Risk and Return of Investments in Online Peer-to-Peer Lending
(Extended Abstract)

Harpreet Singh[a], Ram Gopal[b], Xinxin Li[b]
[a] School of Management, University of Texas at Dallas, Richardson, Texas 75083-0688
[b]School of Business, University of Connecticut, Storrs, CT 06269
harpreet@utdallas.edu, {ram.gopal, xinxin.li}@business.uconn.edu

Abstract
Online peer to peer (P2P) lending has received great coverage in media but little attention from academic researchers. In this study, we focus on risk and return of investments on Prosper (P2P lending website). We find that on average, loans through Prosper provide negative return compared to risk free alternatives such as Treasury Bills. We then use decision tree analysis to segment loans in term of different return and risk profiles. We further determine the efficient frontier for investments on Prosper and calculate the efficiency of loans in various segments. We find that (1) within each credit grade, there exist subgroups which give positive return and for these subgroups risk is aligned with return, and (2) the groups of loans with lower credit grades are more efficient in terms of risk and return alignment than those with higher credit grades. Our study provides investment guidelines for lenders and design implications for online peer to peer lending websites.

1. Introduction
Online peer-to-peer (P2P) lending is a new e-commerce phenomenon where individual lenders provide unsecured loans directly to individual borrowers without the traditional intermediaries such as banks. Since 2006, many internet-based lending services have emerged; online consumer lending has attracted millions of investors and has become an increasingly popular alternative channel for consumers to get personal loans. This demonstrates how web 2.0 can facilitate the creation of new markets competing with traditional providers such as credit card issuers leading to disintermediation of the financial services industry. According to a Wall Street Journal article (Kim 2007), the amount of new P2P loans issued in these lending websites was around $100 million in 2007 and will increase to as much as $1 billion by 2010.

As P2P lending bypasses the intermediaries and thus the associated cost, it can provide investors an opportunity to earn return that may be higher than that available from traditional investments. It also provides loans to borrowers who are not able to get loans from traditional financial markets or who get loans at higher rates in traditional markets. However, as the individual lenders are short of experience and monitoring capabilities compared to banks, these benefits are accompanied with high uncertainty and risk. Also, since most borrowers in these P2P lending sites are one-time borrowers, there is no reputation mechanism that can be built up over time to convey borrowers' credibility like the one used for sellers on eBay. It is therefore essential for lenders to draw similarities on loans across borrowers based on available information to judge on expected risk and return of a new loan request.

In this research we focus on risk and return associated with loans on P2P lending sites. In particular, how can we segment loans in term of different return and risk profiles based on available loan and borrower characteristics? How is the return aligned with the associated risk? What are the factors that influence the efficiency of different loan segments in terms of return and associated risk? Answers to these questions will contribute to the understanding of this new and rapidly growing lending market. Our findings can direct lenders towards ways to optimize their investment strategy and help these online lending service providers improve consumer welfare.

As a relatively new phenomenon, online P2P lending has received great coverage in media, but little attention from academic researchers. Two recent studies (Freedman and Jin 2008; Lin, Prabhala and Viswanathan 2008) studied the impact of social connections on the likelihood of loan requests being successfully funded, the resulted interest rates and loan default rates. Herzenstein et al. (2008) further compared the effects of demographic attributes on the likelihood of funding success for online lending to those documented for the practices of traditional financial institutions. While these three studies examined this market from the borrower's point of view, we set out our analysis from the lender's point of view and focus on risk and return alignment in P2P lending.

2. Background and Data

In this paper we did our analysis using data collected from Prosper.com (an online peer to peer lending website) market place. Prosper was opened to the public on February 13, 2006 and as of August 1, 2008, Prosper had registered 750,000 members and originated 26,273 loans that totaled over 164 million US dollars. All prosper loans are unsecured and fully amortized with a fixed interest rate and a three-year term. Borrowers create listings on Prosper and specify the amount of loan requested (from $1000 to $25,000) and the maximum interest rates they are willing to pay. Lenders have access to information related to borrowers' credit history provided by Prosper and additional voluntary information (pictures, purpose of the loan etc.) provided by borrowers. The auction process is similar to proxy bidding on eBay. A lender can bid on any listing by specifying the lowest interest rate he is willing to accept and the amount he wants to contribute (any amount above $50). A listing gets fully funded if the total amount from bids exceeds the amount requested by the borrower. The minimum interest rate specified by the first lender excluded from funding the loan becomes the contractual rate for the loan.

Our study utilizes the publicly available data downloaded from Prosper.com on August 1, 2008. This data set includes listings that began on or after June 1, 2006 and ended on or before July 2008. We focus on all the loans (generated from successfully funded listings) that were listed from February 2007 to July 2007, which leaves out loans from the first few months of the market which are likely to be unstable and ensures that we have at least two years of payment data for each loan for analysis. The final dataset has 4,504 loans.

3. Methodology

3.1 Return on Investment (ROI) Calculations

To compare the performance of different loans, we first calculate return on investment (ROI) based on monthly discounted payments over a two-year period for each loan in our dataset. Since we do not have data on actual payments and can only observe principal balances, we use the following information from Prosper and some assumptions for the payment calculations:

1. As such, interest accrues on each loan's principal balance on a daily basis.
2. For simplicity, we assume that all payments for loans with status being "current" are paid at the scheduled day.
3. If there is no change in principal balance in a particular period (the loan status will change from "current" to "late"), we assume no payment has been made in that period.
4. On Prosper any time a borrower makes a loan payment, it is applied to the loan balance in this order: (1) late fee (2) accrued interest (3) loan principal balance. If change in principal balance is positive and the status changes from "n-month late" to "current", we assume late fee and

interest accrued are paid in full. Late fee charged by prosper is calculated as the maximum of \$15 or 5% of the outstanding unpaid payment and is passed on to lenders.

5. The service fee is charged to the lenders based on the current outstanding principal balance and its rate varies from 0.5% to 1% depending on the lenders' credit grades. This policy has been changing over time. For simplicity we use 1% as the service fee rate for all the loans. Service fee is charged only in those periods when lenders receive payments.

The return on investment is given as:

$$ROI = \frac{CDP - L}{L}$$

where L is the loan amount and CDP is the cumulative discounted payment. The CDP is calculated as follows:

$$CDP = \frac{P_1}{\left(1+\frac{d}{12}\right)^1} + \frac{P_2}{\left(1+\frac{d}{12}\right)^2} + \cdots \frac{P_{24}}{\left(1+\frac{d}{12}\right)^n} + F * PB_{24}$$

$$P_n = \Delta PB_n + PB_{n-1} * \frac{l_n}{365} * r + L_n - S_n$$

where P_n is the net payment lenders receive at the end of n^{th} month, PB_n is the remaining principal balance at the end of n^{th} month and d is the discount factor. If a loan does not get defaulted or paid earlier, then n is 24; otherwise, n is the month in which the loan gets defaulted or paid in full. r is the interest rate for the loan, l_n is the length of time period in days (e.g. l is 31 for January), L_n is the late payment and S_n is the service fee charged in period n.

3.2 Decision Trees to Group Loans Based on ROI Distribution

Our calculations in previous section suggest that Prosper loans vary significantly in their return. While this may have attracted attention from both academic researchers and practitioners, what has been missing is whether the significantly varied return is aligned with the associated risk. To study this, we need to group loans with similar characteristics together (to create the distribution of ROI) and calculate both the return (the mean ROI) and risk (the variance of ROI) within each group to see if they are aligned when compared across groups. In this study, we first rely on the decision tree methodology to divide loans into groups such that loans within the same group have similar characteristics and distribution of ROI. When applying the decision tree methodology to divide loans into different groups, we use variables[1] related to borrower and loan characteristics which lenders can observe before making their investment decisions as explanatory variables and use ROI as the dependent variable. We then calculate risk and return for each group and examine their positions in the composition of optimal investment portfolios.

For decision tree analysis, we adopt the conditional inference tree approach. The focus of this study is to find out parameters that can be used to group loans which are similar to each other in terms of both mean and variance. Conventional classification and regression trees (CART) can only detect mean shift in the data and hence will give us the groups which are similar to each other in terms of mean value of dependent variable only. The second issue with the CART is that they are biased towards the variables that are continuous or with a higher number of categories. The modeling approach of conditional inference trees uses a unified framework of embedding recursive binary partitioning into well-defined theory of permutation tests developed by Strasser and Weber (1999). Conditional inference trees select the node split based on how good the association of dependent variable is with the independent variable. Unified tests for independence are constructed

[1] For brevity the list of variables used and their definitions are not provided. This list is available upon request from the authors.

by means of the conditional distribution of linear statistics in the permutation test framework developed by Strasser and Weber (1999). Therefore it not only removes the bias due to categories but also chooses variable that are informative (Hothron et al. 2006). The other advantage of this algorithm is that it can handle multivariate response. A suitable multivariate transformation of the response can be used to detect the shift in mean or variance of the dependent variable. More detailed explanation of the algorithm is provided in Hothron et al. (2006).

3.3 Efficient Frontier

The groups identified using decision trees have different distributions of ROI and represent different investment opportunities. Each group will give different return but will also have different risk associated with it. The mean of ROI within each group will give the expected return from the investments in that group and the variance of ROI will give the associated risk. For an efficient market the risk and return should be aligned i.e. the loans with high risk should give high return. Lenders can combine loans from different groups as portfolios to achieve optimal return based on the amount of risk they are willing to take. To assess the efficiency of each group, we first calculate the efficient set which gives maximum return for a given risk. The expected return and variance for an investment portfolio containing loans from different groups are given as:

$$E\left(ROI_p\right) = \ w_1E(ROI_1) + \ w_2E(ROI_2) + \cdots\cdots w_nE(ROI_n)$$
$$V\left(ROI_p\right) = \ (w_1)^2\, V(ROI_1) + \ (w_2)^2\, V(ROI_2) + \cdots\cdots (w_n)^2\, V(ROI_n)$$

where $E\left(ROI_p\right)$ and $V\left(ROI_p\right)$ are the return (mean value of the ROI) and risk (variance of ROI) obtained by investing $\mathrm{w_n}$ fraction of total investment in the nth group. $E(ROI_n)$ and $V(ROI_n)$ are the mean and variance of ROI for the nth group. Since the groups are disjoint (in terms of decision rules), we assume that the covariance between ROIs of different groups is zero. To get the maximum return for a given level of risk, we solve the following optimization problem:

$$max_{w_1,w_1,\cdots w_n}\, E\left(ROI_p\right)$$

subject to

$$V\left(ROI_p\right) \ \le K \qquad (1)$$
$$\textstyle\sum_{i=1}^{n} w_i \le 1 \qquad (2)$$

The first constraint ensures that the risk of the investment portfolio is less than the given level of risk (*K*) and the second constraint is a budget constraint (it ensures that the sum of the proportion of total money invested in different groups is not over 1). After solving this optimization problem we are able to calculate the maximum return for a given risk. Hence for each group we calculate the efficiency measure as follows:

$$Eff_n = \frac{E(ROI_n)}{E^*\left(ROI_p\right)},$$

where $\mathrm{E(ROI_n)}$ is the return (mean value of the ROI) for the nth group and $\mathrm{E^*\left(ROI_p\right)}$ is the maximum return for the level of risk (variance of ROI) associated with the nth group.

4. Results

We find that on average, loans through Prosper provide negative return compared to risk free alternatives such as Treasury Bills. The result is different from those in previous studies (Freedman and Jin 2008) and Prosper website which found a positive estimated rate of return on Prosper. In our study we use only those loans which have at least 24 months of loan history to calculate return whereas other studies have used loans of all ages including loans which have only a few months of history. Since on Prosper a loan becomes defaulted only if it is more than 4 months late, including the loans with only a few months of history to calculate the overall return may give biased results.

Credit grade has been the most used variable to analyze loan performance on Prosper both by lenders and by Prosper. However, our results suggest that it is borrower maximum rate (bmr) instead of credit grade (cg) that is the primary variable to identify groups with similar risk and return profile. In addition, the average return for loans is negative within each credit grade, even for credit grade AA (credit score greater than 760). Figure 1 gives the detail of the final decision tree. Table 1 provides the borrower and loan characteristics for the subgroups of loans yielding positive expected return. Accordingly, lenders on Prosper should make their investment decisions not only based on credit grades but also taking other variables into account.

We further find that for groups which give positive return, risk is aligned with return, i.e. as return increases, risk also increases. We then derive the efficient frontier for investments on Prosper and determine the optimal portfolio of investments for a given level of risk. The maximum possible return for a given level of risk is summarized in Table 1. Table 2 summarizes the optimal % investment in each subgroup to maximize return at various levels of risk. The key finding of this analysis is that the subgroups of loans with higher risk are more efficient in terms of risk and return alignment, i.e., they provide more return for a given risk as compared to loans with lower risk. Our results also suggest that lenders, who are not willing to take high risk, should diversify their portfolio (i.e. invest in subgroups of loans of all credit grades) to maximize return, whereas borrowers who are willing to take more risk can maximize their return by investing in only the subgroups of loans with lower risk.

5. Conclusions

We summarize below the key results and conclusions from the current analysis:

- On average, loans offered through prosper provide negative returns compared to risk free alternatives such as Treasures Bills.
- Borrower maximum rate instead of credit grade is the most important variable for identifying risk and return for loans offered through Prosper.
- The expected average return for loans with very higher borrower maximum rate is usually negative. For loans with borrower maximum rate less than 0.19, there exist subgroups which provide positive return and for these subgroups risk is aligned with return. These groups can be identified using borrower maximum rate as the primary variable and credit grade, inquiries in last six months and bank card utilization as secondary variables.
- The subgroups of loans with higher risk are more efficient in terms of return per unit risk.
- Risk adverse lenders need to diversify more to maximize return compared to risk taking lenders.

References

Freedman, S. and Jin, G. Z. Do Social Networks Solve Information Problems for Peer-to-Peer Lending? Evidence from Prosper.com, Working Paper, University of Maryland.

Herzenstein, M., Andres, R. L., Dholakia, U. M. and Lyandres, E. The Democratization of Personal Consumer Loans? Determinants of Success in Online Peer-To-Peer Lending Communities," Working Paper, University of Delaware.

Hof, R. "Prosper: The eBay of Loans?" Business Week, February 13, 2006.

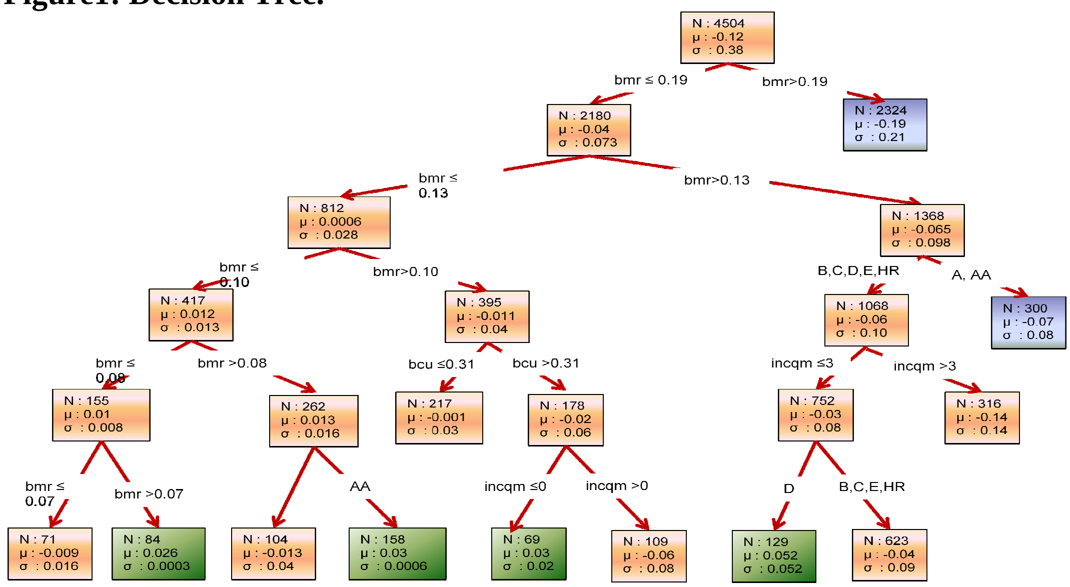
Hothron, T., Hornik, K., and Adam, Z. 2006. Unbiased recursive partitioning: a conditional inference framework. *Journal of computational and Graphical Statistics*, **15** 651-654.

Lin, M., Prabhala, N. R. and Viswanathan, S. "Online Peer to Peer Lending," Working Paper, University of Maryland.

Kim, J. J. "Options Grow for Investors to Lend Online," The Wall Street Journal, July 18, 2007.

Strasser, H., and Weber, C. 1999. On the asymptotic theory of permutation statistics. *Mathematical methods of statistics* **8** 220-250.

Figure1: Decision Tree.



Note: N is the total number of loans, μ is the mean of ROI and σ is the variance.

Table 1: Subgroups with Positive ROI

Group	Splitting Criterion			Mean of ROI (Return)	Variance of ROI (RISK)	Number of Loans	Max ROI for given RISK	Efficiency
	Criterion 1	Criterion 2	Criterion 3					
Group 1	0.7<bmr<=0.8			0.026	0.0003	84	0.029	0.884
Group 2	0.7<bmr<=0.8	CG=AA		0.031	0.00067	158	0.032	0.984
Group 3	0.7<bmr<=0.8	bcu>0.31	Inq6m<=0	0.03	0.019	69	0.046	0.653
Group 4	0.7<bmr<=0.8	Inq6m<=3	CG=D	0.0524	0.052	129	0.052	1.0

Note: bcu is bank card utilization and inq6m is the number of inquiries in last 6 months.

Table 2. Optimal Investment for Given Risk

Return	Risk	Group 1	Group 2	Group 3	Group 4
0.026655	0.000201	67.83%	27.91%	1.31%	0.69%
0.02914	0.000301	41.65%	53.43%	1.76%	3.16%
0.029858	0.000401	30.44%	63.10%	2.10%	4.36%
0.031964	0.000901	0.00%	88.75%	2.92%	8.32%
0.03786	0.006501	0.00%	63.16%	2.18%	34.67%
0.039833	0.010001	0.00%	54.95%	1.56%	43.49%
0.04408	0.0203	0.00%	36.51%	1.03%	62.46%
0.047122	0.03	0.00%	22.95%	1.00%	76.06%
0.049765	0.04	0.00%	11.17%	0.97%	87.87%
0.05211	0.050099	0.00%	0.72%	0.94%	98.34%
0.052459	0.0517	0.00%	0.00%	0.10%	99.90%
0.052481	0.051798	0.00%	0.00%	0.00%	100.00%