Card APRs in 2015

Card Type Type	Low	High	Overall
Travel Rewards Cards	15.15	22.19	17.89
Airline	14.74	22.34	18.09
Hotel	14.99	20.60	17.09
Business Credit Cards	13.33	20.08	16.00
Cash Back Credit Cards	14.01	21.67	17.48
Student Credit Cards	15.21	23.33	18.69

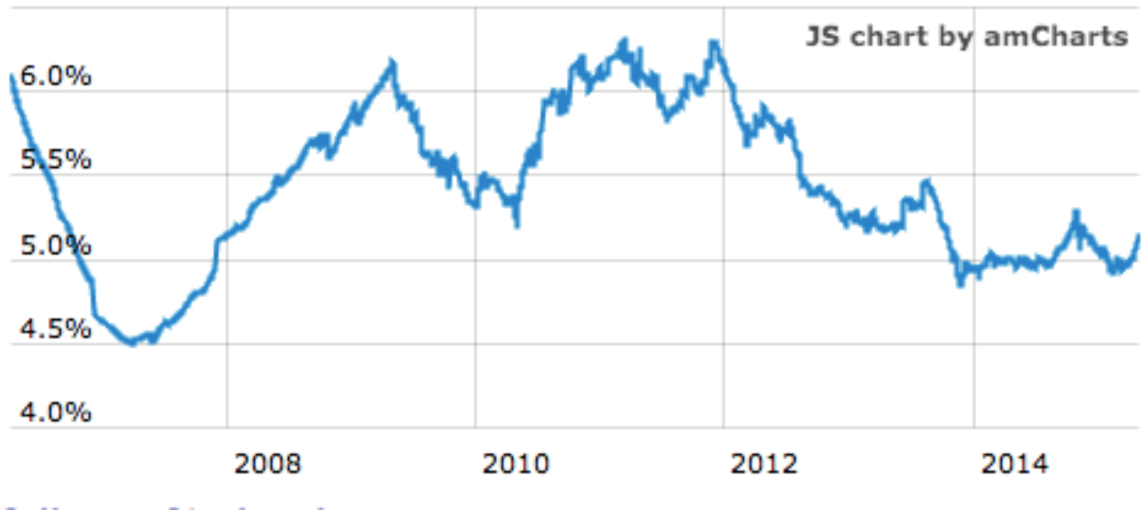
Source: Value Pengui.com

Chart 11: Lending Club Average Interest Rate



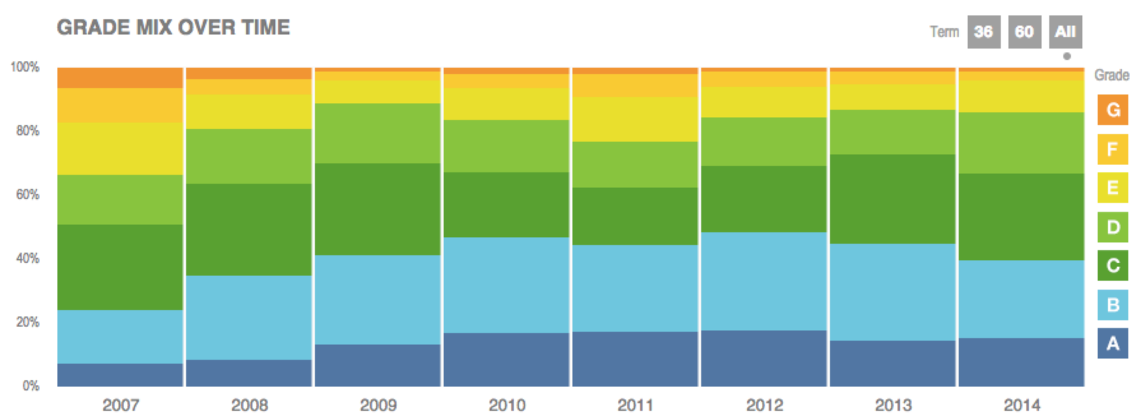
Source: Lending Club.

Chart 12: Average Annual Interest Rate of P2P Loans in the U.K.



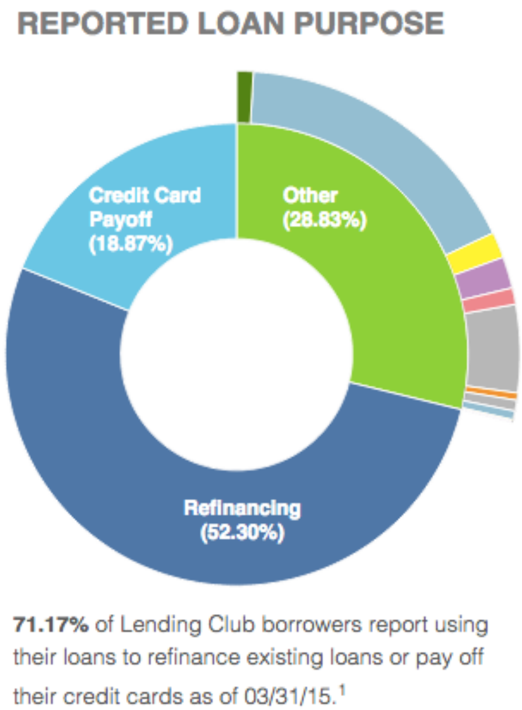
Source: altfi DATA.

Chart 13: Grade Mix of P2P Loans at Lending Club



Source: Lending Club.

Chart 14: Reported Loan Purpose of Borrowers at Lending Club



Source: Lending Club.

Chart 15: Prosper and Lending Club Loans Performance during the Financial Crisis

Prosper				LendingClub			
Vintage	ROI	Age(m)	Loans	Vintage	ROI	Age(m)	Loans
2014	9.26%	11.1	110174	2014	8.72%	13.3	161227
2013	9.01%	21.1	33674	2013	8.11%	22.2	134756
2012	6.57%	33.7	19405	2012	6.7%	33.3	53367
2011	8.06%	38.4	11148	2011	6.28%	41.9	21721
2010	9.85%	36.8	5631	2010	6.21%	42.6	12537
2009	8.88%	36	2034	2009	4.91%	36	5281
2008	-2.29%	36	11471	2008	0.58%	36	2393
2007	-6.68%	36	11386	2007	-1.6%	36	603
2006	-4.2%	36	5872	--	--	--	--

Source: Lendstats.com

TABLE

Table 1: Variables and Definitions

LOAN_AMNT	The listed amount of the loan applied for by the borrower.
TERM	The number of payments on the loan. Values are in months and can be either 36 or 60
INS_RATE	Interest rate charged on the loan.
INSTALLMENT	The monthly payment owed by the borrower if the loan originates.
GRADE	Lending Club assigned loan grade.
EMP_LENGTH	Employment length in years. Possible values are 0~10.
HOME_OWNER	Whether a borrower owns a house or not.
RENT	Whether a borrower rents a house or not.
MORTGAGE	Whether a borrower is now paying mortgage or not.
ANNUAL_INC	The annual income provided by the borrower during registration.
VERIFICATION	Whether the income source provided by the borrower is verified or not.
LOAN_STATUS	Current status of the loan.
DESC_CHAR	The number of characters in the borrower's loan description.
DTI	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested Lending Club loan, divided by the borrower's self-reported monthly income.
DELINQ_2YRS	The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years.
INQ_LAST_6MTHS	The number of inquiries by creditors during the past 6 months.
PUB_REC	The number of derogatory public records.
REVOL_UTIL	Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
TOTAL_ACC	The total number of credit lines currently in the borrower's credit file.

Source: Lending Club Data Dictionary and Author's Own Measurement.

Table 2: Summary Statistics

	Mean	Median	Maximum	Minimum	Std. Dev.	Observations
LOAN_STATUS	0.15	0	1	0	0.36	36117
ANNUAL_INC	68430.91	58163.50	6000000	4000.00	61322.12	
DELINQ_2YRS	0.15	0	11	0	0.49	
DESC_CHAR	288.01	144	3988	0	424.33	
DTI	13.21	13.28	29.99	0	6.68	
EMP_LENGTH	4.77	4	10	0	3.59	
GRADE	2.50	2	7	1	1.36	
HOME_OWNER	0.08	0	1	0	0.27	
INQ_LAST_6MTHS	0.87	1	8	0	1.07	
INSTALLMENT	322.41	277.16	1305.19	15.69	209.69	
INT_RATE	0.12	0.12	0.24	0.05	0.04	
LOAN_AMNT	10835.68	9500	35000	500.00	7245.88	
MORTGAGE	0.44	0	1	0	0.50	
PUB_REC	0.06	0	4	0	0.24	
RENT	0.48	0	1	0	0.50	
REVOL_UTIL	0.49	0.49	1	0	0.28	
TERM	40.74	36	60	36	9.55	
TOTAL_ACC	22.01	20	90	2	11.44	
VERIFICATION	0.56	1	1	0	0.50	

Table 3: Correlations

	TERM	REVOL_UTIL	INT_RATE
GRADE	0.42	0.44	0.95
	LOAN_AMNT		
INSTALLMENT	0.94		

Table 4: Basic Estimation of the Linear Probability Model

Dependent Variable: LOAN_STATUS

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.393	0.044	8.985	0.000**
DELINQ_2YRS	0.002	0.004	0.448	0.654
DESC_CHAR	-5.19E-06	4.32E-06	-1.20	0.230
DTI	8.24E-05	0.001	0.271	0.787
EMP_LENGTH	0.001	0.001	1.948	0.051
HOME_OWNER	-0.001	0.007	-0.080	0.936
INQ_LAST_6MTHS	0.018	0.002	10.415	0.000**
INSTALLMENT	2.74E-05	1.05E-05	2.619	0.009**
INT_RATE	1.260	0.070	17.879	0.000**
LOG_INC	-0.066	0.004	-16.130	0.000**
MORTGAGE	-0.001	0.004	-0.132	0.895
PUB_REC	0.048	0.007	6.265	0.000**
REVOL_UTIL	0.050	0.008	6.335	0.000**
TERM	0.007	0.000	31.294	0.000**
TOTAL_ACC	-4.86E-06	0.000	-0.025	0.980
VERIFICATION_STATUS	-0.0005	0.004	-0.120	0.905
R-squared	0.092	Adjusted R-squared		0.092
F-statistic	244.004	Durbin-Watson stat		1.995

** and * represent statistical significance at the 99% confidence level and the 95% confidence level, respectively.

Table 5: Revised Estimation of the Linear Probability Model

Dependent Variable: LOAN_STATUS

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.393	0.040	9.950	0.000**
EMP_LENGTH	0.001	0.001	2.515	0.012*
INQ_LAST_6MTHS	0.018	0.002	10.545	0.000**
INSTALLMENT	2.06E-05	1.01E-05	2.050	0.040*
INT_RATE	1.311	0.067	19.617	0.000**
LOG_INC	-0.066	0.004	-18.021	0.000**
PUB_REC	-0.0002	3.21E-05	-5.122	0.000**
REVOL_UTIL	0.050	0.007	6.747	0.000**
TERM	0.007	0.000	32.023	0.000**
R-squared	0.091	Adjusted R-squared		0.091
F-statistic	451.977	Durbin-Watson stat		1.994

** and * represent statistical significance at the 99% confidence level and the 95% confidence level, respectively.

Table 6: Estimation of the Logistic Regression Model

Dependent Variable: LOAN_STATUS

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.510	0.186	2.737	0.006**
EMP_LENGTH	0.006	0.002	2.398	0.017*
INQ_LAST_6MTHS	0.086	0.008	11.08	0.000**
INSTALLMENT	0.0001	4.72E-05	2.232	0.026*
INT_RATE	6.058	0.307	19.706	0.000**
LOG_INC	-0.327	0.018	-18.413	0.000**
PUB_REC	-0.001	0.001	-2.214	0.027*
REVOL_UTIL	0.263	0.035	7.619	0.000**
TERM	0.024	0.001	26.991	0.000**
McFadden R-squared	0.101			
LR statistic	3126.228			

** and * represent statistical significance at the 99% confidence level and the 95% confidence level, respectively.