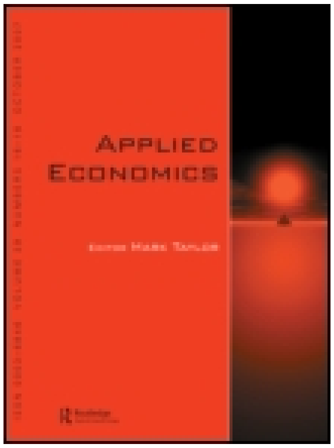

CrossMark
Click for updates

Applied Economics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/raec20

Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending

Riza Emekter[a], Yanbin Tu[b], Benjamas Jirasakuldech[c] & Min Lu[d]

[a] Department of Finance, School of Business, Robert Morris University, Moon Township, PA 15108, USA

[b] Department of Marketing, School of Business, Robert Morris University, Moon Township, PA 15108, USA

[c] Department of Finance, School of Business, Slippery Rock University of Pennsylvania, Slippery Rock, PA 16057, USA

[d] Department of Economics, School of Business, Robert Morris University, Moon Township, PA 15108, USA

Published online: 13 Oct 2014.

To cite this article: Riza Emekter, Yanbin Tu, Benjamas Jirasakuldech & Min Lu (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

To link to this article: http://dx.doi.org/10.1080/00036846.2014.962222

Applied Economics, 2015
Vol. 47, No. 1, 54–70, http://dx.doi.org/10.1080/00036846.2014.962222

Routledge
Taylor & Francis Group

Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending

Riza Emekter[a], Yanbin Tu[b], Benjamas Jirasakuldech[c,*] and Min Lu[d]

[a]*Department of Finance, School of Business, Robert Morris University, Moon Township, PA 15108, USA*
[b]*Department of Marketing, School of Business, Robert Morris University, Moon Township, PA 15108, USA*
[c]*Department of Finance, School of Business, Slippery Rock University of Pennsylvania, Slippery Rock, PA 16057, USA*
[d]*Department of Economics, School of Business, Robert Morris University, Moon Township, PA 15108, USA*

Online Peer-to-Peer (P2P) lending has emerged recently. This micro loan market could offer certain benefits to both borrowers and lenders. Using data from the Lending Club, which is one of the popular online P2P lending houses, this article explores the P2P loan characteristics, evaluates their credit risk and measures loan performances. We find that credit grade, debt-to-income ratio, FICO score and revolving line utilization play an important role in loan defaults. Loans with lower credit grade and longer duration are associated with high mortality rate. The result is consistent with the Cox Proportional Hazard test which suggests that the hazard rate or the likelihood of the loan default increases with the credit risk of the borrowers. Finally, we find that higher interest rates charged on the high-risk borrowers are not enough to compensate for higher probability of the loan default. The Lending Club must find ways to attract high FICO score and high-income borrowers in order to sustain their businesses.

Keywords: Peer-to-Peer lending; credit grade; FICO score; default risk

JEL Classification: D12; E41; E44; G20

I. Introduction

With the advent of Web 2.0, it has become easy to create online markets and virtual communities with convenient accessibility and strong collaboration. One of the emerging Web 2.0 applications is the online Peer-to-Peer (P2P) lending marketplaces, where both lenders and borrowers can virtually meet for loan transactions. Such marketplaces provide a platform service of introducing borrowers to

*Corresponding author. E-mail: b.jirasakuldech@sru.edu

lenders, which can offer some advantages for both borrowers and lenders. Borrowers can get micro loans directly from lenders, and might pay lower rates than commercial credit alternatives. On the other hand, lenders can earn higher rates of return compared to any other type of lending such as corporate bonds, bank deposits or certificate of deposits.

One of the problems in online P2P lending is information asymmetry between the borrower and the lender. That is, the lender does not know the borrower's credibility as well as the borrower does. Such information asymmetry might result in adverse selection (Akerlof, 1970) and moral hazard (Stiglitz and Weiss, 1981). Theoretically, some of these problems can be alleviated by regular monitoring, but this approach poses a challenge in the online environment because the borrowers and the buyers do not physically meet.[1] Fostering and enhancing the lender's trust in the borrower can also be implemented to mitigate adverse selection and moral hazard problems. In the traditional bank-lending markets, banks can use collateral, certified accounts, regular reporting, and even presence of the board of directors to enhance the trust in the borrower. However, such mechanisms are difficult to implement in the online environment which will incur a significant transaction cost.

To reduce lending risks associated with information asymmetry, current online P2P lending has the following arrangements. First, the Lending Club screens out any potential high-risk borrowers based on the FICO score. The minimum FICO score to be able to participate is 640.[2] Second, the typical size of the loans produced in this market is small, which is under $35 000 at the Lending Club. Therefore, these loans are essentially microloans which pose a relatively small loss in case of default. Third, the market maker offers matchmaking systems which can be used to generate portfolio recommendations and minimize lending risks. Fourth, if a borrower fails to pay, the market maker will report the case to a credit agency and hire a collection agency to collect the funds on behalf of the lender. Although there are certain structures imposed in the online P2P that help to minimize the risk, this form of lending is inherently associated with greater amount of risk compared to the traditional lending.

The purpose of this article is to evaluate the credit risk of borrowers from one of the largest P2P platforms in the United States provided by the Lending Club, which help lenders to make more informed decisions about the risk and return efficiency of loans based on the borrowers' grade. There are two related research questions this article will address: (1) What are some of the borrowers' characteristics that help determine the default risk? and (2) Is the higher return generated from the riskier borrower large enough to compensate for the incremental risk? Lenders can allocate their investments more efficiently if they know what characteristics of the borrower affect the default risk. Each borrower is classified by credit grade with corresponding borrowing rate assigned by the Lending Club. To make an efficient allocation, a lender should know whether the higher interest rates set for high-risk borrowers are sufficient to compensate the lenders for the higher probabilities of a potential loss.

This study contributes to the literature in this new and fast growing P2P marketplace in many ways. While there are few studies which explored credit screening problem in the P2P lending platforms, this research differs from the prior research in various aspects (see, for example, Iyer *et al.*, 2009 and Lin *et al.*, 2013). First, this research extends risk analysis research in the online P2P lending by utilizing the new data from the Lending Club, which is contrast to many prior studies which utilize the data from one of the biggest P2P platform (Prosper). Second, this study estimates the default risk of loan applicants based on their significant demographic and characteristic factors, which enables the potential lenders to determine an optimal allocation strategy. Third, this research addresses the issue of selection bias by examining whether there is a significant difference in the default risk of the borrowers from the whole US population and the Lending Club, which yields an important implication for risk minimization for the Lending Club. Finally, this research relates the default risk of borrowers with the returns generated by the lenders by comparing the calculated theoretical interest rate with the actual interest rate charged by the Lending Club for each credit grade category. This provides important information regarding the risk and return efficiency of the Lending Club.

[1] There are many studies investigating online trust in electronic markets such as online auction in information system literature (Gefen *et al.*, 2008).
[2] The market maker authenticates all market participants and provides a credit rating to each borrower.

Our findings suggest that borrowers with high FICO score, high credit grade, low revolving line utilization and low debt-to-income ratio are associated with low default risk. This finding is consistent with the studies by Duarte *et al.* (2012) who report that borrowers with a trustworthy characteristic will have better credit scores but low probability of default. This result also suggests that besides the loan applicants' social ties and friendship as reported by Freedman and Jin (2014) and Lin *et al.* (2013), the four factors discussed above are also important in explaining the default risk. When comparing with US national borrowers, the results show that the Lending Club should continue to screen out the borrowers with lower FICO score and attract the highest FICO score borrowers in order to significantly reduce the default risk. In relating the risk to the return, it shows that higher interest rate charged for the riskier borrower is not significant enough to justify the higher default probability. Our finding here is consistent with the study by Berkovich (2011) who reports that high quality loans offer excess return.

The remainder of the article is organized as follows. In the next section, we review the literature for online P2P lending. Section III describes our data and summarizes the descriptive statistics of online P2P from the Lending Club. In Section IV, we present the descriptions of methodologies and empirical results for evaluating the credit risk and measuring the risk and return efficiency for the Lending Club. The issue of selection bias is also addressed in this section. Section V offers some concluding remarks.

II. Literature Review

Three main streams of research have emerged in response to the growing popularity of P2P lending. The first stream of research examines the reasons for the emergence of online P2P lending. The second stream of research focuses on determining the factors that explain the funding success and default risk. The last stream of research investigates the performance of online P2P loan for a given level of the risk.

Peer group lending has been emerging in local communities and has attracted the research in this area. Conlin (1999) develops a model to explain the existence of peer group micro-lending programmes in the United States and Canada. He finds that peer groups enable fixed costs to be imposed on the entrepreneurs while minimizing the programme's overhead costs. Ashta and Assadi (2008) investigate whether Web 2.0 techniques are integrated to support the advanced social interactions and associations with lower costs for P2P lending. Hulme and Wright (2006) study a case of online P2P lending house, Zopa, in the United Kingdom. They suggest that the emergence of online P2P lending is a direct response to social trends and a demand for new forms of relationship in financial sector under the new information age.

There is extant literature that identifies the factors determining the funding success and default risk. Using the Canadian micro-credit data, Gomez and Santor (2003) find that group lending offers lower default rates than conventional individual lending does. Study by Iyer *et al.* (2009) shows that lenders can evaluate one third of credit risk using both hard and soft data about the borrower. Lin *et al.* (2013) analyse the role of social connections in evaluating credit risk and discover that strong social networking relationship is an important factor that determines the borrowing success and lower default risk. Lin *et al.* (2013) further report that applicants' friendship could increase the probability of successful funding, lower interest rates on funded loans, and these borrowers are associated with lower ex post default rates at Prosper. The importance of social ties in determining loans funded is also examined by Freedman and Jin (2014). The result shows that borrowers with social ties are more likely to have their loans funded and receive lower interest rates. However, they also find evidence of risks to lenders regarding borrower participation in social networks.

Several other studies examine whether certain borrowers' characteristics and personal information determine the success of loan funding and default risk. Herzenstein *et al.* (2008) show that borrowers' financial strength, their listing and publicizing efforts, and demographic attributes affect likelihood of funding success. Study by Duarte *et al.* (2012) further argues that borrowers who appear more trustworthy have better credit score with higher probabilities of having their loans funded and default less often. Larrimore *et al.* (2011) demonstrate that borrowers who use extended narratives, concrete descriptions and quantitative words have positive impact on funding success. However, humanizing personal details or loan justifications have negative influences on funding success. Qiu *et al.* (2012)

further reveal that in addition to personal information and social capital, other variables, including loan amount, acceptable maximum interest rate and loan period set by borrowers, significantly influence the funding success or failure.

Galak *et al.* (2011) further show that lenders tend to favour individual over group borrowers and borrowers who are socially proximate to themselves. They also find that lenders prefer the borrowers who are more like themselves in terms of gender, occupation and first name initial. More interestingly, Gonzalez and Loureiro (2014) have similar findings: (1) when perceived age represents competence, attractiveness has no effect on loan success; (2) when lenders and borrowers are of the same gender, attractiveness might lead to a loan failure (i.e., the 'beauty is beastly' effect) and (3) loan success is sensitive to the relative age and attractiveness of lenders and borrowers. Herzenstein *et al.* (2011) find that herding in the loan auction is positively related to its subsequent performance, that is whether borrowers pay the money back on time.

Research that relates the performance of the online P2P and its risks is very limited. Berkovich's (2011) research is the sole study that investigates this area and finds that high quality loans offer excess return. Therefore, our study that relates the default risk of borrowers with the returns generated by the lenders by comparing the calculated theoretical interest rate with the actual rate charged by the Lending Club will shed some important lights on the risk and return efficiency of the online P2P Lending Club.

III. Data

In this section, the loan applicants' data is first described, followed by loan distribution based on loan purposes, credit grade and loan status and it ends with the detailed descriptive statistics of the loan applicants. This study uses 61 451 loan applications in the Lending Club from May 2007 to June 2012 obtained from www.lendingclub.com. Over the study period, the Lending Club lent about $713 million to borrowers. To address the borrowers' behaviour in online P2P lending, we first examine the main reasons for borrowing money from others. Table 1 lists the borrowers' self-claimed reasons summarized in the Lending Club. Almost 70% of loan requested are related to debt consolidation or credit card debts with a total loan amount requested of approximately $387 million and $108 million, respectively. The number of loan applications for education, renewable energy and vacation contribute less than 1% of total loans with the total loan requested ranging from 1 to 3 million. The borrowers state that their preferences to borrow from the Lending Club are lower borrowing rate and inability to borrow enough money from credit cards. The second purpose for borrowing is to pay home mortgage or to re-model home.

Table 1. Loan distributions by loan purpose (May 2007–June 2012)

Loan purpose	Number of loans	Per cent	Amount	Per cent
Debt consolidation	29 903	48.66	$387 333 325	54.32
Credit card	9072	14.76	$107 904 275	15.13
Others	5643	9.18	$47 217 050	6.62
Home improvement	4365	7.10	$51 392 300	7.21
Major purchase	2890	4.70	$23 836 175	3.34
Small business	2695	4.39	$38 192 450	5.36
Car purchase	2034	3.31	$13 951 525	1.96
Wedding	1340	2.18	$13 573 375	1.90
Medical	992	1.61	$8 346 400	1.17
Moving	799	1.30	$5 527 175	0.78
House	630	1.03	$8 883 800	1.25
Vacation	531	0.86	$3 005 350	0.42
Educational	422	0.69	$2 752 775	0.39
Renewable energy	135	0.22	$1 171 575	0.16
Total	61 451	100.00	$713 087 550	100.00

Notes: The data is obtained from 61 451 loan applicants in the Lending Club, www.lendingclub.com, from May 2007 to June 2012.
The loan-seeking persons are asked to provide the reasons for requesting loans.

The Lending Club uses the borrower's FICO credit scores along with other information to assign a loan credit grade ranging from A1 to G5 in descending credit ranks to each loan. The detailed procedure is as follows: after assigning a base score based on FICO ratings, the Lending Club makes some adjustments depending on requested loan amount, number of recent credit inquiries, credit history length, total open credit account, currently open credit accounts and revolving line utilization to determine the final grade, which in turn determines the interest rate on the loan.[3]

Table 2 reports the loan distribution by credit grade. The majority of borrowing requests have

Table 2. Loans distribution by credit grades (May 2007–June 2012)

Credit grade	Grade	Number of loans	Per cent	Amount of loan provided	Per cent
7 (Lowest risk)	A1	1904	3.10	$14 996 125	2.10
	A2	2340	3.81	$18 704 000	2.62
	A3	2627	4.27	$22 432 650	3.15
	A4	4298	6.99	$43 185 675	6.06
	A5	3846	6.26	$38 652 950	5.42
6	B1	2863	4.66	$29 529 375	4.14
	B2	3202	5.21	$34 522 475	4.84
	B3	4833	7.86	$56 153 575	7.87
	B4	3818	6.21	$44 104 550	6.19
	B5	3991	6.49	$46 491 200	6.52
5	C1	3502	5.70	$40 533 100	5.68
	C2	3272	5.32	$38 785 100	5.44
	C3	2151	3.50	$23 674 600	3.32
	C4	1857	3.02	$19 827 900	2.78
	C5	1763	2.87	$19 430 600	2.72
4	D1	1533	2.49	$16 433 475	2.30
	D2	2029	3.30	$24 214 425	3.40
	D3	1824	2.97	$24 078 700	3.38
	D4	1570	2.55	$21 628 625	3.03
	D5	1361	2.21	$19 467 525	2.73
3	E1	1148	1.87	$17 305 775	2.43
	E2	1050	1.71	$16 252 225	2.28
	E3	878	1.43	$13 846 150	1.94
	E4	764	1.24	$12 870 650	1.80
	E5	673	1.10	$12 147 025	1.70
2	F1	527	0.86	$9 837 575	1.38
	F2	421	0.69	$7 777 750	1.09
	F3	321	0.52	$5 948 050	0.83
	F4	267	0.43	$4 941 875	0.69
	F5	210	0.34	$4 310 225	0.60
1 (Highest risk)	G1	187	0.30	$3 670 900	0.51
	G2	132	0.21	$2 537 325	0.36
	G3	87	0.14	$1 522 825	0.21
	G4	106	0.17	$1 886 525	0.26
	G5	96	0.16	$1 386 050	0.19
	Total	61 451	100.00	$713 087 550	100.00

Notes: The Lending Club uses the borrowers' FICO credit scores along with other information to classify a loan from Grade A1 to G5 in descending credit risk. Therefore, A1 credit grade represents the highest credit quality/low-risk borrowers, whereas G5 credit grade represents the lowest credit quality/ high-risk borrowers. Total amount of loans requested as a percentage of total loan is 19.35% for credit grade group 'A', 29.56% for 'B', 19.94% for 'C', 14.84% for 'D', 10.15% for 'E', 4.59% for 'F' and 1.53% for 'G'.

[3] For more information about grading please refer to: www.lendingclub.com

Table 3. Loan distribution by the loan status (May 2007–June 2012)

Panel A: All loans	Number of loans	Per cent	Amount	Per cent
Charged-off	2848	4.6346	$29 254 975	4.1026
Late (31–120 days)	620	1.0089	$7 478 450	1.0487
Performing payment plan	231	0.3759	$3 045 675	0.4271
Late (16–30 days)	172	0.2799	$1 927 650	0.2703
In grace period	488	0.7941	$6 125 325	0.8590
Current	46 039	74.9199	$557 351 825	78.1604
Fully paid	11 053	17.9867	$107 903 650	15.1319
Total	61 451	100.0000	$713 087 550	100.0000

Panel B: Matured loans	Number of loans	Per cent	Amount	Per cent	Unpaid principal
Charged-off	914	18.6378	$8 764 700	20.6742	$5 526 339.51
Late	29	0.5914	$329 325	0.7768	$27 140.64
Fully paid	3961	80.7708	$33 300 425	78.5490	$0.00
Total	4904	100.0000	$42 394 450	100.0000	$5 553 480.15

grades between A1 and E5. The Highest loan amounts requested are from borrowers with ‘B’ credit grade, which contribute 29.56% of total amount of loans requested. The total number of applicants for this ‘B’ credit grade group is 18 707, which represents total loans of approximately $210 million. The lowest loan amounts requested are from borrowers with the lowest ‘G’ credit grade which accounts for 1.53% of total loans. There are only 608 loan applicants for this lowest credit rating ‘G’ group and it represents approximately $11 million in total loan value. According to the Lending Club’s policy, a loan credit grade is used to determine the interest rate and the maximum amount of money that a borrower can request. The higher the loan grade, the lower the interest rate. A borrowing request with a low grade renders a higher interest rate as a compensation for a high risk held by lenders.

Finally, Panel A of Table 3 shows the loan status for all the loan requests on 20 July 2012. Overall, the default rate is 4.60% with total losses of approximately $29 million.[4] Another 2.45% of total loan requests which constitute $18.6 million could be potentially lost because the borrowers are late in making payment within 30 days or 120 days and not paying the normal instalments. 17.98% of the loans are fully paid with an approximate value of $108 million. The $557 million loans are in current status account for 74.91% of total loans. Naturally, loans with a lower grade demonstrate a higher default rate. Therefore, study on risk management on P2P lending is relevant for the lenders to optimize their investment portfolios. Panel B of Table 3 reports the loan status for the matured loans. The overall loss rate is much higher for matured loans. Among 4904 matured loans, 914 loans are charged-off, which represent 18.6%. The total loss is $5.5 million which represents 13% of all matured loans amount. Less than 1% of the matured loans are late in terms of making payment with the unpaid balance of approximately $27 000. 80.77% or $33 million of matured loans are fully paid.

Table 4 reports the general characteristics and credit history of the online P2P loan applicants from the Lending Club. Based on our sample of 61 451 loan applicants, the average monthly interest charged on a loan is 12.34%. On average, 471 days passed from the issue date of the loan. The average credit grade of a borrower is 25, which corresponds to credit category between B and C. The average size of a typical loan is $11 604 and the average monthly payment is $351. The borrower in general pays back $4384 a month and has $7873 left to be paid. The average ratio of the remaining balance to total loans is 63%.

Examining the borrowers’ characteristics, it shows that the mean income of a borrower from the Lending

[4] Remaining principals to be paid, potential interest is not included in this figure.

Table 4. Descriptive statistics (May 2007–June 2012)

Variables	*N*	Minimum	Maximum	Mean	SD	Skewness	SE	Kurtosis	SE
Interest rate	61 451	0.0542	0.25	0.1240	0.0393	0.3228	0.0099	−0.4025	0.0198
Days passed	61 451	2	1856.00	470.7700	395.5795	1.1407	0.0099	0.8571	0.0198
Credit grade	61 451	1	35.00	24.9175	6.9261	−0.8055	0.0099	0.1729	0.0198
Monthly payment	61 451	15.67	1379.23	350.7500	225.4390	1.1292	0.0099	1.3022	0.0198
Total number funded	61 451	500	35 000.00	11 604.2000	7575.7465	1.0407	0.0099	0.7312	0.0198
Debts to income ratio	61 451	0	0.35	0.1381	0.0677	0.0023	0.0099	−0.7063	0.0198
Payment to date	61 451	0	40 255.97	4384.2200	5056.7450	2.2491	0.0099	6.3427	0.0198
Remaining principal	61 451	0	35 000.00	7872.8600	7659.3242	1.2146	0.0099	1.1474	0.0198
To be paid to the total	61 451	0	1.00	0.6308	0.3598	−0.8226	0.0099	−0.8846	0.0198
Monthly income	61 451	0	500 000.00	5796.9500	5212.4929	24.8027	0.0099	1697.8614	0.0198
FICO score	61 451	0	6.00	3.4821	1.0860	0.4055	0.0099	−0.2455	0.0198
Open credit lines	61 422	1	49.00	9.5593	4.4500	0.9946	0.0099	1.8110	0.0198
Total credit lines	61 422	1	90.00	22.2256	11.3375	0.8292	0.0099	0.7032	0.0198
Revolving credit balance	61 422	0	1 207 359.00	14 315.6000	19 741.3993	10.8822	0.0099	372.6138	0.0198
Revolving line utilized	61 338	0	1.19	0.5156	0.2778	−0.1777	0.0099	−1.0479	0.0198
Inquiries in six months	61 422	0	33.00	0.9914	1.3923	3.4019	0.0099	32.2391	0.0198
Delinquent amount	61 422	0	6053.00	0.0990	24.4238	247.8272	0.0099	61 419.5558	0.0198
Delinquencies two years	61 422	0	13.00	0.1469	0.5107	5.6647	0.0099	53.4673	0.0198
Months since last delinquency	21 749	0	120.00	36.0016	22.1773	0.2819	0.0166	−0.8758	0.0332
Months last received	4226	0	129.00	63.9141	46.0820	−0.4696	0.0377	−1.5230	0.0753

Notes: Credit Grade is the grade assigned by the Lending Club based on the FICOrano credit rating information along with other information. Credit Grade ‘1’ is the loan category of ‘G’ which is the riskiest class of loans. Credit Grade ‘7’ is the loan category of ‘A’ which is the lowest risk borrowers. FICOrano is the credit rating of the borrowers rated by credit card companies. FICO 6 corresponds to borrowers with the FICO score above 780, FICO 5 corresponds to FICO score between 750–779, FICO 4 = 714–749, FICO 3 = 679–713, FICO 2 = 660–678 and FICO 1 = 640–659, respectively.

Club is \$5796[5] with the debts to income ratio of 0.1381. On average, a borrower has 9.56 open credit lines and 22 total credit lines, carries \$14 315 average revolving credit balance and almost half (51.6%) of his or her credit limit. In the last six months, there is 1 credit inquiry requested by an average borrower. Average FICO score category of a typical borrower is 3.48, which corresponds to a FICO score between 680 and 750.

IV. Empirical Results

Evaluation of credit risks

From a lender's perspective the most important concern is whether a borrower will default or not. A lender will benefit if borrowers' characteristics help to predict whether a particular borrower is more or less likely to default. Based on the description of the loan distribution by loan status reported earlier, in the past 5 years the Lending Club provided 61 451 loans. 2848 of these loans are not paid back (fully or partially). Although these charged-off loans translate to a default rate of 4.6%, this is biased downward since the default rate is generally increasing with the maturity. The default rate for the matured loans as indicated earlier is 18.60%.

In this section, we examine the factors that determine the likelihood of the loan default. We first implement the nonparametric tests to examine if there is a significant difference in the variables between defaulted loans and good status loans. Then, we model the default risk of the loan applicants by employing a binary logit regression.

Defaulted loans are loans that are charged-off and late in payment. Good status loans are loans that are fully paid or current in payment schedule. Table 5 reports the results of the nonparametric test and summarizes the differences between defaulted and good loans. These two groups are significantly different in terms of loan and borrower characteristics. The chi-square statistic values of Kruskal Wallis show that interest rate, credit grade, home ownership, FICO score, revolving line utilization, total funds and monthly income between the two groups are statistically different at 1% level. Specifically, we find that the interest rate on a defaulted loan is higher and the amount borrowed is lower. The borrowers of such defaulted loan tend to have low FICO score and low credit grade but higher revolving line utilization. In addition, they have lower monthly income and are less likely to own a home.

Table 5. Nonparametric test of differences between good loans and defaulted loans (May 2007–June 2012)

	Variables	Defaulted loans	Good loans
1	Interest rate	Higher	Lower
2	Credit grade	Lower	Higher
3	Home ownership	Lower	Higher
4	FICOrano	Lower	Higher
5	Revolving line utilized	Higher	Lower
6	Total funded	Lower	Higher
7	Monthly income	Lower	Higher

Notes: Credit Grade is the grade assigned by the Lending Club based on the FICOrano credit rating information along with other information. Credit Grade '1' is the loan category of 'G' which is the riskiest class of loans. Credit Grade '7' is the loan category of 'A' which is the lowest risk borrowers. FICO borrowers with the FICO score above 780, FICO 5 corresponds to FICO score between 750–779, FICO 4 = 714–749, FICO 3 = 679–713, FICO 2 = 660–678 and FICO 1 = 640–659, respectively. All the variables between these two groups are significantly different at the 1% level based on Chi-Square statistic value of Kruskal Wallis Test.

To further determine the precise effect of each variable on the odds of a loan default, we employed the binary logistic regressions that include all the variables under investigation. Let us assume that d_i is an unobserved continuous number representing the likelihood of a default. Therefore, a higher d_i value is an indicative of higher probability of default. In a binary logistic regression, a dependent variable is the probability of the event to occur, in our case it is a default. To convert this number into a number between zero and one, the following transformation is used:

$$p_i = \frac{1}{1 + e^{-d_i}} \tag{1}$$

where p_i is the probability that default will occur. It is further assumed that n independent variables in the binary logistic regressions are linearly related to d_i and the model can described as follows:

[5] There are some outliers in the data with the maximum monthly income of \$500 000. These outliers pull the average up to an unusual high number.

$$d_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_n x_{in} + \varepsilon_i \qquad (2)$$

where x_i is the independent variable i and n is the number of covariates.[6]

Table 6 reports the results of the binary logistic regression. The binary logistic regression model is first estimated with forward stepwise iterative maximum likelihood method. The analysis is repeated with backward stepwise iterative maximum likelihood method. The final results of both methods are similar. Out of the 13 variables included in the binary regression model, only four variables (credit grade, debt-to-income ratio, FICO score and revolving credit line utilization) significantly affect the outcome of the loan.[7] All the estimated coefficients are significant at the 1% level, with the exception of debt-to-income ratio and FICO score of 4 which are significant at the 5% level. Hosmer and Lemeshow's (2000) test shows that the model is adequate in explaining the status of loans with a chi-square value of 4.37. Multi-collinearity is not significant since all SEs of coefficient estimates are much smaller than 2. Final R^2 for the binary regression model is 6.50%.

Based on the binary logistic regression result, the probability of a default for a typical loan can be determined using the following model and estimated coefficients as reported in Panel A of Table 6.

Table 6. Binary logistic regression results (May 2007–June 2012)

Panel A: All loans	β	SE	Wald test statistics	Significance	Exp(β)
Credit grade			519.3233	0.0000	
Credit Grade 1 (highest risk)	2.5730***	0.1597	259.4451	0.0000	13.1058
Credit Grade 2	2.2127***	0.1249	313.9126	0.0000	9.1413
Credit Grade 3	1.8993***	0.1061	320.2410	0.0000	6.6812
Credit Grade 4	1.8099***	0.0971	347.4312	0.0000	6.1100
Credit Grade 5	1.5194***	0.0902	284.0318	0.0000	4.5697
Credit Grade 6	0.8831***	0.0844	109.3658	0.0000	2.4185
Debt to income ratio	−0.7883**	0.3183	6.1357	0.0132	0.4546
Revolving line utilized	−0.4964***	0.0860	33.3146	0.0000	0.6087
FICO score			148.5536	0.0000	
FICO 1	2.4107***	0.5868	16.8760	0.0000	11.1428
FICO 2	1.6426***	0.2075	62.6809	0.0000	5.1688
FICO 3	0.2711	0.1705	2.5291	0.1118	1.3115
FICO 4	0.3234**	0.1644	3.8705	0.0491	1.3818
FICO 5	0.5175***	0.1601	10.4504	0.0012	1.6779
FICO 6	0.5483***	0.1658	10.9408	0.0009	1.7305
Constant	−3.9884***	0.1543	667.7677	0.0000	0.0185
Hosmer and Lemeshow's Test: Chi-square = 4.37					

Panel B: Matured loans only	β	SE	Wald test statistics	Significance	Exp(β)	Probability of default	Actual loan default ratios
Credit grade			101.5271	0.0000			
Credit Grade 1 (highest risk)	1.7732***	0.2829	39.2977	0.0000	5.8899	0.3034	0.3615
Credit Grade 2	1.9000***	0.2466	59.3694	0.0000	6.6865	0.3308	0.3650
Credit Grade 3	1.6752***	0.1997	70.3652	0.0000	5.3401	0.2831	0.2614
Credit Grade 4	1.3528***	0.1911	50.1183	0.0000	3.8684	0.2224	0.2383
Credit Grade 5	1.1180***	0.1859	36.1722	0.0000	3.0590	0.1845	0.1949
Credit Grade 6	0.9507***	0.1901	25.0028	0.0000	2.5877	0.1606	0.1570
Constant	−2.6045***	0.1639	252.6639	0.0000	0.0739	0.0054	0.0659

Notes: The base value of model for credit grade is the highest credit grade (Grade 7). Therefore, the probability of the default of credit grade '7' is solely determined by the constant of the model −2.6045. This is also true for the other categorical variable FICO score on panel A. The highest grade is represented by the constant term. *** indicates significance at the 1% level, and ** indicates significance at the 5% level.

[6] For full details of the derivation of the model, refer to Hosmer and Lemeshow (2000).

[7] These variables are credit grade, debts to income ratio, monthly income, FICO score, open credit lines, total credit lines, revolving credit balance, revolving credit line utilization, inquires six months, delinquent amount, delinquencies two years, months since last delinquency and month since last received.

Table 7. Loan defaults classified by credit grades (May 2007–June 2012)

Category	Defaulted loans	Total loans	Ratio of defaulted loans to total loans
Credit grade 1 (highest risk)	131	608	0.2155
Credit Grade 2	249	1746	0.1426
Credit Grade 3	505	4513	0.1119
Credit Grade 4	808	8317	0.0972
Credit Grade 5	983	12 545	0.0784
Credit Grade 6	865	18 707	0.0462
Credit Grade 7 (lowest risk)	330	15 015	0.0220
Total	3871	61 451	0.0630

Notes: Credit Grade '1' is the loan category of G which is the riskiest class of loans. Credit Grade '7' is the loan category of 'A' with the lowest risk borrowers. Within each credit grade, the Lending Club uses the borrowers' FICO credit scores along with other information to assign a loan from Grade A1 to G5 in descending credit risk. Therefore, A1 credit grade represents the highest credit quality/low-risk borrowers, whereas G5 credit grade represents the lowest credit quality/ high-risk borrowers.

$$\begin{aligned} d_i = b_0 &+ b_1(\text{credit grade}) \\ &+ b_2(\text{debt to income retio}) \\ &+ b_3(\text{FICO score}) \\ &+ b_4(\text{Resolving line utilization}) + e_i \end{aligned} \tag{3}$$

To illustrate, for a risky borrower with a credit grade of 1, debt-to-income ratio of 10%, FICO score of 2 and revolving credit line utilization of 90%, the probability of default is 42.60%. ($d_i = -3.988 + 2.573 - (0.788 \times 0.1) + 1.643 - (0.496 \times 0.9) = -0.2972$, $p_i = \frac{1}{1+e^{-(-0.2972)}} = 0.4260$). For a relatively safe borrower with the highest credit category of 7 assuming that other variables remain the same, the probability of default would only be 5.36%.

The default probabilities for matured loans using binary logistic regression are also estimated and the results are reported in Panel B of Table 6. For these matured loans, only credit grade variables remain in the binary regression model. The estimated coefficients for each credit grade category are significant at the 1% level. All other variables are not significant in explaining the probability of loan default. The results based on the binary regression model show that as the risk of the borrowers increases, so does the probability of the default. The default probability for the matured loans predicted for borrowers with lowest Credit Grade 1 is about 30.34%, and they are 33.08%, 28.31%, 22.24%, 18.45% and 16.06% for Credit Grades 2, 3, 4, 5 and 6, respectively.

Out of the four aforementioned significant variables, credit score is the most important factor determining the outcome of the loan. The results from the binary regression model are consistent with the overall loan default ratio for each credit grade. Table 7 summarizes the ratio of default loans to the total loans provided to each type of borrowers. Out of 608 total loans provided to borrowers with highest risk, 131 loans were defaulted. This translates to a 21.55% defaulted ratio. It is reasonable to conjecture that the riskier borrowers are more likely to default on loan. Our finding supports this conjecture in which the ratio of default loan is declining as the credit grade is increasing. They are 14.26%, 11.19%, 9.72%, 8.74%, 4.62% and 2.20% for borrowers with credit grade of 2, 3, 4, 5 and 6, respectively.

Selection bias

To address the issue of selection bias, we further examine whether there is a significant difference in the default risk of the borrowers from the whole US population and the Lending Club. We compared the two samples. The first sample (Group 1: US National Consumers) consists of 32 410 consumer surveys in 2010 from the Survey of Consumer Finances. The second sample (Group 2: Lending Club Borrowers) consists of 61 451 borrowers from the Lending Club. We first test whether the average debt-to-income ratio and monthly income are significantly different between two groups. Based on the average debt-to-income ratio, the new default risk is calculated under the best case (highest FICO score, highest credit grade and zero revolving line utilization) and worst case scenarios (lowest FICO score and credit grade, and high revolving line utilization) and differences in default risk between two samples are examined. Finally, we examine whether there is a significant difference in the default risk between the two samples based on each particular FICO score.

According to the result reported earlier, debt-to-income ratio (monthly debt payments excluding the mortgage to the monthly total household income) and monthly income are the two variables that affect the default probability. Therefore, we compare the group statistics based on these two variables and the results show that the average debt-to-income ratio of national consumers is 11.66% while this ratio is 13.81% for the Lending Club borrowers. Average monthly income for national consumers is \$6528 while it is \$5797 for Lending club borrowers. These figures suggest that the Lending Club borrowers make less money but have higher debt-to-income ratio compared to the national average. When testing for the differences in the means of both debts to income ratio and monthly income between the two groups, they are statistically significant at the 1% level with the *F*-statistics of 53.21 and 1233.65, respectively.[8] This result is not so surprising given the fact that while the ultra-rich consumers would increase the national average income significantly, these individuals will not probably attempt to borrow from the Lending Club.

We further test whether the samples drawn from two different groups show different effect on the default probabilities. According to binary logistic regression results reported in Table 6, debt-to-income ratio significantly affects the default risk. To measure the impact of the higher debt-to-income ratio of Lending Club borrowers on the default risk, we calculated new default risk using average debt ratios of 0.1166 from the whole nation and 0.1381 from the Lending Club.

To calculate the minimum default risk for each group, we assumed the best case scenario of borrowers with the highest FICO score, highest credit grade, zero revolving line utilization ratio and average debt-to-income ratio. For the maximum default risk we assumed the worst case scenario for all variables (i.e., lowest FICO score, lowest credit grade, highest line utilization ratio and average debt-to-income ratio). The default probability as shown in Table 8 is not significantly different between the two groups. The difference in the default probability between the two groups is very marginal under the best (0.04%) and worst case scenarios (0.41%).

Table 8. Comparison of minimum and maximum default risk between US national consumers versus the Lending Club borrowers

Sample group	Minimum default risk (best case scenario) (%)	Maximum default risk (worst case scenario) (%)
Group 1 US national consumers	2.81	59.63
Group 2 Lending Club borrowers	2.84	60.04
Differences in default risk	0.03	0.41

Notes: Minimum default risk for each group is calculated based on Equation 3 under the best case scenario where borrowers are with the highest FICO score, highest credit grade, zero revolving line utilization and average debt-to-income ratio. Maximum default risk for each group is calculated based on Equation 3 under the worst case scenario for all variables (i.e., lowest FICO score, lowest credit grade, highest line utilization ratio and average debt-to-income ratio).

According to the nonparametric test results reported in Table 5, monthly income affects the default risk, but this variable is not significant in the binary regression as shown in Table 6. Therefore, it is reasonable to omit the impact of the income differences on the default risk between two groups. Additionally, debt-to-income ratio and monthly income figures are self-reported numbers by individual borrower. Hence, the reliability and accuracy of these data might be questionable. It would be ideal to compare these two variables which are compiled by the third party. The FICO score comparison would be a good candidate since it is reported by a third party. There are three main companies that compute FICO scores: Equifax, Experian and TransUnion. Unfortunately, our efforts have been fruitless so far to get a sample of borrowers with their respective FICO scores from one of these companies. We were able to obtain only aggregate data of the number of borrowers in a particular range of FICO score from both the Lending Club and Equifax. Therefore, we would be able to offer only some anecdotal evidence of the selection bias with FICO scores.

[8] To conserve the space, the results are not reported here but are available from the authors upon request.

Panel A of Table 9 shows the distribution of the US national borrowers and the Lending Club borrowers for a particular range of FICO score. To be able to qualify for a loan from the Lending Club, the borrowers need to have a credit score of at least 640. Comparing with the US national borrowers, it indicates that the Lending Club screens out roughly 34.33% of extreme-risk borrowers. However, when comparing the higher FICO score borrowers between the two groups, a different picture emerges. Only 16.9% of the Lending Club borrowers have FICO scores of 750 and above, while this figure is 37.73% for the US national average. Given that the Lending Club refuses 1/3 of the potential high-risk consumers, these numbers should be higher than the national average. This result suggests that high-income consumers with excellent credit scores do not borrow from the Lending Club. The majority of the Lending Club borrowers (82.33%) have FICO scores between 660 and 749, compared to 27.95% of the US national average.

Table 9. FICO score, distribution and default risks of borrowers (US national borrowers versus The Lending Club borrowers)

Panel A: The distribution of borrowers for each category of FICO scores

Group 1 US national borrowers		Group 2 The Lending Club borrowers	
FICO score	Per cent	FICO score	Per cent
300–499	6.80		
500–549	8.52		
550–599	9.32	640–659	0.77
600–649	9.70	660–678	17.19
650–699	12.03	679–713	36.88
700–749	15.92	714–749	28.26
750–799	19.45	750–779	11.93
800–850	18.28	780+	4.97

Panel B: The default risk of borrowers for each particular category of FICO score

FICO scores	Minimum default risk (Best case scenarios)	Maximum default risk (Worst case scenario)
640–659	0.6286	0.5963
660–678	0.3646	0.3234
679–713	0.1175	0.0814
714–749	0.0950	0.0787
750–779	0.0632	0.0720
780+	0.0280	0.0406

Notes: The distribution of the US national borrowers for each particular FICO score range is obtained from Equifax. Default risk is calculated based on the coefficient estimated from the binary regression reported in Table 6. Best case scenario is when the borrower has low debt-to-income ratio, high credit grade and zero revolving line utilization. Worst case scenario is when the borrower has high debt-to-income ratio, low credit grade and high revolving line utilization.

If we can separate general US population into three categories according to their FICO scores (Low (below 650), medium (650–750) and high (higher than 750)), most of the Lending Club borrowers would fall into medium FICO score range (about 80.00% of all the borrowers). The Lending Club eliminates 1/3 of the risky population and other 1/3 avoids borrowing from the Lending Club. Hence, the FICO score associated with each individual borrower should have some implications on the default risk.

There are four significant variables that determine the default risk: FICO score, the Lending Club credit grade, debt-to-income ratio and revolving line utilization. To examine the impact of FICO score on the default risk, we change the other three variables based on the best case and worst case scenarios and calculate the default risk using coefficients estimated from the binary regression reported in Table 6. The results reported in Panel B of Table 9 indicate that FICO score shows a significant impact on the default risk. For a borrower with high credit grade, low debt-to-income ratio, zero revolving line utilization and a FICO score between 750 and 779, default risk is 6.30% while this risk is only 2.80% for a borrower with the same characteristic but a higher FICO score of 780 or above. The difference in default risk on the lower FICO scores is even more striking. The default risk for the FICO score ranging from 660 to 678 is 36.46% while the default risk for the next lower FICO ranging from 640 to 659 is significantly increased to 62.86%.

Therefore, the Lending Club's decision to eliminate riskiest borrower seems logical since the borrowers at the lower FICO scores categories have significantly higher default risk. On the other hand, top one third of the consumers with highest FICO scores do not use the Lending Club, which increases the default risk substantially. If the Lending Club could replace some of its borrowers who have a

FICO score between 714 and 749 with borrowers who have FICO scores above 780, the potential default risk could be decreased from 9.50% to 2.80%.

Risk and return efficiency

Since default probability is an important variable for lenders along with the estimated loan recovery rate given the default occurred, a lender probably would like to know whether the high interest rate earned from a riskier loan is high enough to justify the higher default risk. To answer this question, we first determine the average default month. We then calculate the theoretical interest rate and compare with the actual rates for each loan categories based on the credit grade. To further empirically examine whether the loan with longer duration is more likely to default, we utilize the Cox Proportional Hazard model. Finally, based on all these findings, we compare the pay-offs between the risky loans and safe loans by comparing the ratio of the present value (PV) of loan payments for duration of the loan discounted at the base rate to the amount borrowed.

Based on the matured loan sample of 4904 loans, if a borrower defaults, this happens on average in the 15th month from the issue date of the loan. Given the probability of default as calculated from binary logistic regression of a loan with a certain credit grade, and the average default month, the interest rate risk premiums that should be charged for each loan based on credit ratings is then determined. The interest rates based on these calculations are compared with the actual rates that the Lending Club used. The theoretical interest rates are calculated using the following formula for each credit category:

$$(V_d \times p) + [V_{no} \times (1 - p)] = V_{id} + RP \qquad (4)$$

where V_d is the PV of the loan payments if it is defaulted at a certain time and calculated with the average default month of each loan category, p is the default probability from Equation 1, V_{no} is the PV of loan payments if there is no default and V_{id} is the PV of the loan with the best credit rating, which represents total amount of loans provided. RP is the risk premium that lenders may require to lend to riskier borrowers. The theoretical loan rate that satisfies Equation 4 for each loan category is then calculated. We assumed that loan payments will be reinvested at the safest loan; in other words, all the payments are discounted at the base rate. The minimum expected PV of the payments should be equal to the money lent, that is, where the risk premium is zero.

Table 10 provides the results of the calculated theoretical interest rate and compares with the actual interest rate charged by the Lending Club for each credit grade category. The actual interest rate charged for borrowers with the highest risk is only 10.81%[9] when compared with the calculated theoretical interest rate of 13.00%. The actual interest rate is significantly lower compared to theoretical interest rates for almost all loan grade categories except the highest credit grade category. In other words, the spread which is the difference between actual and theoretical interest rate is negative. For the lowest risk borrower, the actual premium charged on the loans is 1.23%,

Table 10. Loan defaults and risk premium (May 2007–June 2012)

Category	Average default month	Probability of default	Theoretical rate	Actual average rate
Credit Grade 1 (highest risk)	15.134	0.3034	0.1300	0.1081
Credit Grade 2	13.103	0.3308	0.1600	0.0909
Credit Grade 3	15.166	0.2831	0.1200	0.0775
Credit Grade 4	14.726	0.2224	0.0930	0.0644
Credit Grade 5	15.024	0.1845	0.0750	0.0496
Credit Grade 6	15.765	0.1606	0.0620	0.0347
Credit Grade 7 (lowest risk)	19.376	0.0054	0.0016	0.0123

Notes: Credit Grade 1 is the loan category of G which is the riskiest class of loans. Credit Grade 7 is the loan category of A with the lowest risk borrowers. Probability of default is calculated using the binary logistic regression results for matured loans as described in the article. The theoretical rate is calculated based on Equation 4. The actual average rate for each loan category is the actual interest rate charged on loan of a certain grade.

[9] This is the average spread between the interest rate on the highest risk loan (Credit Grade 1) and the interest on the lowest risk borrower (Credit Grade 7).

Table 11. Mortality rates for matured loans (May 2007–June 2012)

	Months after issuance						
Credit grade	0–6	6–12	12–18	18–24	24–30	30–36	>36
1 (Highest risk)	0.0000	0.7642	10.7846	15.2409	21.9088	28.8746	37.6721
2	0.2504	1.8709	6.2511	14.1812	15.5589	30.3942	42.9558
3	0.0000	2.1780	5.3378	9.2254	14.1093	19.9185	25.8008
4	0.0566	1.5545	5.2544	7.5283	12.2122	14.5866	25.1319
5	0.0000	1.1871	5.3518	6.2809	9.8983	13.0214	19.2568
6	0.0523	0.8290	1.8296	5.2652	6.3042	9.8127	15.1573
7 (Lowest risk)	0.0124	0.2898	1.1179	1.6513	1.8097	5.0934	7.7896

Notes: Credit Grade is the grade assigned by the Lending Club based on the FICO credit rating information along with other information. Credit Grade '1' is the loan category of 'G' which is the riskiest class of loans. Credit Grade '7' is the loan category of 'A' which is the lowest risk borrowers.

which is significantly higher than the theoretical premium of 0.16%. This suggests that the lenders earn a premium on the low-risk loans.

To further evaluate the credit risks of the borrowers, it is important to determine the relationship between the duration of the loan and the default probability of the loan. Before attempting to fit a formal model, a mortality matrix of just matured loan is derived and the results are presented in Table 11.

For all 4904 matured loans, the mortality rates increase as the duration of loans increases. This holds true for all types of credit category, with the exception of loans with credit grade of '2'. In addition, the mortality rate is higher for loans with lower credit grade when compared with loans of higher grade. For illustration purpose, the mortality or default rate for loans of Credit Grade '1' with the age between 6–12 months is 7.64%. The mortality rate increases to 10.78%, 15.24%, 21.90%, 28.78% for loans that are 12–18, 18–24, 24–36 months old, respectively. 37.67% of loans with the lowest credit grade of '1' and older than 36 months were default, compared with only 7.79% default rate of loans with highest credit grade of '7'. This finding is consistent with the earlier evidence found on the relationship among the covariates and the event of default based on the binary regression.

We further employ the Cox Proportional Hazard Model (Cox, 1972) to examine the relationship between the duration of the loans and the default rate. The Cox Proportional Hazard Test was first used in modelling of patient survivorship in many disease studies.[10] This technique has been widely implemented in many problems in financial markets including bank failure (Henebry, 1996), hedged fund survival rates (Gregorious, 2002) and default rate on high-yield bonds (Moeller and Molina, 2003).

The Cox Proportional Hazard regression technique allows us to measure the potential for the event to occur at a particular time given that the event did not happen yet. At a given time most loans are current. The model can be expressed as[11]

$$h(t) = [h_0(t)]e^{b_0+b_1x_1+b_2x_2+\cdots+b_nx_n} \tag{5}$$

where $h(t)$ is the hazard rate at time t; in our case, it is the likelihood that the loans will default. $h_0(t)$ is the baseline hazard at time t, n is the number of independent variables, b_j is the jth regression coefficients and x_j is the independent variable j.

In this model, the time-dependent variable was the number of months passed from the issuance date of the loan until the current date (end of July 2012). This date would be replaced with the default date if the loan is defaulted and with maturity date if the loan is fully paid off. The model is estimated with forward stepwise regression method. Out of 13 variables included in the stepwise regression, only credit grade and FICO score are significant in explaining the hazard rate on the loans. The results of Cox Proportional Hazard Model are presented in Table 12.

All the estimated coefficients are significant at the 1% level. The hazard rate or the likelihood of the loan being default increases with the credit risk of the

[10] In this particular case, this technique was used to determine whether certain variables affect the survivorship of the patients.
[11] This model is adapted from the user manual of SPSS.

Table 12. Cox proportional hazard regression results (May 2007–June 2012)

	β	SE	Wald test statistics	Degree of freedom	Significance	Exp(β)
Credit grade			463.5616	6	0.0000	
Credit Grade 1 (highest risk)	1.9648***	0.1201	267.7062	1	0.0000	7.1336
Credit Grade 2	1.8026***	0.0996	327.6165	1	0.0000	6.0658
Credit Grade 3	1.5052***	0.0867	301.4451	1	0.0000	4.5051
Credit Grade 4	1.3965***	0.0802	303.5087	1	0.0000	4.0411
Credit Grade 5	1.1958***	0.0751	253.4287	1	0.0000	3.3063
Credit Grade 6	0.7408***	0.0696	113.1414	1	0.0000	2.0977
FICO score			31.0000	6	0.0000	
FICO 1 (highest risk)	1.1192***	0.3782	8.7604	1	0.0031	3.0626
FICO 2	0.6958***	0.1539	20.4442	1	0.0000	2.0054
FICO 3	0.3517***	0.1294	7.3920	1	0.0066	1.4216
FICO 4	0.3506***	0.1243	7.9554	1	0.0048	1.4200
FICO 5	0.4162***	0.1217	11.7039	1	0.0006	1.5162
FICO 6	0.3230**	0.1294	6.2314	1	0.0126	1.3813

Notes: Credit Grade 1 is the loan category of G which is the riskiest class of loans. Credit Grade 7 is the loan category of A with the lowest risk borrowers. FICO 6 corresponds to the borrowers with the FICO score above 780, FICO 5 corresponds to FICO score between 750–779, 714–749, 679–713, 660–678 and 640–659 for categories 4, 3, 2 and 1, respectively. *** indicates significance at the 1% level and ** indicates significance at the 5% level. Exp(β) indicates the likelihood of a loan with a particular grade being default compared with the loans of highest credit grade.

borrowers. The hazard rate increases from 2.10 to 3.30, 4.04, 4.50, 6.06 and 7.13 for a unit decrease in credit grade. The Cox regression results show that for a borrower with the lowest credit category of ‘G’, for a given time they are 7.13 times more likely to default compared to borrowers with the highest credit category of ‘A’. On the other hand, borrowers with the lowest FICO score are 3.06 times more likely to default compared to the borrowers with the highest FICO score.

To further evaluate the risk and return pay-off to the lenders, we examine whether the high interest rate charged on the risky loan is sufficient to compensate for the high default risk. The Lending Club uses a minimum FICO score of 640 to screen borrowers. Using the interest rate charged for the lowest risk loan, we calculated the ideal PV of each loan assuming that all the payments will be received on time and reinvested at the same rate until the full maturity of the loan and compared with actual PV of each loan. The ratio of actual PV of loan to the ideal PV of the loan is then calculated. If the loan has the minimum risk and fully paid back, the ratio will be one. If the loan is risky and fully paid, the ratio will be greater than 1, and if the loan is defaulted the ratio will be less than one.

To test whether the high interest rate charged for riskier loans is large enough to justify for their higher risk, we used a simple linear regression

$$R = G \times A + \epsilon \tag{6}$$

where R is the $n \times 1$ vector of ratios calculated for each of 4904 matured loan, G is the $n \times 7$ matrix of dummy variables representing the credit ratings for each loan, A is the 7×1 vector of estimated coefficients and ε is the $n \times 1$ vector of error terms. The regression results are summarized in Table 13.

The estimated coefficient for the lowest Credit Grade ‘1’ is 0.9318. That is, for every dollar lent to a riskiest borrower, on average the PV of all the money paid back is only 93.18 cents. This figure is higher for the lowest risk borrowers, which shows a payment of 98.83 cents for one dollar of loan provided. The estimated coefficients are smaller than 1 for all different credit grades and exhibit an increasing pattern from 0.9317 to 0.9883 as credit grade improves from ‘1’ to ‘7’.[12] We further test the null hypothesis that the estimated coefficients are equal to 1 and the null hypothesis is rejected for all categories. All lower credit grade loans from ‘1’ to ‘6’ underperformed the loans with the highest credit grade. The expected PVs of all loans with lower grades are less

[12] The estimated coefficients are 0.9317, 0.8827, 0.9580, 0.9494, 0.9611, 0.9651 and 0.9883 for each credit grade from ‘1’ to ‘7’, respectively.

Table 13. Regression results (May 2007–June 2012)

Dependent variable: SCALEDDIFF
Method: Least squares
Sample 1: 4904
Included observations: 4904
White heteroscedasticity-consistent SEs and covariance

Variables	Coefficient	SE	t-Statistic	Prob.
Credit Grade 1	0.9317***	0.0311	29.9288	0.0000
Credit Grade 2	0.8827***	0.0271	32.5776	0.0000
Credit Grade 3	0.9580***	0.0135	71.1132	0.0000
Credit Grade 4	0.9494***	0.0102	92.6491	0.0000
Credit Grade 5	0.9611***	0.0075	127.9227	0.0000
Credit Grade 6	0.9651***	0.0069	139.1995	0.0000
Credit Grade 7	0.9883***	0.0045	218.0884	0.0000
R^2	0.0062	Mean-dependent variable		0.9603
Adjusted R^2	0.0050	SD-dependent variable		0.2618
SE of regression	0.2612	Akaike info criterion		0.1540
Sum squared residual	333.9776	Schwarz criterion		0.1633
Log likelihood	−370.6061	Hannan-Quinn criterion		0.1573
Durbin-Watson statistic	0.4130			

Notes: Credit Grade 1 is the loan category of G which is the riskiest class of loans. Credit Grade 7 is the loan category of A with the lowest risk borrowers. Scaled difference (dependent variable) is the ratio of present value (PV) of loan payments for the duration of the loan discounted with the base rate to the amount borrowed. *** indicates significance at the 1% level.

than the expected PV of the loans extended to the borrowers with the highest grade of ‘7’. This implies that the higher interest rates charged for the lower grade loans are not high enough to overcome the greater risk that the lenders take. The same conclusion was reached using the binary logit model above. The Lending Club should continue to screen borrowers but should definitely need to reconsider to increase the risk premium on the loans. However, increasing the interest rates on loans may potentially increase the default rates and may fuel adverse selection problems further. Therefore, it may not solve the problem. But lenders should feel comfortable to lend to the highest grade borrowers since they provide good compensation for the risk.

V. Conclusions

Credit risk is an important concern for the P2P loans. This study employs the data from the Lending Club to evaluate the credit risk of the P2P online loans. We find that credit score, debt-to-income ratio, FICO score and revolving line utilization play an important role in determining loan default. The credit categorization used by the Lending Club successfully predicts the default probability with one exception of next lowest credit grade ‘F’. In general, higher credit grade loan is associated with lower default risk.

The mortality risk also increases with the maturity of the loans. Loans with lower credit grade and longer duration are associated with high mortality rate. The Cox Proportional Hazard Test results show that as the credit risk of the borrowers increases, so does the likelihood of loan being default. However, the higher interest rate currently charged for the riskier borrower is not significant enough to justify the higher default probability. This suggests that the lenders would be better off to lend only to the safest borrowers in the highest grade category of 7 or Grade A. Increasing spreads on riskier borrower may lead to a more severe adverse selection resulting in higher default risk.

The Lending Club lenders should either extend credits only to the highest grade borrower or try to find more creative ways to lower the default rate among current borrowers. When comparing with the US national consumers, borrowers with relatively higher income and potentially higher FICO scores do not participate in the P2P market. Creating incentives to attract these types of borrowers would have a significant potential to decrease the default risk in this market.

References

Akerlof, G. (1970) The market for lemons: quality uncertainty and the market mechanism, *Quarterly Journal of Economics*, **84**, 488–500. doi:10.2307/1879431

Ashta, A. and Assadi, D. (2008) Do social cause and social technology meet? Impact of Web 2.0 technologies on peer-to-peer lending transactions, Working Paper, Available at SSRN http://ssrn.com/abstract=1281373 (accessed 30 September 2014).

Berkovich, E. (2011) Search and herding effects in peer-to-peer lending: evidence from prosper.com, *Annals of Finance*, **7**, 389–405. doi:10.1007/s10436-011-0178-6

Conlin, M. (1999) Peer group micro-lending programs in Canada and the United States, *Journal of Development Economics*, **60**, 249–69. doi:10.1016/S0304-3878(99)00043-7

Cox, D. (1972) Regression models and life tables with discussion, *Journal of Statistics Society Series*, **34**, 187–220.

Duarte, J., Siegel, S. and Young, L. (2012) Trust and credit: the role of appearance in peer-to-peer lending, *Review of Financial Studies*, **25**, 2455–84. doi:10.1093/rfs/hhs071

Freedman, S. and 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 2014).

Galak, J., Small, D. and Stephen, A. (2011) Microfinance decision making: a field study of prosocial lending, *Journal of Marketing Research*, **48**, S130–7. doi:10.1509/jmkr.48.SPL.S130

Gefen, D., Benbasat, I. and Pavlou, P. (2008) A research agenda for trust in online environments, *Journal of Management Information Systems*, **24**, 275–86. doi:10.2753/MIS0742-1222240411

Gomez, R. and Santor, E. (2003) Do peer group members outperform individual borrowers? A test of peer group lending using Canadian micro-credit data, Working Paper 2003-33, Bank of Canada, Ottawa, ON.

Gonzalez, L. and Loureiro, Y. (2014) When can a photo increase credit? The impact of lender and borrower profiles on online peer-to-peer loans, *Journal of Behavioral and Experimental Finance*, **2**, 44–58. doi:10.1016/j.jbef.2014.04.002

Gregorious, G. (2002) Hedge fund survival lifetimes, *Journal of Asset Management*, **3**, 237–52. doi:10.1057/palgrave.jam.2240078

Henebry, K. L. (1996) Do cash flow variables improve the predictive accuracy of a cox proportional hazards model for bank failure?, *The Quarterly Review of Economics and Finance*, **36**, 395–409. doi:10.1016/S1062-9769(96)90023-X

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).

Herzenstein, M., Dholakia, U. and Andrewsc, R. (2011) Strategic herding behavior in peer-to-peer loan auctions, *Journal of Interactive Marketing*, **25**, 27–36. doi:10.1016/j.intmar.2010.07.001

Hosmer, D. and Lemeshow, S. (2000) *Applied Logistic Regression*, 2nd edn, John Wiley, New York.

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

Iyer, R., Khwaja, A., Luttmer, E. *et al.* (2009) Screening in new credit markets: can individual lenders infer borrower creditworthiness in peer-to-peer lending?, NBER Working Paper No. 15242, NBER, Cambridge, MA.

Larrimore, L., Jiang, L., Larrimore, J.*et al.* (2011) Peer to peer lending: the relationship between language features, trustworthiness, and persuasion success, *Journal of Applied Communication Research*, **39**, 19–37. doi:10.1080/00909882.2010.536844

Lin, M., Prabhala, N. and Viswanathan, S. (2013) Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending, *Management Science*, **59**, 17–35. doi:10.1287/mnsc.1120.1560

Moeller, T. and Molina, C. (2003) Survival and default of original issue of high-yield bonds, *Financial Management*, **32**, 83–107. doi:10.2307/3666205

Qiu, J., Lin, Z. and Luo, B. (2012) Effects of borrower-defined conditions in the online peer-to-peer lending market. E-life: web-enabled convergence of commerce, work, and social life, *Lecture Notes in Business Information Processing*, **108**, 167–79. doi:10.1007/978-3-642-29873-8_16

Stiglitz, J. and Weiss, A. (1981) Credit rationing in markets with imperfect information, *American Economic Review*, **71**, 393–419.