A Comparison of Credit Loan Service Between Banks and P2P Platform

Written by
Jeffers Yin, Vicky Xu, Jin Sun, Shunli Lu

Introduction

Technology involvement in the banking industry is prevailing. Besides the digitization process of traditional banks, new players are taking more responsibility in terms of credit assessment and funding brokerage, with peer-to-peer lending marketplaces being the most aggressive one.

This study focuses on the differences between traditional banks and peer-to-peer lending marketplaces in terms of their core competency, which is credit risk assessment, or more specifically, loan pricing on an individual level.

We build up a comparison metrics from both the borrowers' perspective and lenders' (or investors') perspective to evaluate the differences in pricing strategies, which reveal the credit risk assessment strategy, between two different kinds of institutions. We collected a dataset from Prosper, the first peer-to-peer lending company in the US, which contains more than 110,000 individual-level loan data.

For the borrowers' perspective, we are targeting at finding possible gaps in terms of loan interest rate to reveal possible risk ignorance, certain discrimination or strategy diversity. Certain borrowers' rates priced by peer-to-peer lending marketplaces could be directly found in our dataset since the dataset is collected from Prosper. However, due to confidential issues, banks do not disclose individual-level information on their cash loans. Thus, we must use an alternative approach to get bank rates before moving to the comparing stage. Based on Hongfeng Peng's paper[1], we developed a setting where banks use the risk-neutral pricing method to determine the loan interest rates. Based on everyone's default probability predicted by logistic regression, we can give each individual borrower a risk-neutral-priced interest rate which represents the rate he or she could get from a traditional bank. In this way, we could compare rates from peer-to-peer lending marketplaces and traditional banks on the same basis.

For the investor's perspective, we develop a utility-maximizing metrics based on risk-return preference survey. Ultimately our metrics would be able to assign each investor the very portfolio, consisting of various financial products, which could maximize their investment utility. This portfolio establishing process would be powered by machine learning models, or specifically, classifiers, which are trained using survey data that we currently do not have access to. So currently, our metrics would be a prototype that have no actual functions since no survey could be conducted now. But once the survey is conducted and data is collected, our prototype would be fully prepared for utilizing investors' utility by selecting ideal financial products for them.

General Research for Bank and P2P Lending Platform

1. Industry Overview

"As of Q3 2018, the Federal Reserve Bank estimated the value of consumer loans, at all commercial banks, to be $1.49 trillion in the United States"[2] This is an incredibly large

[1] 彭红枫,徐瑞峰. P2P 网络借贷平台的利率定价合理吗?——基于"人人贷"的经验证据[J]. 金融论坛,2018,23(09):61-80.

[2] https://www.supermoney.com/studies/personal-loans-industry-study/

amount of money in the current lending market for personal loans. As it continues growing (**Exhibits 1**), the demand for loans catches investors' eyes. P2P Platforms and Banks are two major personal loan providers in the market. As traditional institutional--banks, it provided loans as interest rate range from 5% to 36%[3]. Interest rates vary based on individual information.

While traditional banks have rather a long history from the origin, peer-to-peer lending industry's history in the US started from the launch of Prosper Marketplace, followed by Lending Club. Adverse selection and high default rate were the early stage features of peer-to-peer lending marketplaces since the screening process was poorly developed. What is more, the growth requirement of lending marketplaces shareholders drives these companies to take bold, or even risky, expansion strategies.

2. Business Model

Banks are the financial institutions that collect deposits from customers and providing loans to individuals and enterprise. In addition to that bank provides multiple investment products such as stocks, bonds, ETF, etc. There are three ways banks make profits: By charging interest for money they lend, by charging transaction fees for merchandise they service and by trading financial instruments in the markets. For this research project, we mainly focus on the interest they charge for lending. You can see interest rates provided by different banks in (**Exhibit 2**) and interest rates based on credit score in (**Exhibit 3**)

When a customer applies for a loan, banks need to select qualify customers and give them appropriate interest rates based on their pricing strategy. Banks need to evaluate each customer carefully to avoid loss from defaulting. There are five most fundamental characteristics banks need to look at:

1. Credit history
   Credit history is a record indicate how a person managed his credit in the past, including total debt loan, payment history and number of credit lines. Fico score is a good indicator of a person's credit score.

2. Cash flow history and projection for the business
   Cash flow helps a business owner to know how to spend money, what to spend money on and how much money comes in. Cash flow projection is a way to analyze the business in order to predict any weakness or to evaluate performance for business.

3. Collateral available to secure the loan
   Collateral is a property or other asset that a borrower offers that as a way for a lender to secure their loan. It could be house or automobile. If one has bad credit and bank won't provide loan to him, he can use collateral to apply for a loan.

4. Character
   Character indicate as person's purchasing and payment behavior. How much do you spend per month? Does one pay in full in certain amount of time? What is probability of default?

3 https://www.bankrate.com/loans/personal-loans/rates/#apply

5. Loan documentation that includes business and personal financial statements, income tax returns, a business plan that essential sums up the provides evidence for the first four items listed.

For peer-to-peer lending marketplaces, they business model is quite simple since they do no more than just brokerage. By collecting service fee from both borrowers and lenders and possible rate spread in between, marketplace companies could be able to generate steady cash flow regardless of the underlying risk issued loans may contain.[4] Due to their sole business model, the only thing that affects their revenue would be the loan volume. To maintain a growing loan trading volume, marketplaces need to attract both lenders and borrowers by setting up proper interest rates for each loan. The rate should be not only too low to compensate the involved risk or to appear as attractive for investors, but also too high to discourage borrowers from applying. Thus, rate pricing strategy is the key competency in terms of competing with peers.[5]

3. Regulation

Regulation F is set by the Federal Reserve for banks must institute internal rules that control the amount of risk that they can take in their business proceedings with other institutions. It also limits the amount of credit exposure between banks to 25% of capital, in most cases.[6] The legal lending limit for national banks was set under the United States Code and is overseen by the FDIC and the OCC. Details are reported in U.S.C. Title 12, Part32.3. The code on lending limits states that a financial institution cannot issue a loan to a single borrower for more than 15% of the institution's capital and surplus.[7] These two regulations allow bank to have enough liquidity and prevent bank from default. Bank is not allowed to disclose individual level data, we will further discuss the bank pricing strategy based on the bank level data in the paper.

In 2008, the U.S. Securities and Exchange Commission (SEC) required that peer-to-peer companies register their offerings as securities, pursuant to the Securities Act of 1933.[8] The registration process was an arduous one; Prosper and Lending Club had to temporarily suspend offering new loans, while others, such as the U.K.-based Zopa Ltd., exited the U.S. market entirely. Both Lending Club and Prosper gained approval from the SEC to offer investors notes backed by payments received on the loans. Prosper amended its filing to allow banks to sell previously funded loans on the Prosper platform. Both Lending Club and Prosper formed partnerships with FOLIOfn to create a secondary market for their notes, providing liquidity to investors. Lending Club had a voluntary registration at this time, whereas Prosper had mandatory registration for all members.[9]

Borrowers' view

1. Comparison Design

4 https://www.investopedia.com/terms/p/peertopeer-p2p-economy.asp
5 https://en.wikipedia.org/wiki/Peer-to-peer_lending
6 Regulation F:https://www.investopedia.com/terms/r/regulation-f.asp
7 Legal Lending Limit:https://www.investopedia.com/terms/l/legal-lending-limit.asp
8 https://en.wikipedia.org/wiki/Peer-to-peer_lending
9 https://www.researchgate.net/publication/327836640_Peer-to-Peer_Lending_Business_Model_Analysis_and_the_Platform_Dilemma_in_International_Journal_of_Finance_Economics_and_Trade_IJFET_submitted_August_1st_2018_Accepted_Sept_24th

a. **Overview**

Our comparison metrics focus on the interest rate gap between peer-to-peer lending marketplaces and traditional banks. We hold that the interest rate gap on the same applicant could be a strong indicator of institutes' pricing strategy and risk assessment strategy, therefore a place where we could possibly discover risk ignorance, certain discrimination or strategy diversity.

b. **Data Description**

We use Prosper as our objective data set for this project. The data set contains 113,937 loans with 78 variables, including loan amount, borrower rate, interest rate, occupations, current loan status, borrower income, borrower employment status, borrower credit history, and the latest payment information. Each loan-data slot has been desensitized with a unique data-key to protect borrowers' privacy. Besides the borrower rate and interest rate, all other variables, such as isBorrowerHomeOwener, GroupKey, etc., are left side parameters used to predict the borrower rate, interest rate, and default rate. In addition, the timeframe of the data set is from 2009 to 2016, as the accession of the current dataset is denied without the investor permit.

Although we have limited accession to the bank data, we will use prosper data set to derive the default rate. By doing so, we can also use machine learning tacit to derive the risk neutral interest rate, one rate that is most important benchmark for bank.

2. Model Development

a. Design

We successfully collect more than 100,000 individual-level loan data from Prosper. Each observation contains all kinds of attributes describing the loan borrower and the loan itself including the interest rate. However, due to confidential issues, we are not able to acquire such individual-level data from banks. Thus, we take an alternative approach in order to reach the setting where we can do comparison.

To begin with, we make two key assumptions. Firstly, individual default risk does not vary no matter the loan is issued by banks or peer-to-peer marketplaces. The only matter that affects default risk is the attributes of the borrower. Secondly, bank issue loans based on a risk-neutral basis, where risk-neutral default probability is the only thing affect rates.

Then we firstly predict the default probability P of each borrower using the dataset acquired from Prosper via machine learning models like logistic regression. Based on our first assumption, this predicted probability would be the default probability if this applicant got the loan from banks. Next, using this probability and risk-neutral pricing functions defined in Hongfeng Peng's paper, we come up with the estimated risk-neutral rate that each specific borrower should take if he or she turned to banks for financing.

By comparing the actual rate from peer-to-peer marketplaces and the estimated risk-neutral rate from banks, we can achieve our goals of comparison in this empirical setting.

b. Model Structure

1) Default Rate: P

1. Reasons of using Logistic regression

The logic behind logistic regression is consistent with the result we want to get P. Default rate is a possibility that is within the boundary between 0 and 1. At the same time, logistic regression also gives the result with that boundary that stands for the possibility.

2. Sifting irrelevant factors

Our dataset contains more than 50 categorical variables. With that large number of categorical variables, setting dummies will even exacerbate the situation in which logistic regression cannot be directly used to estimate P. Also, among these categorical variables, only a few of them are heavily impact on the P. In order to make logistic model run through the data and make the analysis more effectively, we use recursive feature elimination to choose variables that affect P most as our data set columns to estimate P. 36 variables are selected by the recursive feature elimination, including occupations, locations, income verifiable, loan categories, etc.

3. Prediction analysis

The logit regression has a 68 percent accuracy to predict whether borrowers have a chance to be default, indicating that our prediction model does have some implications on the default rate, and we will use the result to predict the interest rate for the further research. **(Exhibit 5)**

We also make a logit regression analysis to have a clear look of each factor impact on the default rate. Among these variables, verifiable income has the highest coefficient, 3.4551, indicating that people whose incomes are verified have a much higher rate to not default than those whose income are not verified. It is rational result because verified incomes indicate the borrowers have proofs of their incomes, and those, who do not have ones, have higher possibility of deception in the first place. In addition, the default rate also fluctuated by different time periods. In the 2013 Q4 and 2014 Q1, the default rate is much higher than those in other time period, possibly implying that people who borrow money on the platform have systematic risks happened that drag their financial abilities. Also, different states and occupations play an important role to determine the default rate; people who live in the more developed states or have a higher-income job are less likely to default. The analysis, though will not be used for generation of interest rate, gives us a glance about how different factors will impact on the default rate. In general speaking, people who, from the logit regression analysis, have a high-income job, live in the more developed states, and have evidence to support their income, are less likely to be default. **(Exhibit 6)**

2) Risk-Neutral Model

Then we build our risk neutral model to estimate the risk-neutral interest rate, which only depends on customers' default probability. Here we have the following variables:

(1) V_{d1}: Using risk-neutral interest rate, the present value of all payments to investor before default

(2) V_d: Using risk-neutral interest rate, the present value of all payments to investor if default

(3) V_{no}: Using risk-neutral interest rate, the present value of all payments to investor without default

(4) V_{id}: The present value of payment to investor without default using the interest rate of 7.93%, which is the average rate of AA level credit loan

(5) P: default probability of the loan

(6) RP: risk premium, which should be 0 in our risk neutral model

(7) MR_{no} the monthly payment using risk-neutral interest rate

(8) MR_{id}: the monthly payment using AA level interest rate 11.78%

(9) r: risk-free rate, which is represented by 3-year treasury rate 1.6%

(10) term: the number of repayment terms (month)

(11) Bid: Total debt amount

(12) Rate: risk-neutral rate

(13) $Rate_{id}$: average rate of AA level credit loan, which is 7.93%.

(14) DT: average portion of regular repayment terms on total number of terms, which is 0.7 in our dataset.

The key logic used in our analysis is risk neutral theory. In our dataset, AA level debt has the lowest probability of default, so we use AA level debt as baseline to analyze the risk premium of credit loans.

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

[10]

In formula (1), on the left hand is the expectation of all payments using risk-neutral rate. The debt has probability p for default and the total present value of repayment V_d if default. At the same time, it has probability (1-p) for non-default and the present value of all payment V_{no} without default. Combined the two situations together is the expectation of all repayment of the loan. On the right hand is the present value of regular repayment for AA level loans, which has average annual rate of 7.93%

For formula (1), the variables that we need to figure out is V_d, V_{no} and V_{id}. In order to get those present value, we need to firstly calculate the monthly repayment of loans.

10 彭红枫,徐瑞峰. P2P 网络借贷平台的利率定价合理吗?——基于“人人贷”的经验证据[J]. 金融论坛,2018,23(09):61-80.

$$MR_{no}=\frac{INT\left[\frac{Bid\times\frac{Rate}{12}\times\left(1+\frac{Rate}{12}\right)^{term}}{\left(1+\frac{Rate}{12}\right)^{term}-1}\times 100\right]}{100}$$ [11] (2)

$$MR_{id}=\frac{INT\left[\frac{Bid\times\frac{Rate_{id}}{12}\times\left(1+\frac{Rate_{id}}{12}\right)^{term}}{\left(1+\frac{Rate_{id}}{12}\right)^{term}-1}\times 100\right]}{100} \quad (3)$$

In formula (2), we get the monthly repayment amount for loans using their risk neutral interest rate. In formula (3), we get the average monthly repayment amount for AA level credit loans using their average interest rate 7.93%

$$V_{no}=MR_{no}\times\frac{1-\frac{1}{(1+r)^{term}}}{r} \quad (4)$$

$$V_{id}=MR_{id}\times\frac{1-\frac{1}{(1+r)^{term}}}{r} \quad (5)$$

In formula (4), we calculated the present value of all regular repayments of loans when default does not happen. In formula (5), we calculated the present value of all repayment for AA level credit loan, which will not be default.

$$V_{d1}=MR_{no}\frac{1-\frac{1}{(1+r)^{INT\left[DT\times term-1\right]}}}{r} \quad (6)$$

In Formula (6), we calculated the present value of all repayments before default. $(DT\times term)$ represents the number of terms that the borrower regularly repaid the loan. We assume that there is no insurance protecting the investors from the loss when default happens so that $V_d$should also equel V_{d1}.

$$V_d=V_{d1} \quad (7)$$

[11] Formulas(2)~(7) are all taken from article “彭红枫,徐瑞峰. P2P 网络借贷平台的利率定价合理吗?——基于“人人贷”的经验证据[J]. 金融论坛,2018,23(09):61-80.”

When we put the 7 formulas together, the only variable that we do not know the Rate, which is the risk-neutral interest rate that we are interested in. Here we use the package “rootSolve” in R to solve the non-linear formula and we get the right interest rate that borrowers should face under risk neutral conditions. Even though right now there are hot debates on whether credit loan investors prefer what kind of risk level, analyzing the correctness of valuation on credit loans can provide some business insights for public when they are choosing from Peer-to-Peer platform or traditional banks. Later we use compare the risk-neutral rate and real interest rate for different groups to see if they are preferred or discriminated from the pricing model. If the risk-neutral rate is higher than the real interest rate, then those loans are undervalued. If the neutral rate is lower, the loans are overvalued.

3. Comparison results

a. Results

We have tried two directions to do the comparison. Firstly, we would like to see if there is price undervalued or overvalued among different occupations. Then we compare different geographical groups to see if the state location will affect the pricing of credit loan, which will also indicate the effect of population composition on credit loan interest rate.

(1) Comparison among applicants’ occupation

Computer programmers and Architects are shown to have lower real rate than their risk-neutral rate, indicating that they may be preferred during the price matching. However, Laborer and Fireman tends to have higher real rate than their “right” risk neutral rate, indicating that those jobs might be discriminated. Other occupations have a little bit higher real rate but the differences are not so significant.

Occupation	Risk-neutral Rate	Real Average Rate	Difference
IT	19.7%	17.0%	-2.7%
Executive	16.5%	17.4%	0.9%
Professional	18.3%	18.9%	0.6%
Food Service	20.7%	21.5%	0.8%
Laborer	18.4%	21.1%	2.7%
Fireman	17.6%	19.3%	1.7%
Analyst	18.3%	17.9%	-0.4%
Architect	19.7%	16.6%	-3.1%
Teacher	19.0%	19.5%	0.5%
Waiter / Waitress	21.3%	21.7%	0.4%

(2) Comparison among applicants’ location

In general, Texas and California show a relatively high real rate while the other states does not show great preference or discrimination.

State	Risk-neutral Rate	Real Average Rate	Difference
IL	18.8%	18.7%	-0.1%

NY	18.6%	19.1%	0.5%
OH	19.6%	20.0%	0.4%
TX	17.4%	18.9%	1.5%
CO	18.1%	18.7%	0.6%
MO	20.5%	20.1%	-0.4%
LA	20.2%	20.1%	-0.1%
CA	17.5%	18.7%	1.2%
MD	19.4%	19.8%	0.4%

b. Insights/Recommendations

As for occupation, the mismatching of loan interest rate may be caused by public's different confidence on different groups. For example, computer programmer and architect are regarded as promising and stable groups in terms of their solid professional skills while laborer fireman might be seen as unstable jobs, which means their repayment might not be guaranteed. What surprises us is that executive and professional are also high-income group but they are not preferred in credit loan selection. This might be caused by the unstable financial status of the world financial and business markets, leading to the loss of confidence on those group of borrowers.

From the ten states we got, we can clearly see that Texas and California bear higher interest rate than what they should do. We come up with two possible reasons that can explain this difference. Firstly, this difference might be caused by public's bias on their dominant industries in the two states. There are tons of start-ups in Texas and California, and people might regard those start-ups as highly risky companies, leading to higher expectation on interest return to make up for higher risk. Another explanation could be the automatic market adjustment on economic development. To be specific, the booth of start-ups makes the economy of Texas and California grow so rapid that exceeds their normal speed. Invisible market adjustment might occur to reach the balance and that kind of adjustment can be seen as relatively higher interest rate on credit loan.

All our explanations are based on business sense, without empirical confirmation. The next step of study on this topic could be to design some empirical tests to verify whether those explanations are reliable or not.

Those in discrimination status in terms of credit loan pricing should realize this fact and those people can pay more attention to other sources of loan application, like traditional banks, instead of only focusing on peer to peer platform.

Investor's view

1. Design Logic

When it comes to investors' point of view, we want to utilize machine learning models to provide an optimal investment strategy. Here is the main design logic: there are two main factors that need to consider in investment activities: risk tolerance and expected rate of return. Usually the higher risk tolerance means the higher expected rate of return they will gain. However, loans with higher risk are more likely to default. Based on this investment strategy, we would like to design our utility-maximizing metrics followed by the following steps:

a. Design a survey to collect data from professional investment managers with different levels of risk tolerance and their investment choice

b. Use data to train classification models through machine learning models such as gradient boost random forest, logistic classification etc.

c. Implement model to make the best investment decisions across various financial products

2. Metrics (survey)

Most risk tolerance assessment questions involved in a survey provided by financial institution are as following[12] , this metrics is ideal model and will be modified completed in the future:

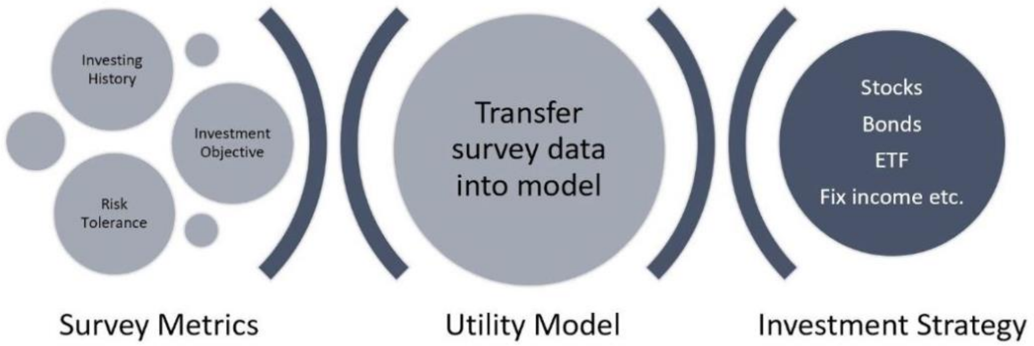
Investing Experience	1.Select the investment you currently have: 2.How much have you spend on previous investment? 3. How many years of experience do you have for investment?
Investment Objective	1.What is your investment attitude? 2.In how many years will you begin making withdrawals from your investments? 3. Once you begin to make withdrawals, how many years will you be making withdrawals? 4. What is your expected rate of return?
Capital	1.What is your income level? 2.What percentage of your income will you make your investment?
Risk Tolerance	1.Protecting my portfolio is more important to me than high returns? 2. Which of the following statements best describes your investment philosophy? 3. If you want to invest in stocks, which one would choose? 4.What do you expect to be your next major expenditure? 5.How long are you looking to invest your money for?

3. Ideal Model Example

In this part, we will give a small demo about how to utilize our metrics to evaluate which financial product will be the best fit for each investor. For everyone who wants to invest on

[12] Survey Questions: https://www.questionpro.com/blog/risk-tolerance-questionnaire/

financial product, we will give them a survey to access their risk tolerance. For example, Bank of America assigns each customer with different levels of risk tolerance for different portfolios of products. (**Exhibit 4)**. The way we are doing is based on the model we trained through data provided by the professional investment manager, we will get an optimal investment strategy to maximize customers' utility, providing higher expected return and minimizing default rate based on customers' preference. As a result, an optimal investment strategy will be a combination of deposit, stocks, bonds, ETF, fixed income etc.



In addition (demand/supply influence on Interest rate)

Beside our major study, we also shed some light on the question: What is determining loan interest rate? We dive into this problem from both the supplier side and demander side. We investigate what attributes of loan supplier and what attributes of loan demander is affecting the real interest rate of a specific loan via machine learning. For the supplier's perspective, we use bank-level performance data gained from Federal Deposit Insurance Corporation (FDIC), while for the demander's perspective, we used the same dataset collected from Prosper.

a) FDIC

i. Design

The dataset acquired from FDIC website contains financial ratios and statistical performance results of each bank and their overall interest rate. We are targeting to pinpoint those most important performance attributes that affect the overall loan product interest rate the most.

ii. Model Development

Random forest is an effective tool in terms of variable selection. By feeding the cleaned dataset to a random forest model, the model could be able to filter out all those irrelevant attributes when determining interest rate level, leaving all 'powerful' and 'directive' attributes.

iii. Bank-level Insights (FDIC)

The left side shows ranked attributes in a ranking manner of reducing the total Mean Squared Error, while the right side shows ranked attributes in a ranking manner of reducing the total model complexity. The ranking results are different from each other. However, by comparing the first several attributes on both sides, we do find out that the two ranking methods still have some degree of "overlap". Five attributes are pinpointed in the overlap area as significant attributes that affects interest rates of banks.

'elnatry': The annualized loan and lease losses as a percent of total assets.
'nonixay': Non-interest expense as a percentage of average total assets.
'lnlsntv': Net loan and lease as a percent of total assets.
'idlncorr': Net loan and lease as a percent of core deposits.
'roaptx': Annualized pre-tax net income as a percent of total assets.

Thus, we can conclude that there are five important performance ratios that has\ve an impact on the average effective interest rate of commercial banks. The most significant one is annualized loss on loans and leases to total asset ratio. Non-interest expense ratio and pre-tax net income ratio also play important roles in setting up interest rates. The amount of loans and leases outstanding, as intuitively expected, also functions in determining the interest rate. At this stage we are only confident to state as above rather than claiming underlying mechanisms of those observed phenomenon. Further causal inference research may be conducted in order to shed more light on the causation relationship and underlying mechanisms. (**Exhibit 7**)

b) P2P Lending

iv. Description

Dataset collected from Prosper contains attributes of the loan borrower and the loan itself. We are targeting at pinpointing those important demographic or non-demographic attributes that determine interest rate. Or in other words, we are looking for key risk indicators from a loan borrower's profile.

v. Model Development

We first plan to use three different types of machine learning methods, including linear methods, tree methods and neural networking, to examine the interest rate of prosper. Linear methods we used are basic linear regression, ridge, and lasso. By squeezing the parameters of useless variables to nearly zero, Ridge and Lasso are very effective methods to testify the data set which, at most of the time, contains irrelevant variables and even multi-collinearity. Without using Ridge and Lasso, we first need to use other variable selection method to get rid of irrelevant variables or we will get multicollinearity problem.

We use random forest as our tree method to estimate the interest rate. Random forest is also a very helpful method to get results by determining interest rate level, leaving all 'powerful' and 'directive' attributes. In addition, we also use the neutral networking to test the result.

After we tested these three different methods, neutral networking is least accurate to test the risk factors of interest rate with nearly zero percent of prediction accuracy. In addition, although lasso and ridge give a general pattern of interest rate composition, the MSE of both methods are fallen into the range of 0.3-0.5, which indicates that both methods are error based. Therefore, we use random forest to show the result of analysis of risk factors of interest rates.

vi. Results (**Exhibit 8**)

As we testify the interest rate that customers got from prosper, we find that credit score is the most important factor to determine the interest rate from prosper. The lower credit scores a borrower has, the higher interest rate investor will receive. It is understandable that there is a negative correlation between borrower's credit score and interest rate because investors who invested the project targeted to the borrower have to bear higher risk with default. At the same time, borrowers who have low credit score also have higher interest rates because they are very likely to default.

In addition, the loan origination quarter is another factor that heavily impact on the interest rate. The possible reason of the time period of issuing loan is strongly correlated with interest rate is that Fed set different interest rate according to the national economic situation at different time period. The Fed rate, the baseline of every rate in the USA, strongly impacts on the interest rate from prosper. As the Fed rate goes high, the interest rate from prosper also goes high.

Besides credit score and the loan origination quarter, other factors such as Available Bankcard Credit, Loan Original Amount, Term, and Debt to income ratio also plays a large role to determine the interest rate for both borrowers and investors.

Conclusion

To compare the credit loan service in traditional banks and peer-to-peer marketplace, we build two-dimension metrics, with borrows' and investors' point of view. Besides, we also have some additional findings in terms of pricing model for credit loan from the perspective of capital supply and demand.

Borrower point of view:

Using risk neutral model, we calculated the risk neutral interest rate that people should be faced with, given their default probability. Then we compare the actual interest rate and their theoretical interest rate, to see if there is any price discrimination for certain groups. From our result, we found that laborer and fireman had relatively higher interest rate, which might be caused by the lack of public confidence in their work stability. In addition, we also found that some states like Texas and California had relatively higher interest rate, which might come from the instability of start-ups or the automatic market adjustment. We suggested that those "discriminated" groups should be aware of those facts and might consider other sources of loans instead of focusing on peer-to-peer platform.

Investor point of view:

Each investor wants to gain high expected rate of return and to avoid large likelihood of default. Our designed utility-maximizing model will generate optimal investment strategy for them. Investors need to go through survey metrics, and we collect results. After putting results into our utility-maximizing model, investor will get an investment strategy that is best fit for them.

In the additional part of our research, we take an empirical view of variables that have an impact on interest rate from both the fund supplier side, which are banks, and the fund demander side, which are borrowers. We found various significant influencing variables for both the supplier and the demander. By using FDIC bank-level data, we find out that annualized loss on loans and leases to total assets ratio is the most significant variable when determining the interest rate level of a bank. Relative amount of loans and leases and relative size of pre-tax net income are also functioning importantly. By using Prosper's individual-level data, we find out that average credit score and loan issuing quarter are the two most significant variables of borrowers when determining their rates. So far, we have only described the relationships from a predicting perspective. Further studies of causal inference would be critical in terms of finding the underlying mechanisms.

Citations

[1] *彭红枫,徐瑞峰. P2P* 网络借贷平台的利率定价合理吗?——基于"人人贷"的经验证据[J]. 金融论坛,2018,23(09):61-80.

[2] Latham, Andrew. "*2019 Personal Loans Industry Study*", 23 Oct 2019, www.supermoney.com/studies/personal-loans-industry-study/

[3] "*Best Personal Loan Tates for November 2019*", Bankrate, 29 Oct 2019, https://www.bankrate.com/loans/personal-loans/rates/#apply

[4] Chappelow Jim, "*Peer-to-Peer Economy*", 1 Apr 2018, www.investopedia.com/terms/p/peertopeer-p2p-economy.asp

[5] Kenton Will, "*Regulation F*", 29 Jun 2018, https://www.investopedia.com/terms/r/regulation-f.asp

[6] Kagan Julia, "*Legal Lending Limit*", 28 Feb 2018, https://www.investopedia.com/terms/l/legal-lending-limit.asp

[7] Omarini Anna, "*Peer-to-Peer lending: business model analysis and the platform Dilemma*", Sep 2018, www.researchgate.net/publication/327836640_Peer-to-Peer_Lending_Business_Model_Analysis_and_the_Platform_Dilemma_in_International_Journal_of_Finance_Economics_and_Trade_IJFET_submitted_August_1st_2018_Accepted_Sept_24th

[8] Bhat Adi, "10 must-have questions in a Risk tolérance questionnaire", 10 Oct 2019, www.questionpro.com/blog/risk-tolerance-questionnaire/

Appendix

Exihibit 1

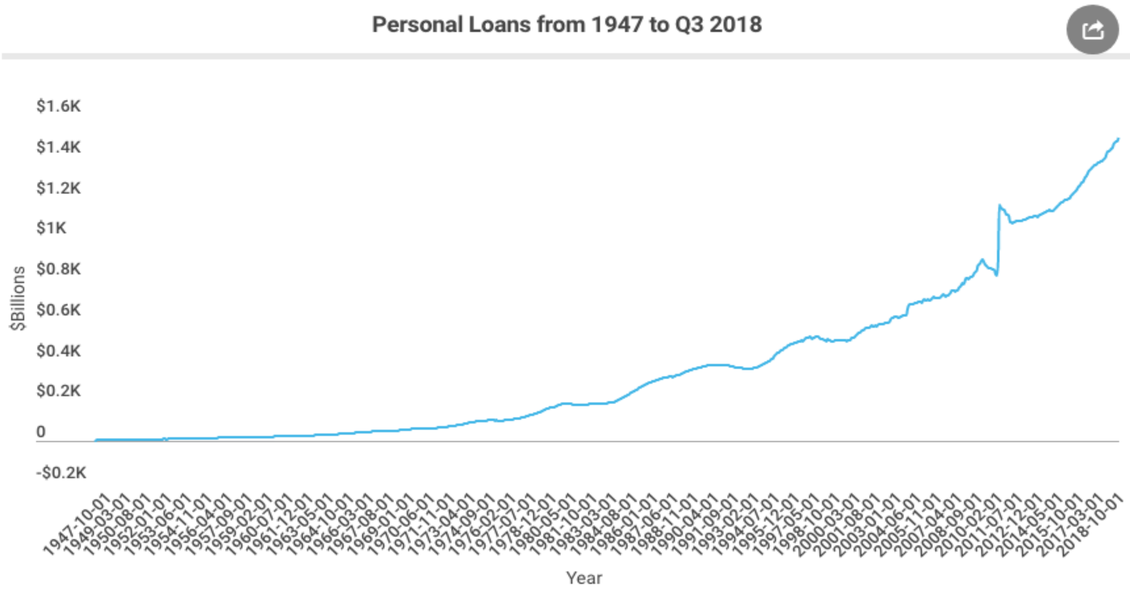


Exhibit 2

Lender	Best For	Est. APR	Min. Credit Score	Learn More
AMERICAN EXPRESS American Express NerdWallet rating	Bank loans for debt consolidation	6.90-19.98%	None	SEE MY RATES on NerdWallet's secure website
DISCOVER PERSONAL LOANS Discover Personal Loan NerdWallet rating	Bank loans for debt consolidation	6.99-24.99%	660	SEE MY RATES on NerdWallet's secure website
Marcus: BY GOLDMAN SACHS Marcus by Goldman Sachs NerdWallet rating	Bank loans for flexible payments, rate discounts	6.99-28.99%	660	CHECK RATE on Goldman Sachs's website
PNC PNC Bank Personal Loan NerdWallet rating	Bank loans for flexible payments, rate discounts	5.99-29.49%	None	SEE MY RATES on NerdWallet's secure website
HSBC HSBC Personal Loan NerdWallet rating	Bank loans for flexible payments, rate discounts	5.49-17.29%	None	SEE MY RATES on NerdWallet's secure website
LIGHTSTREAM LightStream NerdWallet rating	Bank loans for home improvement	5.49-17.29%	660	CHECK RATE on LightStream's website
WELLS FARGO Wells Fargo Personal Loan	Bank loans for home improvement	5.24-22.74%	600	SEE MY RATES on NerdWallet's secure

Exhibit 3

Average Personal Loan Rates by Credit Rating

CREDIT RATING	SCORE RANGE	AVERAGE PERSONAL LOAN INTEREST RATE
Excellent	720 to 850	13.9%
Good	690 to 719	18.0%
Average	630 to 689	21.8%
Bad	300 to 629	27.2%

Rates as of 10/16/2019

Exhibit 4

Exhibit 5

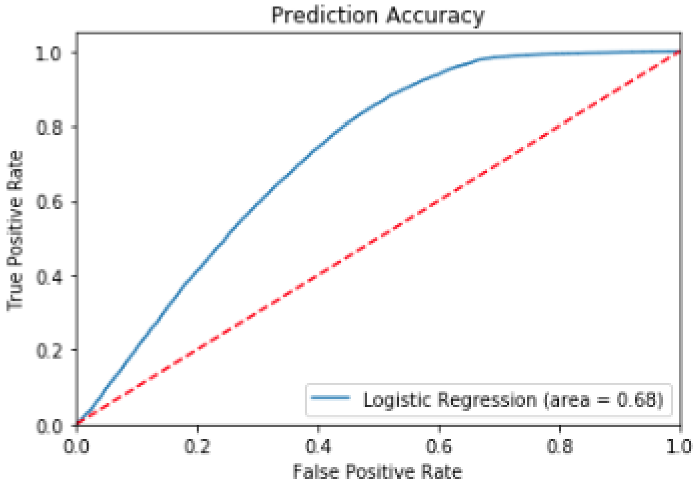

Prediction Accuracy
True Positive Rate
False Positive Rate
Logistic Regression (area = 0.68)
0.0
0.2
0.4
0.6
0.8
1.0

Exhibit 6

```
Optimization terminated successfully.
         Current function value: 0.577377
         Iterations 8
                                Results: Logit
===============================================================================
Model:                    Logit                Pseudo R-squared:     0.167
Dependent Variable:       y                    AIC:                  108036.7871
Date:                     2019-11-14 23:21     BIC:                  108376.8314
No. Observations:         93496                Log-Likelihood:       -53982.
Df Model:                 35                   LL-Null:              -64806.
Df Residuals:             93460                LLR p-value:          0.0000
Converged:                1.0000               Scale:                1.0000
No. Iterations:           8.0000
-------------------------------------------------------------------------------
                                Coef.  Std.Err.    z     P>|z|   [0.025  0.975]
-------------------------------------------------------------------------------
IsBorrowerHomeowner            -0.1446   0.0165  -8.7558 0.0000 -0.1770 -0.1122
CurrentDelinquencies            0.0921   0.0067  13.7670 0.0000  0.0790  0.1052
BankcardUtilization            -0.0377   0.0247  -1.5304 0.1259 -0.0861  0.0106
DebtToIncomeRatio               0.2998   0.0335   8.9454 0.0000  0.2341  0.3655
IncomeVerifiable                3.4551   0.3153  10.9570 0.0000  2.8370  4.0731
PercentFunded                  -2.7217   0.3119  -8.7252 0.0000 -3.3331 -2.1103
Occupation-Computer Programmer -0.3397   0.0431  -7.8733 0.0000 -0.4242 -0.2551
Occupation-Executive           -0.2477   0.0441  -5.6235 0.0000 -0.3341 -0.1614
Occupation-Professional        -0.1420   0.0247  -5.7388 0.0000 -0.1905 -0.0935
Occupation-Food Service         0.2203   0.0805   2.7376 0.0062  0.0626  0.3781
Occupation-Laborer              0.4906   0.0644   7.6137 0.0000  0.3643  0.6169
Occupation-Fireman             -0.4613   0.1274  -3.6200 0.0003 -0.7111 -0.2115
Occupation-Analyst             -0.2443   0.0455  -5.3678 0.0000 -0.3334 -0.1551
Occupation-Architect           -0.8144   0.2174  -3.7469 0.0002 -1.2404 -0.3884
Occupation-Teacher              0.0767   0.0431   1.7800 0.0751 -0.0078  0.1611
Occupation-Waiter/Waitress      0.4534   0.1236   3.6689 0.0002  0.2112  0.6955
BorrowerState-IL               -0.2294   0.0371  -6.1910 0.0000 -0.3020 -0.1568
BorrowerState-NY               -0.1334   0.0332  -4.0241 0.0001 -0.1984 -0.0684
BorrowerState-OH                0.0487   0.0406   1.1988 0.2306 -0.0309  0.1283
BorrowerState-TX               -0.3290   0.0347  -9.4696 0.0000 -0.3971 -0.2609
BorrowerState-CO               -0.5612   0.0633  -8.8711 0.0000 -0.6851 -0.4372
BorrowerState-MO                0.2580   0.0536   4.8098 0.0000  0.1529  0.3631
BorrowerState-LA                0.4607   0.0755   6.1044 0.0000  0.3128  0.6086
BorrowerState-CA               -0.2213   0.0261  -8.4880 0.0000 -0.2723 -0.1702
BorrowerState-MD                0.1680   0.0480   3.4974 0.0005  0.0738  0.2621
EmploymentStatus-Employed      -0.1808   0.0378  -4.7815 0.0000 -0.2549 -0.1067
EmploymentStatus-Full-time     -0.1227   0.0431  -2.8454 0.0044 -0.2072 -0.0382
LoanOriginationQuarter-Q1 2014 -3.8268   0.0782 -48.9512 0.0000 -3.9801 -3.6736
LoanOriginationQuarter-Q3 2012  0.4362   0.0282  15.4690 0.0000  0.3809  0.4915
LoanOriginationQuarter-Q4 2012  0.1034   0.0314   3.2957 0.0010  0.0419  0.1649
LoanOriginationQuarter-Q4 2013 -2.0919   0.0328 -63.7371 0.0000 -2.1562 -2.0275
ListingCategory_numeric-1      -0.0728   0.0201  -3.6325 0.0003 -0.1121 -0.0335
ListingCategory_numeric-3       0.5514   0.0374  14.7262 0.0000  0.4780  0.6247
ListingCategory_numeric-7       0.2496   0.0278   8.9800 0.0000  0.1951  0.3041
Term                            0.0047   0.0008   6.1298 0.0000  0.0032  0.0062
LoanOriginalAmount             -0.0001   0.0000 -33.2682 0.0000 -0.0001 -0.0001
===============================================================================
```

Exhibit 7

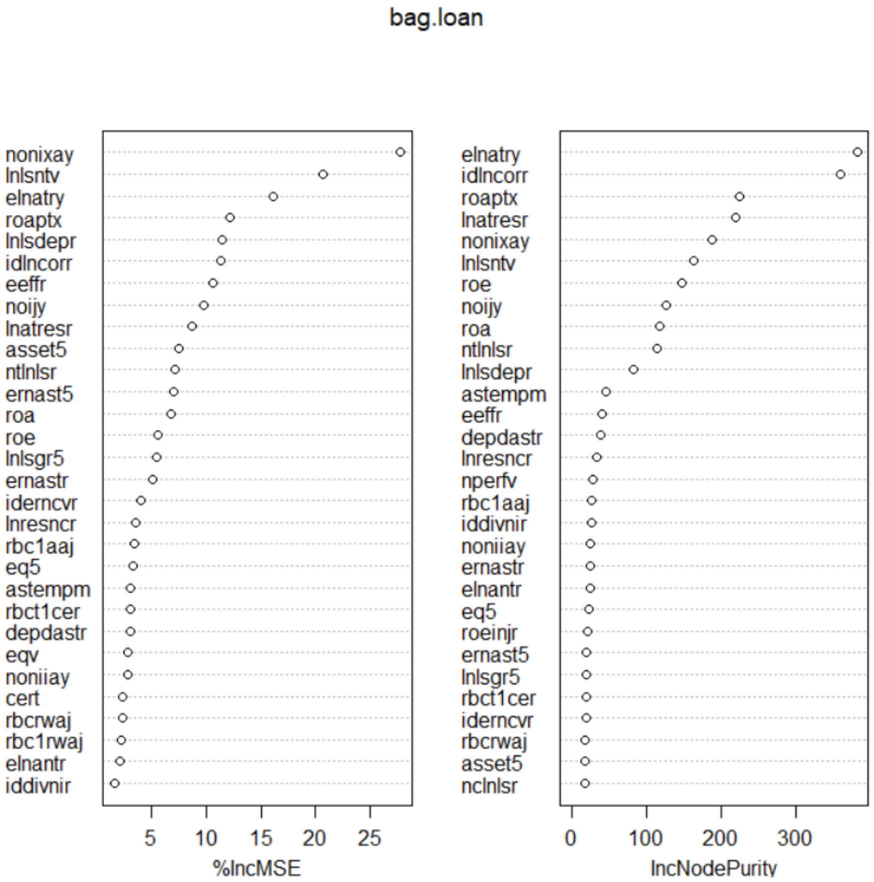


Exhibit 8