

## Proposing a New Research Framework for Loan Allocation Strategies in P2P Lending

(Work in Progress)

Yuting Rong\*, Xi'an Jiaotong University, China, jetryt@163.com  
Wei W. Huang, Xi'an Jiaotong University, China, waynehw@qq.com  
Xin Lai, Xi'an Jiaotong University, China, xlai@xjtu.edu.cn  
Cheng J. Jiang\*, Temple University, USA, cheng.jiang@temple.edu  
Yangzi Peng, Xi'an Jiaotong University, China, 512668920@qq.com  
Linzi Liu, Xi'an Jiaotong University, China, 15829358973@163.com

### ABSTRACT

One of the frontier Web 2.0 applications is online peer-to-peer (P2P) lending marketplace, where individual lenders and borrowers can virtually meet for loan transactions. From a lender's perspective, she not only wants to lower investment risk but also to gain as much return as possible. However, P2P lenders possess the inherent problem of information asymmetry that they don't really know if a borrower has capability to pay the loan or is truthfully willing to pay it in due time, leading them to a disadvantaged situation when making the decision of lending money to the borrower. This study intends to consider the loan allocation as an optimization research problem using the research framework based upon modern portfolio theory with the aim of helping lenders achieve the two goals of gaining high return and lowering risk at the same time.

The expected results of this research are twofold: 1) compared to a logistic regression based credit scoring method, we expect to make more profits for lenders with risk level unchanged, and 2) compared to a linear regression based profit scoring method, we expect to lower risk without lowering return. Our proposed new model could offer insights into how individual lenders can optimize their loan allocation strategies when considering return and risk simultaneously.

*Keywords:* Peer-to-peer lending, Loan allocation, Credit scoring, Profit scoring, Modern portfolio theory.

---

\*Corresponding author

### INTRODUCTION

The arrival of Web 2.0 has brought us the incoming of online peer-to-peer (P2P) lending marketplace, which is a virtual place for individual lenders and borrowers to conduct loan transactions directly without intermediaries (Berger & Gleisner, 2009; Emekter et al., 2015). Since the foundation of the first online peer-to-peer lending platform Zopa in 2006 (Dorflleitner et al., 2016), an increasing number of similar platforms have grown and evolved (Feller et al., 2016), such as Lending Club in American and PPDai.com in China. Specifically, Lending Club, as one of the largest P2P lending platforms in U.S., has attracted more than 2 million customers and funded more than 24 billion due to the end of 2017. As a result of the elimination of a traditional financial intermediary, P2P lending is supposed to lower the transaction cost and increase market efficiency compared to traditional bank loans (Guo et al., 2016; Serrano-Cinca et al., 2015).

However, as lenders and borrowers perform their transactions in a virtual environment (Lin et al., 2013) without the involvement of financial intermediaries, they have insufficient information about each other. Particularly, as borrowers are better informed than lenders of their ability and willingness to repay, information asymmetries (Akerlof, 1970) arises, leading lenders to a disadvantage situation when making decisions. This subsequently results in adverse selection (Spence, 2002) and moral hazard (Petersen & Rajan, 2002) problems. As information asymmetry has become a severe problem for lenders in P2P lending, platforms always provide basic evaluation of loan risk for lenders to make investment decisions, which is often credit-rating based. For example, Lending Club has set up a 7-grade credit-rating system (Xia et al., 2017), and each grade contains 5 sub-grades.

Similarly, traditional credit scoring is usually a dichotomous classification problem and assigns "0" to default loans and "1" to non-default loans (Serrano-Cinca & Gutiérrez-Nieto, 2016). The goal of such methods is to distinguish good (i.e., those supposed to be fully paid) from bad (i.e., those supposed to default) loan applicants or to estimate the probability of default (Verbraken et al., 2014), and then those that the model predicts have a low likelihood of default are accepted, while those with a high likelihood are rejected (Finlay, 2010). As a result, loans with high default probability are always discarded, although some of them are actually with high profitability. However, besides seeking to minimize the percentage of consumers who default, P2P lending platforms also hope to know who are the most profitable (Thomas, 2000; Crook, Edelman & Thomas, 2007), as well as lenders in this market. Thus, there should be a trade-off between risk and return considering the effectiveness of a loan portfolio.

As Markowitz (1952) proposed, investors should consider expected return and risk simultaneously when making investment decisions, the hypothesis (or maxim) that the investor does (or should) maximize discounted return must be rejected. Similarly, rating-based credit risk assessment models, which are practical and have been widely utilized by financial institutions for issuing loans (Guo et al., 2016) should be considered more carefully. These models do not concern about the high profitability of some high-risk loans. In this paper, we are trying to formulate loan allocation as a portfolio optimization problem in the framework of modern portfolio theory, which combines return and risk simultaneously for decision making.

This paper intends to compare linear programming method based on modern portfolio theory with logistic regression based credit scoring and linear regression based profit scoring methods. Our study, therefore, extends the growing research on peer-to-peer lending. More specifically, our work contributes to the research literatures on investment recommendation based on modern portfolio theory in the context of online peer-to-peer lending. For instance, Guo et al. (2016) compared linear programming based on modern portfolio theory with rating-based models, whereas Xia et al. (2017) compared it with credit score and rating-based models. Compared to the methods in Serrano-Cinca and Gutiérrez-Nieto (2016), we try to show that 1) for logistic regression based credit scoring method, our framework derives more profits for investors without improving the risk level, and 2) for linear regression based profit scoring method, our framework lowers risk for investors with the profit unchanged. Given the rapid increasing in the prevalence of online peer-to-peer lending, this study may have important implications not only for researchers, but also for lenders and P2P platforms in marketplace.

The remainder of this paper is organized as follows. The next section describes the research framework of this study, mainly the variables used. We show in Section 3 how to develop the model. Section 4 presents the expected results in twofold, and discusses the possible implications and limitations of this research.

### RESEARCH FRAMEWORK

Independent variables of our proposed new research framework are divided into five categories, including borrower assessment, loan characteristics, borrower characteristics, credit history, and borrower indebtedness. Borrower assessment category includes borrower's credit information related variables. Loan characteristics category contains interest rate, loan amount and loan purpose. Borrower characteristics category describes the basic information of the borrower, including annual income, home ownership and employment length. Borrower's past credit history information is included in credit history category, like number of derogatory public records, the number of inquiries in past 6 months, etc. Moreover, borrower indebtedness is mainly depicted by three ratios, which evaluate the ability of the borrower to pay back the loan, included variables are loan amount to annual income ratio, annual installment to annual income ratio, and debt to income ratio.

As we consider both credit scoring and profit scoring, this study contains two dependent variables. For credit scoring, we calculate the probability of default of each loan by a logistic regression. Among the many models, logistic regression has been the most widely used in the literature (Guo et al., 2016). For profit scoring, we estimate the rank of internal rate of return by a multivariate regression, which is a suitable variable for capturing the profitability of a P2P lending loan (Serrano-Cinca and Gutiérrez-Nieto, 2016).

### RESEARCH DESIGN

Following Serrano-Cinca and Gutiérrez-Nieto (2016), this study applies regression analysis to calculate the probability of default and the rank of the internal rate of return (IRR) of funded loans. We estimate the following two specifications:

$$probability(default=1|x)=\alpha_0+aBorrowerAssessment+\beta LoanCharacteristics +\gamma BorrowerIndebtedness+\delta BorrowerCharacteristics+\phi CreditHistory+\varepsilon \quad (1)$$

$$E(rIRR|x)=\gamma_0+\gamma BorrowerAssessment+\eta LoanCharacteristics +\kappa BorrowerIndebtedness+\lambda BorrowerCharacteristics+\omega CreditHistory+\varepsilon \quad (2)$$

where default and rIRR are the dependent variables which measure the default status and profitability of the funded loan, respectively. As default is a binary variable, we use probit regression to estimate the default probability of a funded loan; whereas rIRR is continuous, we use multivariate regression to estimate Eq. (2). In either equation, all five categories of variables that determine the default probability and profitability of a funded loan are included.

Moreover, for predicting risk of each loan, we find all the loans within the same rating sub-grade in the training set as the target loan and calculate the standard deviation of the internal rate of return within that sub-grade as the predicted risk of the target loan, as Eq. (3) shows:

$$\sigma_i^2=\frac{1}{n-1}\sum_{k=1}^n(IRR_k-\overline{IRR})^2 \quad (3)$$

where  $\sigma_i^2$  represents the risk of loan  $i$ ,  $n$  is the number of loans within the sub-grade,  $IRR_k$  is the internal rate of return of each loan within the sub-grade, and  $\overline{IRR}$  is the mean internal rate of return of that sub-grade.

Now that we have predicted the default probability, expected return and risk of each loan, we will formulate the portfolio selection problem in the framework of Markowitz's (1952) work in our next step of research.

In order to show the effectiveness of our proposed model, we will compare it with two traditional baseline models: 1) a logistic regression based credit scoring model, and 2) a linear regression based profit scoring method.

### RESEARCH DATA COLLECTION AND EXPECTED RESEARCH FINDINGS

The dataset used in this empirical study is officially provided by one of the state-of-the-art P2P lending platforms, Lending Club in the United States. All borrower information and loan information are available on the Lending Club website, as well as payments made during the loan period, which make it easy for us to measure the internal rate of return (IRR) of each loan.

For this paper, we are trying to compare the proposed loan allocation framework with logistic regression based credit scoring method and linear regression based profit scoring method, which are used in Serrano-Cinca and Gutiérrez-Nieto (2016). The expected results are twofold: 1) the proposed method outperforms credit scoring method in that it derives more profits without improving risk, and 2) it outperforms profit scoring method as it lowers the risk with profits unchanged.

Our study could not only provide suggestions for investors on how to allocate their portfolios, but also for P2P lending platforms and financial banks running lending and investment business.

However, there are limitations in this study. For example, more accurate risk prediction of each loan is needed, as we assumed that loans within the same sub-grade had equal risk for lenders; secondly, future research could improve IRR prediction by considering late fees and loss recovery.

### ACKNOWLEDGMENT

This work is partially supported by grants 71731009, 71742005, 71331005 and 71628103 of the National Natural Science Foundation of China.

### REFERENCES

- [1] Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3), 488-500.
- [2] Berger, S., & Gleisner, F. (2009). Emergence of financial intermediaries in electronic markets: The case of online P2P lending. *Business Research*, 2(1), 39-65.
- [3] Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, 183(3), 1447-1465.
- [4] Dorfleitner, G., Priberny, C., Schuster, S., Stoiber, J., Weber, M., Castro, I. D., & Kammler, J. (2016). Description related soft information in peer-to-peer lending - Evidence from two leading European platforms. *Journal of Banking & Finance*, 64, 169-187.
- [5] Emekter, R., Tu, Y., Jirasakuldech, B., & Lu, M. (2015). Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. *Applied Economics*, 47(1), 54-70.
- [6] Feller, J., Gleasure, R., & Treacy, S. (2017). Information sharing and user behavior in internet-enabled peer-to-peer lending systems: An empirical study. *Journal of Information Technology*, 32(2), 127-146.
- [7] Finlay, S. (2010). Credit scoring for profitability objectives. *European Journal of Operational Research*, 202(2), 528-537.
- [8] Guo, Y., Zhou, W., Luo, C., Liu, C., & Xiong, H. (2016). Instance-based credit risk assessment for investment decisions in P2P lending. *European Journal of Operational Research*, 249, 417-426.
- [9] Lin, M., Prabhala, N. R., & Viswanathan, S. (2013). Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. *Management Science*, 59(1), 17-35.
- [10] Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1), 77-91.
- [11] Peterson, M. A., & Rajan, R. G. (2002). Does distance still matter? The information revolution in small business lending. *Journal of Finance*, 57(6), 2533-2570.
- [12] Serrano-Cinca, C., & Gutiérrez-Nieto, B. (2016). The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending. *Decision Support Systems*, 89, 113-122.
- [13] Serrano-Cinca, C., Gutiérrez-Nieto, B., & Lopez-Palacios, L. (2015). Determinants of default in P2P lending. *Plos One*, 10(10), e01394.
- [14] Spence, M. (2002). Signaling in retrospect and the informational structure of markets. *American Economic Review*, 92(3), 434-459.
- [15] Thomas, L. C. (2000). A survey of credit and behavioral scoring: Forecasting financial risk of lending to consumers. *International Journal of Forecasting*, 16(2), 149-172.

- [16] Verbraken, T., Bravo, C., Weber, R., & Baesens, B. (2014). Development and application of consumer credit scoring models using profit-based classification measures. *European Journal of Operational Research*, 238(2), 505-513.
- [17] Xia, Y., Liu, C., & Liu, N. (2017). Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending. *Electronic Commerce Research and Applications*, 24, 30-49.