International Journal of Innovation in Science and Mathematics
Volume 11, Issue 1, ISSN (Online): 2347–9051

Loan Default Prediction Based on XGBoost

Xiaoqian Peng[1], Jingyi Zhang[1], Xu Cheng[1], Yifan Huang[2] and Zhouqiang Zheng[3*]
[1]School of Finance and Taxation, Zhejiang University of Finance and Economics Dongfang College, Hangzhou, China.
[2]School of Finance, Zhejiang University of Finance and Economics, Hangzhou, China.
[3]Experimental Center of Economics and Management, Zhejiang University of Finance and Economics, Hangzhou, China.
*Corresponding author email id: shaw0805@163.com

Date of publication (dd/mm/yyyy): 23/02/2023

***Abstract* – Aiming at the credit risk loss of commercial banks caused by loan default, this paper establishes an integrated learning model to predict customer default based on the loan record data of credit platform to reduce credit risk. According to the characteristics of unbalanced loan data categories and high feature dimensions, this paper cleans the data and uses IV information to screen the factors, so as to select the factors with strong predictive ability. Then, on the basis of comparing various models, this paper chooses the best XGboost model to construct the loan default prediction model and evaluate the model. The analysis found that the indicators of XGboost are high, indicating that the performance of the model is good and can be used to accurately predict loan defaults. This provides a reference for commercial banks and other loan platforms to provide credit products to borrowers.**

***Keywords* – Loan Default Prediction, IV Information, Model Comparison, XGboost.**

I. Introduction

The advanced consumption consciousness promotes the innovation and development of personal credit. The personalization and diversification of people's consumption concepts, the improvement of income levels and the pursuit of high-quality life have made the awareness of advanced consumption more clear and promoted the rapid development of Internet finance. At present, a series of online credit products have emerged. The emergence of various financial and online credit products has intensified the competition in the financial market, and the number of credit intermediaries is also increasing, which allows banks to obtain borrowers' information at a lower price to determine whether to provide loan services. In this context, commercial banks began to change the previous business model, is no longer limited to corporate loans, housing loans, auto consumer loans and other products, and gradually introduced a bank similar to the original. Online credit loans such as e-loan have greatly facilitated and satisfied people's consumption and demand. Personal credit has made great progress in resident credit in China.

The increase in the amount and quantity of personal loans also means more default risks to some extent, especially under the impact of the epidemic, loan defaults and overdue situations are more prominent, and non-performing loan incidents increase. According to the data released by the China Banking and Insurance Regulatory Commission, at the end of the fourth quarter of last year, the total amount of non-performing loans of commercial banks was 2.8 trillion yuan, an increase of 13.5 billion yuan from the end of the last quarter. The non-performing loan ratio of commercial banks was 1.73%, 0.02 percentage points lower than the end of the last quarter. This year, the total amount of bad debts of banks has reached 3.1 trillion yuan, which has exceeded 3 trillion yuan again after last year. The emergence of non-performing assets will inevitably have a significant economic loss to the bank, so that the bank faces greater pressure on bad debts and must be taken seriously. Among all the risk attention evaluations, the attention score of credit risk has increased compared with last year, far exceeding other types of risks.

In order to reduce the risk of default, commercial banks have set up corresponding risk control management departments. In addition to traditional risk control such as manual audit of borrower conditions, big data technology has also been applied to the financial field. Under the traditional risk control mode of banks, the credit risk control process generally relies on manual operation, while manual audit often takes a long time, the process is complex and the work efficiency is low. Nowadays, in the face of the explosive growth of business online and the number of online users, the difficulty of risk control is increased. Traditional manual audit and backward model tools cannot meet the needs of risk control, and it is urgent to transform to digital. With the continuous improvement and development of machine learning algorithms, its application scope is also expanding, such as whether goods are purchased, operator 's customer churn prediction, enterprise illegal fund-raising identification, etc. For commercial banks, machine learning algorithms can also be applied to personal credit default prediction, through various algorithms to identify default users. It is a long-term task and challenge for commercial banks to realize intelligent risk control and improve the ability of default risk identification by means of digital transformation.

II. Related Work

A. *Study of the Factors Affecting Non-Compliance*

Back in 2004, Ma uses the logit model to explore the impact of the financial status of the loan enterprise, the significant characteristics of the lender, and the local economic development on the loan default of SMEs [1].

In 2010, Chou took 273 small and medium-sized listed companies in Shenzhen from 2004 to 2008 as the research object, used the relevant report information disclosed in the annual report of listed companies, and used the unbalanced panel data model to study the influencing factors of credit default risk of small and medium-sized listed companies in China. The results show that there is a positive correlation between the concentration of the first shareholder of listed companies in China and the risk of credit default. There is also a significant negative correlation between the proportion of tradable shares and the growth of enterprises and the risk of credit default [2].

In 2015, based on the theoretical analysis of the influencing factors of loan credit risk of rural commercial banks, Li collected the actual loan data of a rural commercial bank, used the random forest algorithm and Logistic model to empirically analyze the loan data, and obtained several factors affecting the loan default risk of rural commercial banks. It provides a reference for the rural commercial banks to reduce the risk of loan default [3]. Based on the data of a commercial bank in Qingdao from 2003 to 2014, Yang used Logistic regression model to study the influencing factors of housing mortgage loan default. The results show that the macroeconomic environment measured by loan amount, GDP per capital, real estate climate index and other indicators has a significant impact on the default rate of housing mortgage loans [4].

In 2018, Li described the development history and current situation of e-commerce consumer credit based on the data from the questionnaire, using chart analysis to make the phenomenon more intuitive and obvious. Focused on the description of the current situation of e-commerce consumer credit risk, which for the default rate and risk concentration is also based on survey data for data analysis summary. Then, based on the FICO model, the influencing factors of e-commerce consumer credit risk are analyzed. On the basis of introducing the FICO model, the application of the FICO model in China's current e-commerce is also studied. The crosscutting analysis of the influencing factors of risk is carried out and the results of the impact are obtained [5].

In 2020, Zhao chose the borrower's legal representative age, gender, education level, marital status, company rating, loan amount, enterprise scale, income status, loan term, investment industry, loan balance, loan guarantee method, and region, a total of 13 credit risk factors. Using a combination of quantitative analysis and qualitative analysis, and using Logistic regression analysis method, a comprehensive understanding of the influencing factors of default probability, aiming to be able to grasp the influencing factors of credit risk and how to control these risks to a minimum ; finally, on this basis, respectively, from the optimization of customer management, risk supervision, improve the internal control mechanism and other different angles, for financial institutions to do a good job of risk prevention, put forward targeted countermeasures [6].

In 2022, Yao used the monthly data of corporate bonds listed and traded in 2010-2014 from 2015 to 2019 as a sample, and used the propensity score matching method, the KMV model, the principal component analysis method, the Nelson-Siegel interest rate term structure model and the panel fixed effect estimation method for empirical analysis to gradually study the impact of risk factors and risk interweaving on corporate bond credit spreads. The empirical results show that the default risk has a low and insignificant contribution to the credit spread of corporate bonds, and the interest rate risk has a higher contribution to the credit spread of corporate bonds than the liquidity risk. Risk interweaving is significant in the credit spread of corporate bonds [7].

B. *Research on Default Prediction Model*

In 2010, Miao used factor analysis, discriminant analysis, cluster analysis and Logistic model to conduct an empirical study on 1690 student loan samples. The study found that the important factors affecting the risk of student loan default include : tuition debt, employment and age, family, repayment and time, and repayment period [8].

In 2015, Hu selected 24 early warning indicators based on the three-level index system of customer loan default warning and the panel Logit model of bank customer loan default risk warning. According to these indicators, bank customer loan default warning is carried out to improve risk management [9].

In 2020, Nie analyzed and explained the key technologies of data mining and related mining analysis methods. Using existing data to compare multiple models, a relatively accurate modeling method was obtained. Finally, the value of this method to society was briefly reviewed [10]. According to the user's basic attribute data and the numerous data of downloading APP types, Li realizes feature extraction and data weighting processing, and then uses the linear regression with penalty to construct the prediction model, which improves the accuracy of the default judgment of insolvent customers, realizes local optimization, thus improving the hidden risk prediction and control of customer commercial bank loans, and greatly reducing them [11].

In 2022, Zhang optimized the tree model in the form of reinforcement learning, and learned strong rule information with different statistical characteristics through several reinforcement learners in the form of combined logic tree. On this basis, the idea of mixed learning is combined with the characteristics of tree model and linear model to neutralize the defects of each model and prevent over-fitting of small and medium sample data. In the process of model implementation, Zhang adopts the method of grid search tuning, random search tuning and Bayesian optimization tuning to determine the fitting control parameters of the designed model in order to achieve the best model effect [12]. Cao discusses the risk and motivation of personal credit customer default in commercial banks, and discusses the basic ideas of identifying related risks from the perspective of bank risk control. Based on the gradient boosting decision tree, the prediction method of personal credit customer default in commercial banks is designed, and its application method is explained in order to provide reference for commercial banks [13].

In the same year, Xie uses the public credit default data set to model, uses the KS index and the AUC index to test the model effect, and uses Bayesian optimization to search the model hyper parameters. After preprocessing and feature construction of the data set, it is input into the TabNet model. The model performs well, with KS index of 0.408 and AUC index of 0.771, indicating that the model has certain practical value [14].

III. Empirical Analysis

A. *Data Source*

The data used in this paper comes from the loan records of a credit platform. The total amount of data is about 10,000, including 47 columns of variable information, such as Amnt, term, interest rate, installment, grade, sub grade, loan title, employment length, of which 15 columns are anonymous variables, which are the processing of some lenders ' behavior count characteristics.

Table. 1. Data display (first five lines).

Loan Amnt	Interest Rate	Installment	grade	...	Earlies Credit Line	Title
35000	19.52	917.97	E	...	Aug-01	1
18000	18.49	461.9	D	...	May-02	1723
12000	16.99	298.17	D	...	May-06	0
11000	7.26	340.96	A	...	May-99	4
3000	12.99	101.07	C	...	Aug-77	11

B. *Data Preprocessing*

Through preliminary data exploration, it is found that there may be some problems in the data set that need to be addressed. In the study of practical problems, the real data we extract are often messy, which may cause problems such as data format errors, data missing, and outliers. It needs to be processed before modeling operations. If the previous data cleaning problems, will directly affect the later establishment of statistical models and thus affect the quality of data analysis. Therefore, this paper uses Lagrange interpolation method to interpolate missing values ; absolute median method is used to process outliers ; the minmax method was used to standardize the data and eliminate the dimensional difference between the data.

C. *Factor Screening*

Loan default is the result of a variety of factors, so it is very important to screen candidate factors in the construction of default prediction model. The more candidate factors, the more comprehensive the analysis of lenders will be, and the better the effect of the constructed prediction model will be. Therefore, the selection of candidate factors must consider enough dimensions, but also have some explanatory. Based on the previous scholars' research on the influencing factors of loan default and considering the factor characteristics, this paper uses IV information to screen the factors.

The full name of IV value is information value, which is the amount of information or information value. It mainly deals with the selection of independent variables when the dependent variable is a binary variable. IV value is to measure the predictive power of independent variables, similar to the size of the information gain, Gini coefficient. Suppose that a discrete independent variable has n different values, the amount of IV information can be expressed as: $IV = \sum_{i=1}^{n} (Yes_i - No_i)\ln(\frac{Yes_i}{No_i})$

Among them, Yes_i said that in the sample with the i th value on the index, the proportion of defaulters to lenders in the sample ; correspondingly, No_i measures the proportion of non-default.

The greater the IV information of the index, the stronger the predictive ability of the index to the dependent variable, and the more it should be retained. According to the criterion of IV information, the factor with IV information greater than 0.1 has strong predictive ability, while the factor with IV information greater than 0.5 is doubtful, so the factor with IV information between 0.1 and 0.5 is retained.

Since the factors are continuous variables, the box-dividing process should be carried out before calculating IV information content. The variable box-dividing principle in this paper is similar to the binary decision tree, but the objective function is IV information content. The main idea of compartment is as follows :

1. Initialize the dataset D = D0 to be all data. Step 2.
2. For D, the data is sorted from small to large and divided into 10 parts according to the number, recording the division points. Calculate IV0, step 3, when still not divided.
3. Each partition point is traversed to calculate the time-sharing IV. If the maximum IV > IV0 * (1 + alpha) (User given, default 0.01): then divide, and the corresponding maximum IV is determined as the partition point. It divides D into left and right nodes, and the data sets are DL, DR, and step 4. Otherwise: stop.
4. Let D = DL, D = DR, repeat step 2.

Then use the formula to calculate the IV information of each factor. The results are as follows:

Table. 2. IV result table of each factor.

var_name	split_list	positive_sample_num	negative_sample_num	sub_total_num_percentage	positive_rate_in_sub_total	woe_list	iv_list	iv
term	(-INF,3.0]	911	4940	0.7590	0.1557	-0.26	0.04	0.16
term	(3.0,5.0]	582	1276	0.2410	0.3132	0.64	0.11	0.16
Interest Rate	(-INF,6.9]	21	474	0.0642	0.0424	-1.69	0.10	0.44
Interest Rate	(6.92,7.89]	25	384	0.0531	0.0611	-1.30	0.05	0.44
......	......	......	......	......	......	......	......	......
n14	(4.0,+INF)	171	393	0.0732	0.3032	0.59	0.03	0.05
Issue Date DT	(-INF,2100.0]	97	524	0.0806	0.1562	-0.26	0.00	0.00
Issue Date DT	(2100.0,+INF)	1396	5692	0.9194	0.1970	0.02	0.00	0.00

According to the criterion of IV information amount, six factors with IV information amount between 0.1 and 0.5 are reserved, that is, term (Loan term), interestRate (Loan interest rate), grade (Loan grade), subGrade

(Sub-level of loan grade). Fico Range Low (The minimum range to which the borrower's fico is subject at the time the loan is issued). Fico Range High (The maximum range to which the borrower's fico is subject at the time the loan is issued). Since these factors are selected from the range of factors that are important to the model, it is reasonable to assume that the selected factors are valid factors that have both explanatory power and are not strongly linearly dependent. Therefore, it is considered that factor screening has the purpose of effective dimension reduction under the premise of ensuring little change in prediction ability.

D. *Model Construction*

In this paper, all data are disrupted and divided into training set and test set according to the ratio of 7:3. The training set is used to train the model and the test set is used to evaluate the model. Accuracy, Recall, Precision, f1-score and AUC were used as the evaluation criteria for the prediction model.

Firstly, weak classifiers such as decision tree and logistic regression are used to construct the model, and it is found that the classification effect is poor. Therefore, considering the ensemble algorithm, the bagging and random forest based on decision tree in the ensemble learning bagging algorithm and the XGBoost and Catboost in the boosting algorithm are established respectively, and three models under the default parameters are established respectively. According to the evaluation index, the effect of these models is compared, as shown in the following table, and the appropriate model is selected as the prediction model.

Table 3. Effect comparison of model.

Type	Classifier	Accuracy	Precision	Recall rate	F1-score	AUC
Weak Classifier	SVM (linear)	0.6440	0.6238	0.7101	0.6641	0.6641
	SVM (SGD)	0.6383	0.6778	0.5154	0.5856	0.5856
	Native Bayesian	0.6346	0.6123	0.7166	0.6604	0.6604
	Logistic Regression	0.6488	0.6453	0.6474	0.6463	0.6463
	Decision Tree	0.6646	0.6795	0.6122	0.6441	0.6441
	KNN	0.6820	0.6744	0.6933	0.6837	0.6837
Bagging	Bagging	0.6670	0.6511	0.7074	0.6781	0.6781
	Random Forest	0.6716	0.6548	0.7139	0.6831	0.6831
Boosting	Xgboost	0.7791	0.8045	0.7323	0.7667	0.7667
	Cat Boost	0.6539	0.6440	0.6750	0.6591	0.6591

Because this paper studies the default situation, the purpose is to be able to predict the default user, so although the accuracy rate can represent the overall accuracy rate, it has certain reference value, but it is not a suitable index for the content of this paper, so this index is not considered for the time being. For the rest of the indicators, the recall rate and the precision rate are the relationship between the rise and fall. The recall rate can be understood as the proportion of the actual default customers that can be predicted by the model. The precision rate refers to the proportion of actual defaults in the customers that the model predicts as defaults. Often, we pay more attention to whether default customers can be identified. If the model is very accurate in identifying default users, it can avoid the loss and harm caused by the occurrence of most non-performing loans to banks. For the situation that the bank refuses to issue loans due to the prediction of default but the customer does not actually default, the bank will also have certain losses. For example, the decrease in business volume leads to a decrease in income, but the two are compared. Often the former loss is greater, so in the precision rate, recall rate and F1 score of these three indicators, that the recall rate is relatively more important, is the focus of attention indicators. The AUC value is often not affected when the proportion of positive and negative samples changes, while other indicators will be greatly affected, so the AUC value should also be focused on.

It can be seen from the data in the table that in the weak classifier, KNN has a better effect, and the return rate and AUC value are the highest, indicating that KNN has the best prediction effect in the weak classifier. In Bagging, the indicators of Bagging based on decision tree and Random Forest are similar. In Boosting, XGboost is better than Catboost. On the whole, XGboost has the best recall rate and AUC value in all models. Therefore, this paper finally uses XGboost to construct a loan default prediction model.

The accuracy of the model based on the default parameters is 77.91 %. On this basis, using sklearn ' s grid search for parameter adjustment. The specific process is as follows :

Step 1: Determine the number of estimators for learning rate and tree _ based parameter tuning

The max _ depth is generally selected between 3-10, and the starting value is 5.

min _ child _ weight Choose a smaller value, starting at 1, Initial gamma = 0.

subsample, colsample _ bytree generally takes 0.5-0.9, the starting value is 0.8, scale _ pos _ weight = 0.

First, find the optimal number of decision trees required based on the default learning rate of 0.1, increasing from 100 to 900. The optimal n _ estimators is 800.

Step 2: max _ depth and min _ child _ weight parameter tuning

These two parameters are optimized first because they have a great influence on the final result. First coarse-tune the parameters from a wide range, then fine-tune them from a small range. After outputting the results, the optimal max _ depth is 7, min _ child _ weight is 1.

Step 3: gamma parameter tuning

Searching for gamma from [0, 0.1, 0.2, 0.3, 0.4], the optimal gamma is 0.1.

Step 4: Adjust the subsample and colsample _ bytree parameters

The parameters of subsample and colsample _ bytree increase from 0.1, 0.6 to 0.9. The ideal values for the subsample and colsample _ bytree parameters are 0.8, 0.8. We then take a step of 0.05, around this value. The ideal values for the subsample and colsample _ bytree parameters are still 0.8, 0.8. Then take them as the final ideal value.

Step 5: Regularization parameter tuning

The next step is to reduce overfitting by regularization, where the reg _ alpha parameter is adjusted. The ideal value here is 0.1.

Step 6: Reduce learning efficiency

Reduce the learning rate by ten times to 0.001, increase the number of trees to 5000, and increase the score.

The final tuning results are as follows :

Table. 4. Table of tuning results.

Parameter Abbreviation	Parameter Meaning	Initial Value Parameter	Adjustment Results
max_depth	Maximum depth	None	7
min_child_weight	Decision tree	1	1
gamma	Number of	0	0.1
subsample	Minimum sample weight of leaf nodes	1	0.8
colsample_bytree	Random sample	1	0.8
reg_alpha	Spanning Tree Column Sampling	0	0.1
learning_rate	L1 of weights	0.1	0.001
n_estimators	Regularization term	60	5000

Finally, this paper uses accuracy, precision, recall and F1-score to evaluate the XGboost model. The evaluation results are as follows :

Table. 5. Table of tuning results.

Classification Report	Precision	Recall	F1-Score	Support
Not Default	0.80	0.88	0.84	1881
Default	0.86	0.78	0.82	1849
accuracy			0.83	3730
macro avg	0.83	0.83	0.83	3730
weighted avg	0.83	0.83	0.83	3730

The accuracy rate, precision rate, recall rate, F1-score and AUC value all reach 83 %, indicating that the model has good performance and can be used to accurately predict loan defaults. It can be concluded that the integrated learning algorithm improves the prediction ability of the model. In addition, ensemble learning has a strong advantage in the stability of the model. It can be inferred that the XGboost model has a huge advantage in predicting loan defaults.

IV. Conclusion

With the rapid development of personal loan business, loan defaults and overdue situations are more prominent. Although banks have established regulatory authorities for credit business, credit default risk is still a major problem faced by financial institutions such as banks. In this paper, the personal credit behavior data is preprocessed, and the IV information is used to screen the factors. By comparing a variety of machine learning models, the XGBoost model in ensemble learning is finally established to predict the default behavior, and the corresponding indicators are used to evaluate the effect of the model. The following conclusions are obtained:

(1) Most of the factors screened by IV information are related to the credit product itself, such as term (Loan term), interest rate (Loan interest rate), grade (Loan grade), subGrade (Sub-level of loan grade), which indicates that the default probability of a loan can be inferred from the credit product selected by the lender.

In addition, the amount of IV information is very suitable for binary classification problems, and the quality of the selected factors is high.

(2) When comparing the modeling effects of the weak classifier and the ensemble model, it can be found that when the amount of data is large, the prediction ability of the weak classifier is poor and unstable, while the performance of the ensemble model is better.

(3) Because in the actual loan situation, the defaulters generally account for a relatively small proportion, that is, the data set has the problem of imbalance between positive and negative samples. At this time, the accuracy rate, accuracy rate, and F1 score will be affected, while the recall rate and AUC value are almost unaffected. Therefore, the XGboost model is selected as the final default risk prediction model to predict whether the borrower will default.

In summary, the loan default prediction model based on XGboost constructed in this paper can provide reference for commercial banks and other loan platforms to provide credit products to borrowers.

ACKNOWLEDGMENT

This work was supported in part by the Collaborative education project of industry university cooperation of the Ministry of Education of China under Grant 202102576042, and in part by the Zhejiang Education Science Planning Project under Grant 2019SCG175. (Corresponding author: Zhouqiang Zheng).

REFERENCES

[1] Ma Jiujie, Guo Yuhui, Zhu Yong. County small and medium-sized enterprise loan default behavior and credit risk empirical analysis. Management world Bound, 2004, (05) : 58-66 + 87.
[2] Qiu Rongguo, Zhang Jianhua. Research on the determinants of credit default risk of China 's Small and Medium-sized Listed Companies-an empirical test based on 5-year panel data of 273 small and medium-sized listed companies in Shenzhen [J]. Qiu, 2010 (04) : 28-30.
[3] Li Yiheng. Research on the influencing factors of loan default of Rural Commercial Banks [J]. SME Management and Technology (Midterm), 2015 (02) :72-73.
[4] Yang Wenjing, Jiang Wei. Logistic-based study on influencing factors of housing mortgage loan default: a case study of a commercial bank in Qingdao [J]. Journal of Qingdao University (Natural Science Edition), 2015, 28 (02) : 89-92.
[5] Li Shanshan.E-commerce consumer credit risk and its influencing factors [D]. Tianjin University of Commerce, 2018.
[6] Zhao Guangbin. Research on the influencing factors of credit business risk of Agricultural Development Bank of China T Branch [D]. Sichuan Agricultural University, 2020.
[7] Yao Dingjun, Luo Liang. Research on the influencing factors of credit spreads of Corporate Bonds under Multiple Risks [J]. Financial Theory and Practice, 2022 (03) : 87-96.
[8] Liao Maozhong. An empirical study on the influencing factors of student loan default risk [J]. Fudan Education Forum, 2010, 8 (01) : 62-66 + 74.
[9] Hu Yi, Wang Jue, Yang Xiaoguang. Research on the early warning of bank customer loan default risk based on panel Logit model. System engineer Cheng Theory and Practice, 2015, 35 (07) : 1752-1759.
[10] Nie Jiyu. Research on the prediction method of commercial loan default based on credit score and data mining [J].National Circulation Economy, 2020 (04) : 144-147.
[11] Li Tianyang. Research on financial risk prediction algorithm based on big data [D]. Northern University of Technology, 2020.DOI : 10.26926.
[12] Zhang Haodong. Research on credit default prediction model based on credit data of reinforcement learning and blended learning [D]. Wuhan Institute of Posts and Telecommunications Sciences, 2022.
[13] Cao Yujia. Research on the default motivation and prediction scheme of individual credit customers of commercial banks [J].Financial circles, 2022 (26) : 51-53.DOI : 10.19887.
[14] Xie. Research on customer credit default prediction based on deep learning [D]. East China Normal University, 2022. DOI : 10.27149.

AUTHOR'S PROFILE

First Author
Xiaoqian Peng, School of Finance and Taxation, Zhejiang University of Finance and Economics Dongfang College, Hangzhou, China.

Second Author
Jingyi Zhang, School of Finance and Taxation, Zhejiang University of Finance and Economics Dongfang College, Hangzhou, China.

Third Author
Xu Cheng, School of Finance and Taxation, Zhejiang University of Finance and Economics Dongfang College, Hangzhou, China.

Fourth Author
Yifan Huang, School of Finance, Zhejiang University of Finance and Economics, Hangzhou, China.

Fifth Author
Zhouqiang Zheng, Experimental Center of Economics and Management, Zhejiang University of Finance and Economics, Hangzhou, China.