WILEY | Hindawi

Research Article

Study on Effect of Consumer Information in Personal Credit Risk Evaluation

Hongmei Wen (iD),[1] **Xin Sui** (iD),[1] **and Shaopeng Lu**[2]

[1]*School of Finance, Harbin University of Commerce, Harbin 150028, China*
[2]*School of Economics, Nankai University, Tianjin 300072, China*

Correspondence should be addressed to Xin Sui; 513119931@qq.com

This study selects Chinese borrowers' information from a platform that has both online shopping and consumer loan service as sample, studies the effect of consumer information in personal credit risk evaluation, and uses the lLogistic regression model, light gradient boosting machine (LightGBM) algorithm, and Shapley Additive Explanation (SHAP). The results show that the information of all consumer loan groups cannot be covered by traditional credit information. Consumer information can help predict the behavior of borrower's repayment and provide support for personal credit risk evaluation effective. Adding consumption information to the personal credit risk evaluation model can improve the accuracy of the model effectively. The model variables are ranked by feature importance, and there are 5 consumption indicators in the first 5 indicators of feature importance, which further verifies the value and effect of consumption information in personal credit risk evaluation. This study not only reveals the effect and value of consumer information in personal credit risk evaluation effectively, but also provides new ideas for the development of consumer financial market.

1. Introduction

At present, a new round of global scientific and technological revolution continues to deepen, financial technology is rising rapidly, and new business forms, new models, and new products emerge one after another. It not only changes the operation mode of traditional financial services, but also promotes the development of Internet personal consumption loans. At the same time, the vigorous development of personal consumer loans will have an impact on customers' intertemporal consumption behavior. Under this background, the personal credit risk evaluation index system is becoming more and more abundant, and multi-dimensional and massive data are fused and processed. Big data credit investigation is an inevitable trend of personal credit risk evaluation. Nevertheless, some data are missing and the amount of data is lacking, so it is difficult to train [1]. The PBC credit investigation system cannot cover all borrowers' information, which requires financial institutions to explore indicators beyond traditional personal credit risk evaluation index system. The traditional evaluation index system rarely uses consumption information as the evaluation index of personal credit risk. Can consumer information be used for personal credit risk assessment? What is the value and function of consumer information in personal credit risk evaluation? This is the question to be answered in this paper. This study uses the Chinese borrower's data of a platform that has both online shopping and consumer loan to explore this problem.

Driven by financial technology, Internet consumer loans came into being and have made great progress. In early 2014, JD finance launched the "JD white note" to open the chapter of Internet consumer credit. Ant group followed closely with the launch of "Ant Huabei." Since 2015, commercial banks, consumer finance companies, and other financial institutions have vigorously deployed Internet consumer finance business. Up to now, the scale of Internet consumer credit has expanded rapidly, with a scale of 18.7 billion RMB at the beginning of 2014 and 15.4 trillion RMB at the end of 2020, with a compound annual growth rate of 206%. By the end of

2020, there were 778 million credit cards and debit-credit cards used in China, and the number of cards per capita was 0.56. According to the questionnaire analysis of 110 cities in 30 provinces and cities in China by China Institute of Economic Thought and Practice of Tsinghua University, residents' acceptance and satisfaction with consumer finance companies are increasing, and the annual growth rate of adults willing to accept consumer finance services is more than 10%. Compared with the blowout-type development of Internet loan market, the improvement of personal credit risk management ability is insufficient slightly. The most important factor restricting the in-depth study of consumer finance is the lack of micro-data and continuous micro-household sampling [2]. The use of utilizing new information beyond traditional credit investigation methods by financial science and technology innovation helps to reduce information asymmetry in the consumer credit market. It provides an important channel for marginal groups of traditional financial services to obtain high-quality credit services and establish personal credit records [3].

Personal consumer loans have the characteristics of "small amount, unsecured and flexible," and the default risk cannot be quantitatively monitored through the traditional postloan behavior score. Therefore, preloan review has become an important measure to prevent credit risk. In recent years, the asset quality of personal consumer loans has been moving downward, and the nonperforming rate has continued to grow, which has brought new requirements and challenges to the credit risk management of consumer loans. At present, the personal credit-related information comes from the personal credit information of the People's Bank of China. The mainstream three-party credit service institutions such as Baihang credit and Bairong Inc. also provide effective data for personal credit risk evaluation. However, the personal credit reporting system of the people's Bank of China does not cover enough dimensions in terms of data comprehensiveness, hierarchy, and timeliness. The personal credit information mastered by the third-party data company is different from the traditional credit information, mainly from the customer's loan information and fraud score on P2P network lending platforms, microfinance companies, and other platforms. The introduction of third-party data has confirmed that it has great practical significance to apply information other than the credit investigation information of the People's Bank of China to the credit risk evaluation of personal consumer loans. Different from third-party data, consumer information is highly available, but it cannot reflect the solvency and willingness to repay. Therefore, the value and effect of consumer information in personal credit risk evaluation are worthy of in-depth discussion.

Different from the existing research, this paper has made some progress in the following two aspects: first of all, this paper deeply analyzes the differences between consumer information and traditional credit information in personal credit risk evaluation. Second, it proves that consumer information can be an effective supplement to traditional credit information in the practice of personal credit risk evaluation.

2. Previous Research

For a long time, information asymmetry is the main factor affecting personal credit risk. Stiglitz and Weiss [4] first proposed that although the borrower's credit history information (such as credit rating and historical performance) is open to all investors, network anonymity may aggravate the typical information asymmetry of online lending [4]. Information asymmetry also exists in the online lending market [5, 6]. The existence of information asymmetry will lead to two kinds of effects of market shrinkage: one is the "loan sparing"—the effective supply of the market is shrinking; the second is the "crowding out"—the effective demand of the market is shrinking [7].

In the research on the influencing factors of personal loan default, scholars found that there are internal and external factors, which can also be reflected in "hard information" and "soft information" [8, 9]. "Hard information" refers to personal information with strong authenticity and objective existence, such as age, gender, job grade, and credit score, and "soft information" refers to the borrower's group, the number of friends in the social network, consumption ability level, macro-information, and other information [10, 11]. Su and Cheng [12] found that "soft information" is also valuable in their research on the factors affecting the default behavior of online lending borrowers [12]. Chi et al. [13] showed that macroeconomic factors have a certain impact on the borrower's repayment [13].

For a long time, establishing an effective personal credit evaluation system is the core content of personal credit risk management. Scholars' research on personal credit risk evaluation mainly includes two aspects: one is the construction of personal credit risk evaluation index system, and the other is the methods of personal credit risk evaluation. Most of the existing evaluation index systems take the indicators in customers' historical credit transactions as the main indicators [14–16]. In the early stage, the credit risk evaluation methods mainly used logistic regression and expert discrimination, and the logistic regression model in the regression analysis model is the most widely used personal credit risk evaluation model so far [17–25]. The core model of FICO score is logistic regression algorithm. At present, China's commercial banks mainly use analytic hierarchy process and fuzzy evaluation method to evaluate the credit risk of borrowing customers [26]. With the development of financial technology and artificial intelligence, machine learning algorithm can predict the default risk of credit subjects more efficiently and accurately. Verleysen and Francois [27] first applied the decision tree method in personal credit evaluation and achieved good evaluation results [27]. Houle et al. [28] improved the decision tree method based on boosting algorithm on the basis of considering the sample attributes of personal credit risk assessment [28].

However, the mainstream credit evaluation mostly uses the borrower's historical loan information to build the evaluation model, which cannot cover all customer groups. This leads to some customers being shut out of mainstream

credit services, unable to obtain financial support and subject to serious financial constraints. In order to break this constraint, it is particularly important to explore whether indicators beyond traditional credit evaluation system can identify credit risk and be used in credit risk evaluation. Bertrand and Kamenica [29] document that owning an IOS device is one of the best predictors for being in the top quartile of the income distribution [29]. Belenzon et al. [30], and Guzman and Stern [31] have documented that customers having their names in the e-mail address are 30% less likely to default [30, 31]. Digital footprints can facilitate access to credit when credit bureau scores do not exist, thereby fostering financial inclusion and lowering inequality [32–36].

It can be seen that information beyond mainstream credit evaluation indicators can also achieve significant results in credit risk evaluation. Continuously mining information beyond mainstream credit evaluation indicators becomes more and more important to reduce the credit risk caused by information asymmetry and can also continuously improve the ability of risk identification [1, 3]. This study focuses on the effect of consumer information in personal credit risk evaluation. Consumer information has the characteristics of high frequency and timeliness, which is different from the traditional credit indicators. This study attempts to use empirical analysis to reveal the effect of consumption information in personal credit risk evaluation.

3. Data and Variable Description

3.1. Sample Selection. The data of this paper come from a commercial bank. Platforms like ANT and JD have certain channels to obtain customers, but they do not have enough funds, while commercial banks have lacked large-scale channels to obtain customers but have funds. Therefore, the joint loan between the platform and the bank came into being in China. We obtained the borrowers' information data from a Chinese platform which cooperated in joint loan from a commercial bank. As a large life service platform, the platform involves food, hotel, tourism, film, and group purchase. As of December 31, 2020, the number of active merchants and annual transaction users of the platform had increased to 6.8 million people and 510 million RMB, respectively. Therefore, it is very persuasive to select the borrowers' information data of the platform as a sample for empirical research. This study constructs an evaluation index system by obtaining the borrowers' behavior of repayment, consumption information of the platform, transaction information of the platform, information of credit card use, credit information of credit card, personal information, and loan information. According to the borrowers' repayment performance on the platform, the customer is defined as "good" or "bad" to explore the effect of consumption information in personal credit evaluation. We selected 185600 customer information data of the credit granted from April 2020 to August 2020, and 50000 pieces of customer information were randomly selected as samples for analysis.

3.2. Variable Description. The core variables involved in the empirical design of this study are as follows.

3.2.1. Behavior of Repayment. Through the vintage analysis of the behavior of repayment of customers on the platform, the results show that when the borrower has overdue for more than 30 days, there is a 93.75% probability that it will continue to be overdue. Therefore, if the overdue days of the borrower are more than 30 days, the customer is defined as "bad customer" (default = 1). On the contrary, if the customer has no overdue performance or the overdue days are not more than 30 days, it is defined as "good customer" (default = 0).

3.2.2. Consumption Information of the Platform. The consumption information of the platform comes from the consumption records of the borrower on the platform. The following three indicators are formed for the processing of the borrower's consumption records on the platform within one year: consumption ability, consumption frequency, and consumption scene.

3.2.3. Transaction Information of the Platform. The transaction information of the platform comes from the borrower's transaction records on the platform within 90 days, including days of transaction, amount of successful transaction, number of successful transaction, and active days.

3.2.4. Information of Credit Card Use. The information of credit card use comes from the borrower's personal credit information, including average quota utilization and maximum quota utilization.

3.2.5. Credit Information of Credit Card. The credit information of credit card also comes from the borrower's personal credit information, including average quota, number, and total amount.

3.2.6. Control Variable. There are two types of control variables: borrower's personal information and loan information. Personal information includes age, gender, and state of marriage and job; loan information includes amount and rate.

The definitions of main variables and descriptive statistics are shown in Tables 1 and 2 respectively.

4. Evaluation Model and Empirical Results

4.1. Construction of Model

4.1.1. The Effect of Consumer Information in Personal Credit Risk Evaluation. In order to explore the effect of consumer information in personal credit risk evaluation deeply, we use logistic regression model for empirical analysis. The dependent variable is customer behavior of repayment. Independent variables include platform consumption

TABLE 1: Definition of main variables.

Variable type	Variable name	Variable description
Behavior of repayment	Default	If the customer is overdue for more than 30 days, assign 1, and conversely, assign 0
Consumption information of the platform	Consumption ability	Consumption capacity level within one year, consumption amount <30 RMB, assign 1; $30 \leq$ consumption amount < 70 RMB, assign 2; $70 \leq$ consumption amount < 90 RMB, assign 3; consumption amount $\geq$90 RMB, assign 4; no consumption record, assign 0
	Consumption frequency	Consumption frequency grade within one year, consumption frequency <3, assign 1; $3 \leq$ consumption frequency < 6, assign 2; $6 \leq$ consumption frequency < 10, assign 3; consumption frequency $\geq$10, assign 4; no consumption record, assign 0
	Consumption scene	Number of consumption scenarios within a year, *consumption scene* <2, assign 1; $2 \leq$ *consumption scene* < 4, assign 2; *consumption scene* $\geq$4, assign 3; no consumption record, assign 0
Transaction information of the platform	Days of transaction	Days of transaction $\leq$1, assign 1; $1 \leq$ days of transaction < 4, assign 2; days of transaction $\geq$4, assign 3; no found, assign 0
	Amount of successful transaction	Amount of successful transaction $\leq$25, assign 1; $25 \leq$ amount of successful transaction < 200, assign 2; amount of successful transaction $\geq$200, assign 3; no found, assign 0
	Number of successful transaction	Number of successful transactions $\leq$1, assign 1; $1 \leq$ number of successful transactions < 4, assign 2; number of successful transactions $\geq$4, assign 3; no found, assign 0
	Active days	Active days $\leq$4, assign 1; $4 \leq$ active days < 19, assign 2; active days $\geq$19, assign 3; no found, assign 0
Information of credit card use	Average quota utilization	Average quota utilization $\leq$0.35, assign 1; $0.35 \leq$ average quota utilization < 0.75, assign 2; $0.75 \leq$ average quota utilization < 0.85, assign 3; average quota utilization $\geq$0.85, assign 4; no found, assign 0
	Maximum quota utilization	Maximum quota utilization $\leq$0.1, assign 1; $0.1 \leq$ maximum quota utilization < 0.3, assign 2; $0.3 \leq$ maximum quota utilization < 0.5, assign 3; maximum quota utilization $\geq$ 0.5, assign 4; no found, assign 0
Credit information of credit card	Average quota	Average quota $\leq$10000 RMB, assign 1; $10000 \leq$ average quota < 30000 RMB, assign 2; $30000 \leq$ average quota < 50000 RMB, assign 3; $50000 \leq$ average quota < 100000 RMB, assign 4; average quota$\geq$100000 RMB, assign 5; no found, assign 0
	Number	number = 0, assign 1; $1\leq$number< 3, assign 2; $3\leq$number <7; $7\leq$number <12, assign 4; number $\geq$12, assign 5; no found, assign 0
	Total amount	Total amount $\leq$10000 RMB, assign 1; $10000 \leq$ total amount < 30000 RMB, assign 2; $30000 \leq$ total amount < 50000 RMB, assign 3; $50000 \leq$ total amount < 100000 RMB, assign 4; total amount $\geq$100000 RMB, assign 5; no found, assign 0
Personal information	Age	Under 22 years old, assign 1; 22–40 years old, assign 2; beyond 40 years old, assign 3
	Gender	Male, assign 0; female, assign 1
	State of marriage	Unmarried, assign 1; married, assign 2; divorced and other, assign 3; no found, assign 0
	Job	Agriculture, forestry, animal husbandry, and fishery, assign 1; manufacturing, assign 2; social services, assign 3; public service, assign 4; real estate, assign 5; wholesale, assign 6; financial, assign 7; other, assign 8; unknown, assign 9; no found, assign 0
Loan information	Amount	Below 5000 RMB, assign 1; 5000–10000 RMB, assign 2; 10000–20000 RMB, assign 3; 20000–30000 RMB, assign 4; 30000–40000 RMB, assign 5; above 40000 RMB, assign 6
	Rate	Interest rate

TABLE 2: Descriptive statistics.

Variable	Mean	Min	Max	Standard deviation	Observed value
Default	0.04328	0	1	7275.8	50000
Consumption ability	0.196896	0	4	0.17342	50000
Consumption frequency	1.75604	0	4	0.496516	50000
Consumption scene	1.82742	0	4	0.452636	50000
Days of transaction	2.51908	0	3	0.8303	50000
Amount of successful transaction	2.58016	0	3	0.7904	50000
Number of successful transaction	2.54074	0	3	0.82542	50000
Active days	2.41376	0	3	0.87513	50000
Average quota utilization	1.46706	0	4	0.77997	50000
Maximum quota utilization	2.28846	0	4	1.40822	50000
Average quota	1.57802	0	5	0.7446	50000
Number	2.33848	0	4	1.136	50000
Total amount	2.51658	0	5	1.60343	50000
Age	1.95218	1	3	0.404592	50000
Gender	0.43438	0	1	0.46568	50000
State of marriage	1.61786	0	3	0.6959	50000
Job	7.30406	0	9	2.67282	50000
Amount	2.94646	1	6	2.29651	50000
Rate	16.611588	10.8	23.4	3.23299	50000

information, platform trading information, and credit card usage information; credit card credit information and control variables include the borrower's personal information and loan information.

In order to study the recognition ability and effect of consumer information in personal credit risk evaluation, we constructed five logistic regression models. Model 1 discusses the relationship between traditional credit information and borrower's credit risk. The independent variables include credit card credit information and control variables. Model 2 discusses the relationship between credit card use information and borrower's credit risk. The independent variables include credit card use information and control variables. Model 3 discusses the relationship between platform consumption and transaction information and borrower's credit risk. The independent variables include platform consumption information, platform transaction information, and control variables. Model 4 discusses the relationship between all consumption information obtained in this study and the borrower's credit risk. The independent variables include platform consumption and transaction information, credit card use information, and control variables. Model 5 includes all independent variables. The formula is expressed as follows:

$$
\begin{aligned}
&\text{Logistic}(\text{default}_i) = \alpha_1 \text{cc\_credit}_i + \alpha_2 \text{control}_i + \varepsilon, \\
&\text{Logistic}(\text{default}_i) = \alpha_1 \text{cc\_use}_i + \alpha_2 \text{control}_i + \varepsilon, \\
&\text{Logistic}(\text{default}_i) = \alpha_1 \text{consumption}_i + \alpha_2 \text{trading}_i + \alpha_3 \text{control}_i + \varepsilon, \\
&\text{Logistic}(\text{default}_i) = \alpha_1 \text{consumption}_i + \alpha_2 \text{trading}_i + \alpha_3 \text{cc\_use}_i + \alpha_4 \text{control}_i + \varepsilon, \\
&\text{Logistic}(\text{default}_i) = \alpha_1 \text{consumption}_i + \alpha_2 \text{trading}_i + \alpha_3 \text{cc\_use}_i + \alpha_4 \text{cc\_credit}_i + \alpha_5 \text{control}_i + \varepsilon.
\end{aligned}
\tag{1}
$$

Among them, default_i is the dependent variable, α is the coefficient, consumption_i, trading_i, cc_use_i, and cc_credit_i are the explanatory variable, and ε is a random error term. The logistic regression model was implemented by s as 14.3.

4.1.2. Model for Personal Credit Risk Prediction.
LightGBM is an open-source, fast, and efficient lifting framework based on a decision tree algorithm, which supports efficient parallel training and can greatly shorten the training time. LightGBM algorithm is a kind of boosting algorithm. Boosting algorithm learns multiple classifiers by changing the weight of training samples and improving the performance of classifiers through linear combination. Boosting formula can be expressed as follows:

$$
f(x) = W_O + \sum_{M=1}^{M} Wm\phi m(X).
\tag{2}
$$

The objective function is

$$
\text{Obj}(\theta) = L(\theta) + \Omega(\theta).
\tag{3}
$$

Among them, $L(\theta)$ is the error function and $\Omega(\theta)$ is the regular term.

LightGBM also seeks the optimal solution by combining the error function with the regular term by constructing the objective function. The objective function is

$$L(j) = \sum_i l(y_i, \widehat{y}_i) + \sum_k \Omega(f_k). \quad (4)$$

The complexity term of the tree in $\Omega(f_k) = \gamma T + 1/2\lambda\|W\|^2$ algorithm includes the regular term of the total number of leaf nodes and the score of leaf nodes, which can produce the phenomenon of skin overfitting. The regular term is

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda\|W\|^2. \quad (5)$$

LightGBM performs quadratic Taylor expansion on the cost function. Unlike GBDT, LightGBM can not only define the cost function, but also obtain the derivatives of the first and second orders at the same time. Then, the loss function of t is

$$L^{(t)} = \sum_{i=1}^{n} l\left(y_i, \widehat{y}_i^{(t-1)} + f(x_i)_t\right) + \Omega(f_t). \quad (6)$$

The second-order Taylor expansion of (6) is carried out to obtain (7), and the first derivative and second derivative are (8) and (9), respectively:

$$L^{(t)} \approx \sum_{i=1}^{n} l\left(y_i, y_i^{(t-1)} + g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)\right) + \Omega(f_t), \quad (7)$$

$$g_i = \delta_{\widehat{y}^{(t-1)}} L\left(y_i, \widehat{y}_i^{(t-1)}\right), \quad (8)$$

$$h_i = \delta^2_{\widehat{y}^{(t-1)}} L\left(y_i, \widehat{y}_i^{(t-1)}\right). \quad (9)$$

LightGBM divides the eigenvalues into buckets and then constructs the histogram for splitting. The calculation formula is as follows:

$$\text{gain} = \max\left(\text{gain}, \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{G}{H + \lambda}\right). \quad (10)$$

The samples processed by bucket division not only improve the training speed, but also reduce the complexity of calculation and the occupation of computer memory. Due to the advantages of LightGBM algorithm, it is widely used in many fields. We use LightGBM algorithm to predict personal credit risk.

4.2. Empirical Analysis. By analyzing the real performance of the Chinese borrower's credit risk of a platform integrating both online shopping and consumer loan, we measure the traditional credit information with the borrower's credit card credit information, measure the consumption information with the borrower's consumption information, transaction information, and credit card use information on the platform, and further explore the effect of consumer information in personal credit risk evaluation.

4.2.1. The Effect of Consumer Information on the Borrower's Credit Risk. The results of logistic regression are shown in Table 3. Model 1 discusses the relationship between

traditional credit information and borrower's credit risk. The regression coefficient of the average quota is −0.13, which is significant at the 5% significant level. It indicates that the larger the average quota of the borrower, the smaller the credit risk. The regression coefficients of the amount of credit cards and total amount of credit cards are 0.08 and 0.11, respectively, which are significant at the significant level of 5% and 1%, respectively. It indicates that the more credit cards, the greater the total amount of credit card and the greater the credit risk. Model 2 discusses the relationship between credit card use information and borrower's credit risk. The regression coefficient of average quota utilization is 0.22, which is significant at the 1% significant level. It shows that the greater the average quota utilization, the greater the credit risk. The maximum quota utilization of credit card has no significant impact on credit risk. Model 3 discusses the relationship between consumption information of the platform, transaction information of the platform, and borrower's credit risk. Consumption ability, days of transaction, amount of successful transaction, and number of successful transaction have no significant impact on credit risk. The regression coefficients of platform consumption frequency, consumption scene, and active days are −0.14, −0.20, and −0.20, respectively, which are significant at the 1% significant level. It shows that the larger the three variables, the smaller the credit risk. Model 4 explores the relationship between all consumption information obtained and the borrower's credit risk, and the empirical results are consistent with the above analysis. Model 5 discusses the relationship between all independent variables and the borrower's credit risk. State of marriage and rate are positively correlated with the borrower's credit risk at a significant level of 1%. Gender is negatively correlated with the borrower's credit risk at a significant level of 1%-free network, and the peak of the associated credit risk appears between $t = 5 - 10$, which reduces the transmission speed of the associated credit risk. It can be seen that the existence of exposed subjects has a significant delaying effect on the arrival of the peak of the related credit risk infection. When the coefficient of latent transformation $\alpha = 0.2$ and $\alpha = 0.5$, the contagion scale of the related credit risk is less than 0.6. The smaller the coefficient of latent transformation is, the smaller the contagion scale is. It can be seen that the existence of latent subjects has a significant inhibitory effect on the contagion scale of the related credit risk. The smaller the coefficient of latent transformation, the stronger the inhibition. In conclusion, when there are exposed subjects in the network, the latent transformation coefficient has a significant delaying and inhibiting effect on the peak period and the scale of infection of the risk of the associated credit.

4.2.2. Robust Test. In the above empirical analysis, customers overdue for more than 30 days are selected as the definition condition of "bad" customers. In order to verify whether the above empirical analysis results depend on the definition conditions of dependent variables, a robust test will be carried out. Taking "overdue days more than 5 days"

TABLE 3: Results of logistic regression.

Dependent variable	(1)	(2)	(3)	(4)	(5)
Consumption ability			0.5692 (0.1902)	0.0915 (0.1956)	0.1285 (0.1990)
Consumption frequency			−0.1414*** (0.0575)	−0.1747*** (0.0578)	−0.0376** (0.0658)
Consumption scene			−01989*** (0.0663)	−0.2550*** (0.0667)	−0.1554*** (0.0682)
Days of transaction			0.0619 (0.1138)		0.0498 (0.1152)
Amount of successful transaction			0.0352 (0.0575)		0.0182 (0.0592)
Number of successful transaction			−0.1812 (0.1155)		−0.1371 (0.1170)
Active days			−0.1951*** (0.0341)		−0.1929*** (0.0355)
Average quota utilization		0.2203*** (0.0333)		0.2076*** (0.0333)	0.2004*** (0.0332)
Maximum quota utilization		0.0273 (0.0289)		0.0322 (0.0290)	0.0333 (0.0290)
Average quota	−0.1331** (0.0493)			−0.1262** (0.0493)	−0.1263** (0.0494)
Number	0.0813** (0.0401)			0.0829** (0.0401)	0.0875** (0.0402)
Total amount	0.1137*** (0.0365)			0.1126*** (0.0366)	0.1139*** (0.0367)
Age	0.0897 (0.0586)	0.1030* (0.0580)	0.00083(0.0609)	0.0712 (0.0614)	−0.0757 (0.0614)
Gender	−0.3602*** (0.0476)	−0.3221*** (0.0471)	−0.0360*** (0.0473)	−0.3578*** (0.0480)	−0.3358*** (0.0481)
State of marriage	0.2462*** (0.0369)	0.2028*** (0.0338)	0.0772** (0.0353)	−0.2550*** (0.0387)	0.1029*** (0.0390)
Job	0.0192*** (0.00884)	−0.0221** (0.00837)	−0.0272(0.00831)	−0.00755 (0.00908)	0.00409 (0.00915)
Amount	0.4068* (0.308)	0.3927* (0.2731)	0.4048* (0.2901)	0.3917* (0.2721)	0.3927* (0.2705)
Rate	0.1320*** (0.00733)	0.1320*** (0.00731)	0.1219*** (0.00734)	0.1205*** (0.00737)	0.1218*** (0.00741)
Observations	50000	50000	50000	50000	50000
Pseudo R^2	0.032	0.033	0.033	0.032	0.030

Note.***,**, and* represent the significance levels of 1%, 5%, and 10%, respectively.

and "overdue days more than 60 days" as dependent variables, empirical tests are carried out for the above models 1, 4, and 5, respectively.

Table 4 shows the logistic regression results of different dependent variables. Consistent with the results of the above main regression, platform consumption information, platform transaction information, and credit card use information are significantly correlated with the borrower's credit risk. Credit card credit information is also significantly related to the borrower's credit risk. Based on the above analysis, the introduction of borrower consumption information on the basis of traditional credit information can better identify credit risk.

4.2.3. The Promotion Effect of Consumption Information on the Model. LightGBM algorithm can predict personal credit risk effectively and rank the importance of indicators. ROC

curve can predict the accuracy of the model effectively [37–40]. AUC is the area under the ROC curve, which can explain the prediction accuracy of the model. In order to compare the fitting degree of the two models, we use AUC index to evaluate the model quality [41, 42]. In this study, LightGBM algorithm is used to predict the model results of the above models 1 and 5, respectively, and draw the ROC curve. The AUC index results are shown in Figures 1 and 2. Figure 1 shows the prediction results of model 1 test samples. When there is only traditional credit information, the AUC value is 0.70. Figure 2 shows the prediction results of model 5 test samples, with AUC value of 0.81. It can be seen that the introduction of consumption information on the basis of traditional credit information can improve the accuracy of the model effectively, and the addition of consumption information improves the prediction ability of the model significantly. LightGBM algorithm is implemented by Python 3.7.

TABLE 4: Results of robust test.

Dependent variable	od5	od5	od5	od60	od60	od60
Consumption ability			0.1929 (0.1854)		0.2792 (0.2025)	0.1387 (0.2124)
Consumption frequency			−0.0335 (0.0600)		−0.1464** (0.0630)	−0.0538 (0.0722)
Consumption scene			−0.1096*** (0.0629)		−0.2543*** (0.0715)	−0.2077*** (0.0736)
Days of transaction			−0.0211 (0.1038)			0.1716 (0.1290)
Amount of successful transaction			0.0204 (0.0540)			0.0164 (0.0647)
Number of successful transaction			0.0204 (0.0540)			0.2911 (0.1305)
Active days			−0.1913*** (0.0324)			−0.1889*** (0.0388)
Average quota utilization		0.1946*** (0.0307)	0.1946*** (0.0307)		0.2030*** (0.0362)	0.1802*** (0.0361)
Maximum quota utilization		0.0526** (0.0262)	0.0526** (0.0262)		0.0311 (0.0318)	0.0385 (0.0319)
Average quota	−0.0827* (0.0449)		−0.0827* (0.0449)		−0.1055** (0.0537)	−0.0983* (0.0538)
Number	0.0876** (0.0368)		0.0876** (0.0368)		0.0952** (0.0439)	0.1013** (0.0441)
Total amount	0.0661** (0.0335)		0.0661** (0.0335)		0.1099*** (0.0400)	0.1099*** (0.0402)
Age	0.0838 (0.0534)	0.00529 (0.0553)	0.0616 (0.0557)	0.1158* (0.0645)	0.0194 (0.0671)	0.0629 (0.0677)
Gender	−0.3545*** (0.0433)	−0.3609*** (0.0430)	−0.3323*** (0.0438)	−0.3558*** (0.0525)	−0.3327*** (0.0521)	−0.3303*** (0.0531)
State of marriage	0.2344*** (0.0337)	0.0922* (0.0321)	−0.1096* (0.0629)	0.2537*** (0.0405)	0.0667* (0.0387)	0.1074** (0.0429)
Job	0.0208*** (0.00806)	−0.0254*** (0.00759)	−0.006*** (0.00831)	−0.0292*** (0.0096)	−0.0405 (0.00902)	−0.0128*** (0.00997)
Amount	0.4027* (0.301)	0.4053* (0.231)	0.4058* (0.2581)	0.3902* (0.2625)	0.3927* (0.2705)	0.4058* (0.2581)
Rate	0.1209*** (0.00665)	0.1109*** (0.00666)	0.1109*** (0.00673)	0.1524*** (0.00812)	0.1430*** (0.00816)	0.1428*** (0.00823)
Observations	50000	50000	50000	50000	50000	50000
Pseudo R2	0.0313	0.0311	0.03105	0.0312	0.0301	0.03109

Note.***,**, and* represent the significance levels of 1%, 5%, and 10%, respectively.

4.2.4. Feature Importance and SHAP Value of Consumption Information. In order to explore the importance of borrower consumption information and traditional credit information in personal credit risk evaluation, we use LightGBM algorithm and Shapley additive explanation to build a model to rank the feature importance and the impact of consumption information of the platform, transaction information of the platform, and information of credit card. The 12 variables are sorted by importance to generate an importance SHAP value plot, as shown in Figure 3.

According to the ranking results of feature importance and SHAP value, among the top five indicators in the ranking of importance, five consumption information indicators are in 5 places. This also confirms the value and effect of consumer information in personal credit risk evaluation. In the SHAP value, we can see consumption ability and average quota utilization have a positive impact on credit risk, and consumption ability has a strong correlation with credit risk. Active days, days of transaction, and maximum quota utilization have a negative impact on credit risk, and maximum quota utilization has a strong correlation with credit risk. It not only supplements the

indicators beyond the traditional credit information, but also improves the accuracy of personal credit risk evaluation.

5. Discussion

In this study, we investigate the previous studies on personal credit risk evaluation. We also categorize the evaluation of personal credit risk from the perspective of indicators and models. The findings are as follows. First, in many evaluation methods literature, personal credit risk evaluation is considered as a modeling problem. Second, in the selection of personal credit risk evaluation indicators, personal consumption information is rarely or often ignored. Thirdly, in many cases, machine learning method can be used as a supplement to logistic regression.

Our findings also suggest the effect of consumer information in personal credit risk assessment. First, personal consumer loans will affect customers' intertemporal consumption behavior. Second, "soft information," which is different from customer's basic information, is equally important for identifying personal credit risk. Finally, through the empirical evaluation of personal credit risk model, we can find
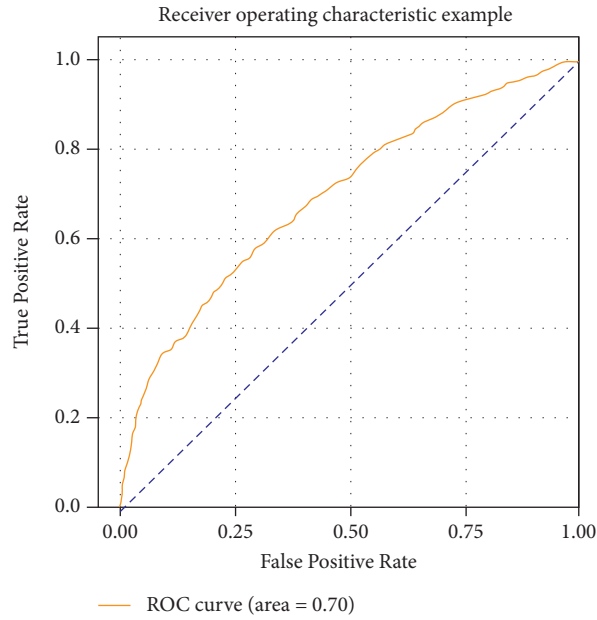
Receiver operating characteristic example

FIGURE 1: Prediction results of model 1.
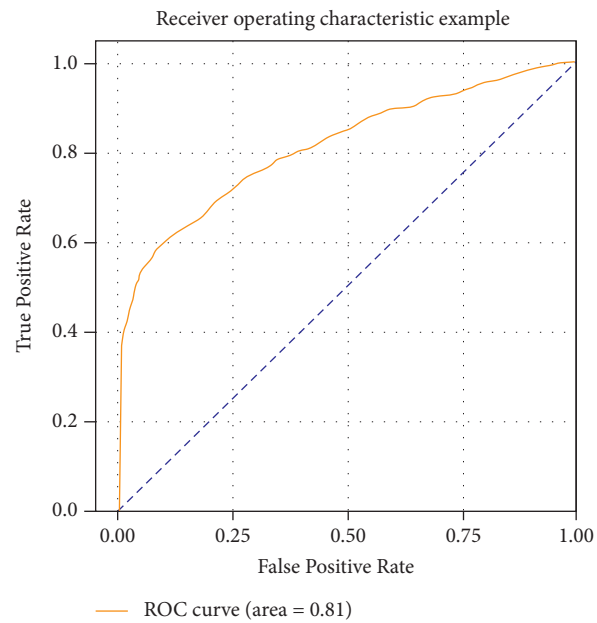
Receiver operating characteristic example

FIGURE 2: Prediction results of model 5.

more accurate consumer credit value. Of course, the prevention of personal credit risk not only stays in the evaluation, but should be able to provide clues on how to avoid default.

This study has several limitations. We do not cover all the consumer information of borrowers. Information that may affect customer consumption is not considered. Although we sort out the literature on personal credit risk assessment, we do not cover all the literature, but introduce the representative research in each aspect. Through this concentration, we come to meaningful findings and insights related to this topic.
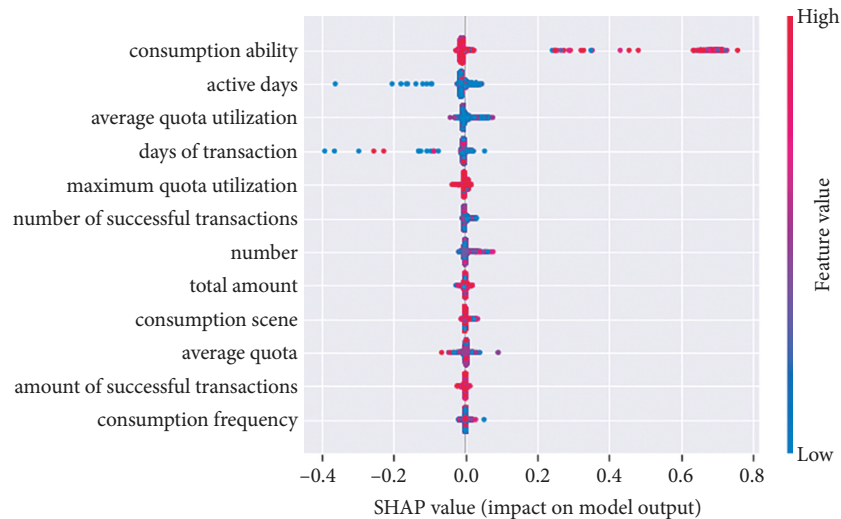
FIGURE 3: Feature importance and SHAP value plot.

6. Conclusions

Through the empirical analysis of personal credit risk evaluation, this paper describes the characteristics and influencing factors of consumption information and proves the value and effect of consumption information in personal credit risk evaluation. In order to alleviate the long-term credit constraints, it is necessary to promote the development of the credit market on the premise of meeting the loan needs of tail customers. With the strong promotion of financial technology, Internet credit products came into being. However, the current credit investigation system cannot cover all the information of loan demanders, and the credit risk evaluation results are inaccurate. Therefore, how to evaluate credit risk effectively has become an urgent problem to be solved. This requires financial institutions to seek information that can widely cover all loan demanders objectively. Different from the traditional credit information, consumer information has the characteristics of easy access and prominent preference, which can be used as an effective supplement to the traditional credit information. Evaluation of the effect of consumer information effective has become the focus of attention in personal credit risk evaluation.

This paper selects the borrower information from a Chinese platform that has both online shopping and consumer loan as a sample. We use logistic regression model, LightGBM algorithm, and Shapley additive explanation to analyze the value and effect of traditional credit information and consumption information in personal credit risk evaluation. The conclusions are as follows: the information of credit card use can predict the borrower's repayment behavior effectively and provide effective support for personal credit risk evaluation. The consumption information of the platform and the transaction information of the platform can predict the borrower's repayment behavior effectively and provide effective support for personal credit risk evaluation. The predictive model accuracy of ROC curve has proved that adding consumption information to the model can improve the accuracy of the model effectively. The importance of model variables is ranked, and there are 5

consumption indicators in the first 5 indicators, which further verifies the value and effect of consumption information in personal credit risk evaluation. Shapley addition explanation also confirmed the influence and contribution of consumption information to personal credit risk. This study not only reveals the effect and value of consumer information in personal credit risk evaluation effectively, but also provides new ideas for the development of consumer financial market.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] J. Duarte, S. Siegel, and L. Young, "Trust and credit: the role of appearance in peer-to-peer lending," *Review of Financial Studies*, vol. 25, no. 8, pp. 2455–2484, 2012.

[2] M. Y. Song, "A literature review on the latest research of consumer finance in China. Contemp," *Contemporary Economic Management*, vol. 37, no. 4, pp. 6–12, 2015.

[3] Z. Wang, C. Zhou, L. Liao, and W. Zhang, "Informational content of consumption behavior in consumer credit risk evaluation," *Economic Research Journal*, vol. 55, no. 1, pp. 149–163, 2020.

[4] J. E. Stiglitz and A. Weiss, "Credit rationing in markets with imperfect information," *The American Economic Review*, vol. 71, no. 3, pp. 393–410, 1981.

[5] A. Sufi, "Information asymmetry and financing arrangements: evidence from syndicated loans," *The Journal of Finance*, vol. 62, no. 2, pp. 629–668, 2007.

[6] M. Klafft, "Peer to peer lending: auctioning microcredits over the internet," in *Proceedings of the International Conference on Information Systems*, Technology and Management, Dubai, March 2008.

[7] G. Zhou, D. Ruan, and C. Fan, "Loan-reduction and crowding-out effects in asymmetric-information credit markets," *Finance Forum*, vol. 26, no. 1, pp. 25–36, 2021.

[8] J. C. Stein, "Information production and capital allocation: decentralized versus hierarchical firms," *The Journal of Finance*, vol. 57, no. 5, pp. 1891–1921, 2002.

[9] S. Li and X. Ji, "The application of the Lasso-Cox model in personal credit risk assessment," *Resource Development & Market*, vol. 37, no. 2, pp. 129–135, 2021.

[10] T. Hildebrand, M. Puri, and J. Rocholl, "Adverse incentives in crowdfunding," *Management Science*, vol. 63, no. 3, pp. 587–608, 2017.

[11] M. Lin, N. R. Prabhala, and S. Viswanathan, "Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending," *Management Science*, vol. 59, no. 1, pp. 17–35, 2013.

[12] Y. Su and C. Cheng, "An empirical study on the influencing factors of P2P online borrowers' default Behavior," *Journal of Financial Development Research*, vol. 1, pp. 70–76, 2017.

[13] G. Chi, Y. Zhang, and B. Shi, "The debt rating for small enterprises based on Probit regression," *Journal of Management Sciences in China*, vol. 19, no. 6, pp. 136–156, 2016.

[14] R. Emekter, Y. Tu, B. Jirasakuldech, and M. Lu, "Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending," *Applied Economics*, vol. 47, no. 1, pp. 54–70, 2015.

[15] M. J. Garmaise and G. Natividad, "Consumer default, credit reporting, and borrowing constraints," *The Journal of Finance*, vol. 72, no. 5, pp. 2331–2368, 2017.

[16] Y. Q. Xu and M. Q. Pan, "Application of analytic hierarchy process and support vector machine in personal credit evaluation," *Chinese Journal of Management Science*, vol. 24, no. S1, pp. 106–112, 2016.

[17] J. A. Ohlson, "Financial ratios and the probabilistic prediction of bankruptcy," *Journal of Accounting Research*, vol. 18, no. 1, pp. 109–131, 1980.

[18] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, Wiley, New Jersey, USA, 2nd ed edition, 2000.

[19] G. Kordas, "Credit scoring using binary quantile regression," in *Statistical Data Analysis Based on the L1-Norm and Related Methods,"*, Y. Dodge, Ed., Springer, Berlin, Germany, pp. 125–137, 2002.

[20] M. Du, L. Li, and X. Zhang, "Research on P2P credit risk prediction based on two-step subsampling algorithm," *Journal of Systems Science and Mathematical Sciences*, vol. 41, no. 2, pp. 566–576, 2021.

[21] X. Liu and W. Wang, "A comparative study on three models for credit risk identification in Chinese commercial banks," *Economic Survey*, vol. 32, no. 6, pp. 132–137, 2015.

[22] Z. Khemais, D. Nesrine, and M. Mohamed, "Credit scoring and default risk prediction: a comparative study between discriminant analysis & logistic regression," *International Journal of Economics and Finance*, vol. 8, no. 4, pp. 39–53, 2016.

[23] T. Cheng and Y. Wen, "Probability of default measurement study for individual client of margin trading and securities lending business," *Journal of Financial Research*, vol. 41, no. 4, pp. 174–189, 2016.

[24] F. Luo and X. Chen, "Credit risk assessment of personal small loan based on logistic regression model and its application," *The Theory and Practice of Finance and Economics*, vol. 38, no. 1, pp. 30–35, 2017.

[25] S. Hu, H. Lei, and H Hu, "Research on credit risk measurement of China's real estate enterprises based on logistic model," *China Soft Science*, vol. 12, pp. 157–164, 2018.

[26] M. Jiang, P. Xu, X. Ren, and K. Che, "Research on algorithms development and optimization for personal credit scoring," *Journal of Harbin Institute of Technology*, vol. 47, no. 5, pp. 40–45, 2015.

[27] M. Verleysen and D. Francois, "The curse of dimensionality in data mining and time series prediction," in *Proceedings of the Neural Networks*, pp. 758–770, Barcelona, Spain, June 2005.

[28] M. E. Houle, H.-P. Kriegel, P. Kröger, E. Schubert, and A. Zimek, "Can shared-neighbor distances defeat the curse of dimensionality," in *Proceedings of the international Conference on Scientific and Statistical Database Management*, pp. 482–500, SSDBM), Heidelberg, Germany, January 2010.

[29] M. Bertrand and E. Kamenica, "Coming apart? cultural distances in the United States over time," in *Proceedings of the NBER Working Pap*, Chicago, IL, USA, June 2018.

[30] S. Belenzon, A. K. Chatterji, and B. Daley, "Eponymous entrepreneurs," *The American Economic Review*, vol. 107, no. 6, pp. 1638–1655, 2017.

[31] J. Guzman and S. Stern, *The State of American Entrepreneurship: New Estimates of the Quantity and Quality of Entrepreneurship for 15 US States, 1988-2014*, NBER Working Pap, Cambridge USA, 2016.

[32] M. Pagano and T. Jappelli, "Information sharing in credit markets," *The Journal of Finance*, vol. 48, no. 5, pp. 1693–1718, 1993.

[33] S. Djankov, C. McLiesh, and A. Shleifer, "Private credit in 129 countries☆," *Journal of Financial Economics*, vol. 84, no. 2, pp. 299–329, 2007.

[34] T. Beck, A. Demirguc-Kunt, and P. Honohan, "Access to financial services: measurement, impact, and policies," *The World Bank Research Observer*, vol. 24, no. 1, pp. 119–145, 2009.

[35] M. Brown, T. Jappelli, and M. Pagano, "Information sharing and credit: firm-level evidence from transition countries," *Journal of Financial Intermediation*, vol. 18, no. 2, pp. 151–172, 2009.

[36] T. Berg, V. Burg, A. Gombović, and M. Puri, "On the rise of fintechs – credit scoring using digital footprints," *Review of Financial Studies*, vol. 33, no. 7, 2018.

[37] H. Noh, T. Roh, and I. Han, "Prognostic personal credit risk model considering censored information," *Expert Systems with Applications*, vol. 28, no. 4, pp. 753–762, 2005.

[38] S. Agarwal, P. M. Skiba, and J. Tobacman, "Payday loans and credit cards: new liquidity and credit scoring puzzles?" *The American Economic Review*, vol. 99, no. 2, pp. 412–417, 2009.

[39] Z. Mo, C. F. Zhang, W. Wei, D. C. You, and S. Zhang, "Research on personal credit risk assessment method based on bagging integration," *Systems Engineering*, vol. 37, no. 1, pp. 143–151, 2019.

[40] J. Mallick, S. Alqadhi, S. Talukdar et al., "Risk assessment of resources exposed to rainfall induced landslide with the development of GIS and RS based ensemble metaheuristic machine learning algorithms," *Sustainability*, vol. 13, no. 2, p. 457, 2021.

[41] K. Ye, "Research on personal credit risk assessment based on machine learning algorithm," *International Journal of Intelligence Science*, vol. 6, no. 1, pp. 146–156, 2020.

[42] M. Wang and H. Yang, "Research on personal credit risk assessment model based on instance-based transfer learning," *International Journal of Intelligence Science*, vol. 11, no. 01, pp. 44–55, 2021.