

RESEARCH

Open Access



# A predictive indicator using lender composition for loan evaluation in P2P lending

Yanhong Guo<sup>1\*</sup> , Shuai Jiang<sup>1</sup>, Wenjun Zhou<sup>2</sup>, Chunyu Luo<sup>1</sup> and Hui Xiong<sup>3</sup>

\*Correspondence:

guoyh@dlut.edu.cn

<sup>1</sup> School of Economics and Management, Dalian University of Technology, Linggong Road, Dalian, China  
Full list of author information is available at the end of the article

## Abstract

Most loan evaluation methods in peer-to-peer (P2P) lending mainly exploit the borrowers' credit information. However, the present study presents the maturity-based lender composition score, which exploits the investment capability of a group of lenders who fund the same loan, to enhance the P2P loan evaluation. More specifically, we extract lenders' profiles in terms of performance, risk, and experience by quantifying their investment history and develop our loan evaluation indicator by aggregating the profiles of lenders in the composition. To measure the ability of a lender for continuous improvement in P2P investment, we introduce lender maturity to capture this evolution and incorporate it into the aggregation process. Our empirical study demonstrates that the maturity-based lender composition score can serve as an effective indicator for identifying loan quality and be included in other commonly used loan evaluation models for accuracy improvement.

**Keywords:** P2P lending, Maturity assessment, Lender composition, Loan evaluation

## Introduction

With the exciting developments in information technology, peer-to-peer (P2P) lending, which has emerged as a new way of financing and investing, has undergone rapid growth in recent years. Notably, P2P lending platforms are internet-based lending intermediaries among individual users who may participate as borrowers or lenders. In this marketplace, borrowers submit applications for loans, referred to as listings, by providing many details about the loan request, such as the purpose and the total amount of funds needed. Lenders are then allowed to partially fund these loan requests by specifying their respective investment amounts. If the requested total dollar amount of the listing is fulfilled within a prespecified time, the transaction proceeds and the listing becomes a loan. Tremendous efforts have been made in both practice and academic research to better understand this economic phenomenon and trading system (Wang et al. 2015a; Liu et al. 2020; Du et al. 2020).

In P2P lending, the lender decides on which loans to invest in by considering the return and associated risk. Moreover, if the borrowers default on their payment obligations, the lenders absorb a loss. Thus, it is crucial for lenders to evaluate each loan by determining whether it may present a significant default risk (Ge et al. 2016;

Wang et al. 2020). Consequently, loan evaluation models have attracted widespread attention. As a core decision support tool, credit scoring is used extensively in financial institutions to predict borrower repayment behavior and provide accurate credit risk estimations (Shen et al. 2020). This type of model scores loans based on their profit potential such that the lenders can identify good quality loans that have higher scores. A core problem entails the prediction of a loan's profitability, which mainly involves a credit risk assessment. While most existing methods primarily utilize credit information from the borrowers (see, e.g., Guo et al. 2016; Thomas et al. 2005), information gleaned from lenders has been found to be highly effective for identifying profitable loans in P2P lending (Luo et al. 2011; Guo et al. 2021). Indeed, loans in P2P lending systems are typically funded collectively by many lenders. Diversification is made possible by allowing individual lenders to spread their money across many different loans. Lenders make different investment decisions based upon differences in experience, knowledge, sources of information, judgment, personal preferences, and so on. In other words, the lender composition of a loan—the group of lenders who invest in this loan—can reveal useful information that indicates successful loan evaluation.

In the present study, we propose a maturity-based lender composition score to exploit the potential of lender information for loan evaluation in P2P lending. First, we perform a quantitative analysis of the expertise of lenders and build lender profiles, including various segments of lenders with different characteristics. In this process, we model the investment relationship between loans and lenders in a straightforward way. The past performance, risk preference, and experience of each lender are summarized. Second, by leveraging the statistical theory of observed confidence intervals, we quantify the lender maturity for measuring the ability of a lender for continuous improvement in P2P investment. This mechanism ensures that, with the same average performance, higher maturity lies in lenders who have made more investments over time as opposed to newer lenders with less lending experience. Finally, the expertise of the lenders who invest in a loan is aggregated to formulate the maturity-based lender composition score. We demonstrate the effectiveness of the maturity-based lender composition score through extensive experimentation on a real-world P2P lending dataset. The results show that our maturity-based lender composition score could serve as an effective indicator for identifying loan quality and be incorporated into other commonly used loan evaluation models for performance improvement, including logistic regression (LR), support vector machine (SVM), and random forests (RFs).

The rest of this paper is organized as follows: in the next section, we present a literature review. Thereafter, we describe the dataset to provide a general background of how P2P lending works and the research context. This dataset is utilized to demonstrate our research methodology. The “[Methodology](#)” section provides details about how we construct the maturity-based lender composition score for loan evaluation in P2P lending. The “[Experimental design](#)” section outlines the structure, strategy, and rationale of our experiments. In the “[Results](#)” section, we report the experimental results for validating the effectiveness of the proposed maturity-based lender composition score. Finally, concluding remarks are given in the last section.

## Literature review

P2P systems attract much attention for sharing information and facilitating collaboration among the participating members (Chen et al. 2014; Kantere et al. 2009; Dewan and Dasgupta 2010). As a novel economic model, P2P lending has been introduced as a new e-commerce phenomenon in recent years (Hulme and Wright 2006). To benefit new players in this market, several studies have investigated factors that improve the chances of converting a listing to a funded loan and have provided decision support for borrowers in composing their requests (Wang et al. 2015b; Herzenstein et al. 2011b; Puro et al. 2011).

Credit risks are uncertainties associated with financing, and risk analysis can help lenders to detect risks in advance, take appropriate actions to minimize the defaults, and support decision-making (Kou et al. 2014). Similar to other financing marketplaces, lenders in P2P lending need to evaluate each loan's default risk. Traditional scorecard modeling is still applicable as the information about the loans and borrowers in P2P lending has the same structure as traditional loans. Tao et al. (2017) reveal that the credit profile of borrowers could significantly affect their loan requests' fulfillment likelihood and also predict the default probability. Compared with traditional lending, P2P lending has some novel features that can be explored; these include language features of request (Herzenstein et al. 2011b), social signals (Greiner et al. 2009; Freedman and Jin 2008; Lin et al. 2013), and even photos of the borrower (Ravina 2012; Pope and Sydnor 2011). This implies that feature selection (Kou et al. 2021) and feature construction have significant potential in this field. Novel methods, such as semi-supervised SVM (Li et al. 2017), profit scoring system (Serrano-Cinca and Gutiérrez-Nieto 2016), instance-based kernel regression (Guo et al. 2016, 2021), misclassification cost matrix credit grading (Wang et al. 2021), and cost-sensitive boosted tree model (Xia et al. 2017), are also being developed to assist lenders' investment decision-making.

While most of these methods utilize information from the borrowers, the information from lenders has been under-explored. Identifying and following successful lenders may be an effective investment strategy that enables tapping into the wisdom of the crowd (James 2004), wherein aggregating information from a group is often ultimately better than information from individuals of the group. For example, in stock investment, Hill and Ready-Campbell (2011) find that investment picks from experts in the crowd should be weighted more than novices in the crowd. The same strategy has also emerged in various social trading scenarios, both among traditional mutual fund managers (Jiang and Verardo 2018) and newer platforms that deploy copy trading (Doering et al. 2015). We are interested if similar trends or predictions could be found in the P2P lending market to provide a possible solution based on such evidence. This idea motivates our study.

Lenders following others in a decision-making system may result in herding, which is defined as a greater likelihood of participation in auctions with more existing bids. Herzenstein et al. (2011a) study the herding behavior in P2P loan auctions and report a positive relationship between partial funding and additional bids. Even though herding behaviors manifest irrationality (i.e., investment decision-making based on information irrelevant to the quality of the loan), it has been established that herding effects are associated with better loan performances. This phenomenon is also reported and visualized by Ceyhan et al. (2011), who utilize the loan fulfillment information for outcome

prediction of the loan. They find that owing to the herding effect, listings with partial funds are more likely to become loans, and loans based on herding behaviors in the past are more likely to be repaid. Zhang and Liu (2012) separates herding behaviors into irrational and rational herding by investigating whether the herding effect is moderated by observable listing attributes. It has been noted that such a distinction is critical as rational herding could be beneficial. The above-mentioned works indicate that following other people in P2P lending might be a viable strategy in selecting loan opportunities.

By viewing a loan as a portfolio of lenders, our study is related to works in composition analysis. Composition analysis, or structure analysis, has been widely used in finance and economics. For example, the portfolio composition (Markowitz 1991; Kolm et al. 2014) evaluates a portfolio of assets, each having different risks and returns, to reduce overall risk to achieve a given level of return. Similarly, capital structure has been considered an important indicator of a firm's financial position (Modigliani and Miller 1958). Furthermore, firms' board composition has been considered to be an important factor that affects firms' overall performance (Linck et al. 2008).

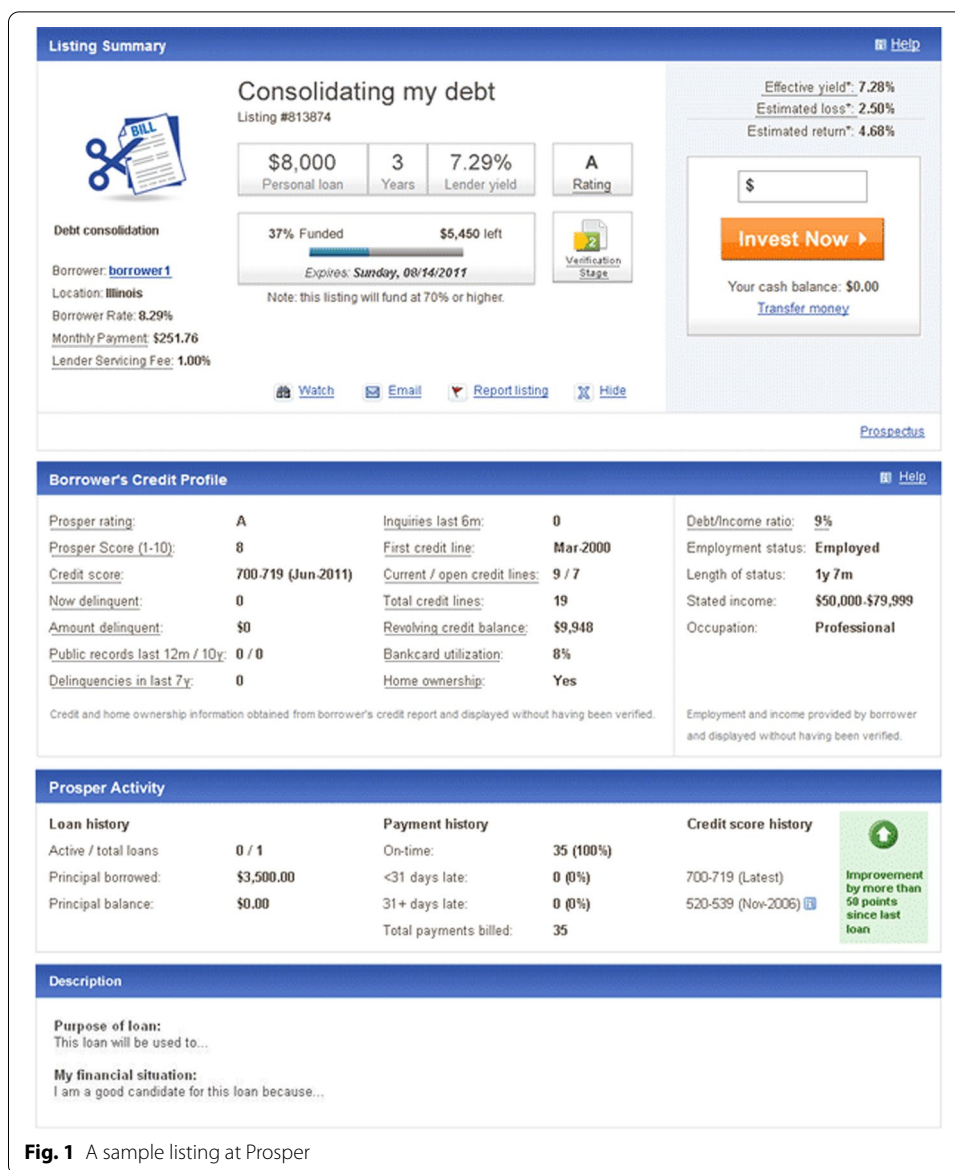
Finally, our study is relevant to works in crowdfunding research. Crowdfunding platforms, such as Kickstarter and Indiegogo, follow a crowd-based funding model (Rakesh et al. 2015; Hong et al. 2018). Researchers usually analyze the inherent factors and predict the participants' behaviors (Chen et al. 2016; Fan-Osuala et al. 2018; Zhao et al. 2019). Despite the focus on startup projects rather than personal lending, the general idea of lender composition analysis may help develop newer venture ideas or discover interesting ideas in which to invest.

## Data and background

Our study is based on a dataset from Prosper, one of the world's largest P2P lending marketplaces. It has more than two million members and has funded over two billion USD in loans. Fig. 1 provides a screenshot of a typical listing viewed by a prospective lender. As we can see from this figure, there are four sections of information: the listing summary, the borrower's credit profile, the borrower's activity history on Prosper, and the description of the loan.

Our dataset includes several relational data tables. The `Members` data table contains the users' basic registration information. The `Credit Profile` table contains the borrowers' personal credit information. The `Listings` table contains information about the loan requests. The `Loans` table contains all information about the loans, such as loan terms, amount, interest rate, and payment status (e.g., paid, late, or default). This table is the most important to evaluate the performance of a loan. The `Bids` table contains the bidding time and the dollar amount contributed by each lender on each loan. Given this information, we can learn about each lender's number of investments and the amount invested on each loan. Based on the data, we can also know whether each loan has been fulfilled. This information provides the basis to build an entire investment network.

We collected a total of 17,407 loans from Prosper that originated between June 2007 and December 2008 and involved 34,155 lenders. All these loans followed a term of 36 months; thus, the payment data of these loans started in June 2007 and ended in



**Fig. 1** A sample listing at Prosper

December 2011. Our sample only consists of loans that are closed at the time of data collection. In other words, either the loans were paid in full or the borrower defaulted.

The loan-level attributes are a combination of loan, borrower, and lender characteristics. The loan attributes include loan purpose, request amount, interest rate, and origination date. Borrower attributes can be divided into two groups: credit risk and individual information. Credit risk features include the borrower's FICO credit score, the number of credit inquiries made upon the borrower's credit score in the last 6 months, and the number of delinquencies the borrower has amassed over the last 7 years. The borrowers' characteristics include whether they are a homeowner, the state in which they reside, employment status, income range, and the debt-to-income ratio. Lender information includes identifiers of lenders that funded the loan and their

specific participation amount. Additionally, certain unstructured data, in the form of pictures or text, are available.

### Methodology

In this section, we describe the maturity-based lender composition score for loan evaluation in P2P lending. We first construct lender profiles by outlining three basic metrics to summarize their historical investment records. Next, we introduce the lender maturity and discuss its methodological foundations. Finally, we provide details of the maturity-based lender composition score.

#### Lender profile

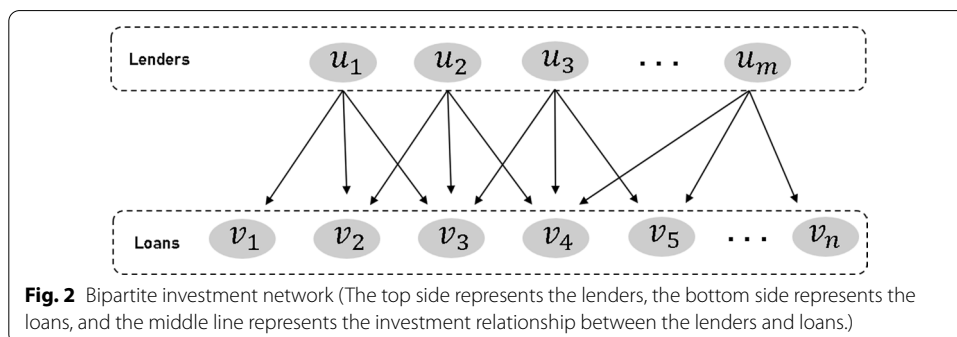
Bipartite networks are widely used to model the relationship between two types of entities (Holme et al. 2003). In P2P lending, we consider the following two types of entities: *lenders*, those who provide capital to selective loans; and *loans*, which are the listings that are funded or to be funded. In the following, we describe the construction of the bipartite investment network with basic notations. Suppose there is a set of  $m$  lenders  $U = \{u_1, u_2, u_3, \dots, u_m\}$  and a set of  $n$  loans  $V = \{v_1, v_2, v_3, \dots, v_n\}$ . We can build a bipartite investment network  $G = \{U, V, E\}$ , where  $U$  and  $V$  are the two types of entities mentioned above, and  $E = (e_{ij})_{m \times n}$  are the edges connecting them. Each edge,  $e_{ij}$ , represents the amount lender  $u_i$  has lent to loan  $v_j$ . Note that  $e_{ij} = 0$  if lender  $u_i$  has never lent to loan  $v_j$ .

As shown in Fig. 2, in P2P lending, we can construct a whole bipartite investment network, based on lender investment data and loan performance data, which provides the basis to quantitatively analyze lender profiles and lender composition.

Based on the lending amount matrix  $E$ , we further define the loan weight matrix  $\Omega = (\omega_{ij})_{m \times n}$ , where  $\omega_{ij}$  is the ratio of lender  $u_i$ 's investment on loan  $v_j$  to the total amount of all of  $u_i$ 's investments in the investment network  $G$ , calculated as:

$$\omega_{ij} = \frac{e_{ij}}{\sum_{j=1}^n e_{ij}}. \tag{1}$$

Similarly, we define the lender weight matrix  $\Theta = (\theta_{ij})_{m \times n}$  where  $\theta_{ij}$  is the ratio of lender  $u_i$ 's investment in loan  $v_j$  to the total amount that loan  $v_j$  receives from all lenders in the investment network  $G$ , calculated as:



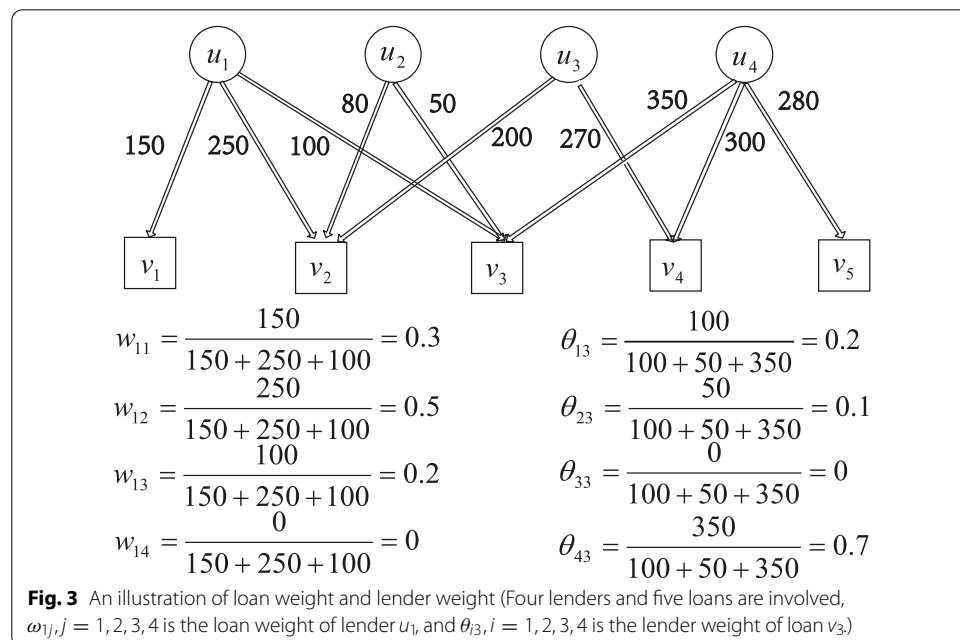
$$\theta_{ij} = \frac{e_{ij}}{\sum_{i=1}^m e_{ij}}. \tag{2}$$

A mini example of the P2P lending investment network is provided in Fig. 3. In this example, it is evident that there are four lender nodes ( $u_1, u_2, u_3,$  and  $u_4$ ), shown in ovals; and five loan nodes ( $v_1, v_2, v_3, v_4,$  and  $v_5$ ), shown in rectangles. The links between lender and loan nodes represent investment relations, with investment amounts labeled on the edges.

Considering that lenders' investment ability is not time-invariant (i.e., the expertise of lenders is constantly updated over time, especially by learning from their experiences with several investments made), recent investments may be more representative of their ability than those made a long while ago. To better quantify the lender's current expertise, higher weights should be given to the lender's more recent investments. Thus, we introduce a time-decaying weight to formulate the importance of the lender's past investments. Consistent with Newton's Law of Cooling (Davidzon 2012), we assume that the time decay rate of the importance is proportional to its value. Hence, we employ the exponential decay function, which is mostly used to formulate the decay effect (Baucells and Bellezza 2017), to evaluate the importance of all the lender's historical investment record. Specifically, when using lender  $u_i$ 's historical investment on loan  $v_k$  to estimate a new target investment  $v_j$ , the time-decaying weight would be:

$$t_{jk} = N_0 e^{-\delta(t_j - t_k)}, \tag{3}$$

where  $N_0$  is the initial decay value and the parameter  $\delta$  controls the decay rate;  $t_k$  and  $t_j$  represent the origination dates of loans  $v_k$  and  $v_j$ , respectively; and the part  $t_k - t_j$  is the time difference between  $v_k$ 's and  $v_j$ 's origination dates.



To analyze the lender composition of loans, we should first characterize the lenders. We describe basic quantitative variables, extracted from their past investment histories, to represent lenders' proficiency in making successful investments. Specifically, we consider performance, risk, and experience, and describe the computational equations for synthesizing each of these metrics into an operationalized numerical measure.

The first variable for lenders' profiles is the overall investment performance or return. This variable represents how successful the lender has been in the past. As a lender in the P2P lending marketplace typically has made multiple investments before, each of which may have a different rates of return, the past investment return of lender  $u_i$  can be calculated as a weighted average of investment returns from all previous investments, written as:

$$\bar{y}_i = \sum_{j=1}^n \frac{t_j \cdot \omega_{ij}}{\sum_{j=1}^n t_j \cdot \omega_{ij}} R_j, \tag{4}$$

where  $t_j$  is the time decay between the origination date of loan  $v_j$  and today,  $\omega_{ij}$  is the weight in Eq. 1, and  $R_j$  represents the rate of return of loan  $j$ . Note that as loans could default with partial or no payment, the rate of return  $R_j$  could be negative.

Standard deviation is commonly used to quantify risk (Markowitz 1991). For each lender  $u_i$ , we define investment risk preference,  $s_i$ , as the weighted standard deviation of investment rates of return, written as:

$$s_i = \sqrt{\sum_{j=1}^n \frac{t_j \cdot \omega_{ij}}{\sum_{j=1}^n t_j \cdot \omega_{ij}} (R_j - \bar{y}_i)^2}, \tag{5}$$

which can be naturally interpreted as the variation or stability of performance in past investments.

It is reasonable to believe that the more the past investments, the more the experience a lender has acquired in the P2P lending marketplace. For each lender  $u_i$ , we define investment experience as the number of previous investments. In P2P lending, given an investment network, the degree of a node  $u_i$ , denoted as  $E_i$ , is the number of edges that have one end attached to the node. We can compute investment experience  $k_i$  using Eq.6:

$$k_i = \#\{e_{ij} | e_{ij} > 0, \forall j\} \tag{6}$$

where # represents the cardinality of a set.

Summarily, in this section, we quantitatively analyze past performance, risk preference, and experience. The analysis of lender profiles provides the basis on which we can analyze the loan's lender composition.

**Lender maturity**

Maturity is a measurement of the ability of an organization or individual for continuous improvement in a particular discipline (Vicente 2017). The higher the maturity, the better the chances that decisions made by the organization/individual would lead to improvements either in the quality or use of the resources (Becker et al. 2009). Most



maturity models qualitatively assess people/culture, processes/structures, and objects/technology (Mettler 2011). In project quality management, maturity is used by a business or organization as a benchmark of how mature their processes are, and how well they are embedded in their culture, with respect to service or product quality management (Crosby 1979; Caballero et al. 2008). In personal career planning, maturity is utilized to reflect individuals' readiness to make well-informed, age-appropriate career decisions and to shape their career carefully in the face of existing societal opportunities and constraints (Naidoo 1998; Savickas 2011). In knowledge management, maturity is employed to assess the capability of an organization with respect to the management of its knowledge resources (Kulkarni and Louis 2003; Grundstein 2008).

In the previous subsection, to assess the expertise of lenders, the lender's profile is built based upon that lender's past investment history. Intuitively, lenders with more historical investments have a higher degree of maturity than inexperienced lenders, which consequently leads to increased accuracy when estimating lenders' profiles. For example, we can better predict the performance of a lender with 30 previous investments than that of a lender with only three past investments. Based on this logic, we introduce maturity into profiling lender's expertise, in which the maturity is used to capture the dynamic changes of the accuracy of the lender's investment ability indicators with the accumulation of investment records.

The maturity of lenders with different investment experiences varies significantly. Moreover, lenders can accumulate knowledge through continuous investment, thereby changing from low to high maturity and tending to stability. Hence, to improve accuracy, it is necessary to identify the maturity level of lenders and the way maturity evolves. It is illogical to believe that lenders are mature only if they have had at least a minimum number of past investments. However, lender experience, or the number of past investments, may be an over-simplified measure for building lender profiles because the number of past investments may not necessarily have a linear relationship with the actual maturity of a lender. Moreover, it is difficult to choose a good cut-off point for a minimum number of past investments without considering the distribution of data.

Herein, we propose a way to quantitatively measure lender maturity, which considers their past performance and risk preference, and thus facilitates a comparison among lenders with different numbers of past investments. Specifically, we consider the adequacy of experience (i.e., the number of past investments) as a sample size problem and quantitatively measure lender maturity by estimating the probability of their performance falling into a fixed-width confidence interval.

Suppose that the performance of lenders  $u_i$  is a random variable  $Y_i$  that follows a normal distribution with mean  $\mu_i$  and standard deviation  $\sigma_i$ . The performance of the lender investing in a loan is a sample from that normal distribution. More specifically, each past investment made (as indexed by  $j$ ) is an independent and identically distributed observation from that distribution. Thus, we have

$$Y_{ij} \stackrel{iid}{\sim} N(\mu_i, \sigma_i^2), j = 1, 2, \dots, k_i, \quad (7)$$

where  $k_i$  is the total number of past investments, as seen in Eq. 6.

With  $k_i$  observed investments, the average performance of the lender,  $\bar{Y}_i$ , would also follow a normal distribution, with a smaller standard deviation:

$$\bar{Y}_i \sim N\left(\mu_i, \frac{\sigma_i^2}{k_i}\right). \tag{8}$$

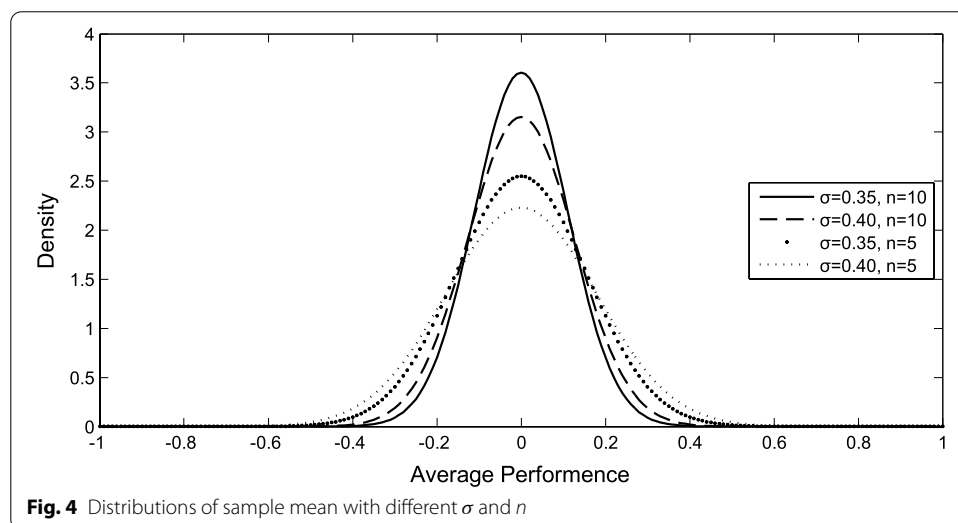
Graphically, we would expect the probability density curve of  $\bar{Y}_i$  to have a thinner and taller peak, if either  $\sigma_i$  is smaller or  $k_i$  is larger, as illustrated in Fig. 4.

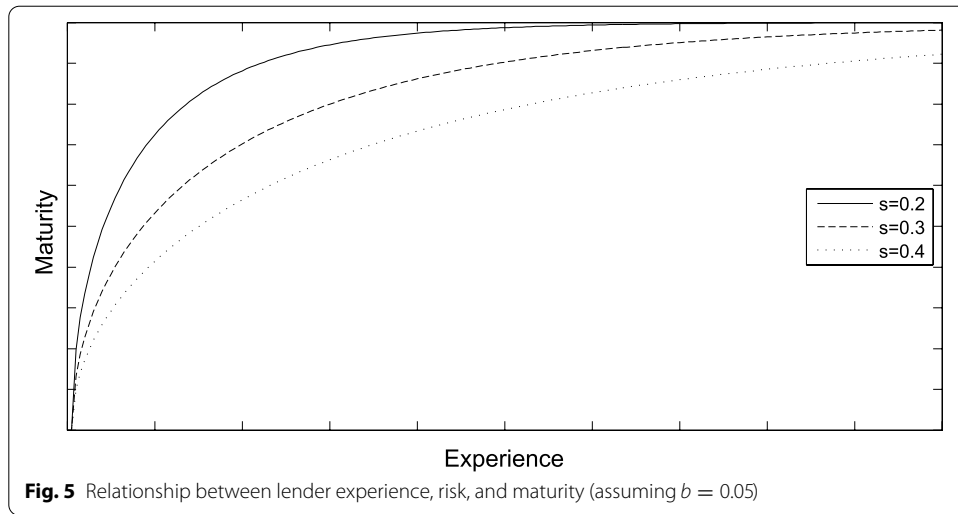
In statistics, confidence intervals are widely used to provide a range of probable values for a given confidence level. With a larger sample size, the variance of the sample mean is reduced, resulting in a narrower confidence interval under the same confidence level because the density curve narrows toward the center. The concept of maturity utilizes the confidence interval mechanism reversely, which we refer to as “interval confidence”. In other words, instead of fixing the confidence level, we fix the margin of error and determine the corresponding confidence value.

As mentioned earlier, lender  $u_i$ 's average performance has a normal distribution, whose mean and standard deviation can be estimated using historical data. Using sample mean  $\bar{Y}_i = \frac{1}{k_i} \sum_{k_i} Y_{ij}$  to estimate  $\mu_i$ , and sample standard deviation  $s_i$  to estimate  $\sigma_i$ , the lender's average performance will follow  $t_{k_i-1}$ , a  $T$  distribution with  $(k_i - 1)$  degrees of freedom, which helps us to establish confidence intervals.

Assuming the same margin of error for all lenders, we evaluate the maturity of a lender using the concept of an observed confidence level (Polansky 2007). Specifically, we set the margin of error,  $b$ , as a global parameter for all lenders. The maturity of a lender  $u_i$  is then defined as the probability of capturing his or her true level of performance, given by mean  $\mu_i$ , within the interval  $[\bar{Y}_i - b, \bar{Y}_i + b]$ . This way, despite their different performances, risks, and number of investments, we could synthesize a numerical value that enables comparison among different lenders.

For lender  $u_i$ , the margin of error can be written as:





**Table 1** Examples of lender maturity calculation

Lender ( $l$ )	Return ( $\bar{y}_l$ )	Risk ( $s_l$ )	Experience ( $k_l$ )	Maturity ( $f_l$ )
1	0.20	0.15	60	0.987
2	0.15	0.20	55	0.931
3	0.10	0.25	50	0.836
4	0.05	0.30	45	0.730

$$b = t_{k_i-1}(\alpha_i) \cdot \frac{s_i}{\sqrt{k_i}} \tag{9}$$

Our goal is to solve  $\alpha_i$ , the percentile that corresponds to a critical value  $\frac{b\sqrt{k_i}}{s_i}$  in a  $T$  distribution with  $(k_i - 1)$  degrees of freedom. Formally, we define

$$f_i = 1 - 2\alpha_i = 2 \cdot F_{t_{(k_i-1)}}\left(\frac{b\sqrt{k_i}}{s_i}\right) - 1 \tag{10}$$

as lender  $u_i$ 's maturity, where  $F_{t_{(k_i-1)}}(x)$  represents the cumulative density function of  $t_{(k_i-1)}$  distribution. This measure, as a probability value, ranged between 0 and 1, provides a suitable maturity measure.

Figure 5 shows the relationship between the maturity score and the actual number of past investments for different levels of lender risks. As we can see, the maturity score increases monotonically with the increase in experience, and the increase levels of when the experience is sufficiently large. Furthermore, the speed of the increase in maturity depends on the level of risk. When the risk is higher, the maturity increases slowly.

Based on the examples shown in Fig. 3, we can come up with simple profiles (including lender maturity calculation) of lenders as shown in Table 1. We can learn

from Table 1 that lower risk and higher experience can be translated into higher maturity of lenders.

#### Lender composition for loan evaluation

In this subsection, we study the quantitative evaluation of loans based upon their lender composition, considering lender maturity to improve the superiority of the loan evaluation model.

In a typical case in a bipartite network, a lender can invest across many different loans, and a loan can be funded by several different lenders, i.e., lender composition. As lenders may have individually contributed different amounts to the total loan, each lender is not equally important. Furthermore, realizing that more weight should be given to the lenders with higher maturity, we adjust the lender weights by considering their maturity. Specifically, we define a maturity-adjusted lender weight,  $\tilde{\theta}_{ij}$ , as:

$$\tilde{\theta}_{ij} = \frac{\theta_{ij}f_i}{\sum_{i=1}^m \theta_{ij}f_i}, \quad (11)$$

where  $\theta_{ij}$  is the unadjusted version in Eq. 2.

For each loan  $v_j$ , the lender performance composition,  $ADJ.CR_j$ , is defined as the weighted average of its lenders' performance. We use the maturity-adjusted lender weights from Eq. 11. Specifically,

$$ADJ.CR_j = \sum_{i=1}^m \tilde{\theta}_{ij}\tilde{y}_i. \quad (12)$$

Similarly, we use the maturity-adjusted lender weights to compute the adjusted lender risk composition,  $ADJ.CP_j$ , as follows:

$$ADJ.CP_j = \sum_{i=1}^m \tilde{\theta}_{ij}s_i. \quad (13)$$

Note that even though the weights are being adjusted, the correlation between lenders is ignored when combining risk (i.e., standard deviation). Assuming independence among lenders is logical as they make their own decisions.

Finally, the composition score,  $ADJ.CS_j$ , is defined as the ratio of return to risk:

$$ADJ.CS_j = \frac{ADJ.CR_j}{ADJ.CP_j}, \quad (14)$$

where  $ADJ.CR_j$  is given in Eq. 12 and  $ADJ.CP_j$  is given in Eq. 13.

The maturity-based lender composition score,  $ADJ.CS_j$ , is a synthesized index that considers both return and risk, and it can be directly applied to the loan evaluation task for indicating the investment value of loans. Each loan can be scored in this way, and all the loans in this marketplace can be ranked according to their maturity-based lender composition score. Evidently, loans with a higher score,  $ADJ.CS_j$ , which implies higher performance composition ( $ADJ.CR_j$ ) or lower risk composition ( $ADJ.CP_j$ ), translate to better investment values.

**Table 2** Description and statistics of variables

Variables	Mean	SD	Description
$X_1$	5525	5258	The monetary amount of the loan
$X_2$	0.2022	0.0903	The Borrower's interest rate for this loan
$X_3$	672.90	75.5151	FICO credit score of the borrower
$X_4$	2.0259	2.8737	Inquires number of borrower in the last 6 months
$X_5$	4.0224	9.9051	Number of borrower's delinquencies in the past 7 years
$X_6$	0.2350	0.2550	The debt to income ratio of the borrower
$X_7$	0.6798	2.1153	Current delinquent of borrower's accounts
$X_8$	3.7121	1.2032	The income range of the borrower
$X_9$	1.4317	2.0168	Loan score calculated by PageRank mechanism
$X_{10}$	-0.3139	0.1254	Loan score without considering lender maturity
$X_{11}$	-0.2805	0.1507	Loan score proposed by this paper
$Y$	0.2985	0.4577	The current status of the loan: 0-Paid; 1-Default

### Experimental design

To find empirical evidence to support the value of the maturity-based lender composition score for loan evaluation in P2P lending, we perform extensive experimental studies using a real-world dataset as described in the third section. This section describes the overall design for the experiments and outlines the choices made at each step of the evaluation process.

### Moving window strategy

The proposed loan evaluation indicator mainly includes two phases: building lenders' profiles and calculating lender composition scores of P2P loans. Accordingly, we divide the loan data into training and testing sets. The training set is utilized to build lender profiles, while the testing set is utilized to calculate the lender composition score. Loans in the profile building phase (i.e., the training set) must precede those in the composition score calculation phase (i.e., the testing set). Therefore, we apply the moving window strategy to divide the training and testing sets. More specifically, at the beginning of the experiment, each loan observation is assigned a unique identifier according to its creation time. Thereafter, the first  $n$  loans are used as the training data, and the next (i.e., the  $(n + 1)$ -th) loan is the test loan. In the subsequent step, the training window moves forward: the previous test case becomes part of the training data and a new test loan is selected. This process continues in this sequential fashion until the full data set is exhausted.

### Benchmark variables

To verify that the maturity-based lender composition score is an effective indicator to identify investment values of loans, we select a range of variables and compare their importance in forecasting the probability of default and return rate. A description of these variables is shown in Table 2.

As illustrated in Table 2,  $X_1, X_2, \dots, X_{11}$  are independent variables, while  $Y$  is the dependent variable. The loan status  $Y$  assumes two possible values that represent a loan's

**Table 3** Variables combinations (The variables in borrower credit are extracted from borrower information, while PageRank, LS, and ADJCS are extracted from lender information.)

Combination	Class	Variables
A	Borrower Credit	$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8$
B	Borrower Credit + PageRank	$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9$
C	Borrower Credit + LS	$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_{10}$
D	Borrower Credit + ADJCS	$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_{11}$

The statistic is considered significant when its *P*-value is less than 0.05

outcome: paid or defaulted. Among the independent variables, attributes  $X_1, X_2, \dots, X_8$  are extracted from borrower information. These variables are commonly used in traditional loan evaluation practices. Therefore, all our predictive models include these variables. In particular, our first benchmark includes only these variables (see Combination A in Table 3).

The last three independent variables,  $X_9, X_{10}$ , and  $X_{11}$ , are derived from the lenders' information;  $X_{11}$  is the maturity-based composition score proposed herein, whereas  $X_9$  and  $X_{10}$  are two benchmark indicators with which we compare  $X_{11}$ . Specifically,  $X_9$  is the score derived from the PageRank algorithm (Brin and Page 1998). As our proposed lender composition score eventually provides a ranked list of loans, PageRank, a state-of-the-art ranking method, is selected to provide a strong benchmark. When applying the PageRank algorithm for loan evaluation, we construct a one-mode loan network, wherein the connections between loans are determined by lenders. Finally, the loan score without considering lender maturity serves as another benchmark.

**Predictive models**

This study aims to investigate whether the predictive model that incorporates a lender composition feature could lead to more accurate and profitable P2P investments than those models that solely consider borrower credit information. Thus, we set four different combinations of independent variables as the input of the predictive model, as shown in Table 3.

In Table 3, feature combination A only contains the borrower's credit information, which is our base model following traditional practice. Combination B adds the variable,  $X_9$ , loan value calculated by the PageRank mechanism, which exploits lenders' information. Combination C adds the variable,  $X_{10}$ , the loan score without considering lender maturity. Combination D uses the proposed loan score,  $X_{11}$ , which introduces lender maturity into the lender composition.

To robustly evaluate the predictive power of lender composition, we use different predictor combinations as model inputs, but additionally use three vastly different predictive models for comparison, defined as:

1. *Logistic regression (LR)*. An individual model that specifies a linear relationship between the response variable and predictors.
2. *Support vector machine (SVM)*. An individual model that specifies a nonlinear relationship between the response variable and predictors.

3. *Random forests (RFs)*. An ensemble model that specifies a nonlinear relationship between the response variable and predictors.

The R software (Version 4.0.0) is used for all experiments. The following packages are used to run the experiments: LR is implemented by the base function named *glm*, SVM using the *e1071* package, and RFs using the *randomForest* package. All the hyper-parameters involved in these algorithms are tuned via the grid search strategy. The radial basis function is adopted as the kernel function in the SVM.

### Results

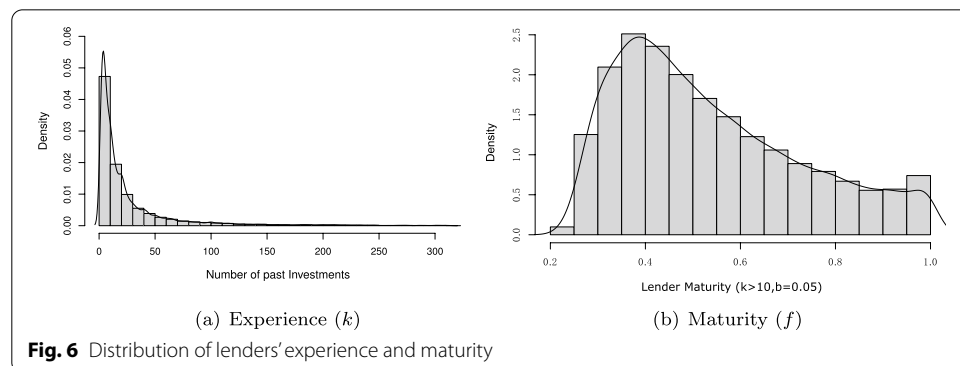
In this section, we first show statistical distributions from the Prosper data. Thereafter, we compare our lender composition indicator with other available variables to study its effectiveness for indicating the potential investment value of a loan. Finally, we demonstrate the model performance to investigate the predictive power of our proposed lender composition score.

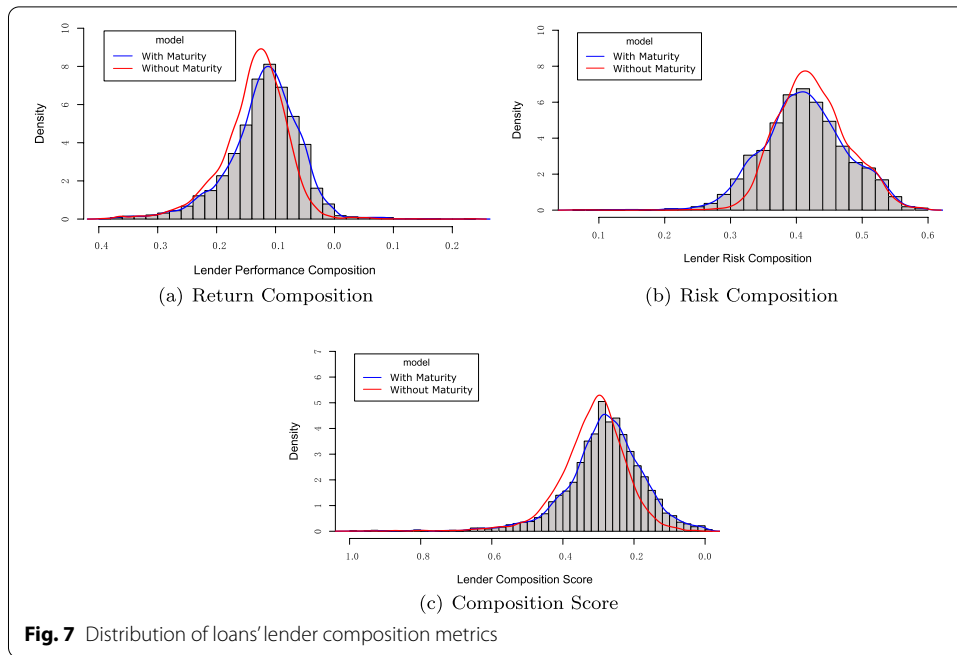
#### Distributions and associations

In this subsection, we show the distributions of the proposed statistics, which are pertinent to lender profiles and lender compositions. We also review the risk and return association in our data.

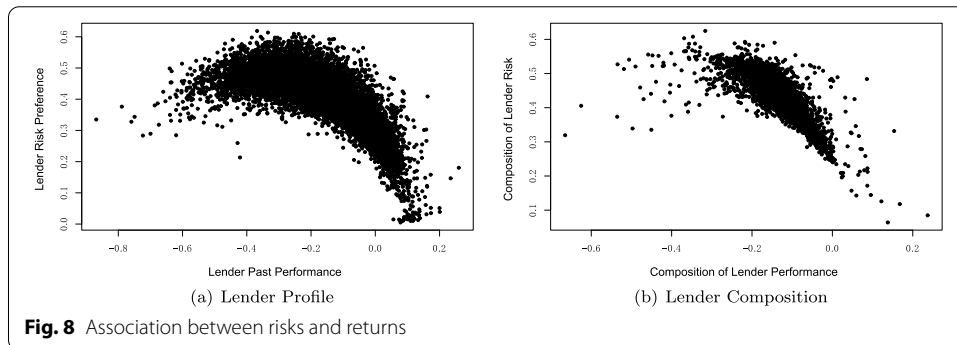
The distribution of lender profiles is shown in Fig. 6. From Fig. 6a, we notice that the distribution of lender experience is extremely skewed. This is well-aligned with our expectation that the number of lenders drops quickly when we require a higher threshold of minimum investments. We choose a cut-off point of 5 as the minimum number of lender experiences required to be included in the training data. The distribution of their maturity is shown in Fig. 6b, and it is unimodal and skews moderately to the right. This indicates that a large group of the lenders possesses lower maturity.

In Fig. 7, we show the distributions of performance composition (CR), risk composition (CP), and the composition score (CS), both with maturity adjustments (gray bars and blue density curve) and without maturity adjustments (red density curve). We could see that despite much similarity and overlap of the density curves, there is a slight shift when adding maturity adjustment. The probability density curves without





**Fig. 7** Distribution of loans' lender composition metrics



**Fig. 8** Association between risks and returns

a maturity adjustment have a thinner and taller peak, which implies that the proposed lender maturity provides a better distinction among loans.

Finally, we consider the association between risk and return in different scenarios. Figure 8a shows the scatter plot of lenders' risk versus lenders' performance, while Fig. 8b shows the scatter plot of each loans' composite risk versus composite return. In both cases, the data points lay on a curve.

**Lender composition as an indicator**

In this subsection, we study the possibility of the proposed maturity-based lender composition score as an indicator for predicting loan outcome (i.e., paid or defaulted). When used individually, we find that the maturity-based lender composition score, compared with a range of other baseline variables, performs as a more powerful predictor for indicating the loan's investment value. For this design, we also group loans



**Table 4** Significance test between groups

Model	Group	Mean	SD	T-Statistic	P value
Without Maturity	Paid	-0.2925	0.0968	20.114	< 0.001
With Maturity	Default	-0.3670	0.1114	<b>22.262</b>	< 0.001
	Paid	-0.2499	0.1184		
	Default	-0.3553	0.1444		

The statistic is considered significant when its P-value is less than 0.05

**Table 5** Correlations between variables and loan default status

Variable	Correlation	Adjusted R <sup>2</sup>	P value
X <sub>1</sub>	0.0279	0.0005	0.0772
X <sub>2</sub>	0.2608	0.0678	< 0.001
X <sub>3</sub>	-0.1766	0.0310	< 0.001
X <sub>4</sub>	0.1797	0.0320	< 0.001
X <sub>5</sub>	0.0517	0.0024	< 0.0011
X <sub>6</sub>	0.0837	0.0068	< 0.001
X <sub>7</sub>	0.0950	0.0088	< 0.001
X <sub>8</sub>	-0.0195	0.0001	0.2202
X <sub>9</sub>	-0.1613	0.0258	< 0.001
X <sub>10</sub>	-0.3425	0.1171	< 0.001
X <sub>11</sub>	<b>-0.3977</b>	<b>0.1580</b>	<b>&lt; 0.001</b>

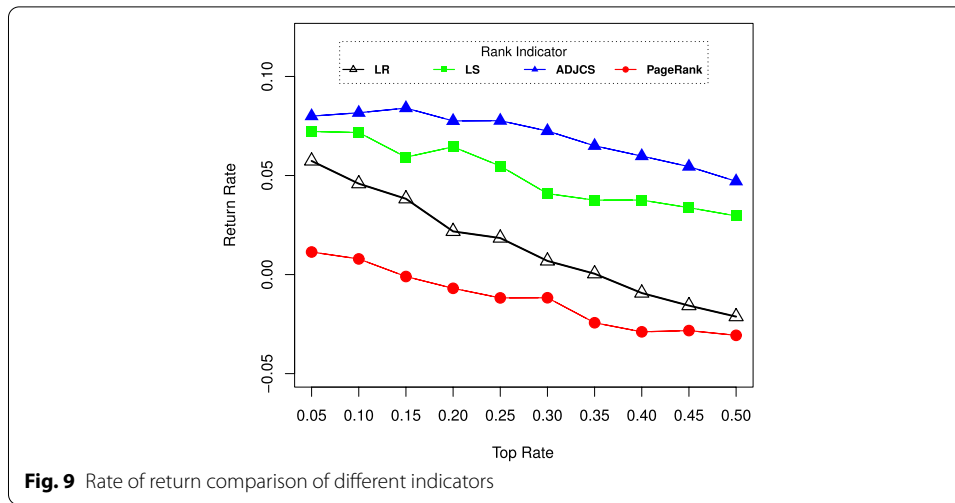
The statistic is considered significant when its P-value is less than 0.05

based on their default status and compare the lender composition score of loans within each group.

First, we compare the lender composition scores between the paid loans versus the defaulted loans, as shown in Table 4. In both with and without maturity cases, the difference between the paid and the defaulted loan groups is highly significant, but the T-statistic for the composition scores with maturity is higher than without it.

The above-mentioned results in Table 4 show that there is a significant difference in the lender composition score between the defaulted and the paid loans, and the lender maturity intensifies this distinction. However, we want to further explore how lender composition score, when considering lender maturity, compared with other available variables when used as a predictor for loan outcome. There are two ways of verifying the effectiveness of the composition score for predicting the response variable. First, we apply a single explanatory variable at a time to predict the loan default status, and the results are listed in Table 5. We see that the maturity-based lender composition score yields better results than any other variable.

Second, we rank loans by using one of the indicators: *LR*, *LS*, *ADJCS* or *PageRank*, where *LR* is the predicted default probability obtained by applying logical regression to the borrower variables of X<sub>1</sub>–X<sub>8</sub> only, which has been the most widely used in the literature; *PageRank* is the score by applying the PageRank algorithm; *LS* is the loan score without considering the lender maturity; and *ADJCS* is the maturity-based lender



**Table 6** Mean square errors of different predictive models

Variables combination	LR	SVM	RFs
A	0.1912	0.1882	0.1992
B	0.1910	0.1842	0.1949
C	0.1826	0.1724	0.1863
D	0.1743	0.1591	0.1765

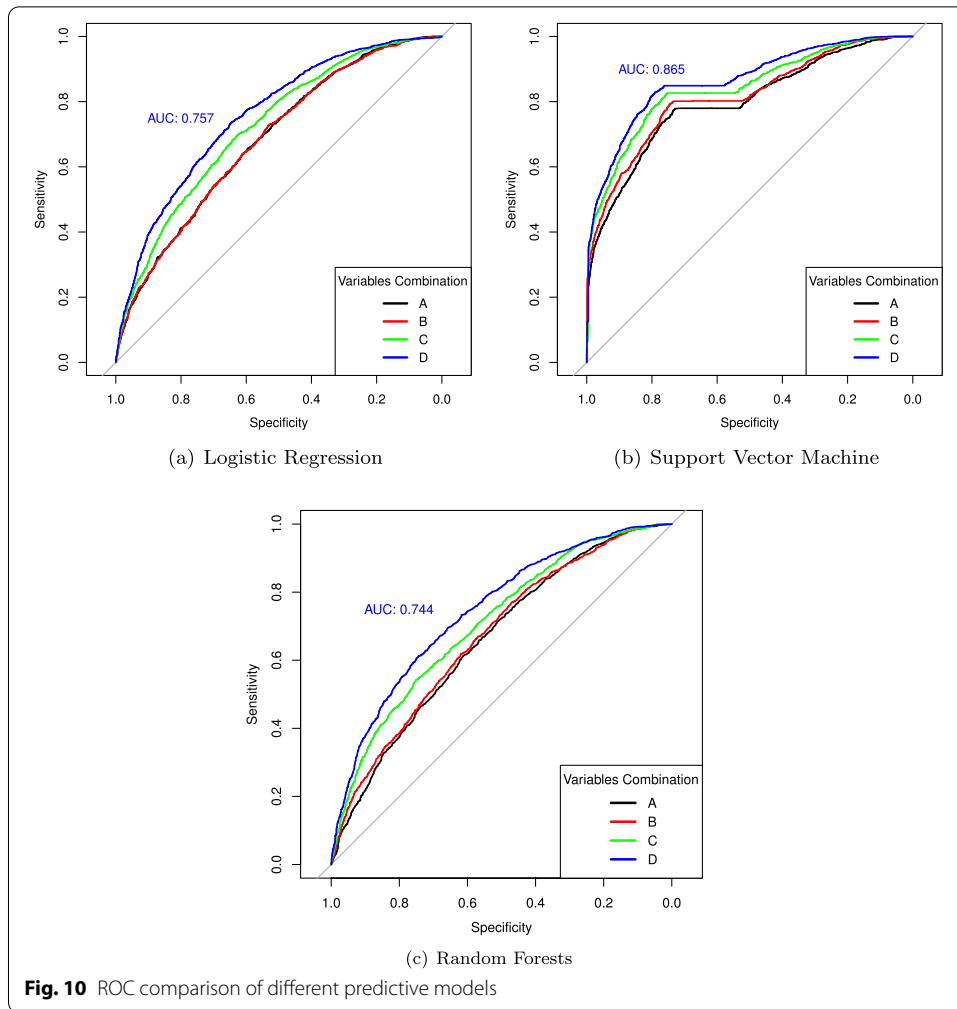
composition score proposed herein. We compare the rate of return by selecting top loans ranked by each of these indicators, and the results are shown in Fig. 9.

In Fig. 9, the x-axis represents the percentage of top loans in which to invest (i.e., top rates), and the y-axis is the corresponding rate of return if we invest in such a set of loans with equal weights. We find that with different top rate thresholds, loans chosen according to *ADJCS* consistently show higher rates of return than the others. Overall, the maturity-based lender composition score computed in the present study is a better indicator of loan quality than other indicators.

**Predictive power of lender composition**

In this subsection, we investigate the predictive power of the maturity-based lender composition score. We train predictive models by combining different variables to predict loans’ default probability, which is an important research area in the P2P loan evaluation literature (Serrano-Cinca and Gutiérrez-Nieto 2016; Tao et al. 2017; Dendramis et al. 2020). Based on the predicted default probability, the classification performance and the outcomes of different models are compared. The prediction algorithms and variables we consider are described in detail in the last section.

First, we compare the mean square error between different predictive models. As shown in Table 6, regardless of the machine learning algorithm used, the mean square error is the smallest when the input is the variable combination *D*. Conversely, the combination *A* produces a larger error. This suggests that our maturity-based lender composition score can significantly reduce the prediction error of the model.



**Table 7** Precision comparison of different predictive models

Observations	Variables combination	LR	SVM	RFs
ALL	A	0.7308	0.7085	0.7299
	B	0.7326	0.7091	0.7376
	C	0.7428	0.7174	0.7587
	D	0.7539	0.7807	0.7747
TOP 50%	A	0.8037	0.8685	0.7953
	B	0.8032	0.8819	0.8008
	C	0.8321	0.8964	0.8157
	D	0.856	0.9098	0.8453
TOP 30%	A	0.8596	0.8787	0.8418
	B	0.8621	0.8895	0.8479
	C	0.8779	0.9153	0.8671
	D	0.9086	0.9377	0.8954

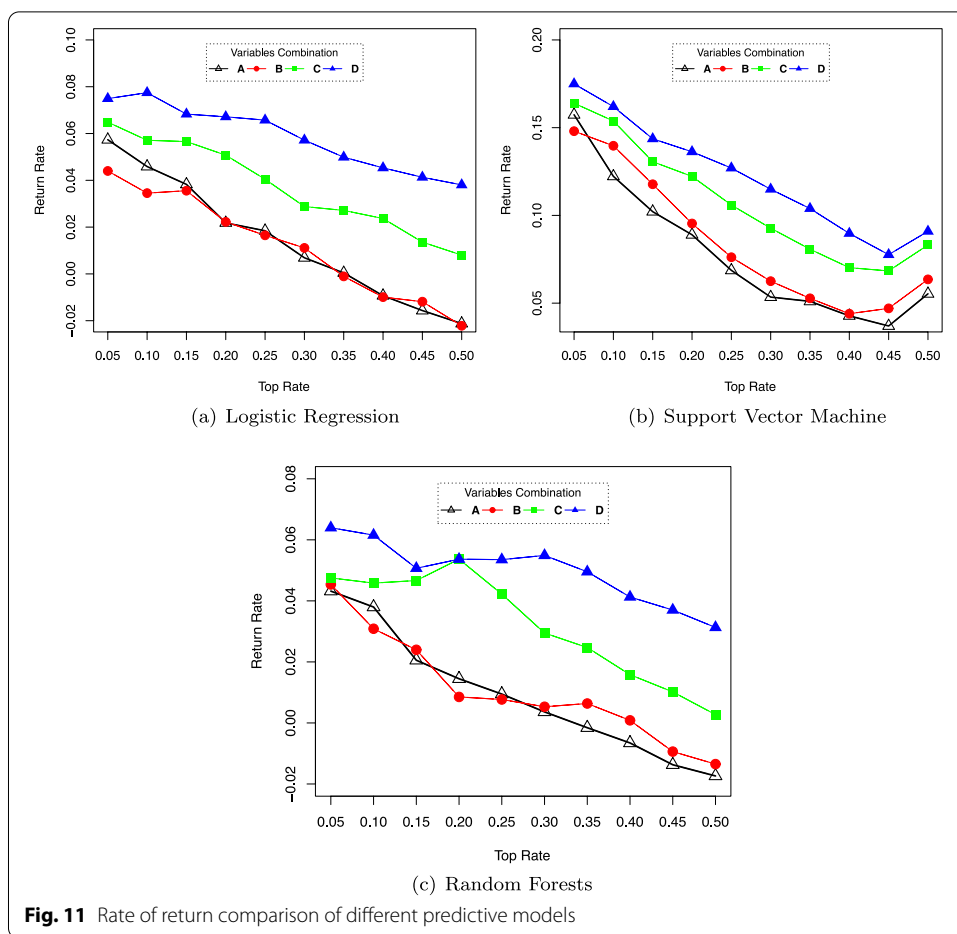
Second, considering the impact of data distribution on the classification results, we draw the receiver operating characteristic curve (ROC) and compute the area under the curve (AUC) based on the prediction results of each model; the results are shown in Fig. 10. In Fig. 10a, the LR algorithm is employed, and we can see that when the input is variable combination  $D$ , the AUC of this predictive model is 0.757, which outperforms the other three combinations. When the learning algorithm is changed to SVM or RFs, as shown in Fig. 10b and c, similar comparison results can still be observed. In other words, when the evaluation measure is the AUC, the proposed indicator still duly works.

Third, in P2P lending, lenders generally pay more attention to loans with lower default probability because such loans usually imply higher profitability. From this perspective, in Table 7, we illustrate the classification precision, which measures the percentage of the predicted non-default loans that are successfully classified by the model, for each of our constructed model. When considering all the observations in our test set, we see in the first row of Table 7 that irrespective of the applied algorithm, the models that include lender information as an independent variable (Combination  $B$ ,  $C$ , and  $D$ ) have better performances than the baseline that does not consider lender information (Combination  $A$ ). Further, Combination  $D$ , which incorporates the maturity-based lender composition score, presents even better results than Combination  $B$  and Combination  $C$ , as it has higher precision. These results show that our indicator could make overall improvements on the loan default prediction process.

While the aforementioned improvement of classification precision is determined by all loans, the return rate of investment depends more on the best quality part of all loans. This is because the lenders, in general, select these best quality loans as their investee rather than all candidates. We suggest that the lower default probability predicted by the model also indicates the higher return rate. Thus, we additionally present the detailed improvement of default prediction by considering the distribution of predicted probability. As shown in the second and third rows of Table 7, we select 50% and 30% of loans with lower default probability as best candidates set, respectively, and compare the classification precision of these best loans set. The conclusions we could draw from the results are consistent with the case of all observations. Moreover, the improvement of classification precision increases with the decrease in default probability. When considering the 30% of loans with the lowest default probability, Combination  $D$  could make an approximately 3% improvement over Combination  $C$ , which reveals that our indicator exhibits more significant improvement on the candidates with a lower predicted probability of default.

To further investigate the profitability value implied in the improvement of the default probability prediction, we evaluate each model's actual return by selecting the top  $\gamma$  percentage of the most attractive loans based on that model's predictions and compare different models' average returns.

In Fig. 11, we show the results that compared Combination  $D$ 's average return rate against the other combinations in the entire test set. In this figure, the x-axis,  $\gamma$ , is the percentage of candidates in which to invest, and the y-axis is the corresponding rate of return if we invest on such a set of loans with equal weights. As evident in Fig. 11a, when the LR algorithm is applied, the models that incorporate the combination of the variables, which include lender composition indicator (Combinations  $C$  and  $D$ ) consistently



have the best return rate performance with different values of  $\gamma$ . Furthermore, Combination *D* has even better results than Combination *C*, as it has higher return rates. The same conclusion holds for SVM and RFs, as is evident in Fig. 11b and c, respectively.

From all these results, we logically conclude that the maturity-based lender composition score proposed herein consistently demonstrates competitive performance compared to several benchmarks, which include a range of borrowers' credit features and two representative lender indicators.

### Concluding remarks

In this paper, we present the maturity-based lender composition score, which exploits lender information, for loan evaluation in P2P lending. First, we build profiles to quantify the lenders' ability to pick high-quality loans. Next, based on these lender profiles, we formulate a lender maturity factor to measure the ability of a lender for continuous improvement in P2P investment. Finally, we develop a maturity-based lender composition score to predict the profit potential of each loan. Our empirical results on a real-world P2P lending dataset reveal that the maturity-based lender composition score improves the efficiency of indicating the investment value, and loan evaluation models

with this lender composition indicator perform significantly better than without it. Summarily, the maturity-based lender composition score proposed herein effectively indicates the investment value of loans and improves loan evaluation accuracy.

Additionally, we design an effective framework to extract the investment expertise of lenders and prove that identifying and following the more mature lenders can lead to better investment performance. By providing a way to acquire knowledge about loans and the implicit behavior of lender crowd, we further contribute to the literature related to the wisdom of crowds. Hence, the maturity-based composition score proposed herein not only provides methodological and theoretical support for loan evaluation in P2P lending but also provides decision support for investors in similar application scenarios, such as the securities market.

From the perspective of P2P lending platforms, all the bidding and profit information of the lenders, which play vital roles in profiling the lender's investment ability and in constructing the lender composition score, can be collected and analyzed. Hence, there is a robust chance that these platforms would employ our model to construct the proposed indicator of loan applications and incorporate it into the credit evaluation system of loans to strengthen risk identification. Furthermore, the lender composition score can be provided to the potential lenders as a value-added service; thus, lending platforms can help lenders to identify the credit risk of loans more comprehensively, which not only reduces the investment risk of lenders but also facilitates the loan credit management of the platform itself.

It should be noted that our loan evaluation method primarily focuses on judging the relative merit of the potential investment value of the loans. Therefore, we do not attempt to optimally assess the expected return and risk of the loans, which serve as necessary inputs for loan portfolio optimization. Conceptualizing an efficient way to build investment decision-making based on lender composition analysis is a promising area to which appropriate risk assessment can significantly contribute. In addition to exploiting lenders' historical investment records in this P2P lending marketplace, it is worthwhile to further strengthen the investor composition analysis for loan evaluation by using investors' social networks, behavioral records, and other external information.

#### **Acknowledgements**

We would like to express our sincere appreciation to editors and the reviewer team in making valuable comments and suggestions to the paper. Their comments have improved the quality of the paper immensely.

#### **Authors' contributions**

GYH conceived of the study, designed the methodology, participated in writing, reviewing and editing the manuscript, acquired the financial support for the project leading to this publication, and took leadership responsibility for the research activity planning and execution. JS implemented the computer code and supporting algorithms, performed the experiments, analyzed the results, and participated in writing, reviewing, and editing the manuscript. ZWJ participated in experimental design and writing, reviewing, and editing the manuscript. LCY participated in the design of the study and performed the statistical analysis. XH participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

#### **Funding**

This research was partially supported by the Natural Science Foundation of China (Nos. 71974031, 71771034), the Chinese Universities Scientific Fund (No. DUT19RW216) and the Economic and Social Development Project of Liaoning Province (No. 20201slktyb-019). It was also supported in part by the National Science Foundation (NSF) via the Grant Number IIS-1648664.

#### **Availability of data and materials**

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

## Declarations

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>School of Economics and Management, Dalian University of Technology, Linggong Road, Dalian, China. <sup>2</sup>Department of Business Analytics and Statistics, University of Tennessee, Knoxville, USA. <sup>3</sup>Management Science and Information Systems Department, Rutgers University, New Brunswick, USA.

Received: 1 August 2020 Accepted: 8 June 2021

Published online: 22 June 2021

## References

- Baucells M, Bellezza S (2017) Temporal profiles of instant utility during anticipation, event, and recall. *Manage Sci* 63(3):729–748
- Becker J, Knackstedt R, Pöppelbuß J (2009) Developing maturity models for it management: a procedure model and its application (bise—research paper). *Bus Inform Syst Eng* 1(3):213–222
- Brin S, Page L (1998) The anatomy of a large-scale hypertextual web search engine. *Comput Netw ISDN Syst* 30(1–1):107–117
- Caballero I, Caro A, Calero C, Piattini M (2008) Iqm 3: information quality management maturity model. *J Univ Comput* 14(22):3658–3685
- Ceyhan S, Shi X, Leskovec J (2011) Dynamics of bidding in a P2P lending service: effects of herding and predicting loan success. In: WWW, pp 547–556
- Chen G, Hu T, Jiang D, Lu P, Tan K, Vo HT, Wu S (2014) Bestpeer++: a peer-to-peer based large-scale data processing platform. *IEEE Trans Knowl Data Eng* 26(6):1316–1331
- Chen L, Huang Z, Liu D (2016) Pure and hybrid crowds in crowdfunding markets. *Financ Innov* 2(1):1–18
- Crosby B (1979) *Quality is free*. New American Library, New York
- Davidzon MI (2012) Newton's law of cooling and its interpretation. *Int J Heat Mass Transf* 55(21–22):5397–5402
- Dendramis Y, Tzavalis E, Varthalitis P, Athanasiou E (2020) Predicting default risk under asymmetric binary link functions. *Int J Forecast* 36(3):1039–1056
- Dewan P, Dasgupta P (2010) P2p reputation management using distributed identities and decentralized recommendation chains. *IEEE Trans Knowl Data Eng* 22(7):1000–1013
- Doering P, Neumann S, Paul S (2015) A primer on social trading networks—institutional aspects and empirical evidence. In: Presented at EFMA annual meetings 2015, Breukelen/Amsterdam
- Du N, Li L, Lu T, Lu X (2020) Prosocial compliance in p2p lending: a natural field experiment. *Manag Sci* 66(1):315–333
- Fan-Osuala O, Zantedeschi D, Jank W (2018) Using past contribution patterns to forecast fundraising outcomes in crowdfunding. *Int J Forecast* 34(1):30–44
- Freedman S, Jin GZ (2008) Do social networks solve information problems for peer-to-peer lending? Evidence from prosper.com. Technical report, NET Institute
- Ge R, Feng J, Gu B (2016) Borrower's default and self-disclosure of social media information in p2p lending. *Financ Innov* 2(1):30
- Greiner ME, Wang H (2009) The role of social capital in people-to-people lending marketplaces. In: Proceedings of the international conference on information systems, p 29
- Grundstein M (2008) Assessing enterprise's knowledge management maturity level. *Int J Knowl Learn* 4:415–426
- Guo Y, Zhou W, Luo C, Liu C, Xiong H (2016) Instance-based credit risk assessment for investment decisions in P2P lending. *Eur J Oper Res* 249(2):417–426
- Guo Y, Jiang S, Qiao H, Chen F, Li Y (2021) A new integrated similarity measure for enhancing instance-based credit assessment in p2p lending. *Expert Syst Appl* 175(114):798
- Herzenstein M, Dholakia UM, Andrews RL (2011a) Strategic herding behavior in peer-to-peer loan auctions. *J Interact Mark* 25(1):27–36
- Herzenstein M, Sonenshein S, Dholakia UM (2011b) Tell me a good story and I may lend you money: the role of narratives in peer-to-peer lending decisions. *J Mark Res* 48(SPL):S138–S149
- Hill S, Ready-Campbell N (2011) Expert stock picker: the wisdom of (experts in) crowds. *Int J Electron Commer* 15(3):73–102
- Holme P, Liljeros F, Edling CR, Kim BJ (2003) Network bipartivity. *Phys Rev E* 68(5):56–107
- Hong Y, Hu Y, Burtch G (2018) Embeddedness, pro-sociality, and social influence: evidence from online crowdfunding. *MIS Q.* <https://srn.com/abstract=3125936>
- Hulme MK, Wright C (2006) Internet based social lending: past, present and future. *Soc Fut Observ* 174(2):1–115
- James S (2004) *The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. Doubleday, New York
- Jiang H, Verardo M (2018) Does herding behavior reveal skill? An analysis of mutual fund performance. *J Financ* 73(5):2229–2269
- Kantero V, Skiadopoulou S, Sellis T (2009) Storing and indexing spatial data in p2p systems. *IEEE Trans Knowl Data Eng* 21(2):287–300
- Kolm PN, Tütüncü R, Fabozzi FJ (2014) 60 years of portfolio optimization: practical challenges and current trends. *Eur J Oper Res* 234(2):356–371
- Kou G, Peng Y, Wang G (2014) Evaluation of clustering algorithms for financial risk analysis using mcdm methods. *Inf Sci* 275:1–12

- Kou G, Xu Y, Peng Y, Shen F, Chen Y, Chang K, Kou S (2021) Bankruptcy prediction for smes using transactional data and two-stage multiobjective feature selection. *Decis Support Syst* 140(113):429
- Kulkarni U, Louis RS (2003) Organizational self assessment of knowledge management maturity. In: *AMCIS 2003 proceedings*, p 332
- Li Z, Tian Y, Li K, Zhou F, Yang W (2017) Reject inference in credit scoring using semi-supervised support vector machines. *Expert Syst Appl* 74:105–114
- Lin M, Prabhala NR, Viswanathan S (2013) Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending. *Manag Sci* 59(1):17–35
- Linck JS, Netter JM, Yang T (2008) The determinants of board structure. *J Financ Econ* 87(2):308–328
- Liu Z, Shang J, Wu S, Chen P (2020) Social collateral, soft information and online peer-to-peer lending: a theoretical model. *Eur J Oper Res* 281(2):428–438
- Luo C, Xiong H, Zhou W, Guo Y, Deng G (2011) Enhancing investment decisions in P2P lending: an investor composition perspective. In: *KDD*, pp 292–300
- Markowitz HM (1991) *Portfolio selection: efficient diversification of investments*, 2nd edn. Wiley, Hoboken
- Mettler T (2011) Maturity assessment models: a design science research approach. *Int J Soc Syst Sci* 3(1–2):81–98
- Modigliani F, Miller MH (1958) The cost of capital, corporation finance and the theory of investment. *Am Econ Rev* 48(3):261–297
- Naidoo AV (1998) Career maturity: a review of four decades of research. ERIC
- Polansky AM (2007) *Observed confidence levels: theory and application*. Chapman & Hall/CRC, Boca Raton
- Pope DG, Sydnor JR (2011) What's in a picture? Evidence of discrimination from prosper.com. *J Hum Resour* 46(1):53–92
- Puro L, Teich JE, Wallenius H, Wallenius J (2011) Bidding strategies for real-life small loan auctions. *Decis Support Syst* 51(1):31–41
- Rakesh V, Choo J, Reddy CK (2015) Project recommendation using heterogeneous traits in crowdfunding. In: *Proceedings of the 9th international AAAI conference on web and social media*
- Ravina E (2012) Love & loans: the effect of beauty and personal characteristics in credit markets
- Savickas ML (2011) Career maturity: the construct and its measurement. *Career Dev Q* 32(4):222–231
- 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. *Decis Support Syst* 89:113–122
- Shen F, Zhao X, Kou G (2020) Three-stage reject inference learning framework for credit scoring using unsupervised transfer learning and three-way decision theory. *Decis Support Syst* 137(113):366
- Tao Q, Dong Y, Lin Z (2017) Who can get money? Evidence from the Chinese peer-to-peer lending platform. *Inf Syst Front* 19(3):1–17
- Thomas LC, Oliver RW, Hand DJ (2005) A survey of the issues in consumer credit modelling research. *J Oper Res Soc* 56(9):1006–1015
- Vicente A (2017) Open information security maturity model
- Wang H, Chen K, Zhu W, Song Z (2015a) A process model on p2p lending. *Financ Innov* 1(1):1–8
- Wang H, Kou G, Peng Y (2021) Multi-class misclassification cost matrix for credit ratings in peer-to-peer lending. *J Oper Res Soc* 72(4):923–934
- Wang P, Zheng H, Chen D, Ding L (2015b) Exploring the critical factors influencing online lending intentions. *Financ Innov* 1:8. <https://doi.org/10.1186/s40854-015-0010-9>
- Wang Z, Jiang C, Zhao H, Ding Y (2020) Mining semantic soft factors for credit risk evaluation in peer-to-peer lending. *J Manag Inf Syst* 37(1):282–308
- Xia Y, Liu C, Liu N (2017) Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending. *Electron Commer Res Appl* 24:30–49
- Zhang J, Liu P (2012) Rational herding in microloan markets. *Manage Sci* 58(5):892–912
- Zhao H, Jin B, Liu Q, Ge Y, Chen E, Zhang X, Xu T (2019) Voice of charity: prospecting the donation recurrence donor retention in crowdfunding. *IEEE Trans Knowl Data Eng*: 1–1

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---