

# A Credit Risk Evaluation Index System Establishment of Petty Loans for Farmers Based on Correlation Analysis and Significant Discriminant

Linpeng Hai

Computer Science and Technology Henan Polytechnic University, Jiaozuo, China  
Email: hailp@hpu.edu.cn

Baofeng Shi

School of Business Management, Dalian University of Technology, Dalian, China  
Email: fengbei\_wuyu@163.com

Guorong Peng

School of Preparatory Education, Hubei University for Nationalities, Hubei, China  
Email: 158824939@qq.com

**Abstract** — By the end of 2011, agricultural population in accounted for 48.73% in China. Also, farmers in China are dispersed and their financial information is incomplete, which leads to credit risk evaluation system of farmers in China is not sound at all. Most Chinese banks even have not established the rating system, so farmer credit risk rating system pressed for solution. This study proposes an index screening method based on correlation analysis and significant discriminant. The proposed method is demonstrated using the 2044 petty loans for farmers of a Chinese state-owned commercial bank's. The results demonstrate that the proposed method can accurately screen the indicators, which can effectively distinguish default customers from non-default ones. Moreover, the results of an empirical study showed that "Loan purpose", "Expenses/ incomes", "Increasing rate of regional GDP" are the key indicators to distinguish default customers from non-default ones.

**Index Terms** — petty loans for farmers, credit risk evaluation, index screening; significant discriminant

## I. INTRODUCTION

By the end of 2011, agricultural population in accounted for 48.73% in China [1]. Also, farmers in China are dispersed and their financial information is incomplete, which leads to credit risk evaluation system of farmers in China is not sound at all. Most Chinese banks even have not established the rating system, so farmers' credit risk rating problems need to be solved urgently.

Research the credit rating system of petty loans for farmers has the significant meanings. Firstly, it provides the basis for measuring default risk of petty loans and

loan pricing based on the credit rating. Secondly, offering better financial support and service for the solution of "Three Rural Issues" including countryside, agriculture, farmers and development of primary industry.

According to the characteristics of petty loans for small private business, the establishment of credit risk evaluation index system of petty loans for farmers is to delete the indicators of information repeated and to reserve the indicators which have an obvious effect on default state.

### A. Present situation of evaluation index system of petty loan for farmers

FICO credit rating model of American Credit Bureau which evaluates customers' credit status consists of 5 aspects, such as historical records of customers' paying credit, the time length of customers building credit records [2]. The customer credit rating "5C principle", which is generally accepted in financial circle, is one of the main methods for credit rating [3]. It rates debtor's repayment will and capacity from 5 aspects: Character, Capital, Capacity, Collateral and Condition of Business. Moody's [4], Standard & Poor's [5] and Fitch Ratings [6] established the enterprise credit rating systems, including the liquidity ratio, return on assets and so on. Citibank built comprehensive credit rating system from the quality of borrowers, operation capacity, management ability and other aspects, and evaluates enterprise's credit [7].

There is the present situation of evaluation index systems of the Chinese authoritative organizations. Agricultural Bank of China established the rating indicators system for farmers which including 9 indicators, such as age, health status, household incomes status and so on [8]. The credit rating system for farmers of Postal Savings Bank of China established the credit

---

Corresponding Author: Baofeng Shi

rating system from 4 aspects, such as family structure, repayment willingness, and solvency [9]. The rating system has 15 indicators including age, credit status, family's per capita net income each year and others. Rural Credit Cooperatives of Sichuan Province established the rating indicators system for farmers including 12 indicators, such as marriage, health, income status and so on [10]. Credit rating indicators system for corporate customers of China Construction Bank including basic quality, credit records and so on [11]. Small business credit risk evaluation index system for corporate customers of Industrial and Commercial Bank of China includes asset-liability ratio and other indicators [12].

There is the present situation of evaluation indicators systems by sorting academic literatures. R. Mahjabeen (2008) obtained the influence of loan purpose, the value of durable goods and other factors on the risk of petty loans [13]. K. Carling (2007) established credit risk evaluation model, consist of macroeconomics factors, such as GDP growth rate and unemployment rate [14]. J. Copestake (2007) got that gender, age, family net assets and other indicators had an obvious effect on credit risk evaluation of petty loans for small private business, by using questionnaire survey to financial institutions [15].

#### *B. Present situation of evaluation index screening and evaluation methods*

Using a weighted least squares SVM, L. Yu et al. (2011) studied the classification of default customers and non-default customers in credit risk evaluation [16]. A. M. Alam et al. (2010) have observed that the kernel canonical correlation analysis performs better than the robust canonical correlation analysis in object recognition and index screening [17]. Y. Xu et al. (2009) saved weighted-more indicators as credit evaluation index system for on-line electronic businesses by using AHP and SPA method [18]. X. Zhou et al. (2010) studied the identification of key indicators about customer defaults or not by using nearest subspace method [19]. Using saliency and clustering analysis, Q. Zhang et al. (2010) studied the object extraction [20]. Wei et al. (2009) studied the risk evaluation model of high-technology based on TOPSIS method [21]. Gong et al. (2011) studied the feature classification applying in a three-dimensional canonical correlation analysis [22].

Although the existing research has made great progress, there are some disadvantages. Credit risk evaluation system of farmers is not sound at all. Most banks in China even have not yet established the rating system. Besides, the existing index systems have drawbacks that having repeated information and being unable to effectively distinguish default customers from non-default ones.

The purpose of this paper is to set up a credit risk evaluation index system of petty loans for farmers, which is based on correlation analysis and significant discriminant, by using a Chinese state-owned commercial bank's 2044 petty loans for farmers.

## II. CONSTRUCTING PRINCIPLE OF CREDIT RISK EVALUATION INDEX SYSTEM OF PETTY LOANS FOR FARMERS

### *A. Characteristics of petty loans for farmers*

Because of the unsoundness of financial information of petty loan for farmers, it is difficult to find the classic indicators, credit rating theory to describe the credit status of farmers.

Large amount of farmers' loans, small amount for single loan, and subjective loan review was time-consuming, so it's necessary to establish a credit rating system of petty loans for farmers.

Regional economic development has significant impact on the solvency of farmers. As the difference in natural environment and national policy, economic development levels are uneven. Then, the economic development in different regions is an important factor in the repayment of farmers.

### *B. The difficulties of this issue*

Farmers in China are dispersed and their financial information is unsound. Credit risk evaluation system of farmers is not complete. Most banks in China even have not established the rating system, so farmers credit risk rating problem need to be solved urgently. Secondly, how can choose the indicators which can distinguish default and non-default significantly? Among indicators of Agricultural Bank of China, household property status, credit records of bank and other indicators can not effectively distinguish default and non-default. Among indicators of Postal Savings Bank of China, property records, business transactions and other indicators can not effectively distinguish default and non-default. The existing indicators can not effectively distinguish default and non-default, and default customers often have a high credit score. So, it brings certain difficulties whether to grant loans.

### *C. The ideas to solve the difficulties*

On the basis of the available indicators from a credit risks evaluation and loan pricing system for petty loans of a Chinese national commercial bank, this paper establishes an extensive index system of petty loans for farmers based on the high-frequency indicators of the authoritative institutions, such as S&P, FICO [2-12]. Then, by using the correlation analysis to delete the indicators of information repeated and the significant discriminant to select the indicators which can effectively distinguish default customers from non-default ones, this paper sets up the index system of petty loans for farmers.

## III. THE ESTABLISHMENT OF INDEX SYSTEM OF PETTY LOANS FOR FARMERS

### *A. The establishment of extensive index system*

This paper selects extensive indicators on the basis of the available indicators from a Chinese national commercial bank [23]. Then, this paper selects 65 indicators of petty loans for farmers, which includes 5

criterion layers, i.e. “Basic information”, “Repayment ability”, “Repayment willing”, “Guarantee and joint guarantee”, “Macro environment”, as shown in Column 1, 2, 5 of Table I. The indicator’s literature sources are listed in Column 3 and 6 of Table I.

On the basis of the principle that data can be observed, we removed 21 unavailable indicators, such as “Area ratio of disposable assets”, “Total property”. Other 44 indicators are left. The deleted indicators are marked with “Unavailability delete” in Column 7 of Table I.

TABLE I.

Extensive Index Set of Farmers Credit Rating

(1) Criterion layers	(2) Indicator layers	(3) References	(4) Screening result	(5) Indicator layers	(6) References	(7) Screening result	
Basic information	Loan purpose	[9][11][13]	Reserved	Supporting population	[9-10]	Deleted by significance	
	Age	[2-6][8-11][15]		Family number/labor force	[9]		
	Value of house owning	[8-12]		Number of members	[8][10]		
	Marital status	[9][11]	Deleted by significance	Number of labor force	[2][10]	Deleted by correlation analysis	
	Gender	[9][11-12][15]		House value	[2][8]		
	Education background	[9][11-12]		...	...		...
Household expenses	[9][11]	Reserved	Area ratio of disposable assets	[9]	Unavailability delete		
Repayment ability	Expenses/ incomes		[2-5][7][9]	Agricultural producing incomes	[8-10]	Deleted by correlation analysis	
	Non-agricultural incomes/total incomes		[8-10]	Net agricultural incomes	[8-10]	Deleted by significance	
	Net income of borrower		[8-10]	Total expenses	[12]		
	Net income each year / per capita GDP		[9][11]	Deleted by significance	Expense of family’s daily life	[9]	Deleted by significance
	Education cost of children each year		[9-10]		...	...	
	Agricultural producing expenses	[9][11-12]	Total property		[15]	Unavailability delete	
Repayment willings	Private loans	[9][11-12]	Reserved	Loaning records of borrower	[2][8-12]	Deleted by significance	
	Residential stability	[9][11][14]		...	...		...
	Residential status	[12]	Deleted by significance	Social reputation status	[9][11]	Unavailability delete	
Guarantee and joint guarantee	Strength of guarantor	[7][9-10]	Reserved	Age of guarantor	[9-10][12]	Deleted by significance	
	Education background of guarantor	[7][9]		...	...		...
	Gender of guarantor	[9-12]	Deleted by significance	Credit status of joint guarantor	[9]	Unavailability delete	
	Marital status of guarantor	[9-11]		Deleted correlation analysis	—		—
Macro environment	Engel's coefficient	[9][11][14]	Reserved	Regional government policy	[10][12]	Deleted by correlation analysis	
	Increasing rate of regional GDP	[9][10][14]		...	...		...
	CPI	[9][10]		Technical support efforts	[8][10]	Unavailability delete	
	Per capita agricultural output value	[9][11]		—	—	—	

1) Data standardization

(1) The standardization of positive indicators

The positive indicators are indicators whose values are the bigger, the better, such as “Value of house owning”.

Let  $v_{ij}$  denote the indicator data of the  $i^{th}$  customer and the  $j^{th}$  indicator. Let  $x_j^i$  denote the standardization of the  $i^{th}$  customer and the  $j^{th}$  indicator. Let  $n$  denote the number of customers. Then [24]

$$x_j^i = \frac{v_{ij} - \min_{1 \leq s \leq n}(v_{sj})}{\max_{1 \leq s \leq n}(v_{sj}) - \min_{1 \leq s \leq n}(v_{sj})} \quad (1)$$

(2) The standardization of negative indicators

The negative indicators are indicators whose values are the smaller, the better, such as “Household expenses /

household incomes”. The standardization of negative indicators is as follows.

$$x_j^i = \frac{\max_{1 \leq s \leq n}(v_{sj}) - v_{ij}}{\max_{1 \leq s \leq n}(v_{sj}) - \min_{1 \leq s \leq n}(v_{sj})} \quad (2)$$

(3) The standardization of interval indicators

The interval indicators are indicators which are reasonable only when they lie in certain intervals, such as “Consumer price index” and “Age”.

The ideal interval of “Age” is [31, 45] “Reference [9]”. The repayment ability and repayment willingness of customers are strong in the interval.

The ideal interval of “Consumer price index” is [101,105] “Reference [24]”. There is neither inflation

nor deflation existing within this interval.

Let  $x_j^i$  denote the standardization of the  $i^{th}$  customer and the  $j^{th}$  indicator. Let  $v_{ij}$  denote the indicator data of the  $i^{th}$  customer and the  $j^{th}$  indicator. Let  $n$  denote the number of customers. Let  $q_1$  denote the left boundary of the ideal interval. Let  $q_2$  denote the right boundary of the ideal interval. Then [24]

$$x_j^i = \begin{cases} 1 - \frac{q_1 - v_{ij}}{\max(q_1 - \min_{1 \leq i \leq n}(v_{ij}), \max_{1 \leq i \leq n}(v_{ij}) - q_2)}, & v_{ij} < q_1 \quad (a) \\ 1 - \frac{v_{ij} - q_2}{\max(q_1 - \min_{1 \leq i \leq n}(v_{ij}), \max_{1 \leq i \leq n}(v_{ij}) - q_2)}, & v_{ij} > q_2 \quad (b) \\ 1 & , q_1 \leq v_{ij} \leq q_2 \quad (c) \end{cases}$$

(4) The standardization of qualitative indicators

The scoring standard of qualitative indicators is obtained using rational analysis for qualitative indicators, which is shown in Column 2 to 6 of Table II.

Firstly, we adopted the 1 and 0 scoring method to score 5 indicators, such as “Private loan” which has only two cases. Secondly, we adopted the equidistant scoring method to score 10 indicators, such as “Number of members”, “Loan purpose”, “Number of labor force”. Thirdly, we adopted the unequal scoring method to score 4 indicators, such as “Education background”, “Marital status”, “Gender”.

TABLE II.

The Scoring Criteria of Qualitative Indicators

(1) Serial number	(2) Criterion layers	(3) Indictors layers	(4) options number	(5) Options	(6) Scoring
1	Basic information	Education background	1	Undergraduate and above	1.00
2			2	Junior college	0.90
3			3	High school and technical secondary school	0.60
4			4	Junior high school	0.40
5			5	Primary school	0.20
6			6	Other	0.00
...	...	...	...	...	...
66	Guarantee and joint guarantee	Guarantee or not	1	Yes	1.00
67			2	No	0.00
...		...	...	...	...
70	Guarantee and joint guarantee	Group membership of Co-guarantee	1	Friendly relations, associating frequently, very familiar, business partners or neighbors	1.00
71			2	Ordinary relations, a little familiar	0.80
72			3	Not know	0.50

2) The first indicator screening using correlation analysis

The aim of correlation analysis is to delete the indicators of large correlation from the whole extensive index system, avoiding repeated information.

Let  $x_j^i$  denote the standardization of the  $i^{th}$  customer and the  $j^{th}$  indicator. Let  $\bar{x}_j$  denote the mean value of the  $j^{th}$  indicator. Let  $r_{jk}$  denote the correlation coefficient between the  $j^{th}$  indicator and the  $k^{th}$  indicator. Then [25]

$$r_{jk} = \frac{\sum_{i=1}^n (x_j^i - \bar{x}_j)(x_k^i - \bar{x}_k)}{\sqrt{\sum_{i=1}^n (x_j^i - \bar{x}_j)^2 (x_k^i - \bar{x}_k)^2}} \quad (4)$$

As a matter of experience, the threshold of correlation coefficient  $M$  equals 0.80 [25]. In other words, if the absolute value of the correlation coefficient  $r_{jk}$  of two indicators is more than 0.8, then the two indicators reflect the repeated information and one of them can be deleted. For example, this paper deleted 4 indicators based on correlation analysis screening in the following empirical study, such as “House value”.

3) The second indicator screening using significant discriminant

The aim of significant discriminant is to delete the indicators of little influence on default status after the first indicator screening, ensuring that the reserved indicators can effectively distinguish default customers from non-default ones.

Let  $a$  and  $c_j$  denote regression coefficients. Let  $m$  denote the number of indicators. Let  $x_j^i$  denote the standardization of the  $i^{th}$  customer and the  $j^{th}$  indicator. Let  $\varepsilon_i$  denote random error. Let  $y_i$  denote the default status of the  $i^{th}$  customer,  $y_i$  equal to 0 denoting the  $i^{th}$  customer a non-default customer and  $y_i$  equal to 1 denoting the  $i^{th}$  customer a default customer. Then, the Probit regression function between default status  $y_i$  and indicators  $x_j^i$  is as follows [25].

$$y_i = a + \sum_{j=1}^m c_j x_j^i + \varepsilon_i \quad (5)$$

Constructing the Probit regression function between evaluation indicators  $x_j^i$  and default state  $y_i$  of farmers, it is obtained that the bilateral probability  $P_j$  of regression coefficient  $c_j$  for every indicator  $x_j^i$ . It can distinguish whether the indicators  $x_j^i$  has an obvious effect on default state of farmers by comparing the bilateral probability  $P_j$

with the given critical probability  $P_0$ .

According to the references [25], the threshold probability  $P_0$  equals 0.05. If  $P_j < P_0=0.05$  [25], refuse the assumption that the true value of regression coefficient  $c_j$  corresponding to the indicator  $x_j^i$ . It indicates that probability of the true value of regression coefficient  $c_j$  not being zero is at least 95%. In other words, the indicator  $x_j^i$  can significantly distinguish the default statuses and it should be reserved. Conversely, if  $P_j \geq P_0=0.05$ , it indicates that the indicator  $x_j^i$  can not significantly distinguish the default statuses and it should be deleted.

For example, this paper deleted 25 indicators based on significant discriminant screening in the following empirical study, such as “Number of labor force”, “Expense of family’s daily life”.

#### 4) *The reasonability judgment of the established index system*

On the basis of the idea that variance reflects the information content of indicators, whether the established index system is reasonable or not can be decided.

Let  $S$  denote the covariance matrix of indicators data. Let  $trS$  denote the trace of the covariance matrix, i.e. the sum of the main diagonal indicator variance in the covariance matrix. Let  $In$  denote the variance contribution rate of the screened  $s$  indicators to the selected  $h$  extensive indicators. Let  $s$  denote the number of the screened indicators. Let  $h$  denote the number of the selected extensive indicators. Then [24]

$$In = \frac{trS_s}{trS_h} \quad (6)$$

The meanings of the Equation (6): Equation (6) is the ratio of the variance sum  $trS_s$  of  $s$  screened indicators to the variance sum  $trS_h$  of  $h$  extensive indicators. It indicates the ratio that the information content of  $s$  screened indicators accounts for of the information content of the  $h$  extensive indicators.

If the established index system can reflect more than 95% original information by using less than 30% indicators, the screened index system is reasonable [24].

### IV. EMPIRICAL STUDY ON THE INDEX SCREENING MODEL BASED ON PETTY LOANS FOR FARMERS IN CHINA

#### A. *Samples and data source*

##### (1) *Samples*

The samples for empirical study are from a Chinese national commercial bank’s 2 044 petty loan samples for farmers [23]. Obviously, the sample covers the main provinces of China, which can ensure the reliability and applicability of this evaluation index system of petty loans for farmers.

##### (2) *Data source*

The data in the first to 44<sup>th</sup> Row and the first to 2044<sup>th</sup> Column in Table III are from the farmers petty credit loan system of a Chinese national commercial bank headquarter [23]. Owing to having no missing data, we used all data without any adjustment while building

credit risk evaluation index system of petty loans for farmers.

The 2044 farmers consist of 1876 non-default customers and 168 default customers. The default status of each customer is shown in the corresponding Column, the 45<sup>th</sup> Row in Table III.

#### B. *The establishment of index screening model*

##### 1) *The standardization of evaluation indicators data*

###### (1) *Scoring the quantify indicators*

According to the indicator type in Column  $d$  of Table III, take the original data of positive indicators  $v_{ij}$  from Column 1 to 2044 of Table III into Equation (1), the original data of negative indicators  $v_{ij}$  into Equation (2) and the original data of interval indicators  $v_{ij}$  into Equation (3), and then the standardized data of indicators  $x_j^i$  are obtained. The results are shown in Column 2045 to 4088 of Table III corresponding with quantitative indicators.

###### (2) *Scoring the qualitative indicators*

According to the indicator type in Column  $d$  of Table III, standardized scores of qualitative indicators can be obtained in Table III based on the scoring criteria of qualitative indicators in Table II. The results are shown in Column 2045 to 4088 of Table III corresponding with qualitative indicators.

##### 2) *The first indicator screening based on correlation analysis*

Substituting the standardized data of indicators  $x_j^i$  in Column 2045 to 4088 in Table III into Equation (4), the correlation coefficients  $r_{ij}$  of 44 indicators are obtained, as shown in Table IV.

As mentioned in “The first indicator screening using correlation analysis”, if the absolute value of the correlation coefficient  $r_{jk}$  of two indicators is more than the threshold 0.8 [25], it is indicated that the two indicators reflect the repeated information and one of them can be deleted.

Known from the third Row and the 12<sup>th</sup> Column in Table IV, the correlation coefficient of “Value of house owing” and “House value” is 0.991. Because 0.991 is more than the threshold 0.8, it indicates that these two indicators are reflecting repeated information. Because “Value of house owing” reveals the basic information of debtors more than “House value”, it is reasonable to delete “House value”.

In the same way, this paper deleted another 4 indicators. They are “Agricultural producing incomes”, “Marital status of guarantor”, “Regional government policy”. All these 4 deleted indicators are marked with “Deleted by correlation analysis” in Column 4 and 7 in Table I.

In a word, the first screening deleted 4 indicators which reflect repeated information and reserved 40 indicators based on correlation analysis.

##### 3) *The second indicator screening based on significant discriminant*

Substituting the 40 indicators reserved after

correlation analysis in Table I into Equation (5), the Probit regression model between default status  $y_i$  and indicators  $x_j^i$  is obtained as follows.

$$y_i = a + c_1x_1^i + \dots + c_{40}x_{40}^i + \varepsilon_i \quad (7)$$

The parameter  $i$  equals 1, 2, ..., 2044 respectively in Equation (7).

Substituting the standardized data of indicators  $x_j^i$  in Column 2045 to 4088 in Table III into Equation (7), the regression coefficients  $c_j$  of 40 indicators and the corresponding bilateral probability  $P_j$  are obtained, as shown in Column 4 and 5 in Table V.

TABLE III.

The Original Data and Standardized Data of Indicators of Petty Loans for Farmers

(a) Serial number	(b) Criterion layers	(c) Indicators	(d) Indicator type	Original data of indicators				Standardized data of indicators $x_j^i$			
				$v_{ij}$				Non-default customers		Default customers	
				(1) Song Mingshan	(2044) Xu Mingjun	(2045) Song Mingshan	(3920)Liu Dachun	(3921)Chen Fugui	(4088)Xu Mingjun		
1	Basic information	Loan purpose	Qualitative	3	3	0.600	0.600	0.600	0.600		
14		House value	Positive	0.130	0.000	0.39	0.33	0.31	0.29		
15	Repayment ability	Expenses/incomes	Positive	18.989	8.124	1.000	0.887	0.000	0.123		
23		Expense of family's daily life	Negative	120	8	0.737	0.534	1.000	0.001		
24	Repayment willings	Private loans	Qualitative	1.12	0.51	0.014	0.025	0.019	0.006		
30		Repayment to net income ratio	Negative	35.28	64.84	0.688	1.000	0.991	0.426		
31	Guarantee and joint guarantee	Strength of guarantor	Positive	0.066	0.009	0.041	0.157	0.347	0.005		
36		Age of guarantor	Interval	45	36	1.000	0.059	0.094	1.000		
37	Macro environment	Engel's coefficient	Negative	0.373	0.399	0.892	0.135	0.246	0.773		
44		Regional government policy	Qualitative	1	3	1.000	0.600	0.000	0.000		
45	—	Default or not $y_i$	—	0	1	0	0	1	1		

TABLE IV.

Correlation Coefficient  $r_{kj}$  of Indicators of Petty Loans for Farmers

(a) Serial number	(b) Indicators	(1) Loan purpose	(3) Value of house owing	(12) House value	(44) Regional government policy
1	Loan purpose	1.000	0.023	0.018	-0.019
3	Value of house owing	0.023	1.000	<b>0.991</b>	-0.025
12	House value	0.018	<b>0.991</b>	1.000	0.409
44	Regional government policy	-0.019	-0.025	0.409	1.000

According to the indicator screening standard in “The second indicator screening using significant discriminant” mentioned above, if  $P_j \geq P_0=0.05$ , the indicator  $x_j^i$  should be deleted. On the contrary, the indicator  $x_j^i$  should be retained. Comparing the critical probability  $P_0=0.05$  with data in the first Row and the 5<sup>th</sup> Column in Table V, the bilateral probability  $P_1$  corresponding to the regression coefficient  $c_1$  of the first indicator “Loan purpose” is less than the critical probability 0.05, i.e.  $P_1=0.024 < P_0=0.05$ . It indicates that the indicator “Loan purpose” can significantly distinguish the default status and should be reserved. The result is marked with “Reserved” in the first Row and 7<sup>th</sup> Column in Table V.

Similarly, comparing the critical probability 0.05 with the other data in the 5<sup>th</sup> Column in Table V, the screening results were listed in the corresponding Row in the 7<sup>th</sup> Column in Table V.

In a word, the second screening deleted 25 indicators on the basis of the reserved 40 indicators from the first screening, such as “Number of labor force”, “Expense of family’s daily life”. Then, the final index system is established, including 15 indicators, which are marked with “Reserved” in Column 4 of Table I.

4) *The reasonability judgment of the established index system*

Substitute the data of the first to 2044<sup>th</sup> Column in

Table III into Equation (6), the variance contribution rate  $In = 95.71%$  can be obtained. In other words, the established index system reflected 95.71% original information by using 23.08% (15/65=23.08%) indicators.

The index system is reasonable according to the standard in “The reasonability judgment of the established index system” mentioned above.

TABLE V.

The Index Screening Based on Significant Discriminant

(1) Serial number	(2) Criterion layers	(3) Indicators	(4) Regression coefficients $c_j$	(5) Bilateral probability $P_j$	(6) Critical probability $P_0$	(7) Screening result
1	Basic information	Loan purpose	0.997	<b>0.024</b>	0.050	<b>Reserved</b>
...		...	...	...		...
14	Repayment ability	Number of labor force	-2.017	0.837		Deleted
15		Expenses/ incomes	0.126	<b>0.017</b>		<b>Reserved</b>
...	Repayment willings	...	...	...		...
23		Expense of family's daily life	0.440	0.959		Deleted
24	Guarantee and joint guarantee	Private loans	0.570	<b>0.004</b>		<b>Reserved</b>
...		...	...	...		...
30	Macro environment	Repayment to net income ratio	-0.171	0.551		Deleted
31		Strength of guarantor	2.235	<b>0.026</b>		<b>Reserved</b>
...	Per capita agricultural output value	...	...	...	...	
36		Age of guarantor	-1.247	0.350	Deleted	
37	Per capita agricultural output value	Engel's coefficient	-0.339	<b>0.019</b>	<b>Reserved</b>	
...		...	...	...	...	
40	...	...	...	...	Deleted	

C. Comparative analysis

1) The comparative analysis of the established index system with the existing index systems

(1) The existing researches lack the key indicators which can effectively distinguish default from non-default. “Loan purpose”, “Expenses/ incomes”, “Increasing rate of regional GDP” and other indicators can distinguish non-default from default ones. Those indicators are reserved by the significant discriminant screening.

This paper deleted several indicators of index system from Postal Savings Bank of China [9], such as “Gender”, “Supporting population”. Those indicators can not effectively distinguish default from non-default, as shown in Column 4 and 7 in Table I.

The paper deleted several indicators of index system from Industrial and Commercial Bank of China [12], such as “Residential status”, “Total expenses”. Those indicators can not effectively distinguish default from

non-default, as shown in Column 4 and 7 in Table I.

(2) The existing index systems have unobservable indicators. This paper deleted several unobservable indicators of the index system from Agricultural Bank of China [8], such as “Area ratio of disposable assets”. The results are shown in Column 7 of Table I.

2) Correspondence between the index system and the characteristics of petty loans for farmers

The corresponding relation between the established index system of petty loans for farmers and characteristics of farmers is shown in Table VII.

(1) Reflecting financial information of the farmers through “Expenses/ incomes”, “Non-agricultural incomes/total incomes” and other indicators.

(2) Reflecting the effect of economic development on repayment capacity of farmers through “Engel's coefficient”, “Increasing rate of regional GDP” and other indicators.

TABLE VII.

The Corresponding Relation between the Established Index System and Characteristics of Farmers

(1) Serial number	(2) The characteristics of farmers	(3) Indicators that reflect the characteristics
1	Financial information is unsound	“Expenses/ incomes”, “Non-agricultural incomes/total incomes” and other indicators
2	Regional economic development has a significant impact on the repayment of farmers	“Engel's coefficient”, “Increasing rate of regional GDP” and other indicators

V. CONCLUSIONS

A. Main work and conclusions

Using correlation analysis to delete the indicators of repeated information, and significant discriminant to select the indicators which can effectively distinguish

default customers from non-default ones, this paper established the credit risk evaluation index system of petty loans for farmers including 15 indicators, which involved “Basic information”, “Repayment ability”, “Repayment willing”, “Guarantee and joint guarantee”, “Macro environment” 5 criterion layers.

The empirical study on 2044 petty loan samples for

farmers from a Chinese national commercial bank showed that “Loan purpose”, “Expenses/ incomes”, “Increasing rate of regional GDP” and other indicators can distinguish non-default from default ones.

### B. Specials and contributions

This study established the credit risk evaluation index system of petty loans for farmers, which reflected the characteristics of farmers. Firstly, the farmers' financial information is reflected through “Expenses/ incomes”, “Non-agricultural incomes/total incomes” and other indicators. Secondly, the effect of economic development on repayment of farmers is reflected through “Engel's coefficient”, “Increasing rate of regional GDP” and other indicators.

### REFERENCES

- [1] National Bureau of Statistics of China, *China Statistics Yearbook – 2012*, Beijing: China Statistics Press, pp. 169–173, September 2012.
- [2] L. Jiang, “U.S. personal credit scoring system and its Implications for China,” *China's financial*, vol. 7, pp. 66-67, April 2006.
- [3] H. Lin, M. Xia, *Enterprise credit rating theory and Practice*, 1<sup>st</sup> ed., Beijing: University of International Business and Economics Press, pp. 79–127, November 2003.
- [4] Moody's Investors Service, *Global Credit Research*, Moody's Investors Service, June 2005.
- [5] Standard & Poor's, *Enterprise Credit Rating*, Standard & Poor's, October 2010.
- [6] Fitch Ratings, *Understanding Credit Ratings-Limitations and Usage*, [Http://www.fitchratings.com/creditdesk/public/ratings\\_definitions/index.cfm](http://www.fitchratings.com/creditdesk/public/ratings_definitions/index.cfm), March 15, 2013.
- [7] J. Chen, “Citibank's business risk and credit policies,” *Global Outlook: modern commercial banks*, vol. 2, pp. 58-60, February 1999.
- [8] Agricultural Bank of China, *Management of farmers credit rating of Agricultural Bank of China*, Agricultural Bank of China, 2008.
- [9] Postal Savings Bank of China, *Merchant credit rating table of Postal Savings Bank of China*, Postal Savings Bank of China, 2009.
- [10] Emei credit cooperatives, *The operating rules and regulations of micro-credit loans to farmers for Rural Credit Cooperatives of Sichuan Province*, [Http://www.emxh.com/article/2007/0803/file\\_142.html](http://www.emxh.com/article/2007/0803/file_142.html), April 10, 2013.
- [11] Construction Bank of China, *Small business customers evaluation approaches of China Construction Bank*, Construction Bank of China, pp. 1-8, 2007.
- [12] Industrial and Commercial Bank of China, *The notice about printing the evaluation method of the small business corporate clients credit rating of China Industrial and Commercial Bank means*, Industrial and Commercial Bank of China, No. 78, 2005.
- [13] R. Mahjabeen, “Microfinancing in Bangladesh: Impact on households, consumption and welfare,” *Journal of Policy Modeling*, vol. 30, No. 6, pp. 1083-1092, November –December 2008.
- [14] K. Carling, T. Jacobson, J. Lindé, and K. Roszbach, “Corporate credit risk modeling and the macro-economy,” *Journal of Banking & Finance*, vol. 31, No. 3, pp. 845-868, March 2007.
- [15] J. Copestake, “Mainstreaming Microfinance: Social Performance Management or Mission Drift?,” *World Development*, vol. 35, No. 10, pp. 1721-1738, October 2007.
- [16] L. Yu, X. Yao, S. Wang, K.K. Lai, “Credit risk evaluation using a weighted least squares SVM classifier with design of experiment for parameter selection,” *Expert Systems with Applications*, vol. 38, No. 12, pp. 15392-15399, November –December 2011.
- [17] A. M. Alam, M. Nasser, and K. Fukumizu, “A Comparative Study of Kernel and Robust Canonical Correlation Analysis,” *Journal of Multimedia*, vol. 5, No. 1, pp. 3-11, February 2010.
- [18] Y. Xu, Y. Zhang, “An online credit evaluation method based on AHP and SPA,” *Communications in Nonlinear Science and Numerical Simulation*, vol. 14, No. 7, pp. 3031-3036, July 2009.
- [19] X. Zhou, W. Jiang, Y. Shi, “Credit risk evaluation by using nearest subspace method,” *Procedia Computer Science*, vol. 1, No. 1, pp. 2449-2455, May 2010.
- [20] Q. Zhang, Y. Zheng, H. Liu, J. Shen, and G. Gu, “Perceptual Object Extraction Based on Saliency and Clustering,” *Journal of Multimedia*, vol. 5, No. 4, pp. 393-400, August 2010.
- [21] Y. Wei, P. Liu, “Risk Evaluation Method of High-technology Based on Uncertain Linguistic Variable and TOPSIS Method,” *Journal of Computers*, vol. 4, No. 3, pp. 276-282, March 2009.
- [22] X. Gong, J. Zhou, H. Wu, G. Lei, X. Li, “Application to three-dimensional canonical correlation analysis for feature fusion in image recognition,” *Journal of Computers*, vol. 6, No. 11, pp. 2427-2433, November 2011.
- [23] Post Savings Bank of China, *Credit Risk Evaluation System of Petty Loan for Farmers*, Post Savings Bank of China, 2011.
- [24] G. Chi, W. Wang, *Comprehensive evaluation theory, method and application based on the scientific development*, 1<sup>st</sup> ed., Beijing: Science Press, pp. 27-79, September 2009.
- [25] A. C. Rencher, G. B. Schaalje, *Linear models in statistics*, New Jersey: John Wiley & Sons, inc., Publication, 2008.

**Linpeng Hai** (Male) was born in 1979. He is interested in network control and risk management. Email: hailp@hpu.edu.cn.

**Baofeng Shi** (Male) was born in 1984. He is a financial engineering doctor graduate student in Dalian University of Technology. He is interested in credit risk management of banks and evaluation of complex systems. Email: fengbei\_wuyu@163.com.

**Guorong Peng** (Female) was born in 1983. She is interested in the establishment of mathematical model. Email: 158824939@qq.com.