

Research on the Influencing Factors of Personal Credit Based on a Risk Management Model in the Background of Big Data

Ximing Lv^{1,2}, Jianbao Li³, Shunkai Zhang³, Yi Li³, Chun Wang³

¹School of Mathematical Sciences, Inner Mongolia University, Hohhot, China

²School of Statistics and Mathematics, Inner Mongolia University of Finance and Economics, Hohhot, China

³School of Finance, Inner Mongolia University of Finance and Economics, Hohhot, China

Email: lvximing@vip.163.com

How to cite this paper: Lv, X.M., Li, J.B., Zhang, S.K., Li, Y. and Wang, C. (2017) Research on the Influencing Factors of Personal Credit Based on a Risk Management Model in the Background of Big Data. *Journal of Applied Mathematics and Physics*, 5, 722-733.

<https://doi.org/10.4236/jamp.2017.53061>

Received: February 24, 2017

Accepted: March 28, 2017

Published: March 31, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative

Commons Attribution International

License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Between states, between enterprises and enterprises, between people, it can be stated that credit is full of every corner of our lives. But the current lack of social credit is fundamental. Credit risk is particularly prominent. In the extensive data generation today, the information on personal credit statistics is very large, but still lack the data system processing and screening. Through the information retrieval of 200 credit information reports, this paper constructs the evaluation system of personal credit by using the basic information of the individual. The basic information of these individuals has great convenience in information collection and information statistics, and this basic information covers all aspects that are likely to result in the breach of contract. Through the use of single factor analysis and logistic model to solve the index system, you can not only find the impact of individual indicators on the degree of personal credit, but also see the overall impact of indicators on the degree of credit, that is, the weight of the indicators. Finally, four different credit ratings are divided by assigning the indicators to the scores. Credit rating can clearly measure the respective credit situation. Through the classification of these levels, measuring the credit line when a person in the individual credit operation, at the same time, it can provide reference and proval to administrative departments, which is benefit for managing credit risks. It has a substantial meaning and value in use. The solution to the rating system cannot only be applied to individuals, but also to the enterprises, with a wide range of versatility.

Keywords

Personal Credit, Information Retrieval, Single Factor Analysis, Logistic Regression Model, Division of Credit Rating

1. Introduction

Confucianism was born in a particular historical atmosphere of the Spring and Autumn period and the Warring States Period [1]. There is a profound culture foundation. Over more than two thousand years of history, Confucianism is widely assigned to the world. Human society has come into the modern society, but to some extent, Confucianism still plays an important role. It has a far-reaching influence on the development of our society. There is truth in Confucianism in the five principles, “benevolence, righteousness, propriety, wisdom, faith”; the real reason of the five principles is all members are in society Rational communication principle, Rapport principle and Harmony principle. This paper will study the five principles in “credit”. Research on personal credit, Credit plays a major role in modern times [2]. A person’s good credit directly impact on people’s survival and development. In this paper, through the information collection of 200 credit reports, we construct the logistic regression model for credit evaluation, and solve the model by Eviews software, then obtain the weight of each index and give scores to each index depending on the weight; according to the size of the score, you can see a person's credit rating.

2. Total Problem Analysis

In order to effectively evaluate the customer’s credit, we extract valuable indicators and data from the 200 credit reports, analyzing the data, determining the data objects and their attributes, and then analyzing the use of appropriate methods for each indicator to give weight. Finally, through each person’s credit score and different credit scores divided according to the credit rating, we can clearly determine the credit status of each customer with a certain practical application value. In summary, the general picture is in **Figure 1**.

3. Establish an Index System

Through statistical analysis of 200 credit reports, we analyze the social roles and social status of the people, according to the basic condition of their personal information and subdivide and find out the credit status of the people in different classes. We establish the index system in **Figure 2**.

Through the understanding of the basic situation of individuals, we can judge the person’s ability of repayment, consumption concept and direction. Through these can reflect a person’s credit status [3].

3.1. The Impact of Gender on Credit

Different gender groups of people have different consumer attitude, there is a big difference in the cost, and there are also some differences in their salaries. Differences in consumption concept directly affect the deadline of repayment.

3.2. The Influence of Marital Status on Credit

There are different uses of funds for the people who are in different marital status. As for unmarried people, their funds are used primarily in the preparation of

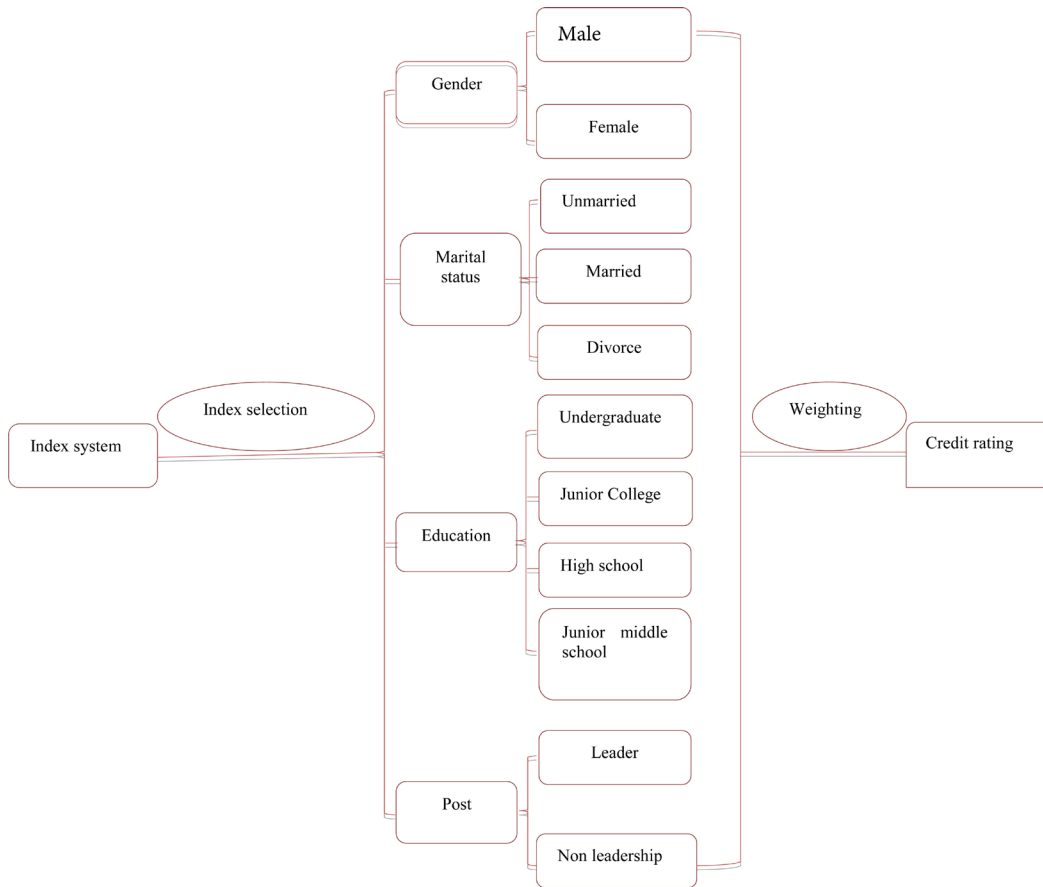


Figure 1. Article total flow chart.

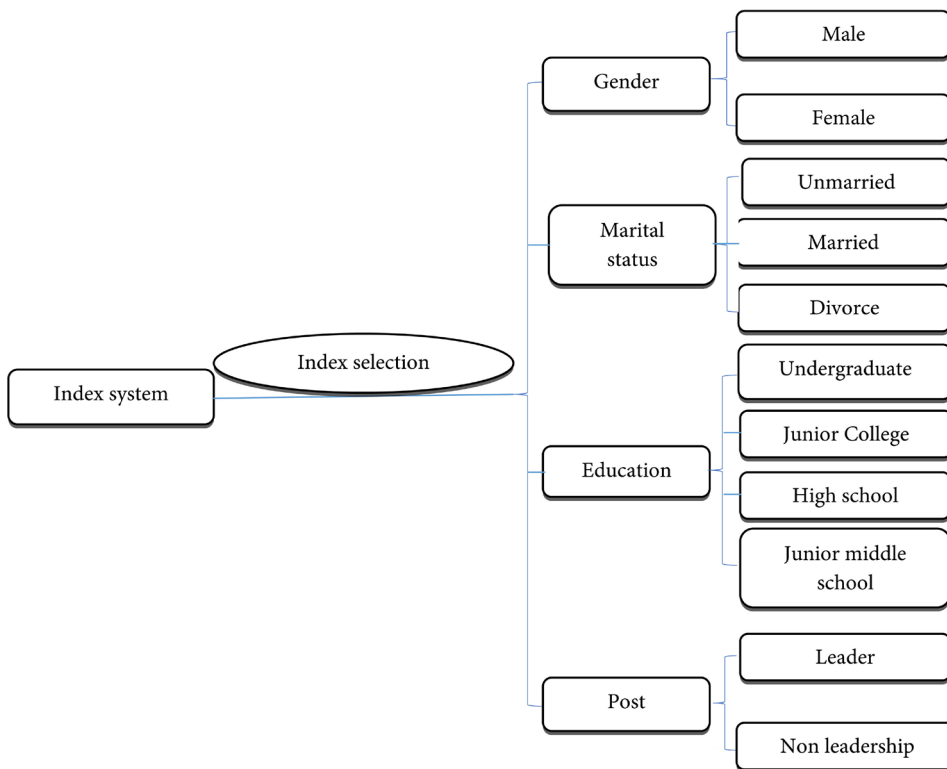


Figure 2. Index system diagram.

future marriage; married people spend money to support their families, but divorce is more complex. There is a close relationship between the direction of the use of funds and credit.

3.3. The Influence of Education on Credit

Educational status reflects a person's education level. To a large extent, education and remuneration has a certain positive correlation. The sufficiency of funds is closely related to credit [4].

3.4. The Influence of Position on Credit

Compared with the non-leadership group, the leadership group has a higher salary, and they have a high degree of concern and attention to credit.

4. The Establishment of Logistic Model

4.1. Single Factor Analysis

We analyze the extraction data and find out that the four factors that have an impact on the overdue are gender (X_1), marital status (X_2), education (X_3), job level (X_4). We analyzed the four factors separately.

4.1.1. Gender Impact on Overdue

We screened out overdue information of customers of different gender through Excel, and selected each customer overdue numbers and the maximum number of overdue months two indicators to do correlation analysis [5].

Through the analysis, we get the proportion of different gender effects on overdue **Table 1**.

Through the above table we made the histogram in **Figure 3**.

From **Figure 3** we can find that the impact of gender on the overdue, the male overdue proportion is higher than the female overdue ratio, indicating that men are lower than women in the degree of credit. By comparing the overdue ratio, it can be seen that both men and women have seriously overdue behavior.

4.1.2. The Impact of Marital Status on Overdue

We screened out the overdue information of the customers in different marital status in Excel and selected the overdue number of each client to do the dependent variable to get the proportion of different marital status in **Table 2**.

Make the proportional histogram as **Figure 4**.

We can find that the overdue proportion of married people is much higher.

Table 1. Relationship between sex and overdue proportion.

Gender	Overdue amount	Overdue proportion	Non overdue quantity	Non overdue proportion
Male	60	0.3	56	0.28
Female	46	0.23	38	0.19

Table 2. The relationship between marital status and overdue proportion.

Marital status	Overdue amount	Overdue proportion	Non overdue quantity	Non overdue proportion
Unmarried	22	0.11	16	0.08
Married	76	0.38	65	0.325
Divorce	15	0.075	6	0.03

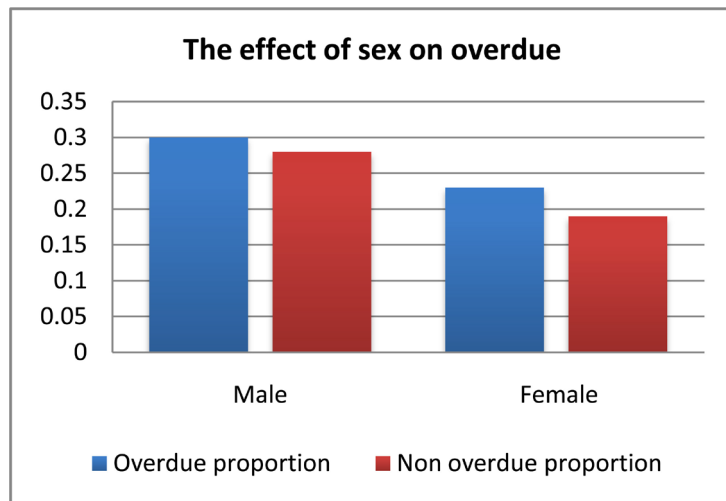


Figure 3. Gender and overdue ratio histogram.

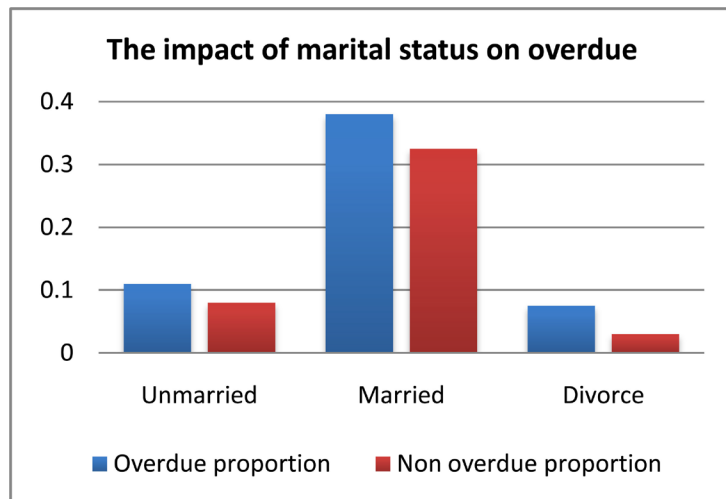


Figure 4. Marital status and overdue ratio histogram.

than that of unmarried and divorced people, indicating that married people have poor credit, which may be due to the enormous cost of their families. Non-overdue proportion of married people is greater than unmarried people and divorced people, which indicates that married people are the main customers of credit loans.

4.1.3. The Impact of Education on Overdue

As before we screened the overdue information of customers with different

educational background in Excel, and selected the overdue number of each customer to do the dependent variable, obtain the proportion of overdue repayment of different education which is showed in the table **Table 3**.

According to the proportion in the table which is shown in the **Figure 5**.

From **Figure 5** we can find that college education population and high school education population overdue proportion is greater than the university education population and junior high school education population, college schooling overdue takes the highest proportion. High school education and college education population are overdue proportion and non-overdue ratio is high, indicating two main consumer groups' credit. The proportion of overdue proportion of the university population is not different from the non-overdue proportion, and the proportion of the overdue proportion of the junior high school education population is greater than the overdue proportion, indicating that it has good credit.

4.1.4. The Impact of Overdue of whether It is Leadership

We screened out the overdue information of customers in different positions in Excel, and selected the overdue number of each customer to do the dependent variable, obtaining the proportion of overdue repayment of leadership and non-leadership which is showed in the following table **Table 4**.

According to the table, the following proportion is shown in the bar chart **Figure 6**.

Table 3. The relationship between educational background and overdue proportion.

Education	Overdue amount	Overdue proportion	Non overdue quantity	Non overdue proportion
Undergraduate	18	0.09	17	0.085
Junior College	42	0.21	30	0.15
High school	32	0.16	27	0.135
Junior middle school	11	0.055	23	0.115

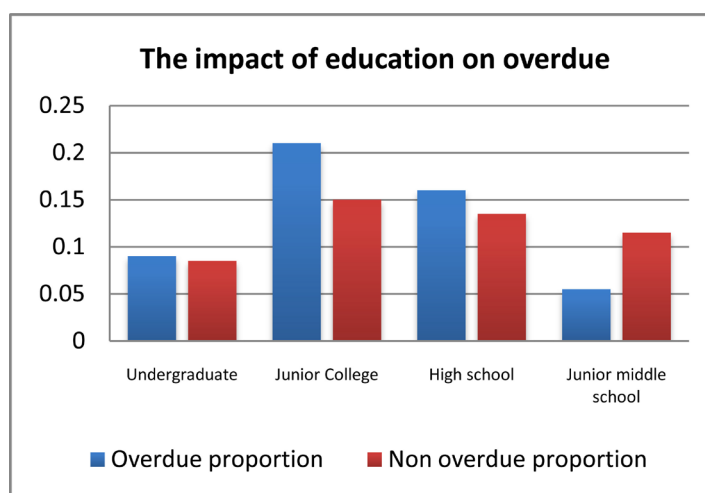


Figure 5. Education status and overdue ratio histogram.

Table 4. The relationship between job position and overdue proportion.

Post	Overdue amount	Overdue proportion	Non overdue quantity	Non overdue proportion
leadership	56	0.28	40	0.2
non-leadership	55	0.275	49	0.245

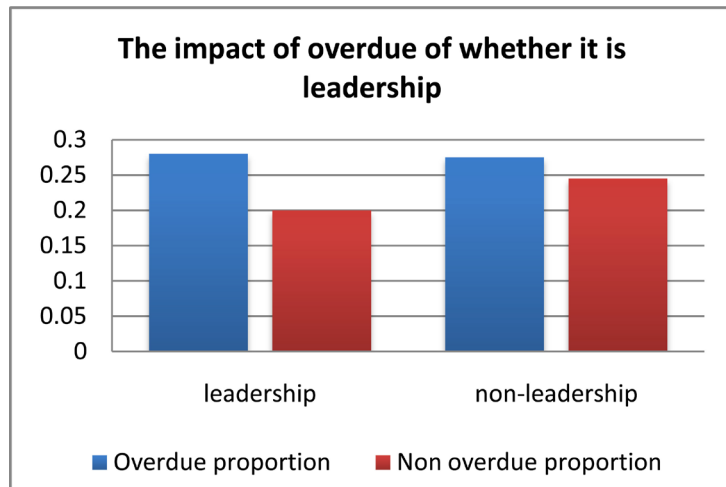


Figure 6. Jobs and overdue ratio histogram.

From **Figure 6** we can find that the proportion of overdue leadership is greater than the non-overdue ratio, and the difference is large. Non-leading overdue ratio is more important than non-overdue ratio, the difference between the two is smaller. On the whole, the credit of the leaders is poor. Regardless of leadership or non-leadership, the overdue ratio is greater, indicating whether the credit is good or not, the leadership is not much relevance.

Next we will build the logistic regression model for credit evaluation.

4.2. Model Introduction

Logistic regression model belongs to probabilistic nonlinear regression; it is a multivariate analysis method to examine the relationship between the results of the two categories and the influence factors [6]. It is easy in structure, but it can handle the effects of discrete anomaly data points when dealing with complex data systems which are composed of multiple metrics. Logistic regression model is utilized extensively. In social sciences such as sociology, psychology, demography, politics, economics, and public health, a large number of observed dependent variables are of two classifications. Logistic regression model can well address the problems in these fields.

4.3. Model Principle

The standard linear regression model is:

$$Y = \vartheta + \beta_1 \times X_1 + \dots + \beta_m \times X_m \tag{1}$$

And by the standard linear regression model we can replace Y with probability

P , and get:

$$P = \vartheta + \beta_1 \times X_1 + \dots + \beta_m \times X_m \quad (2)$$

But this model has a lot of restrictions in the application. Statisticians use logistic transformation to solve this problem. Logistic transformation introduction: The ratio of the probability of occurrence of a result and the probability that the result does not occur is usually called the odd number. This is $\text{Odds} = \frac{\pi}{1-\pi}$. Take the logarithm $\lambda = \ln(\text{Odds}) = \ln\left(\frac{\pi}{1-\pi}\right)$. This is the logistic transformation. By transforming, the range of values $\text{Logit}(\pi)$ is extended to the entire real field centered on 0. This makes it possible to predict the π value at any value of the independent variable. Therefore, we build $\text{Logit}(\pi)$ as the dependent variable, establishment of logistic regression model with p independent variables:

$$\text{Logistic}(P) = \beta_0 + \beta_1 \times X_1 + \dots + \beta_p \times X_p \quad (3)$$

Among them, $\text{Logistic}(P) = \ln\left(\frac{P}{1-P}\right)$. The logistic regression model that fits the two classifiers is transformed into the parameters of the fitted linear model, Among them, $\beta_1, \beta_2, \beta_3, \dots, \beta_p$ is the regression coefficient, which shows the contribution of each influencing factor X_i to P , and β_0 is a constant term.

According to the above equation, we can get the following formular.

$$P = \frac{\exp(\beta_0 + \beta_1 \times X_1 + \dots + \beta_p \times X_p)}{1 + \exp(\beta_0 + \beta_1 \times X_1 + \dots + \beta_p \times X_p)} \quad (4)$$

$$1 - P = \frac{1}{1 + \exp(\beta_0 + \beta_1 \times X_1 + \dots + \beta_p \times X_p)} \quad (5)$$

4.4. Model Application

We choose regression analysis of the effects of independent variables (X_1), marital status (X_2), education (X_3) and job level (X_4) on predictive variables overdue (Y). Which $Y = 0$ when the repayment is not overdue, $Y = 1$ when the repayment overdue X_1 X_2 X_3 and X_4 are deterministic variables associated with Y . So as to obtain the contribution of each factor to the predictor [7], we use Eviews software for simulation. The results are shown in the **Table 5**.

From the results we can see that the four independent variables are significant (X_1), marital status (X_2), education (X_3), job height (X_4), and the final regression equation can be obtained as follows:

$$P = \frac{e^Y}{1 + e^Y} \quad (6)$$

$$Y = -3.08526394277 + 1.11941730358 \times X_1 + 0.149338115176 \times X_2 + 0.215431901503 \times X_3 + 13.0143637733 \times X_4 \quad (7)$$

Transformed:

Table 5. Regression results of logistic model.

Dependent Variable: Y
 Method: ML - Binary Logistic (Quadratic hill climbing)
 Date: 02/19/17 Time: 13:35
 Sample: 1 96
 Included observations: 96
 Convergence achieved after 3 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
C(1)	-3.085264	0.231389	-13.3336	0.0000
C(2)	1.119417	0.629502	1.778259	0.0387
C(3)	0.149338	0.193974	0.769887	0.0434
C(4)	0.215432	0.561374	0.383758	0.0271
C(5)	13.01436	0.859665	15.13888	0.0000
McFadden R-squared	1.000000	Mean dependent var		0.500000
S.D. dependent var	0.502625	S.E. of regression		6.44E-41
Akaike info criterion	0.104167	Sum squared resid		3.78E-79
Schwarz criterion	0.237726	Log likelihood		0.000000
Hannan-Quinn criter.	0.158154	Deviance		0.000000
Restr. deviance	133.0843	Restr. log likelihood		-66.54213
LR statistic	133.0843	Avg. log likelihood		0.000000
Prob(LR statistic)	0.000000			
Obs with Dep = 0	48	Total obs		96

$$\text{Logistic}(P) = -3.08526394277 + 1.11941730358 \times X_1 + 0.149338115176 \times X_2 + 0.215431901503 \times X_3 + 13.0143637733 \times X_4 \tag{8}$$

We can use the Logistic model for credit evaluation. We just need to be gender (X_1), marital status (X_2), education (X_3), position (X_4) score into the model to get the final credit score.

5. Divide Credit Rating

According to Logistic model, we will make each factor corresponding logistic coefficient as credit influence weight. As sex analysis of the weight of men is 1.119417, the same way to get the other three factors is showed in the **Table 6**.

Then we finally come to the formula of the credit index Y :

$$Y = -3.085263 + 1.119417 \times X_1 + 0.149338 \times X_2 + 0.215431 \times X_3 + 13.014363 \times X_4 \tag{9}$$

According to the basic information in the credit report (gender, education, marital status, job level) on the proportion of repayment overdue, we made a detailed regression statistics, then, we put the data into the calculation formula of the credit index Y :

Table 6. The index of share rights.

Gender	1.119417
Marital status	0.149338
Education	0.215432
Post	13.01436

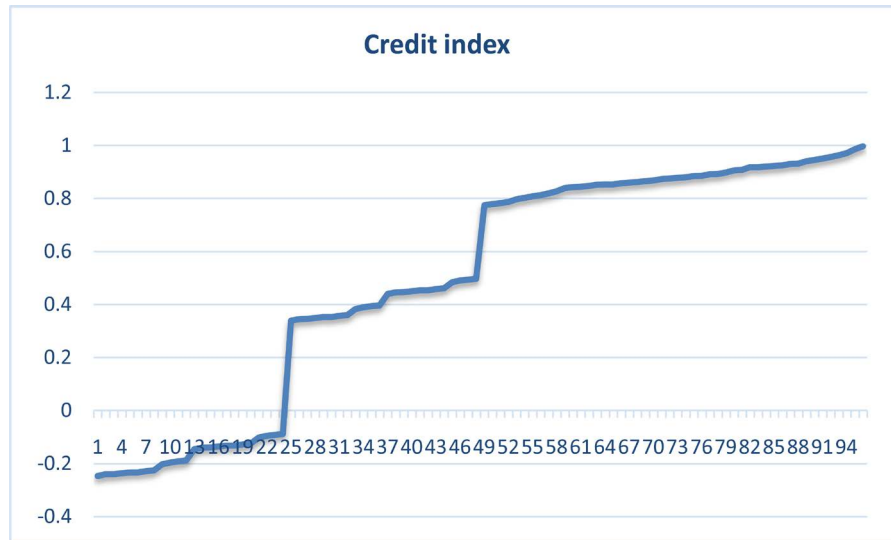


Figure 7. Credit index chart.

$$Y = -3.085263 + 1.119417 \times X_1 + 0.149338 \times X_2 + 0.215431 \times X_3 + 13.014363 \times X_4 \tag{10}$$

We get the corresponding credit index trend graph in **Figure 7**.

Based on the user credit index derived from the regression data, we use the credit index of 0.8, 0.6, and 0.4 as the standard; the customer level is divided into excellent good moderate poor four grades.

Excellent (0.8 - 1): Consumer’s credit condition is the best, there are almost no default risk; financial institutions bear the least risk as the optimal lending standard [8].

Good (0.6 - 0.8): Customer credit situation is better; occasionally there will be a breach of contract; financial institutions need to bear some risks, but still within reasonable limits.

Moderate (0.4 - 0.6): Customer credit status in general, the probability of default is higher than the previous two levels, but it requires financial institutions to take reasonable measures to make up for risk, making the proceed are still greater than the risk. This level is the minimum lending standard.

Poor (0.4 or less): Customer credit is poor; breach of contract often occurs; the pecuniary institutions face enormous risks; they shouldn’t be lent.

6. Results

In the course of the study, we investigated the personal credit information of 200

credit users. They can be viewed in a single factor analysis. The proportion of men overdue > the proportion of women overdue; non-overdue proportion of male > non-overdue proportion of women. Because we are highest in the course of the survey than the female boss, so the men overdue and non-overdue rate is higher than women's situation. In the case of marital status, overdue rankings are married > unmarried > divorce. Non-overdue rankings are married > unmarried > divorce. We can note that the credit situation in the married population is biased towards polarization. We can think that as people enter into marriage, people's mental state has gradually become distinct. People in sound financial condition may have no pressure on repayment. While some people are in poor economic situation, repayment ability is mediocre. In terms of qualifications, overdue rate ranked college > high school > undergraduate > junior high school. Non-overdue rate ranked college > high school > junior high school > undergraduate. We can see high school and college education overdue rate is relatively high. While students in junior high school may not acknowledge the using of credit, so users in this part are small and the rate is low. Undergraduate people may pay more attention to their credit records. In terms of posts, overdue rankings lead greater than non-leadership, non-overdue rankings of non-leaders than leaders. A certain extent reflects the leadership may spend more. In terms of repayment is not good enough.

7. Conclusions

Throughout the full text, the overdue number is greater than the number of non-overdue. To some extent, it reflects the Chinese consumers do not attach importance to personal credit situation. The society has more and more credit consumption, which formed the credit risk system, is more and more complex. The government needs to create a complete set of credit regulatory system in order to maintain the overall credit. We set up a set of evaluation index system in the study of credit status and a logistic regression model to analyze the data; the analysis process is rigorous. We feel that these methods can be used as a reference to establish a credit regulatory system.

For the risk regulatory authorities, according to the individual's credit rating, we set a risk range. Once beyond this risk range, the risk of lending will be a magnifying trend. We should not lend at this time, at the same time, we should make the individuals' credit recording and update it timely. For the emergence of credit overdue individuals, we should adjust their credit level to prevent the emergence overdue behavior again.

Although the management is strict, it just an external factor for constructing the whole credit system; the internal factor is still people's attention to themselves. This requires our respective efforts to maintain our credit. We should pay attention to our personal credit situation, establish the correct concept of consumption and values. When the individual's credit level can be improved, the community's overall credit can be improved. It will be more harmonious between people.

Acknowledgements

This research was carried out with support of National Natural Science Foundation of People's Republic of China (project 71661025 and 11602115).

References

- [1] Fang, H.Q. and Zeng, Y. (2004) Bank Credit Risk Assessment Method Empirical Research and Comparative Analysis. Financial Research, Washington DC.
- [2] Chi, G.T., Pan, M.D. and Qi, F. (2014) Design and Application of a Bank Credit Risk Rating Model Based on Small Sample. Quantitative Economics and Technology Research, Singapore.
- [3] Li, Z.H. and Li, M. (2005) China's Commercial Bank Credit Risk Identification Model and its empirical research. Economic Science, California.
- [4] Deng, J., Qin, T. and Huang, S. (2013) Research on credit risk early warning of listed Companies in China Based on Logistic Model. *Financial Theory and Practice*, **40**, 22-26.
- [5] Shi, Q.Y. and Qin, W.S. (2006) Personal Credit Scoring Model and Its Application. Beijing China Founder Publishing House, Beijing.
- [6] Hong, Y.B. (2015) Logistic Model Coefficient Comparison Problem and Solution Strategy: A Review. *Society*, **35**, 220-241.
- [7] Zhu, J.G. and Lu, Z.F. (2009) Monetary Policy, Corporate Growth and Changes in Cash Holdings. *Management of the World*, **3**, 152-158.
- [8] Shi, X.J. and Li, J. (2009) Alternative Relationship between Commercial Credit and Bank Borrowing and its Countercyclicality. *Finance and Economics Research*, **34**, 4-15.



Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact jamp@scirp.org