Can investors' collective decision-making evolve? Evidence from Peer-to-Peer lending markets

Dongwoo Kim (dwkim0302@gmail.com)
KB Financial Group https://orcid.org/0000-0003-2219-083X

Research Article

Keywords: Investors' Prediction, Prediction Accuracy, Evolution, Repayment Performance, Credit Assessment

Posted Date: April 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-381657/v1

Abstract

This study tries to identify the accuracy of individual investors' capability to predict a borrower's creditworthiness in peer-to-peer lending markets and examine whether their ability is likely to evolve over time. The results of this study show that there is no significant difference between the predictive power of investors' ex-ante funding decision model and that of the ex-post repayment model over a borrower's repayment performance. Furthermore, the predictive power of investors' ex-ante funding decision over a borrower's repayment performance is shown to improve over time . It is also found that the main reason why investors' predictive power improve over time is because investors can assess more accurately the information provided by the platform operator and describe the borrower's characteristics. The results of this study are important as they confirm the possibility of optimizing and streamlining the P2P lending market, through the evolution of investors' decision making.

1. Introduction

Peer-to-peer (P2P) lending is one of the newly emerging e-commerce businesses in which investors and borrowers make loan transactions through an online platform (D. Chen, Lai, & Lin, 2014; Wang, Greiner, & Aronson, 2009). Emerging in the mid-2000s, P2P lending is seen as an alternative to traditional financing in the lending market. Unlike traditional financing, P2P lending can provide higher welfare for individual market participants by eliminating the intermediary function of financial institutions. In other words, a borrower and an investor can make direct money transactions without the intervention of an intermediary; instead of eliminating the transaction margin held by the intermediary in traditional financing, the borrower can take out loans at lower interest rates, and the investors can pursue higher returns.

However, there are important prerequisites. In order for the eliminated brokerage margin of existing financing methods to be linked to minimizing the borrowers' costs and maximizing investors' profits, the P2P lending system must have similar or higher efficiency than traditional financial systems. For this to be possible, the most important factor is the investors' credit assessment capability on the borrower. If the accuracy and efficiency of individual investors' credit assessments are significantly inferior to those of existing financial systems, the credit risk that could be resolved in existing financial institutions could lead to new costs, offsetting the possibility of greater savings from eliminating brokerage margins. In this case, the increase in the welfare of individual participants, which is the advantage of P2P lending over traditional finance, will no longer be valid. Thus, upgrading individual investors' credit assessment capabilities in P2P lending is an important factor that should be secured, in order for P2P lending to grow as a competitive or alternative means to traditional financing.

In particular, the efficiency of P2P lending should be secured at a high level to make it a competitive alternative financing method, given that traditional financing has long pursued cost optimization based on accumulated data, expertise, and experience. However, unlike traditional financing with a professional credit rating system, it is difficult for P2P lending to maximize the accuracy of credit evaluations, given that most investors are regular individuals relatively lacking experience and expertise.

At present, individual investors' decision-making in P2P lending markets and the assessment of their quality has been a matter of both practical and theoretical interest. The investors' decision-making is not that of individual investors, but an aggregated outcome of investor groups. That is, it means collective decision-making. In the P2P lending market, each individual makes a bidding decision, but the actual availability of the loan is determined by the collective results of these individual decisions (Freedman & Jin, 2017; Yum, Lee, & Chae, 2012). Therefore, it is important to focus on the collective decision-making of investor groups, not individual investors' decision-making, when evaluating funding decisions in P2P lending. Meanwhile, several studies have shown that individual investors in the P2P lending markets still lack the ability to evaluate (Tao, Dong, & Lin, 2017; Zhu, 2018). The lack of collective credit evaluation capabilities of investors can be a serious obstacle for P2P lending to grow as an alternative means of financing. However, its potential as an alternative financial instrument could be high if the collective decision-making of investors were to improve as the market gradually matures in the future. Therefore, it is a practical and academically important issue to ensure that investors' collective decisions have room for improvement as market structures evolve over time. In this respect, the study is the first to our knowledge to study this issue.

To do this, it is necessary to first identify the level of collective decision-making by investors. However, so far, while studies related to P2P lending have raised the issue of individual investors' lack of credit evaluation capabilities, there has been no further study on by how much they fall short. Identifying the skill level of a group of inexperienced, non-professional investors to assess a borrower's creditworthiness is important in assessing whether the P2P lending can serve as a competitive alternative market. In this respect, the study seeks to compare the *ex-ante* actual credit evaluation capabilities of investors compared to *ex-post* optimal capabilities. Therefore, this study first presents the following research question.

Research question 1: How different is the predictive power on a borrower's default of the individual investors' ex-ante funding decision model from that of the ex-post repayment model?

After identifying suboptimal levels of investors' ability to make collective funding decisions, this study's primary research purpose is to determine whether the quality of the decision can be improved as the market structure evolves over time. Therefore, we present a second research question as follows.

Research question 2: Do investors' predictive powers improve as market structures evolve over time?

In addition, it will also be important to identify what the specific improvement factors are if investors' collective decision-making evolves. To this end, the study also seeks to further clarify what investor information is improving the judgment among the borrowers' various information. The study also presents the following research questions in relation to it.

Research question 3: If investors' predictive power evolves, what is the reason for it?

While the investors' *ex-ante* funding decision primarily determines whether the funding is approved or not, the fundamental criterion for investors' decision-making is whether or not the borrower will repay the money loaned on time (Duarte, Siegel, & Young, 2012; Greiner & Wang, 2010; Kgoroeadira, Burke, & Stel, 2019; Yoon, Li, & Feng, 2019). That is, the borrower's creditworthiness plays a decisive role in the investor's *ex-ante* funding decision-making. Although the borrower's *ex-post* repayment performance may not necessarily be consistent with his/her creditworthiness as it is determined by a combination of various factors, it could still be the best proxy. Therefore, the closer the proximity of the borrower's creditworthiness, evaluated in advance by investors at the funding stage, to the borrower's *ex-post* repayment result, the higher the predictive power over the borrower's creditworthiness. From this perspective, the better the investors' *ex-ante* decision-making, which predicts the borrower's ex-post repayment performance, the more the investors' credit assessment capabilities evolve. Therefore, it would be important to investigate how well investors' *ex-ante* funding decisions predict the resultant *ex-post* repayment performance. In that respect, the study first seeks to identify the difference between the predictive power of investors' *ex-ante* funding decision model and that of the *ex-post* repayment model on a borrower's actual repayment result. Additionally, as mentioned earlier, in order for P2P lending to firmly establish itself as an alternative financing method that can compete with traditional financing, the investors' credit assessment capabilities need to be enhanced. To this end, this study tries to confirm the potential for the evolution of investors' credit assessment capability, and identify the cause of such evolution, if it is confirmed.

This study theoretically contributes to prior studies by identifying the evolution of investors' *ex-ante* collective decision-making and its detailed rationale for the first time. In a practical perspective, this study emphasizes the platform operators' role in developing the market through providing accurate information and maintaining dynamic communication with investors.

2. Review Of Related Literature

In Figure 1, I have schematized and classified existing studies on lending markets into subtopics to identify factors affecting a borrower's repayment performance or investors' funding decision-making and developing predictive models on repayment performance. Traditionally, identifying factors (i.e., *ex-post* determinants) affecting the borrower's repayment performance in the loan market is an important research topic. First, studies on repayment performance determinants have been conducted mainly on the traditional loan markets such as banks and credit cards. These studies have identified various factors affecting a borrower's loan default, including demographic information such as gender, age, education, credit-related information such as debt-income ratio, annual income, credit score, and loan-related characteristics such as loan amount, interest rate, loan duration (Chiang, Chow, & Liu, 2002; Crook & Banasik, 2004; Roberts & Sepulveda M, 1999; Steenackers & Goovaerts, 1989). Recently, this research has also been actively performed on P2P lending. Some studies have identified factors on repayment performance in the post-loan process and discovered new information through comparative analysis with the *ex-ante* determinants on investors' decision-making. Among these, several have compared and analyzed the effects of the diverse characteristics of the borrower (Hu, Liu, He, & Ma, 2019; Tao et al., 2017), others highlighted particular features including gender (D. Chen, Li, & Lai, 2017; X. Chen, Huang, & Ye, 2020), historical loan performance (Ding, Huanga, Li, & Meng, 2019), appearance (Duarte et al., 2012), social network (Freedman & Jin, 2017), university reputation (J. Li & Hu, 2019), residential district (J. Jiang, Liu, & Lu, 2019), friendship (Lin, Prabhala, & Viswanathan, 2013), race (Pope & Sydnor, 2011), platform guarantee mechanism (Shi, Jin, & He, 2019; Zhu, 2018), lender-borrower communication (Xu & Chau, 2018), and loan pricing structure (Wei & Lin, 2017).

Some studies further identify *ex-post* determinants and suggest predictive models of repayment performance for conventional loan markets. These studies use traditional loan market data to test the predictive accuracy of existing credit rating models (Desai, Crook, & Overstreet Jr., 1996; Lee, Chiu, Chou, & Lu, 2006), to create loan assessment models (Lee, Chiu, Lu, & Chen, 2002; Malhotra & Malhotra, 2002), or to develop new models that predict the credit risk of individual borrowers (Noh, Roh, & Han, 2005). Recently, published studies have also presented models that predict repayment performance with *ex-post* determinants on P2P lending (Guo, Zhou, Luo, Liu, & Xiong, 2016; C. Jiang, Wang, Wang, & Ding, 2018; Z. Li, Li, Yao, & Wen, 2019; Ma, Zhao, Zhou, & Liu, 2018; Malekipirbazari & Aksakalli, 2015; Serrano-Cinca, Gutiérrez-Nieto, & López-Palacios, 2015).

Few studies have been conducted on factors that affect investors' decision-making at the loan funding stage for conventional loan markets. The investors are generally single financial institutions, such as banks and credit card companies, not crowds. The practical need for such research was not significant because financial institutions could make their loan decision-making using the credit rating model based on their accumulated loan transaction data. However, unlike the conventional loan market, the emerging P2P loan market has resulted in many investors participating. The platform operator retains the borrower's information, weakening the investor's access to it. In this regard, studies on factors affecting loan funding have been actively carried out in recent years.

The most actively studied P2P lending-related subject in academic research is the individual investors' decision-making in the loan funding phase. Many scholars have studied factors (i.e., *ex-ante* determinants) that affect investors' decision-making in various ways. Several researchers have focused on determining the impact of the overall factors (Cai, Lin, Xu, & Fu, 2016; X.-h. Chen, Jin, Zhang, & Yang, 2016; Feng, Fan, & Yoon, 2015; Kgoroeadira et al., 2019). Others have focused on specific factors such as linguistic styles (Herzenstein, Sonenshein, & Dholakia, 2011; Kim, Maeng, & Cho, 2018; Larrimore, Jiang, Larrimore, Markowitz, & Gorski, 2011), gender (Barasinska & Schafer, 2014), cultural and geographical differences (Burtch, Ghose, & Wattal, 2014), trust in borrowers or intermediaries (D. Chen et al., 2014; Wan, Chen, & Shi, 2016), punctuation used in loan descriptions (X. Chen, Huang, & Ye, 2018), appearance (Jin, Shang, & Ma, 2019), voluntary information (Han, Chen, Liu, Luo, & Fan, 2018; Michels, 2012), friendship (Liu, Brass, Lu, & Chen, 2015), and herding behavior among investors (Herzenstein, Dholakia, & Andrews, 2011; Y. Jiang, Ho, Yan, & Tan, 2018; Kim, 2020; Luo & Lin, 2013; Yu, Dan, Ma, & Jin, 2018). The related literature is summarized in Table 1.

Previous studies related to predictions in P2P lending markets have focused on presenting an *ex-post* model that can be called an optimal decision model. This is important because it could serve an optimal decision model for ordinary individual investors who lack experience and expertise in credit evaluation, presenting an instruction that can be referenced in future funding decisions. However, it is no less important to assess how accurately the *ex-ante* model, which is a realistic suboptimal decision model for investor groups, predicts the actual *ex-post* repayment results. This is because it can give the investors an opportunity to access to what degree their collective decision-making capability is inferior to the optimal case. This study includes all three aforementioned topics, and further attempts to present a predictive model based on *ex-ante* determinants, and compare and analyze how much this predictive model predicts repayment performance over *ex-post* predictive models. This study is intended to contribute to existing studies as it attempts to determine whether investors' *ex-ante* funding decision model evolves over time and, if so, identify the reasons why.

3. Theoretical Background And Hypothesis Development

3.1. Bounded Rationality

According to the findings of the previous studies, there is a gap between the factors investors consider in their funding decision-making and those that affect actual repayment results of borrowers to some extent. X. Chen et al. (2020) revealed that while female borrowers outperform male borrowers in repayment performance in the Chinese P2P lending market, investors hardly reflect this difference in their funding decisions. Additionally, Freedman and Jin (2017) argued that investors do not fully understand the relationship between the borrower's social ties and his/her creditworthiness and found that although the repayment performance of the borrower with social ties is inferior to that of the borrower without ties, investors tend to prefer them more in their funding decisions. These differences can be explained by the theory of bounded rationality. In P2P lending, as the platform operator has the authority to determine the scope of the information disclosed, investors must make funding decisions with limited accessibility to borrowers' information. Therefore, although they try to make decisions with the best sense of rationality, the decisions are likely to be inferior to the predictive power of the model based on the *ex-post* repayment performance. Based on such logical development, this study presents the following hypothesis to answer research question 1.

Hypothesis 1: The predictive power of the ex-ante funding model for the borrower's actual repayment performance will be inferior to that of the ex-post repayment model.

3.2. Law of Effect

Bounded rationality-based investors' funding decisions can be improved through the process of cumulative trial and error experience. Particularly, psychologist Thorndike has presented the law of effect, one of the most representative principles of learning (Thorndike, 1913). According to the law, people engage in behaviors that have desirable consequences and avoid behaviors that result in undesirable consequences. Therefore, the important outcome of a behavior is the information it provides about behavioral consequences. That is, if specific decision-making produces a satisfying result in a particular situation, and it becomes more likely to recur in the future, decision-makings that result in unpleasant outcomes are less likely to recur. Based on the law of effect to P2P lending, investors' funding decisions can be changed to increase the predictive power on borrowers' repayment performance by using consistent repayment results to modify each predictor's judgment. Based on this logical flow, this study suggests the hypothesis of research question 2 as follows.

Hypothesis 2: The predictive power of investors' funding decision-making on the borrower's repayment performance will improve over time.

Borrower information available to investors on the P2P lending platform can be classified into three types: platform-generated, loan-describing, and borrower-attributed. In previous studies, borrower information was divided into two types of loan-describing and borrower-attributed (Herzenstein, Andrews, Dholakia, & Lyandres, 2008). However, unlike other platforms, in Moneyauction and Popfunding, the credit factors, which were classified as borrower-attributed information in previous studies, are generated internally and not by being provided from external credit agencies. In this regard, they are classified separately as the platform-generated information in this study. Among them, for the platform-generated information, which are the internally-evaluated credit factors, and the loan-describing information covering loan amount, interest rate, duration, and others, investors' judgment can be straightforward. This is because the information's empirical or logical relationships with a borrower's repayment performance are relatively well established. The higher the credit score, the smaller the loan amount, the lower the loan interest rate, and the shorter the loan duration, the better the borrower's repayment performance. However, investors' judgment of borrower-attributed information is unclear compared to the two types of information discussed. In other words, the relationship between the borrower-attributed information, such as gender, age, marital status, region, and occupation with his/her repayment performance is often not statistically significant or varies depending on the circumstances. Given the nature of this information, part of the greater influence of the law of effect on investors' decision-making may be borrower-attributed information rather than platform-generated or loan-describing information. The cumulative trial and error experience is likely to show more tangibly decisional improvement on the borrower-attributed information, which is relatively unclear in the tendency of judgment. Therefore, this study presents the following hypothesis regarding research question 3.

Hypothesis 3-1: The contribution to the predictive power of borrower-attributed information will improve over time.

Hypothesis 3-2: The predictive power based on borrower-attributed information improves more than the predictive power based on platform-generated or loan describing information.

4. Data And Methodology

4.1. Data

The data used in the study is extracted from Korea's oldest market-leading platforms, Moneyauction and Popfunding, through web-scrapping methods. Since the two platforms dominated the Korean P2P lending market for a considerable period of time, the transactions on these platforms represent the market situation and is also superior in terms of data availability compared to other platforms. Therefore, the study uses data from the two platforms for the analysis. Depending on the type of analysis, dependent variables include dichotomous variables such as funding and repayment results. The dependent variable *Funding*, which represents funding status, is set to 1 if 100% of the funding has been achieved and 0 otherwise. Similarly, the dependent variable *Repayment*, which represents repayment status, is 1 when the borrower has repaid 100% of the principal and interest due, and 0 otherwise. Explanatory variables include platform-generated information (credit score, credit rating, repayment quality), loan-describing information (amount, interest rate, period, and etc.), and borrower-attributed information (gender, age, marital status and etc.). As different types and numbers of available information vary by platform, we try to clarify the structure of research datasets. For each platform, the variables used for analysis and their descriptions are listed in Table 2.

Moneyauction data used in the analysis consists of 34,961 funding data and 4,384 repayment data as of the end of November 2016. For each analysis, the above funding data were reclassified into three categories: full period, former period, and latter period. The reason for subsetting the full period into the former and latter periods is to compare the change in predictive power over time and to investigate whether the collective decision-making of investors improves with changes in market structure. According to the primary objective of this study, the analysis using the full period data was named Model I. The full period data was divided into two subsequent datasets: before and after 2011, which was roughly the median year based on the dataset. This is because 2011 was a time when there was a big change in the Korean P2P lending market. Previously, Moneyauction had a monopoly in the market, but it later changed to the more competitive structure of an oligopolistic system with the entry of a second player, namely Popfunding. This market change created a more diverse pool of investors than before. Therefore, this is likely to be an appropriate reference point to investigate whether collective decision-making by investor groups evolves over time. The related analyses were named Models I-1 and I-2. Accordingly, there are 34,961 data used in Model I, and 16,178 and 18,783 data used in Model I-1 and Model I-2, respectively. For the repayment data, the number of data used in Model II, which is the analysis of the full period, is 4,384. For each analysis, the descriptive statistics of the data are shown in Table 3 below.

Popfunding data used in the analysis was taken from January 2011 until the end of December 2019, with 10,794 funding cases and 1,186 repayment cases. The timeframe for analysis based on time changes was around 2014. In 2014, the Korean P2P lending market began to change significantly once again as a number of new second-generation platforms entered the market following the entry of Moneyauction and Popfunding. As platform diversity increased and the market grew, the pool of investors began to expand in earnest. Thus, similar to 2011, it is considered an appropriate reference point to determine whether collective decisions by investor groups evolve over time by comparing the time points before and after 2011. Similar to Moneyauction, the funding data was classified into three types: full period, former period, and latter period. Model III, the analysis using full period data, contains 10,794 cases. Model III-1 containing the former data, and Model III-2 containing the latter data, divided according to the year 2014, contain 7,671 and 3,123 cases, respectively. Model IV, the analysis of repayment, used a total of 1,186 cases. For each analysis, the descriptive statistics of the data are described in Table 4.

4.2. Methodology

This study first adopts a logit equation using stepwise logistic regressions to establish the optimal model for each analysis described. The structure of the empirical equation is as follows:

$$\Pr(Y_i = 1) = \beta_0 + \sum_{i=1}^{i=n} \beta_{1i} X_i + \sum_{i=1}^{i=n} \beta_{2i} C_i + \varepsilon_i \quad (1)$$

Here the dependent variable Y_i is a binary variable equal to 1 if the loans are successfully funded (in case of funding) or the loans are successfully repaid (in case of repayment) and 0 otherwise. The explanatory variables are represented as X_i, with their details described in Table 2. The time intervals of the data used in this study are just under 10 years: from 2007 to 2016 for Moneyauction and from 2011 to 2019 for Popfunding. Therefore, the dependent variables, funding and repayment results, are likely to be affected by macro factors in addition to the explanatory variables. In addition, consideration of the time element is important because the primary purpose of this study is to investigate whether the collective decision-making of investors improves as the market structure evolves over time. Therefore, it should not be excluded from the possibility that macro factors as well as the characteristic elements of a borrower will affect the change in collective decision-making over time. To this end, as in several prior studies (Freedman & Jin, 2017; Yoon et al., 2019), macro factors such as the unemployment rate, banks' average interest rate for unsecured loan, stock market index, and exchange rate were added as control variables. ε_i is the random disturbance term. Eq. (1) is estimated using logistic regression, widely employed to predict binary dependent variables and constructed via the logit transformation (Jaeger, 2008). For each model applied to this study, the composition of the variables is as shown in Table 5.

Stepwise selection with backward elimination of predictors from the full predictor model was applied to find each model's most suitable specification. That is, from a full model including all explanatory variables (i.e., 18 variables for Moneyauction and 21 variables for Popfunding), every variable with a p-value higher than 0.1 was excluded starting with the variable with the highest p-value. The stepwise selection method effectively develops predictive functions because when other independent variables exist in the regression equation, only variables that influence the dependent variables are included in the equation (Lemon, Roy, Clark, Friedmann, & Rakowski, 2003). The model with the lowest AIC (Akaike Information Criterion) value was selected (Christensen, 2006). Next, we adopted relative weight analyses to determine the relative importance of explanatory variables for each model (Tonidandel & LeBreton, 2015; Tonidandel, LeBreton, & Johnson, 2009). A cross validation method was then used to determine how much the predictability for a borrower's actual repayment result of the *ex-ante* model, the investors' funding decision model (Models I and III, for Moneyauction and Popfunding, respectively), differed from that of the *ex-post* repayment model (Models II and IV, for Moneyauction and Popfunding, respectively). For the *ex-ante* models, the predictive model was developed using the funding data, and the consequent prediction was made using the repayment data. For each model, the resulting estimates were classified into one of the dichotomous predicted results of normal or default, based on a cut-off value of 0.5, creating a confusion matrix comparing actual and predicted repayment results. The prediction accuracy was then calculated in average proportion of corrected predictions of all estimates using the values in the confusion matrix.

A total of ten predictions were made for each model, with first the funding data divided into five, then each subgroup again divided into two groups, one being used as a training data to set the model, and the other being used to validate the optimal cut-off value of the prediction. The actual repayment data was then used to make predictions to obtain the results. The above process was repeated once more for each subgroup by switching the training data and the validation data to make an alternative prediction. This created two predictions for each of the five subgroups, and ten were made overall. This is a variant of the 5x2 cross validation presented by Dietterich (1998). For the *ex-post* model, the prediction was made by applying the original Dieterich's 5x2 cross validation method. After dividing the repayment data into five subgroups, one subgroup was designated as test data for actual prediction and the other four subgroups were each divided into two again. Of the two further subgroups, one was used as training data for model setup and the other as a validation data for finding the optimal cut-off value for the prediction. Then, the training and validation data were switched to make another prediction for the same test data. These five changes resulted in two predictions for each dataset, resulting in ten predictions. The above cross validation method is schematic in Figure 2.

To ensure that the resulting predictions from the *ex-ante* predictive models and *ex-post* models for each platform showed significant differences, a paired *t*-test and a Wilcoxon signed rank test (Wilcoxon, Katti, & Wilcox, 1970) were performed. For both tests, we set the same null hypothesis (H_0) as the mean of the two sample groups. In addition, a Shapiro-Wilk normality test (H_0: normality exists) was performed to verify the normality of the distribution of prediction accuracy values (Shapiro & Wilk, 1965).

Next, in the analyses comparing the prediction accuracies with time changes, to determine whether investors' funding decision is evolving for each platform, the predictive accuracies of the *ex-ante* model divided on a time basis (i.e., Models I-1, I-2 for Moneyauciton and Models III-1, III-2 for Popfunding) were calculated and compared using the modified Dieterich 5x2 cross validation method.

Finally, to find the causes of possible differences in the prediction accuracies of *ex-ante* models in each platform, this study compared the differences in prediction accuracy and which type of explanatory variables is considered or not. For this, the types of explanatory variables are divided into three: platform-generated, loan-describing, and borrower-attributed. The platform-generated variables refer to information that the platform operator has evaluated in order to assess the borrower's creditworthiness on his own, including the *Score* variable for Moneyauction, and the *Rating* and *Quality* variables for Popfunding. The loan-describing variables refer to the information provided by the borrower during the load application process, including the *Amount*, *Rate*, *Duration*, *Purpose*, and *Text*. Other details about the borrower are grouped into the borrower-attributed variables.

In the case of Moneyauction, models without the platform-generated variables are Models I-1-1 and I-2-1, depending on time changes, and models without the loan-describing variables are Models I-1-2 and I-2-2. Models without the borrower-attributed variables are Models I-1-3 and I-2-3. Similarly, for Popfunding, models without the platform-generating variables are Models III-1-1 and III-2-1, and models without the loan-describing variables are Models III-1-2 and III-2-2. Models without the borrower-attributed variables are Models III-1-3 and III-2-3. All analyses were performed using the statistical package R (version 3.6.3).

5. Results

5.1. How different is the individual investors' predictive power on a borrower's default ex-ante funding decision model from that of the ex-post repayment model?

Table 6 shows the optimal *ex-ante* and *ex-post* models for each platform, selected through the stepwise logistic regression method. First, the AUC (Area Under the Curve) values, which proxy in-sample classification accuracies for Models I and II, are 0.837 and 0.713, respectively. This can be generally assessed as moderate and good classification power, respectively (Metz, 1978; Sweets, 1988). Model I distinguishes investors' funding decisions well with the AUC value of 0.837, and Model II distinguishes the borrower's repayment results to a moderate level with the AUC value of 0.713. However, a closer look at each predictor shows a difference in its effect between the investors' funding decision and the borrower's repayment performance. For the *Score* variable, the result of the self-assessment of the borrowers' creditworthiness by the Moneyauction operator and its effect on the investors' *ex-ante* funding decision is found to be statistically significant, and has high relative importance of 27.5%. However, it has not had a significant impact on the *ex-post* repayment result. While the collective relative importance of loan-describing variables (*Amount*, *Rate*, *Duration*, *Purpose*, and *Text*) is 56.1% for the investors' *ex-ante* funding decision, it is just 34.5% for *ex-post* repayment performance. In particular, *rate* is the variable found to affect the *ex-ante* funding decision and *ex-post* repayment performance in opposite ways. This means that while investors prefer higher-interest loans in funding decision, lower-interest loans actually show better repayment performance. In addition, among the borrower-attributed variables, the *Residence* and *Failure*, also have opposite effects on the *ex-ante* funding decision and *ex-post* repayment performance.

Next, as seen in Models III and IV, for Popfunding, the classification accuracy of each model is higher than Moneyauction. The AUC value for Model III is 0.945, rated excellently by general criteria, and the AUC value for Model IV is also good at 0.863. However, in the case of Popfunding, there is also a difference in the explanatory variables that make up the *ex-ante* model and *ex-post* model, as in Moneyauction. Especially, for the borrower-attributed variables, far fewer factors affect the *ex-post* repayment performance than the investors' *ex-ante* decision-making. Given that there have been similar findings in several existing studies(Hu et al., 2019; J. Jiang et al., 2019), this study confirms again that there are differences between factors considered by investors at the *ex-ante* funding decision stage and factors affecting the actual *ex-post* repayment results.

The differences between the *ex-ante* and *ex-post* models make it possible to speculate that investors may not be able to properly predict the borrower's repayment performance at the funding decision stage. To confirm this, the results of an out-of-sample prediction with cross-verification of the method described in Figure 2 are provided in Table 7.

The Shapiro-Wilk test confirms that the distribution of both sample groups is normal. Accordingly, a pair of *t*-tests is conducted to check if the difference in the mean between the two groups is significant, resulting in the prediction accuracy of Models I and II being different at a 95% confidence level. However, Wilcoxon signed rank test results show no difference at the 95% confidence level. Even Models III and IV are found to have no difference in prediction accuracy at a 95% confidence level. Therefore, the out-of-sample predictive power of the *ex-ante* and *ex-post* models for each platform is found to have minimal differences. This is a rather surprising result considering that the previous analyses show the differences in the composition of the explanatory variables and their relative importance to each model. In conclusion, although the two models differ in the composition of the explanatory variables and their relative importance, their accuracy of predicting repayment performance is found not to have significant differences. Hypothesis 1 is therefore partially supported as summarized in Table 8.

5.2. Does the investors' predictive power evolve over time?

Table 9 shows the changes seen in the *ex-ante* model with over time for each platform. The in-sample classification accuracies for Moneyacution are considered good for Models I-1 and I-2, with AUC values of 0.803 and 0.861, respectively; in contrast, Popfunding's is considered excellent, at AUC values at 0.961 and 0.918, for Models III-1 and III-2, respectively. When comparing Models I-1 and I-2, there are variables in which the relative importance or the directional orientation affecting the investor's decision making has changed significantly. In addition, the composition of the variables that have a significant impact on the investors' decision-making is also different. Specifically, the relative importance of *Amount* increases significantly over time, while that of *Score*, *Rate*, *Text*, *Success*, and *Insurance* decrease. The variables *Purpose*, *Residence*, and *Work years* have changed the direction of the impact on funding success. For *Purpose*, the probability of funding success for repayment purposes is higher than other purposes as seen in Model I-1, but loan applications for the purpose of repayment have a rather lower probability of funding success as seen in Model I-2. For *Residence*, the funding success probability of the borrower of one's own house is higher in Model I-1, while in Model I-2, the funding success probability of the borrower of a rental house is higher. In the case of the *Work years*, the shorter the working period, the higher the probability of funding success a borrower has as shown in Model I-1; in contrast to the results in Model I-2, which shows that the longer the working period, the higher the probability of funding success. In addition, the composition of the explanatory variables that are significant to the models has also changed. In Model I-1, while the *Cohabitants* variable is significant, and the *Marriage* variable is not, the exact opposite is found in Model I-2.

Similar changes are identified in Popfunding. In Model III-2, compared to Model III-1, the relative importance of the variables *Quality* and *Past loans* increase, while that of *Rate* and *Income* decrease. For *Quality*, even the direction of the effect on the dependent variable, goes opposite ways in the two models. It is found in Model III-1 that borrowers with poorer past repayment performance have a higher funding probability, while in Model III-2, borrowers with better past repayment performance have a higher funding probability instead. In the case of *Income*, the funding probability is higher for those with lower incomes in Model III-1; however, it is the opposite in Model III-2. There is also a difference in the composition of significant explanatory variables. In Model III-1, *Region*, *Resistance*, and *Office* are significant variables, but in Model III-2, these variables become less significant and instead the *Gender*, *Cohabitants*, *Home*, and *Credit events* are the more significant variables.

As time goes by, this study identifies differences in the relative importance and directional effect of variables, and the type of variables that make up the *ex-ante* models for each platform. If that is the case, will these change in a way that increases the out-of-sample predictive power of the model? To answer this, the results of the comparative analyses of predictive power are given in Table 10.

Since the Shapiro-Wilk test confirms that all sample groups exhibit normal distribution, the difference in prediction accuracy between each group is determined by through a paired *t*-test. Even if the normal distribution assumption is not satisfied, the Wilcoxon signed rank test also shows similar results to the paired *t*-test. The prediction accuracies of Models I-1 and I-2 are different at a 99.9% confidence level, although both are likely to be poor levels. Thus, the prediction accuracy of Model I-2 is shown to be 1.2%p higher than that of Model I-1. The prediction accuracies of Models III-1 and III-2 are moderate and good levels, and it they are also different at a 99.9% confidence level. The predictive accuracy of Model III-2 is found to be 6.2%p higher than that of Model III-1. These confirm that, as outlined in Table 11, the predictive power of investors' *ex-ante* funding decision model for both platforms evolves over time. In conclusion, Hypothesis 2 has been verified.

5.3. If the investors' predictive power does evolve, what is the reason for it?

Table 12 shows the differences in predictive power when explanatory variables by type (platform-generated, loan-describing, and, borrower-attributed) are included and excluded from the Moneyauction model. Model I-1 and its derivatives (Models I-1-1, I-1-2, and I-1-3) are for the *ex-ante* model in former period, and Model I-2 and its derivatives (Models I-2-1, I-2-2, and I-2-3) are for the *ex-ante* model in the latter period. First, results show that in the former period, the predictive power is significantly lower when the platform-generated variables are included (i.e., Model I-1), than when they are excluded (i.e., Model I-1-1). On the other hand, for the latter period, the predictive power of the model when the platform-generated variables are included (i.e., Model I-2) is significantly higher compared to when they are included (i.e., Model I-2-1). This means that in the early days, the *Score*, a platform-generated variable, contributed to improving the investors' *ex-ante* predictive power, but over time, it is undermining it. Next, it is shown in both the former and latter periods that even though loan-describing variables are relatively the most important to the investors' *ex-ante* model, the predictive accuracy of the *ex-ante* model is relatively intact whether that type of variable exists or not, compared to the cases of other types of variables. However, both periods show that the predictive accuracy drops by 3.3p% and 4.2%p, for the former period and the latter period, respectively, when borrower-attributed variables are omitted from the model, resulting in a relatively significant decrease in investors' *ex-ante* predictive power. Therefore, the borrower-attributed variables have the greatest impact on the investors' *ex-ante* predictive power over the other two types of variables, particularly in latter period. In conclusion, it is possible that the reason why the investors' prediction accuracy has improved over time on Moneyauction is due to more accurate judgments made about the information.

Table 13 shows the effects of the variables by type on the investors' *ex-ante* predictive power in Popfunding. For platform-generated and borrower-attributed variables in the former period, the predictive power increases by 1.5%p and 2.3%p respectively, when the variables are excluded (as shown on Models III-1-1 and III-1-3 respectively), compared to when they are included (Model III-1). This means that in the former period, investors are inaccurately assessing the effects of the platform-generated and borrower-attributed variables. On the other hand, in the latter period, when the variables are excluded, as shown on Models III-2-1 and III-2-3, the predictive power is lower by 6.8%p and 0.8%p, respectively, compared to when they are not excluded (Model III-2). Thus, over time, it can be seen that investors are more accurately considering the platform-generated and borrower-attributed information in making *ex-ante* funding decisions.

In the latter period, the prediction accuracies are found not to have significant differences when the loan-describing variables are considered (Model III-2) and when they are not (Model III-2-2). Therefore, it can be said that the loan-describing information do not significantly contribute to improving the investors' *ex-ante* predictive power. In conclusion, it is possible that the reason why the investors' *ex-ante* decision-making increases the predictability of repayment performance over time in Popfunding, is due to the investors being able to more accurately evaluate the platform-generated and borrower-attributed information. The investors' proper understanding of the impact of the information provided by the platform operator about a borrower's creditworthiness, highly contributes to the improved prediction accuracy. Therefore, Hypothesis 3.1 is supported in both cases of Moneyauction and Popfunding. In Hypothesis 3.2, it is supported in Moneyauction, but not in Popfunding. This is because in Popfunding, the improvement of the predictive power by platform-generated information is greater than that by borrower-attributed information.

Table 14 summarizes the detailed rationale that improves the predictive power of investors' *ex-ante* decision on a borrower's repayment performance over time in Moneyauction and Popfunding.

6. Discussion

In summary, the results show that there is no significant difference between the predictive power of the investors' *ex-ante* funding decision over a borrower's repayment performance, and that of the *ex-post* repayment model. Second, the results show that the predictive power of investors' *ex-ante* funding decision over a borrower's repayment performance improves over time. Finally, although the detailed rationales may vary between platforms, the results show that the reason why investors' predictive power improves over time is mainly the investors accessing more accurately the borrower's information and characteristics provided by the platform operator. It also shows that the effect of loan-describing information on improving investors' predictive power is likely to be relatively limited.

Based on these results, there are several issues to be considered. First, the information provided by the platform operator plays a very important role in the investors' *ex-ante* prediction over the actual repayment results. Moneyauction investors have relatively lower *ex-ante* prediction accuracy than Popfunding investors, most likely because of the accuracy of the information provided by the platform operator. Table 6 shows in detail how investors consider the *Score* variable that is provided by the Moneyauction operator. It has an average relative importance of 27.5% in their *ex-ante* funding decision, even though it does not play a significant role in the *ex-post* model. This means that while the *Score* variable does not significantly affect a borrower's actual repayment performance, investors are likely miscalculating its importance. Similarly, in Table 6, the *Rating* and *Quality* variables which are information provided by the Popfunding operator, are proven to play an important role in the *ex-post* model, with a relative importance of 61.8%. Investors are found to consider this with a relative importance of 42.2% in their *ex-ante* funding decision. This difference is estimated to have significantly affected the investors' *ex-ante* predictive power on both platforms. In Tables 6 and 7, Moneyauction investors have an *ex-ante* predictive power of only 62.0%, a poor level of accuracy, while Popfunding investors are at 76.5%, a relatively good level of accuracy. Therefore, to increase the predictability of investors' *ex-ante* funding decision in a P2P loan market, it very important for platform operators to make accurate credit assessments of borrowers. Consequently, the role of the platform operator's information analysis and delivery needs to be more emphasized in P2P lending.

In addition, the platform operator needs to continuously update its own credit rating system as the empirical loan transaction data gradually accumulates. We can reference Table 12, which shows the results of the Moneyauction analysis. Initially, when considering the information provided by the platform operator, the predictive power of the investors' *ex-ante* funding decision improves. However, when it comes to the latter period, it rather has an adverse effect of worsening the predictive power. It has been argued in previous papers that there may be problems with the accuracy of the platform's self-assessed credit rating model (Tao et al., 2017; Zhu, 2018). On the other hand, Table 13 shows that the case of Popfunding is the opposite of Moneyauction. In the beginning, information provided by the platform operator undermined the investors' *ex-ante* predictive power, but over time it has contributed to significantly improving the investors' predictive power. As such, it has been proven that the effectiveness of the platform's credit rating system will change over time, and it is important for the platform operator to continuously update the initial credit rating model based on accumulated empirical data.

This study confirms the possibility that the prediction accuracy of investors' *ex-ante* funding decision over a borrower's repayment performance can be improved over time. For example, in Table 9, the Moneyauction investors have made adjustments to lower the relative importance of the *Score* variables, information that is provided by the platform, over time. This may be because they gradually realized the inaccuracy of the *Score* variables. In the case of Popfunding, investors initially considered the *Quality* variable, information that is generated by the platform, as opposed to the actual directional effect it had on actual repayment performance. They also considered it being less important. However, they are observed to have improved judgement, identifying the proper direction and recognizing its importance gradually. In the introduction, the researcher pointed out that cost-effectiveness must be achieved for the P2P lending system to be firmly established as an alternative for traditional financing. Thus, to this end, it is most important to improve the investors' accuracy of judgement. From this point of view, as investors' judgment is actually improving, the P2P lending system can gradually become an alternative to traditional financing by further strengthening its theoretical advantages in the near future.

On the contrary, it is also partially observed that the investors' *ex-ante* decision-making is inaccurate for certain types of information. In Table 9, Moneyauction investors tend to prefer rental borrowers over self-owned borrowers over time in *ex-ante* funding decision. This reflects the opposite effect of the *Residence* variable on actual repayment performance. This phenomenon is also seen in the *Purpose* variable. Investors gradually prefer borrowers who apply for loans for other purposes than for repayment of existing loans, but in reality, borrowers with the purpose of repaying existing loans are found to have better repayment performance. It is not much different in Popfunding either. At the borrower's income level, it has been proven that the repayment performance of low-income borrowers is actually better, but over time, investors have shifted their *ex-ante* funding decision to favor high-income borrowers. As such, investors' decisions may change over time to improve the accuracy of their predictions for certain information, while on the contrary, the accuracy of their decisions for other information may deteriorate. In this respect, to improve the quality of the investors’ decision making, it will be important for platform operators make a constant effort to analyze their accumulated empirical data and give feedback and communicate with investors.

7. Conclusion

7.1. Implication

This study theoretically contributes to existing studies by examining the predictive power of investors' *ex-ante* decision-making model over the borrower’s *ex-post* repayment performance model, which has not been covered by P2P lending related studies yet. It is also meaningful in that the evolution of investors’ *ex-ante* decision-making and its detailed rationale are first identified in this paper. This study confirms the possibility of optimizing and streamlining the P2P lending market through the evolution of investors’ decision making, thus giving hope to the sustainable development of the market as an alternative financing method in the near future. In addition, this study has gone a step further from simply confirming that the investors' *ex-ante* decision-making model differs from the actual *ex-post* repayment model. It also presents a rather unexpected result of finding no significant difference between the two models in terms of predictive performance, despite the differences in each of their predictor composition.

In a practical perspective, this study reaffirms the results of several previous studies that show how the platform's credit assessment capabilities may be inaccurate. Through this, the need for platform operators to continually refine and update their information analysis and its delivery to investors is highlighted. It empirically confirms for the first time that if the platform provides inaccurate information, the predictive power of investors' *ex-ante* decision-making could deteriorate significantly. Another contribution of this study is that, based on empirical evidence, it emphasizes the platform operators’ role in developing the market through providing accurate information and maintaining dynamic communication with investors.

7.2. Limitations

This study analyzes how investors' decision-making evolves by classifying the explanatory variables by type of information (platform-generated, loan-describing, and borrower-attributed information). However, as there may certain limitations in analyzing the detailed cause of the evolutionary change, this topic deserves to the subject of future research.

In addition, it seems necessary to consider not only the micro factors covered in this study but also the macro factors that cause the overall P2P lending market change. In Korea, the P2P lending market first emerged in 2007, not too late compared to the UK, where the market was first developed. Since then, it had plateaued for some time, but market conditions have changed rapidly with the emergence of many platforms around 2014. In fact, the two platforms, Moneyauction and Popfunding, which are the subjects of this study, also showed significant changes in transaction volume and number of loan listings around that time. In this regard, macro factors may have also affected the evolution of investors’ *ex-ante* funding decision. There is a need for further research on these factors to gain a more thorough understanding on the evolutionary change of investors’ decision-making over time.

This study proves that investors' decision-making evolves. However, the results of this study alone may not indicate whether this evolution is only attributed to the improved capabilities of existing investors, or because investors with inferior capabilities are kicked out of the platform and replaced with new investors with superior capabilities. Therefore, further studies that take into account the characteristics of investors, such as their investment experience, could make the outcomes of this study more valuable.

Additionally, all cases except 100%-funded listings were uniformly regarded as unfunded, even if the bidding rates varied from 0% to 100%. Although we treated the dependent variables as two-dimensional categorical variables owing to the data limitation, we would have had more detailed results by identifying them as continuous variables.

Declarations

Conflicts of interests: On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Barasinska, N., & Schafer, D. (2014). Is crowdfunding different? Evidence on the relation between gender and funding success from a German peer-to-peer lending platform. *German Economic Review, 15*(4), 436-452.

Burtch, G., Ghose, A., & Wattal, S. (2014). Cultural differences and geography as determinants of online prosocial lending *Mis Quarterly, 38*(3), 773-794.

Cai, S., Lin, X., Xu, D., & Fu, X. (2016). Judging online peer-to-peer lending behavior: A comparison of first-time and repeated borrowing requests. *Information & Management, 53*(7), 857-867.

Chen, D., Lai, F., & Lin, Z. (2014). A trust model for online peer-to-peer lending: A lender's perspective. *Information Technology and Management, 15*(4), 239-254.

Chen, D., Li, X., & Lai, F. (2017). Gender discrimination in online peer-to-peer credit lending: Evidence from a lending platform in China. *Electronic Commerce Research, 17*(4), 553-583.

Chen, X.-h., Jin, F.-j., Zhang, Q., & Yang, L. (2016). Are investors rational or perceptual in P2P lending? *Information Systems and e-Business Management, 14*(4), 921-944.

Chen, X., Huang, B., & Ye, D. (2018). The role of punctuation in P2P lending: Evidence from China. *Economic Modelling, 68*, 634-643.

Chen, X., Huang, B., & Ye, D. (2020). Gender gap in peer-to-peer lending: Evidence from China. *Journal of Banking & Finance, 112*, 105633.

Chiang, R. C., Chow, Y.-F., & Liu, M. (2002). Residential mortgage lending and borrower risk: The relationship between mortgage spreads and individual characteristics. *Journal of Real Estate Finance and Economics, 25*, 5-32.

Christensen, R. (2006). *Log-linear models and logistic regression* (2nd ed.). New York: Springer Science & Business Media.

Crook, J. N., & Banasik, J. (2004). Does reject inference really improve the performance of application scoring models? *Journal of Banking & Finance, 28*(4), 857-874.

Desai, V. S., Crook, J. N., & Overstreet Jr., G. A. (1996). A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research, 95*(1), 24-37.

Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation, 10*(7), 1895-1923.

Ding, J., Huanga, J., Li, Y., & Meng, M. (2019). Is there an effective reputation mechanism in peer-to-peer lending? Evidence from China. *Finance Research Letters, 30*, 208-215.

Duarte, J., Siegel, S., & Young, L. (2012). Trust and credit: The role of appearance in peer-to-peer lending. *The Review of Financial Studies, 25*(8), 2455-2484.

Feng, Y., Fan, X., & Yoon, Y. (2015). Lenders and borrowers' strategies in online peer-to-peer lending market: An empirical analysis of PPDai.com *Journal of Electronic Commerce Research, 16*(3), 242-260.

Freedman, S., & Jin, G. Z. (2017). The information value of online social networks: Lessons from peer-to-peer lending. *International Journal of Industrial Organization, 51*, 185-222.

Greiner, M. E., & Wang, H. (2010). Building consumer-to-consumer trust in e-finance marketplaces: An empirical analysis. *International Journal of Electronic Commerce, 15*(2), 105-136.

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*(2), 417-426.

Han, J.-T., Chen, Q., Liu, J.-G., Luo, X.-L., & Fan, W. (2018). The persuasion of borrowers' voluntary information in peer to peer lending: An empirical study based on elaboration likelihood model. *Computers in Human Behavior, 78*, 200-214.

Herzenstein, M., Andrews, R. L., Dholakia, U. M., & Lyandres, E. (2008). The democratization of personal consumer loans? Determinants of success in online peer-to-peer loan auctions. *Bulletin of the University of Delaware, 15*(3), 274-277.

Herzenstein, M., Dholakia, U. M., & Andrews, R. L. (2011). Strategic herding behavior in peer-to-peer loan auctions. *Journal of Interactive Marketing, 25*(1), 27-36.

Herzenstein, M., Sonenshein, S., & Dholakia, U. M. (2011). Tell me a good story and I may lend you money: The role of narratives in peer-to-peer lending decisions. *Journal of Marketing Research, 48*(SPL), S138-S149.

Hu, R., Liu, M., He, P., & Ma, Y. (2019). Can investors on P2P lending platforms identify default risk? *International Journal of Electronic Commerce, 23*(1), 63-84.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language, 59*(4), 434-446.

Jiang, C., Wang, Z., Wang, R., & Ding, Y. (2018). Loan default prediction by combining soft information extracted from descriptive text in online peer-to-peer lending. *Annals of Operations Research, 266*(1-2), 511-529.

Jiang, J., Liu, Y.-J., & Lu, R. (2019). Social heterogeneity and local bias in peer-to-peer lending–Evidence from China. *Journal of Comparative Economics*.

Jiang, Y., Ho, Y.-C. C., Yan, X., & Tan, Y. (2018). Investor platform choice: herding, platform attributes, and regulations. *Journal of Management Information Systems, 35*(1), 86-116.

Jin, J., Shang, Q., & Ma, Q. (2019). The role of appearance attractiveness and loan amount in peer-to-peer lending: Evidence from event-related potentials. *Neuroscience letters, 692*, 10-15.

Kgoroeadira, R., Burke, A., & Stel, A. v. (2019). Small business online loan crowdfunding: who gets funded and what determines the rate of interest? *Small Business Economics 52*(1), 67-87.

Kim, D. (2020). The importance of detailed patterns of herding behaviour in a P2P lending market. *Applied Economics Letters, 27*(2), 127-130.

Kim, D., Maeng, K., & Cho, Y. (2018). Study on the determinants of decision-making in peer-to-peer lending in South Korea. *Asia-Pacific Journal of Accounting & Economics*, 1-19.

Larrimore, L., Jiang, L., Larrimore, J., Markowitz, D., & Gorski, S. (2011). Peer to peer lending: The relationship between language features, trustworthiness, and persuasion success. *Journal of Applied Communication Research, 39*(1), 19-37.

Lee, T.-S., Chiu, C.-C., Chou, Y.-C., & Lu, C.-J. (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis, 50*(4), 1113-1130.

Lee, T.-S., Chiu, C.-C., Lu, C.-J., & Chen, I.-F. (2002). Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications, 23*(3), 245-254.

Lemon, S. C., Roy, J., Clark, M. A., Friedmann, P. D., & Rakowski, W. (2003). Classification and regression tree analysis in public health: Methodological review and comparison with logistic regression. *Annals of Behavioral Medicine, 26*(3), 172-181.

Li, J., & Hu, J. (2019). Does university reputation matter? Evidence from peer-to-peer lending. *Finance Research Letters, 31*, 66-77.

Li, Z., Li, K., Yao, X., & Wen, Q. (2019). Predicting prepayment and default risks of unsecured consumer loans in online lending *Emerging Markets Finance and Trade, 55*(1), 118-132.

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.

Liu, D., Brass, D. J., Lu, Y., & Chen, D. (2015). Friendships in online peer-to-peer lending: Pipes, prisms, and relational herding. *Mis Quarterly, 39*(3), 729-742.

Luo, B., & Lin, Z. (2013). A decision tree model for herd behavior and empirical evidence from the online P2P lending market. *Information Systems and e-Business Management, 11*(1), 141-160.

Ma, L., Zhao, X., Zhou, Z., & Liu, Y. (2018). A new aspect on P2P online lending default prediction using meta-level phone usage data in China. *Decision Support Systems, 111*, 60-71.

Malekipirbazari, M., & Aksakalli, V. (2015). Risk assessment in social lending via random forests. *Expert Systems with Applications, 42*(10), 4621-4631.

Malhotra, R., & Malhotra, D. K. (2002). Differentiating between good credits and bad credits using neuro-fuzzy systems. *European Journal of Operational Research, 136*(1), 190–211.

Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in nuclear medicine, 8*(4), 283-298.

Michels, J. (2012). Do unverifiable disclosures matter? Evidence from peer-to-peer lending. *The Accounting Review, 87*(4), 1385-1413.

Noh, H. J., Roh, T. H., & Han, I. (2005). Prognostic personal credit risk model considering censored information. *Expert Systems with Applications, 28*(4), 753-762.

Pope, D. G., & Sydnor, J. R. (2011). What's in a picture? Evidence of discrimination from Prosper. com. *Journal of Human resources, 46*(1), 53-92.

Roberts, J. A., & Sepulveda M, C. J. (1999). Demographics and money attitudes: A test of Yamauchi and Templer's (1982) money attitude scale in Mexico. *Personality and Individual Differences, 27*(1), 19-35.

Serrano-Cinca, C., Gutiérrez-Nieto, B., & López-Palacios, L. (2015). Determinants of default in P2P lending. *PloS one, 10*(10), e0139427.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*, 591-611.

Shi, X., Jin, Q., & He, L. (2019). The bilateral effects of platform-sponsored collateral in Peer-To-Peer (P2P) lending: Evidence from China. *Emerging Markets Finance and Trade*, 1-25.

Steenackers, A., & Goovaerts, M. (1989). A credit scoring model for personal loans. *Insurance: Mathematics & Economics, 8*(1), 31-34.

Sweets, J. F. (1988). Hold that pendulum! Redefining Fascism, collaborationism and resistance in France. *French Historical Studies*, 731-758.

Tao, Q., Dong, Y., & Lin, Z. (2017). Who can get money? Evidence from the Chinese peer-to-peer lending platform. *Information Systems Frontiers, 19*(3), 425-441.

Thorndike, E. L. (1913). *Educational psychology, Vol 1: The original nature of man.*: Teachers College.

Tonidandel, S., & LeBreton, J. M. (2015). RWA web: A free, comprehensive, web-based, and user-friendly tool for relative weight analyses. *Journal of Business and Psychology, 30*(2), 207-216.

Tonidandel, S., LeBreton, J. M., & Johnson, J. W. (2009). Determining the statistical significance of relative weights. *Psychological methods, 14*(4), 387-399.

Wan, Q., Chen, D., & Shi, W. (2016). Online peer-to-peer lending decision making: Model development and testing. *Social Behavior and Personality: an international journal 44*(1), 117-130.

Wang, H., Greiner, M., & Aronson, J. E. (2009). People-to-People lending: The emerging e-commerce transformation of a financial market. In M. L. Nelson, M. J. Shaw, & S. T. J. (Eds.), *Lecture Notes in Business Information Processing. Value Creation in E-Business Management*.

Wei, Z., & Lin, M. (2017). Market mechanisms in online peer-to-peer lending. *Management Science 63*(12), 4236-4257.

Wilcoxon, F., Katti, S. K., & Wilcox, R. A. (1970). Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test. *Selected tables in mathematical statistics, 1*, 171-259.

Xu, J. J., & Chau, M. (2018). Cheap talk? The impact of lender-borrower communication on peer-to-peer lending outcomes. *Journal of Management Information Systems, 35*(1), 53-85.

Yoon, Y., Li, Y., & Feng, Y. (2019). Factors affecting platform default risk in online peer-to-peer (P2P) lending business: an empirical study using Chinese online P2P platform data. *Electronic Commerce Research, 19*(1), 131-158.

Yu, H., Dan, M., Ma, Q., & Jin, J. (2018). They all do it, will you? Event-related potential evidence of herding behavior in online peer-to-peer lending. *Neuroscience letters, 681*, 1-5.

Yum, H., Lee, B., & Chae, M. (2012). From the wisdom of crowds to my own judgment in microfinance through online peer-to-peer lending platforms. *Electronic Commerce Research and Applications, 11*(5), 469-483.

Zhu, Z. (2018). Safety promise, moral hazard and financial supervision: Evidence from peer-to-peer lending. *Finance Research Letters, 27*, 1-5.

Tables

Table 1. Summary of related literature

Category	Literature	Featured determinants or results
Theme 1: Determinants of *ex-post* repayment performance		
Traditional lending	Chiang et al. (2002), Crook and Banasik (2004), Roberts and Sepulveda M (1999), Steenackers and Goovaerts (1989)	Demographic information (e.g., gender, age, education), credit-related information (e.g., debt-income ratio, annual income, credit score), and loan-related characteristics (e.g., loan amount, interest rate, loan duration)
P2P lending	Hu et al. (2019), Tao et al. (2017)	Diverse characteristics of a borrower
	D. Chen et al. (2017), X. Chen et al. (2020)	Gender
	Ding et al. (2019)	Historical loan performance
	Duarte et al. (2012)	Appearance
	Freedman and Jin (2017)	Social network
	J. Li and Hu (2019)	University reputation
	J. Jiang et al. (2019)	Residential district
	Lin et al. (2013)	Friendship
	Pope and Sydnor (2011)	Race
	Shi et al. (2019), Zhu (2018)	Platform guarantee mechanism
	Xu and Chau (2018)	Lender-borrower communication
	Wei and Lin (2017)	Loan pricing structure
Theme 2: Prediction of *ex-post* repayment performance		
Traditional lending	Desai et al. (1996), Lee et al. (2006)	Test the predictive accuracy of existing credit rating models
	Lee et al. (2002), Malhotra and Malhotra (2002)	Create loan assessment models
	Noh et al. (2005)	Develop new models that predict the credit risk of individual borrowers
P2P lending	Guo et al. (2016), C. Jiang et al. (2018), Z. Li et al. (2019) Ma et al. (2018), Malekipirbazari and Aksakalli (2015), Serrano-Cinca et al. (2015)	Present models that predict repayment performance with *ex-post* determinants
Theme 3: Determinants of *ex-ante* funding decision-making		
P2P lending	Cai et al. (2016), X.-h. Chen et al. (2016), Feng et al. (2015), Kgoroeadira et al. (2019)	Overall factors including a borrower's demographic characteristics, loan terms, credit and financial information
	Herzenstein, Sonenshein, et al. (2011), Kim et al. (2018), Larrimore et al. (2011)	Linguistic styles
	Barasinska and Schafer (2014)	Gender
	Burtch et al. (2014)	Cultural and geographical differences
	D. Chen et al. (2014), Wan et al. (2016)	Trust in borrowers or intermediaries
	X. Chen et al. (2018)	Punctuation used in loan descriptions
	Jin et al. (2019)	Appearance
	Han et al. (2018), Michels (2012)	Voluntary information
	Liu et al. (2015)	Friendship
	Herzenstein, Dholakia, et al. (2011), Y. Jiang et al. (2018), Kim (2020), Luo and Lin (2013), Yu et al. (2018)	Herding behavior

Table 2. Variables

Variables		Explanations	Moneyauction	Popfunding
Dependent variables				
Funding		Funding status (1=Success, 0=Failure)	O	O
Repayment		Repayment status (1=Repaid, 0=Default)	O	O
Explanatory variables				
Platform-generated	*Score*	Internal credit score ranging from 0 to 600	O	
	Rating	Number of monthly ordinary repayment		O
	Quality	Repayment quality (2=A, 1=B, 0=C, -1=D)		O
Loan-describing	*Amount*	Loan amount by the ten KRW	O	O
	Rate	Loan interest rate	O	O
	Duration	Loan maturity on a monthly basis	O	O
	Purpose	Loan purpose (1=Refinancing, 0=Others)	O	O
	Text	The number of characters in loan description	O	O
Borrower-attributed	*Gender*	Gender (1=Man, 0=Woman)	O	O
	Age	Age	O	O
	Marriage	Marital status (1=Married, 0=Not married)	O	O
	Region	Residential area (1=Metropolitan, 0=Local)	O	O
	Cohabitants	Number of people living together	O	O
	Residence	Residence condition (1=Own, 0=Rental)	O	O
	Grade	External credit grade ranging from 1 to 10	O	
	Success	Number of previous funding successes	O	
	Failure	Number of previous defaults	O	
	Work years	Number of years a borrower has worked	O	
	Insurance	Social insurance status (1=Insured, 0=Uninsured)	O	
	Mobile	Verification of mobile number (1=Yes, 0=No)		O
	Home	Verification of home number (1=Yes, 0=No)		O
	Office	Verification of office number (1=Yes, 0=No)		O
	Job	Job quality (1=Permanent, 0=Temporary)		O
	Income	Monthly income by the ten KRW		O
	Past loans	Number of previous loans on Popfunding		O
	Credit events	Special credit history (1=Have, 0=Do not have)		O
	Vehicle	Vehicle ownership (1=Have, 0=Do not have)		O
Control variables				
Unemployment rate		Quarterly unemployment rate in Korea	O	O
Bank loan interest rate		Monthly average interest rate of unsecured personal loan of banks in Korea	O	O
KOSFI		Daily composite stock price index in Korea	O	O
FX (USD/KRW)		Daily USD/KRW exchange rate in KRW	O	O

Table 3. Descriptive statistics for Moneyauction

Variables		Model I					Model I-1					Model I-2					Model II				
		N	Max	Min	Mean	S.E.	N	Max	Min	Mean	S.E.	N	Max	Min	Mean	S.E.	N	Max	Min	Mean	S.E.
Dependent variables																					
Funding		34961	1	0	0.258	0.438	16178	1	0	0.198	0.398	18783	1	0	0.311	0.463					
- Success		9030					3197					5833									
- Failure		25931					12981					12950									
Repayment																	4384	1	0	0.654	0.476
- Repaid																	2868				
- Default																	1516				
Explanatory variables																					
Platform-generated	*Score*	34961	485	-105	203.7	0.424	16178	430	-95	179.7	0.627	18783	485	-105	224.4	0.531	4384	485	-75	254.2	1.031
Loan-describing	*Amount (10,000 KRW)*	34961	50000	2	596.2	3.349	16178	2500	100	577.5	3.828	18783	50000	2	612.4	5.287	4384	5000	100	477.8	4.930
	Rate (%)	34961	0.42	0.2	0.284	4.3e-4	16178	0.42	0.02	0.318	0.001	18783	0.36	0.03	0.255	4.9e-4	4384	0.40	0.05	0.304	0.001
	Duration (months)	34961	60	2	20.75	0.038	16178	60	6	21.26	0.066	18783	36	2	20.31	0.041	4384	36	6	20.14	0.102
	Purpose	34961	1	0	0.300	0.002	16178	1	0	0.394	0.004	18783	1	0	0.220	0.003	4384	1	0	0.314	0.007
	- Refinancing	10500					6370					4130					1375				
	- Others	24461					9808					14653					3009				
	Text	34961	15626	0	502.7	2.336	16178	6524	0	514.1	3.544	18783	15626	0	493.0	3.094	4384	5720	0	642.2	7.713
Borrower-attributed	*Gender*	34961	1	0	0.670	0.003	16178	1	0	0.657	0.004	18783	1	0	0.682	0.003	4384	1	0	0.671	0.007
	- Man	23439					10636					12803					2941				
	- Woman	11522					5542					5980					1443				
	Age	34961	82	22	38.18	0.039	16178	82	27	39.24	0.054	18783	74	22	37.27	0.055	4384	68	24	39.13	0.096
	Marriage	34961	1	0	0.371	0.003	16178	1	0	0.348	0.004	18783	1	0	0.390	0.004	4384	1	0	0.391	0.007
	- Married	12960					5626					7334					1713				
	- Not married	22001					10552					11449					2671				
	Region	34961	1	0	0.482	0.003	16178	1	0	0.541	0.004	18783	1	0	0.431	0.004	4384	1	0	0.573	0.007
	- Metropolitan	16842					8746					8096					2512				
	- Local	18119					7432					10687					1872				
	Cohabitants	34961	18	0	2.288	0.009	16178	12	0	2.234	0.013	18783	18	0	2.333	0.012	4384	8	0	2.260	0.023
	Residence	34961	1	0	0.047	0.003	16178	1	0	0.473	0.004	18783	1	0	0.466	0.004	4384	1	0	0.452	0.008
	- Own	16416					7657					8759					1981				
	- Rental	18545					8521					10024					2403				
	Grade	34961	10	1	7.546	0.008	16178	10	1	7.965	0.013	18783	10	1	7.184	0.009	4384	10	1	7.107	0.023
	Success	34961	10	0	0.127	0.003	16178	5	0	0.064	0.003	18783	10	0	0.181	0.005	4384	9	0	0.422	0.014
	Failure	34961	2	0	0.014	0.001	16178	2	0	0.018	0.001	18783	2	0	0.010	0.001	4384	2	0	0.026	0.003
	Work years	34961	40.9	0	2.001	0.019	16178	40.9	0	2.242	0.029	18783	40.1	0	1.794	0.025	4384	40.9	0	2.344	0.055
	Insurance	34961	1	0	0.480	0.003	16178	1	0	0.432	0.004	18783	1	0	0.522	0.004	4384	1	0	0.606	0.007
	- Insured	16792					6984					9808					2655				
	- Uninsured	18169					9194					8975					1729				

Table 4. Descriptive statistics for Popfunding

Variables		Model I					Model I-1					Model I-2					Model II				
		N	Max	Min	Mean	S.E.	N	Max	Min	Mean	S.E.	N	Max	Min	Mean	S.E.	N	Max	Min	Mean	S.E.
Dependent variables																					
Funding		10794	1	0	0.110	0.312	7671	1	0	0.092	0.289	3123	1	0	0.176	0.381					
- Success		1186					783					403									
- Failure		9608					6888					2720									
Repayment																	1186	1	0	0.725	0.446
- Repaid																	861				
- Default																	325				
Explanatory variables																					
Platform-generated	*Rating*	10794	127	0	8.368	0.210	7671	127	0	5.664	0.200	3123	127	0	15.01	0.512	1186	127	0	44.57	0.969
	Quality	10794	2	-2	0.085	0.007	7671	2	-2	0.039	0.007	3123	2	-2	0.198	0.016	1186	2	-2	0.647	0.043
Loan-describing	*Amount (10,000 KRW)*	10794	1000	100	243.6	0.847	7671	1000	100	251.8	0.888	3123	1000	100	223.3	1.907	1186	1000	100	243.6	4.198
	Rate (%)	10794	0.30	0.05	0.274	4.3e-4	7671	0.30	0.05	0.287	4.5e-4	3123	0.25	0.05	0.243	0.001	1186	0.30	0.15	0.278	0.001
	Duration (months)	10794	36	2	16.54	0.059	7671	24	6	16.71	0.069	3123	36	2	16.15	0.109	1186	36	2	14.97	0.177
	Purpose	10794	1	0	0.255	0.004	7671	1	0	0.266	0.005	3123	1	0	0.230	0.008	1186	1	0	0.241	0.012
	- Refinancing	2757					2039					718					286				
	- Others	8037					5632					2405					900				
	Text	10794	5971	0	655.6	5.455	7671	5971	0	648.6	6.405	3123	5155	0	672.6	10.39	1186	5285	0	1061	20.05
Borrower-attributed	*Gender*	10794	1	0	0.551	0.005	7671	1	0	0.541	0.006	3123	1	0	0.577	0.009	1186	1	0	0.508	0.015
	- Man	5951					4148					1803					602				
	- Woman	4843					3523					1320					584				
	Age	10794	84	21	42.52	0.079	7671	77	21	42.72	0.093	3123	84	21	42.02	0.145	1186	76	23	43.69	0.220
	Marriage	10794	1	0	0.494	0.005	7671	1	0	0.516	0.006	3123	1	0	0.441	0.009	1186	1	0	0.540	0.014
	- Married	5337					3961					1376					641				
	- Not married	5457					3710					1747					545				
	Region	10794	1	0	0.537	0.005	7671	1	0	0.545	0.006	3123	1	0	0.518	0.009	1186	1	0	0.521	0.015
	- Metropolitan	5798					4180					1618					618				
	- Local	4996					3491					1505					568				
	Cohabitants	10794	9	0	2.282	0.016	7671	9	0	2.317	0.019	3123	9	0	2.195	0.029	1186	9	0	2.363	0.046
	Residence	10794	1	0	0.266	0.004	7671	1	0	0.261	0.005	3123	1	0	0.278	0.008	1186	1	0	0.235	0.012
	- Own	2871					2002					869					279				
	- Rental	7923					5669					2254					907				
	Mobile	10794	1	0	0.724	0.004	7671	1	0	0.700	0.005	3123	1	0	0.782	0.007	1186	1	0	0.802	0.012
	- Verified	7816					5373					2443					951				
	- Unverified	2978					2298					680					235				
	Home	10794	1	0	0.376	0.005	7671	1	0	0.432	0.006	3123	1	0	0.239	0.008	1186	1	0	0.339	0.014
	- Verified	4058					3312					746					402				
	- Unverified	6736					4359					2377					784				
	Office	10794	1	0	0.698	0.004	7671	1	0	0.749	0.005	3123	1	0	0.571	0.009	1186	1	0	0.682	0.014
	- Verified	7529					5746					1783					809				
	- Unverified	3265					1925					1340					377				
	Job	10794	1	0	0.484	0.005	7671	1	0	0.439	0.006	3123	1	0	0.593	0.009	1186	1	0	0.505	0.015
	- Permanent	5224					3371					1853					599				
	- Temporary	5570					4300					1270					587				
	Income (10,000 KRW)	10794	25000	0	300.0	4.481	7671	25000	0	291.5	5.233	3123	12500	0	320.7	8.627	1186	2000	0	287.8	4.627
	Past loans	10794	13	0	0.630	0.016	7671	12	0	0.438	0.015	3123	13	0	1.102	0.039	1186	13	0	2.988	0.078
	Credit events	10794	1	0	0.432	0.005	7671	1	0	0.426	0.006	3123	1	0	0.448	0.009	1186	1	0	0.605	0.014
	- Have	4666					3267					1399					717				
	- Do not have	6128					4404					1724					469				
	Vehicle	10794	1	0	0.442	0.005	7671	1	0	0.444	0.006	3123	1	0	0.436	0.009	1186	1	0	0.407	0.014
	- Have	4769					3408					1361					483				
	- Do not have	6025					4263					1762					703				

Table 5. Variable composition by models

Models	Dependent variables	Explanatory variables
I, I-1, I-2, III, III-1, III-2	*Funding*	*Platform-generated, Loan-describing, Borrower-attributed*
II, II-1, II-2, IV, IV-1, IV-2	*Repayment*	*Platform-generated, Loan-describing, Borrower-attributed*
I-1-1, I-2-1, III-1-1, III-2-1	*Funding*	*Loan-describing, Borrower-attributed*
I-1-2, I-2-2, III-1-2, III-2-2	*Funding*	*Platform-generated, Borrower-attributed*
I-1-3, I-2-3, III-1-3, III-2-3	*Funding*	*Platform-generated, Loan-describing*

Table 6. Comparisons of *ex-ante* and *ex-post* logistic models

Variables		Model I		Model II		Model III		Model IV	
		β	RW	β	RW	β	RW	β	RW
Platform-generated	*Score*	0.014**	0.275						
	Rating					0.117**	0.367	0.047**	0.255
	Quality					0.048*	0.055	0.731**	0.363
Loan-describing	*Amount*	-0.002**	0.380	-0.001**	0.064	-0.004**	0.051	-0.003**	0.036
	Rate	11.320**	0.125	-7.288**	0.090	7.113**	0.017	26.157**	0.082
	Duration	-0.026**	0.039	-0.054**	0.154	-0.099**	0.113	-0.092**	0.044
	Purpose			0.447**	0.035	-0.285**	0.004	-0.271	0.003
	Text	0.001**	0.017	0.000	0.002	0.000**	0.094		
Borrower-attributed	*Gender*	-0.097**	0.000	-0.152*	0.006				
	Age	-0.001**	0.003						
	Marriage					0.301**	0.003	0.295**	0.003
	Region	0.253**	0.003	0.264**	0.023	-0.070	0.001	-0.071	0.001
	Cohabitants	-0.003**	0.001	-0.038	0.005				
	Residence	-0.011	0.000	0.294**	0.020	-0.163**	0.003		
	Grade	-0.238**	0.054	-0.170**	0.100				
	Success	0.706**	0.072	0.317**	0.104				
	Failure	0.508**	0.003	-2.970**	0.322				
	Work years			0.051**	0.038				
	Insurance	0.352**	0.023	0.349**	0.033				
	Mobile					0.441**	0.012	0.501*	0.024
	Home					-0.075*	0.002		
	Office								
	Job					0.127**	0.001		
	Income					-0.001**	0.025	-0.002**	0.023
	Past Loans					-0.693**	0.220	-0.143	0.136
	Credit events					0.165**	0.018	0.343*	0.025
	Vehicle					-0.284**	0.007	0.263	0.002
Unemployment rate		-0.483**	0.001			-0.273**	0.002		
Bank loan interest rate		0.061**	0.003	0.232**	0.005	0.093**	0.005	0.371**	0.008
KOSFI		0.002**	0.000			0.000**	0.000		
FX (USD/KRW)		0.000**	0.000			0.000**	0.000		
(Intercept)		-8.452**		3.954**		-7.671**		-3.711**	
N		34961		4384		10794		1186	
AUC		0.837		0.713		0.945		0.863	
Pseudo-R^2		0.361		0.171		0.507		0.437	
LR χ^2		9811.53**		571.34**		3235.17**		427.61**	

Note: β and RW stands for the regression coefficient and relative weight, respectively.
** and * indicate statistical significance at the 1% and 5% levels, respectively.

Table 7. Tests on the difference in prediction accuracy between *ex-ante* and *ex-post* models

Prediction	Obs.	Prediction accuracy		Shapiro-Wilk test		Paired t-test			Wilcoxon signed rank test	
		Mean	Variance	W	p-value	df	t-statistics	p-value	W	p-value
Model I	10	0.620	4.736e-6	0.899	0.204	9	-2.533	0.032	25	0.064
Model II	10	0.639	4.171e-4	0.935	0.415					
Model III	10	0.765	2.671e-5	0.929	0.423	9	-1.703	0.123	25	0.064
Model IV	10	0.774	2.273e-4	0.903	0.204					

Table 8. Summary of the differences in prediction accuracy between *ex-ante* and *ex-post* models

Prediction	Average prediction accuracy		Difference
	Ex-ante model	*Ex-post* model	Results
Models I vs. II	62.0%	63.9%	+ 1.9%p*
Models III vs. IV	76.5%	77.4%	+ 0.9%p

Note: ** and * indicate statistical significance at the 1% and 5% levels, respectively.

Table 9. Comparisons of *ex-ante* logistic models over time

Variables		Model I-1		Model I-2		Model III-1		Model III-2	
		β	**RW**	**β**	**RW**	**β**	**RW**	**β**	**RW**
Platform-generated	*Score*	0.015 **	0.250	0.016 **	0.205				
	Rating					0.167 **	0.321	0.079 **	0.332
	Quality					-0.278 **	0.020	0.361 **	0.120
Loan-describing	*Amount*	-0.001 **	0.213	-0.003 **	0.459	-0.004 **	0.050	-0.003 **	0.037
	Rate	9.415 **	0.208	14.140 **	0.169	21.857 **	0.144	7.155 **	0.046
	Duration	-0.026 **	0.038	-0.041 **	0.037	-0.090 **	0.094	-0.113 **	0.102
	Purpose	0.277**	0.008	-0.066 **	0.011	-0.361 **	0.007	-0.183 **	0.001
	Text	0.001 **	0.074	0.000 **	0.004	0.000 **	0.098	0.000 **	0.064
Borrower-attributed	*Gender*	-0.187 **	0.004	-0.011	0.000			-0.135*	0.007
	Age	0.002	0.002	0.003 **	0.002				
	Marriage			-0.084 **	0.002				
	Region	0.118 **	0.004	0.356 **	0.005	-0.131 *	0.001		
	Cohabitants	-0.047 **	0.003					-0.033	0.001
	Residence	0.003 *	0.000	-0.082**	0.000	-0.335 **	0.010		
	Grade	-0.186 **	0.045	-0.407 **	0.041				
	Success	1.013 **	0.097	0.608 **	0.048				
	Failure	0.550 **	0.009	0.142 **	0.000				
	Work years	-0.029 **	0.006	0.023 **	0.003				
	Insurance	0.416 **	0.038	0.253 **	0.009				
	Mobile					0.207 **	0.001	0.675 **	0.024
	Home							-0.126	0.001
	Office					0.295 **	0.001		
	Job					0.009 **	0.002	0.170 **	0.001
	Income					-0.001 **	0.059	0.000 **	0.016
	Past Loans					-1.150 **	0.168	-0.331 **	0.226
	Credit events							0.187 **	0.013
	Vehicle					-0.432 **	0.014		
Unemployment rate		-0.565**	0.001	-0.389**	0.002	-0.173**	0.002	-0.315**	0.002
Bank loan interest rate		0.050*	0.000	0.446**	0.003	0.081**	0.001	0.097**	0.003
KOSFI		0.000**	0.000	0.005**	0.000			0.000**	0.000
FX (USD/KRW)		-0.002**	0.000	0.014**	0.000				
(Intercept)		-1.732**		-2.889**		-9.130**		-4.187**	
N		16178		18783		7671		3123	
AUC		0.803		0.861		0.961		0.918	
Pseudo-R^2		0.285		0.457		0.544		0.489	
LR χ^2		3179.12**		7315.71**		2132.61**		1143.41**	

Note: β and RW stands for the regression coefficient and relative weight, respectively.
** and * indicate statistical significance at the 1% and 5% levels, respectively.

Table 10. Tests on the difference in prediction accuracies of *ex-ante* models over time

Prediction	Obs.	Prediction accuracy		Shapiro-Wilk test		Paired t-test			Wilcoxon signed rank test	
		Mean	Variance	W	p-value	df	t-statistics	p-value	W	p-value
Model I-1	10	0.624	2.053e-5	0.922	0.382	9	-6.114	0.000	1	0.000
Model I-2	10	0.636	6.112e-6	0.932	0.465					
Model III-1	10	0.759	1.243e-4	0.944	0.602	9	-11.563	0.000	0	0.000
Model III-2	10	0.821	7.056e-5	0.965	0.840					

Table 11. Summary of the results of prediction accuracy change over time

Prediction	Average prediction accuracy		Difference
	Former *ex-ante* model	Latter *ex-ante* model	Results
Model I	62.4%	63.6%	+ 1.2%p**
Model III	75.9%	82.1%	+ 6.2%p**

Note: ** and * indicate statistical significance at the 1% and 5% levels, respectively.

Table 12. The effects of types of variables on prediction accuracy on Moneyauction

		Model I-1	Model I-1-1	Model I-1-2	Model I-1-3	Model I-2	Model I-2-1	Model I-2-2	Model I-2-3
Type of variables	Platform-generated		X				X		
	Loan-describing			X				X	
	Borrower-attributed				X				X
Prediction accuracy	Obs.	10	10	10	10	10	10	10	10
	Mean	0.624	0.607	0.618	0.591	0.636	0.642	0.629	0.594
	Variance	2.053e-5	2.867e-5	7.993e-5	4.615e-5	6.112e-6	1.564e-5	3.285e-5	1.754e-5
	Mean difference	Reference	-0.017**	-0.006*	-0.033**	Reference	0.006**	-0.007**	-0.042**

Note: ** and * indicate statistical significance at the 1% and 5% levels, respectively.

Table 13. The effects of types of variables on prediction accuracy on Popfunding

		Model III-1	Model III-1-1	Model III-1-2	Model III-1-3	Model III-2	Model III-2-1	Model III-2-2	Model III-2-3
Type of variables	Platform-generated		X				X		
	Loan-describing			X				X	
	Borrower-attributed				X				X
Prediction accuracy	Obs.	10	10	10	10	10	10	10	10
	Mean	0.759	0.774	0.724	0.782	0.821	0.753	0.816	0.813
	Variance	1.243e-4	2.196e-5	3.467e-4	4.564e-4	7.056e-5	1.465e-4	1.195e-4	6.997e-5
	Mean difference	Reference	0.015**	-0.035**	0.023**	Reference	-0.068**	-0.005	-0.008**

Note: ** and * indicate statistical significance at the 1% and 5% levels, respectively.

Table 14. Summary of the results of prediction accuracy change over time by type of variables

Prediction	Type of variables	Former *ex-ante* model		Latter *ex-ante* model	Results
Model I	Platform-generated	1.7%p ↑	→	0.9%p ↓	Deteriorated
	Loan-describing	0.6%p ↑	→	0.7%p ↑	No significant change
	Borrower-attributed	3.3%p ↑	→	4.2%p ↑	Enhanced
Model III	Platform-generated	1.5%p ↓	→	6.8%p ↑	Enhanced
	Loan-describing	3.5%p ↑	→	···	No significant change
	Borrower-attributed	2.3%p ↓	→	0.8%p ↑	Enhanced

Figures

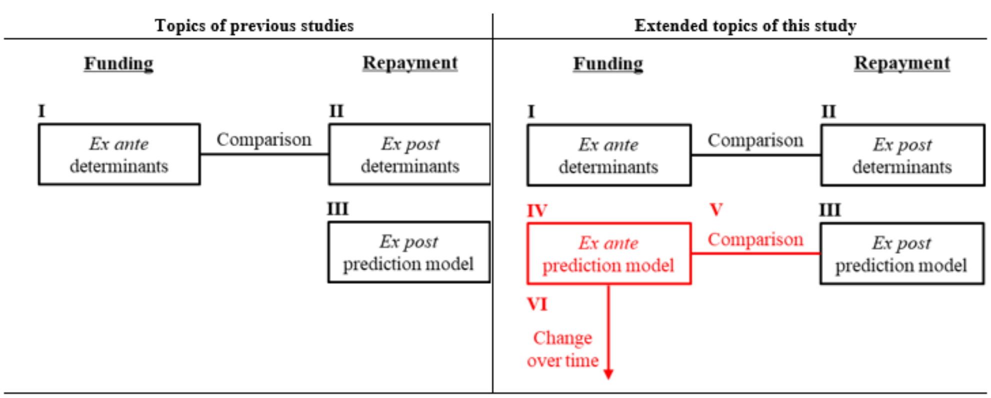

Figure 1

Theoretical contributions of this study

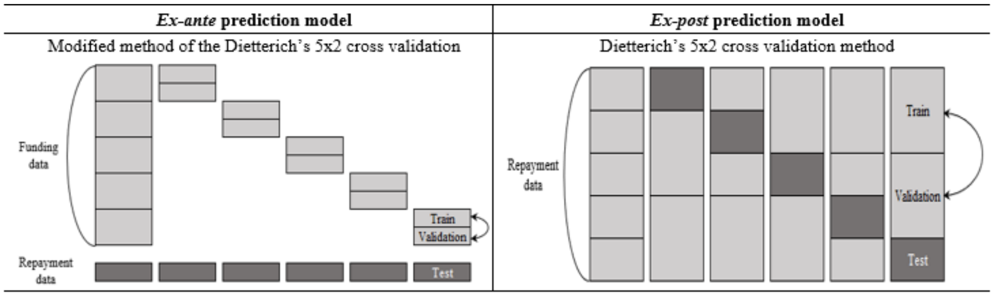

Figure 2

Cross validation methods for measuring prediction accuracy of models