

Research on Personal Credit Evaluation

Miao Wei

Panyapiwat Institute of Management
Nonthaburi, Thailand
mw018@163.com

Abstract—Traditional regional economics and management science may justify the personal credit evaluation system in one way or another, but generally speaking, they are not specific enough. This article sets the personal credit evaluation based on criminal justice data as the focus and uses theories of management, regional economics, sociology, law and criminology to explore its source and inner mechanism of action.

Keywords—regional economics; personal credit evaluation; Criminal justice data

I. INTRODUCTION

The Silk Road Economic Belt and the 21st-Century Maritime Silk Road is very important for the development of Southeast Asian and South Asia. People often focus on the economics of the Belt and Road, but the most important part of the B & R is the credit system of Southeast Asian and South Asia. Personal credit is the foundation for both individual accomplishments and national development. The biggest online payment system Alipay in this area also work successful in both China and US [1], but this system relies heavily on personal credit system, so if we hope the economy develop rapidly and steadily near the Belt and Road in Southeast Asian and South Asia, we must solve personal credit problems. This article sets the personal credit evaluation based on criminal justice data as the focus and uses theories of management, regional economics, sociology, law and criminology to explore its source and inner mechanism of action.

A. Personal credit evaluation in China

Judicial execution more difficult but also destroys the mutual trust between people and restrains the healthy growth of the social economic order. In recent years, Chinese courts at all levels deal with about 2,400,000 new execution cases on average each year, the number reaching 2,820,000 in 2014, 70 percent of which witnessed the property debtors delay the execution, or even evade or violently defy the execution, causing serious waste of judicial resources [2].

Key figures of personal credit are not currently transparent in China, and most personal credit evaluation organizations have not included judicial data in their research. Even the credit information publicity system supported by the judicial power is not without flaws. Today, the inquiry platform for the breach of the person subjected to execution is available for the public to check the information of the breachers in all the execution cases either newly received or unfinished by 3,484 local courts

in China since January 1st, 2007. In November, 2013, the Executive Board of the Supreme People's Court signed a memorandum of cooperation with the Credit Information Center of the People's Bank of China, agreeing on the decision to include the list of the defaulters in the bank's work related to credit information [3], the defaulters' breach of execution to be added to their credit archives and presented in the format of a credit report to financial institutions for reference before they approve a loan in future. So far the courts at all levels have successfully matched with 19 bank systems nationwide. However, only the People's Bank of China is able to realize money online real-time transfer execution, and that is only after a complicated process. What's more, the public is not given access to the information. There is still a lot of space for improvement [4].

B. Personal credit evaluation in US.

The US has both governmental and non-governmental organizations that specialize in collecting personal credit data related with the judicial field. Despite the differences in social and economic systems, China bears similarity to the US in the fundamental economic principles that elicit breach of the credit system [5], and the criminal motives. Combining personal credit evaluation with judicial data will exert three influences on business transactions: (1) It will prevent transaction cheating, improve the security of the transactions and make them more normative by verifying the credit of the transaction parties. (2) It will help to drill into the heads of the businesses and individuals the concept of obeying the law, respecting the contract and valuing one's credit in mind and actions [6]. (3) It will serve as a punishment to those who have defaulted in the judicial execution and restrict their capacity and qualifications in big economic activities.

II. THEORY AND HYPOTHESES

A. Transaction Cost Theory

Traditional economics assumes that the transaction cost is zero, and the pricing mechanism is omnipotent by efficiently organizing and coordinating transactions. However, the new institutional economics (NIE), which has its roots in two articles by Ronald Coase, "The Nature of the Firm" and "The Problem of Social Cost", holds that there is a transaction cost and thus comes the transaction cost theory. Coase proposed that transaction cost was the price paid for attaining precise market information, as well as the cost for negotiation and regular contracts. Specifically speaking [6], the transaction cost

includes the cost for attaining information, signing contracts, negotiating, monitoring and managing, and dealing with defaults. The transaction cost theory gives an explanation to the credibility problems in society from the perspective of the cost: if abandoning one's credibility could not bring as serious economic losses, the defaults will remain, whereas if the cost of abandoning one's credibility proves much above the profits, the defaults will, no longer driven by interest, diminish and even disappear [7]. Therefore a strict punishment mechanism for defaults must be established to reduce the frequency of defaults and warrant a healthy credit environment.

Hypothesis 1: Default cost is closely associated with default frequency

B. Routine Activity Theory

Routine Activity Theory is actually an extension of Rational Choice Theory. Developed by Marcus Felson and Lawrence E. Cohen in 1979, it analyzes and explains the different probabilities of criminal assaults by different groups. The theory states that a criminal is an event, which needs at least three factors to aggregate in time and space to occur: a motivated offender, a suitable target, and the absence of a capable guardian (not only a family guardian, but also social guardians like policemen) [8]. In other words: for a crime to occur, a likely offender must find a suitable target with capable guardians absent. Without anyone of the three, the crime might not be committed. To take a macro-level view, among the three factors, any change to any one may influence the crime probability. Once the three simultaneously appear, the probability will increase progressively. The motivated offender is only going to commit a criminal activity after comprehensively weighing and considering the time, location, the suitable victim and the lack of a capable guardian. Therefore, to avoid a crime, those factors must be reduced to make sure at least one factor is missing [9]. It actually contains punishment deterrence here by making the crime factors incomplete to prevent the offender from taking any chance, to let him or her know that a crime means more risks and so give up the criminal intention. However, the theory is more about inevitability of punishment than its severity.

Hypothesis 2: Judicial supervision and punishment reduces the frequency of defaults

C. Profit Maximization Theory

Profit maximization theory, also a market theory, proposes that a firm aims to get profits by manufacturing and selling its products. If the total revenue exceeds the total cost, there will be a surplus called "profit". According to the relationship between cost and revenue, if the total revenue is equal to the total cost, the firm neither gains nor losses, but only gains normal profits. However, if the total revenue is below the cost, there will be a loss. In 1968, Gary Becker, an American economist, used the basic economic theory of "cost-benefits" in his article "Crime and Punishment: An Economic Approach" to analyze crime and punishment, holding that the theory also applied to the rational offender: only when the benefits of the crime outweigh the cost will the offender commit a criminal action. Thus to curb the offender's motivation, the crime cost must be increased so as to surpass the benefits. Punishment

deterrence may be used here, for it is a process during which the offender compares and calculates the crime cost, or risks, and the benefits. Each criminal is a "rational" calculator, comparing the crime cost and benefits before a final decision to commit a crime is made. In his book "The Economic Approach to Human Behavior", Becker expounds upon the cost of crime and punishment, and provides the best solutions to counter different crimes with different risk tendencies of the offenders [10]. According to Becker, compared with legal acts, criminal acts are more uncertain. So to curb crimes the "subjective expected utility" of legal acts must be strengthened, whereas if the "subjective expected utility" of criminal acts outweighs that of legal acts, the offenders will opt for a crime.

The construction of the credit system must keep an institutional rein on the three factors of crime, lower the risks of defaults, better the objective conditions for both sides to keep promise and use punishment deterrence to prevent anyone taking a chance to default, for if the cost of defaults is far above that of keeping one's credibility, defaulting will be no longer a rational choice.

Hypothesis 3: Rewarding credibility will reduce the default frequency.

III. METHODOLOGY

A. Logistic Regression Model

In studying real problems, people have to find out the relationship between the probability of a random event A and certain factors. P means the probability of A when the independent variable is X_1, X_2, \dots . The logistic regression model is described as:

$$P = \frac{1}{1 + e^{-s}} \quad (1)$$

$$S = \beta_0 + \sum_{i=1}^k \beta_i X_i \quad (2)$$

X_i ($i=1,2,\dots,k$) is the independent variable, meaning the evaluation indicators in personal credit evaluation. β_i ($i=1,2,\dots,k$) is the coefficient. The regression value $p \in (0,1)$ means the outcome of the credit evaluation [11]. The closer it gets to 0, the worse the credit is. Since p falls between 0 and 1, $\text{logit}(p) = \ln p / (1-p)$ can be used to transform the equation above into a linear model:

$$\ln \frac{p}{1-p} = \beta_0 + \sum_{i=1}^k \beta_i X_i \quad (3)$$

B. Lasso Algorithm

Lasso Analysis Method, introduced by Robert Tibshirani in 1996, formulates a penalty function to create a refined model, shrinking large regression coefficients in order to reduce over-fitting and to make the interpretation simpler. It is an estimate to handle with collinear covariates [12]. The basic idea of Lasso is to force the sum of the absolute value of the regression coefficients to be less than a fixed value, minimize the residual sum of squares, set certain coefficients to zero, and effectively choose a simpler model that does not include those coefficients.

Let there be a given data (x^i, y_i) , $i=1,2,\dots,n$. Let $x^i=(x_{i1}, x_{i2}, \dots, x_{ip})^T$ be the predictor variable and y_i be the response variable. In ordinary regression models, all the observed values are presumed to be independent from each other, or the response variables are independent when a fixed predictor variable is given. Meanwhile, consider all the x^i are normalized, namely fitting $\sum_i x_{ip}/N=0$, $\sum_i x_{ip}^2/N=1$, If $\hat{\beta}=(\hat{\beta}_0, \dots, \hat{\beta}_p)^T$,

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_i^n \left[(y_i - \sum_j^p \beta_j x_{ij})^2 + \sum_j |\beta_j| \right] \right\} \quad (4)$$

In (x) of (4), the first measures the goodness of fit while the second is in fact a penalty function which prevents over-fitting. As the harmonic coefficient λ varies, the number of preserved variables also changes. The bigger λ is, the stronger the penalty function is and the fewer variables will be preserved.

C. Lasso Estimate of Logistic Regression Model

Fan and Li proposed to introduce Lasso penalty method into the linear model and generalized parametric model, as well as in proportional hazard model. Zou (2006) posed a framework of making the linear model adapting to Lasso, holding that the sum of squares of residuals can be replaced by likelihood function.

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \{ -1(\beta) + \lambda \sum_j |\beta_j| \} \quad (5)$$

$$l(\beta) = \sum_{i=1}^n P_i = \sum_{i=1}^n \left[y_i \log \frac{p_i}{1-p_i} + \log(1-p_i) \right] \quad (6)$$

$$= \sum_{i=1}^n \{ y_i (\beta_0 + \sum_{i=1}^k \beta_i x_i - \log[1 + \exp(\beta_0 + \sum_{i=1}^k \beta_i x_i)]) \} \quad (7)$$

So the $\hat{\beta}$ in Lasso Estimate of Logistic Regression Model can be described as :

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \{ y_i (\beta_0 + \sum_{i=1}^k \beta_i x_i - \log[1 + \exp(\beta_0 + \sum_{i=1}^k \beta_i x_i)]) + \lambda \sum_j |\beta_j| \} \quad (8)$$

IV. DATA ANALYSIS

A. Source of Data and Variables

The data is from Montgomery County Community Supervision Department of Public Safety and Correctional Services. This article chooses 50000 cases from the data base. From graph 1, there are several variables for each single case

B. Model Evaluation and Variable Selection

Fig. 1. The single case from the Maryland Community Supervision data base.

The evaluation model combining Lasso and logistic regression model in this article uses the glmnet program package of R-studio for cross validation and gets the graph which indicates the model error varies on the Y-axis as λ changes on the X-axis [13]. The X-axis above is the number of variables selected for each λ .

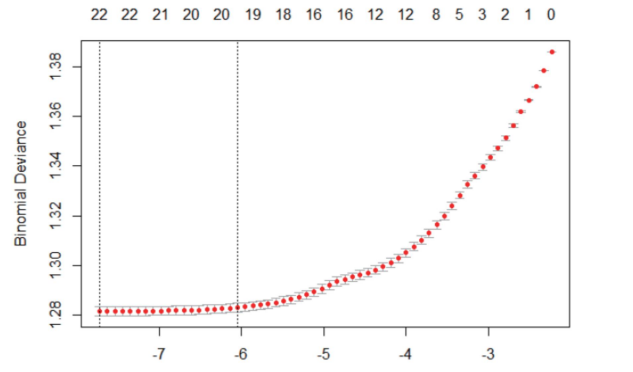


Fig. 2. lambda and the numbers of the variables

In graph 2, the dotted line on the left indicates the value of λ when the error is minimized and the number of variables selected by the model. As can be seen in the graph, the error changes as λ grows in value. So does the shrinking degree of the model [14]. Therefore, the model error and the number of selected variables must be comprehensively considered. Graph 2 indicates that as λ grows in value, the shrinking degree of the model increases and its function to select variables is more obvious. To get as many important variables as possible, the perfect value for λ would be what makes the shrinking degree of the model strongest. According to Tibshirani (1996) [15] in his research, when the log (Lambda) value on the right-side dotted line is chosen and $\lambda=e-6$.

Table 1 lists the parameter estimation results of all variable logistic model, Lasso-logistic model and Ridge-logistic model. As can be seen from the graph, the all variable logistic model, with the significance level being 5%, sifts out X1, X4 and keeps 4 independent variables; Lasso-logistic model abandons X1, X2, X4 and keeps 3 variables; X1, X2, X4 fails to select the variables. After comparison, Lasso-logistic model stands out in reducing the number of variables, making the

model simpler and making it easier to spot the key variables that will influence the final result.

TABLE I. VARIABLE EVALUATION

variables	symbol	All variable logistic model	Lasso-logistic model	Ridge-logistic
<i>Income level</i>	<i>X1</i>	0	0	-1.041
<i>Criminal record</i>	<i>X2</i>	-0.080	0	0.077
Defaults cost	<i>X3</i>	0.171	0.148	0.158
<i>Insurance</i>	<i>X4</i>	0	0	-0.017
Supervision level	<i>X5</i>	0.427	0.466	0.472
<i>Parole and Probation</i>	<i>X6</i>	-0.144	-0.130	-0.137

After analyzing X3, X5 and X6, the conclusion of H1, H2, H3 can be made:

Default cost is closely associated with default frequency

Judicial supervision and punishment reduces the frequency of defaults.

Rewarding credibility will reduce the default frequency.

C. Accuracy Comparison of the Models

TABLE II. PREDICTION ACCURACY

Model	95%	90%	85%	80%
ALL variable logistic	65.15%	66.10 %	66.28 %	65.74 %
Lasso-logistic	65.13%	65.79 %	65.75 %	65.75 %
Ridge-logistic	64.57. %	65.53 %	65.81 %	65.81 %

After the establishment of the evaluation model, this article also tests one by one the prediction accuracy of the three models mentioned above: all variable logistic model, Lasso-logistic model and Ridge-logistic model. Table 2 offers the prediction performance of the three models at different sampling rates. It can be seen that there is little difference between their performances, with all variable logistic model, Lasso-logistic model slightly but not remarkably better than Ridge-logistic model. However, in real prediction, faced with enormous amounts of data, predictors have to take into account the time it takes to predict. That is, both prediction accuracy and efficiency are important. In this aspect, Lasso-logistic model excels for it needs shorter time to produce a prediction upon a huge database than all variable logistic model. In real practice, while processing the sample amount of data in this article, Ridge-logistic model has already shown signs of insufficiency in its running speed, which makes it time-consuming besides its lack of advantage in accuracy. Therefore, comprehensively speaking, Lasso-logistic model is better at prediction than all variable logistic model and Ridge-logistic model.

V. CONCLUSION

This article introduces Lasso analysis method to Logistic Regression Model and brings judicial data into personal credit evaluation. With Lasso analysis method and Logistic Regression Model combined, the function to shrink the model and select the variables is realized. The conclusions are as followed: first, after analysis and comparison, Lasso-logistic model proves to set more regression coefficients to zero and boasts a stronger ability to shrink the model than all variable logistic model and Ridge-logistic model; second, there is little difference between their performance in predicting, with all variable logistic model, Lasso-logistic model slightly but not remarkably better than Ridge-logistic model; third, considering the huge amount of data to be processed in real practice, Lasso-logistic model is obviously better than all variable logistic model and Ridge-logistic model [16]. Therefore to use the logistic model combined with Lasso analysis method as the model for personal credit valuation enables an efficient selection of judicial data evaluation indicators, increases the evaluation efficiency and enhances the effect of personal credit evaluation.

In conclusion, the personal credit system based on judicial credit data will:(1) establish institutional deterrence against defaults (2) make it possible to check the information about the defaulter's property (3) efficiently propel the person subjected to execution to actively fulfill the due obligations (4) prevent financial risks (5) mitigate social transaction risks (6)reduce the waste of social resources caused by invalid lawsuits (7) make the public administrative management more scientific.

A successful credit system is the key of the development of this regional economy, and this is the most important part of The Silk Road Economic Belt and the 21st-Century Maritime Silk Road in Southeast Asian and South Asia. This article try to solve personal credit problems, and sets the personal credit evaluation based on criminal justice data as the focus and uses

theories of management, regional economics, sociology, law and criminology to explore its source and inner mechanism of action, it is the first time to solve this kind of problems and prepare for the establishment of the credit system near the Belt and Road in Southeast Asian and South Asia.

ACKNOWLEDGMENT

This paper is supported by Professor Hua Guowei, Beijing Jiaotong University and Dr. Udamber, Department of Public Safety and Correctional Services, thanks for the information and helps from these professors.

REFERENCES

- [1] A. Saunders, and L.D. Allen, "Credit Risk Management In and Out of the Financial Crisis: New Approaches to Value at Risk and Other Paradigms," John Wiley & Sons, 2010.
- [2] Brunnermeier, Markus, and L. Pederson, "Market Liquidity and Funding Liquidity," *Review of Financial Studies*, 2009, pp. 2201-2238.
- [3] Brunnermeier, Markus, "Deciphering the Liquidity and Credit Crunch 2007-2008," *Journal of Economic Literature*, pp. 77-100, 2009.
- [4] C. Tittle, V. Ekaterina, Botchkovar, and O. Antonaccio, "Criminal Contem Plation, National Context, and Deterrence," *Journal of Quantitative Criminology*, 2010.
- [5] L.D. David, "Credit Risk Modeling: Theory and Applications," New Jersey: Princeton University Press, 2009.
- [6] D. Gang, K.L. Kin, and J. Yen, "Credit scorecard based on logistic regression with random coefficients," *International Conference on Computational Science*, vol. 1, pp. 2463 – 2468, 2010.
- [7] K.L. Manova, "Credit Constraints, Heterogeneous Firms, and International Trade.," *Review of Economic Studies*, vol. 80, pp. 711-744, 2013.
- [8] K. Kuldeep, and D.H. John, "Forecasting Credit Ratings Using An Ann And Statistical Techniques," *International Journal of Business Studies* vol. 11, pp. 91-108, 2003.
- [9] P. Marco, and J. Tullioi, "The European Experience with Credit Information Sharing, Published in Credit Reporting Systems and the International Economy," Edited by M. Miller. Cambridge: MIT Press, 2003.
- [10] G. Mark and K. Nobuhiro, "inancial Intermediation and Credit Policy in Business Cycle Analysis," *Handbook of monetary economics*, vol. 3, pp. 547 – 599, 2010.
- [11] R. Martin, and P. Kelly, "Credit and Classification: The Impact of Industry Boundaries in Nineteenth-century America," *Administrative science quarterly*, vol. 54, 2009, pp. 486-520.
- [12] G. Mason "Money, Credit, and Crisis. *American Journal of Economics and Sociology*, vol. 68, pp. 983 – 1038, 2009.
- [13] G. Mojtaba, "Visceral factors, criminal behavior and deterrence: empirical evidence and policy implications," *European Journal of Law and Economics*, 2012.
- [14] D. Simeon, "Private Credit in 129 Countries," *Journal of Financial Economics*, vol. 2007, 2007, pp. 299-329.
- [15] R. Tibshirani, "Regression shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Vol. 1*, pp. 267-288, 1996.
- [16] D. West, "Neural Network Credit Scoring Models," *Computers and Operational Research*, vol. 11, 2000, p. 1131.