The Comparative Analysis of the Models in Default Warning of the Credit Clients in Commercial Banks

Duan Yun, Zhang Jianying, Cao Lin

School of Management, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China
Email:duanyun_hit@163.com; Telephone: 01186-451-8641-4011

Abstract: Based on the real data sets which involves different periods and different sizes of cities from national commercial banks, this paper applied the methodologies of Multivariate Discrimination Analysis, Logistic Regression, Decision Tree and Neural Network to accomplishing the detailed empirical research. Through the modeling analysis, the results showed that the determinant power of Multivariate Discrimination Analysis was the most ineffective, and Logistic Regression was also unsatisfactory, while Decision Tree and Neural Network performed more effectively, especially the method of Decision Tree, whose results were prefect from several aspects of analysis. As a result of it, this paper recommended the methodologies of Decision Tree and Neural Network for national commercial banks to create models of the warning of defaults of credit clients. [The Journal of American Science. 2007;3(2):30-34]. (ISSN: 1545-1003).

Keywords: Credit Risk, Multivariate Discrimination Analysis, Logistic Regression, Decision Tree, Neural Network

1. Introduction

There exist many risks in the management of commercial banks, such as: credit risk, interest rate risk, market risk etc, where credit risk is the most significant risk without doubt. Modern credit risk measure models can be applied into the management of commercial banks, which is the main trend of future management of commercial banks.

According to the conditions of China, "classification technology" is still one of the most effective methods in credit default warning. Therefore, this paper applied classification technology and financial theories & data mining technology researching the problems of credit default models in commercial banks.

This paper selected 4 typical methodologies of classification technology: Multivariate Discrimination Analysis, Logistic Regression, Decision Tree and Neural Network, by using these models to do empirical research on the real data of commercial banks, whose aid is to compare the powers of these models, and to put forward some helpful conclusions for the management of commercial banks in China.

2. Data Resource

This paper selected two real data sets of financial ratios from the commercial banks in different periods & different sizes, including manufacturing, finance, architecture & some other industries.

Followed is the introduction of these two data sets. (Table 1)

This paper selected 19 financial ratios in for categories. These ratios cover capital structure, liability-paid ability, capital management ability & profit ability to measure the performance of the companies. (Table 2)

There are still many other factors which influence the credit conditions of companies, as a result of it, this paper applied "year" & "industry" variables to solve this problem.

3. The Results of the Empirical Research

The result of Multivariate Discrimination Analysis is as followed (Table 3):

Table 1. Data sets

	Set I	Set II
Data Resource	One of the big size commercial banks	One of the medium size commercial banks
Year	1998- 1999	2000-2002
Sum of the data	2937	133

Table 2. Financial ratios

	Capital Structure	Liability-paid Ability	Capital Management	Profit Ability
Ratios	Liability Ratio	Cash into Current Debt	Cash Turning Rate	Gross Income Ratio
	Current Capital Ratio	Current Ratio	ART rate	Net Profit Ratio
	Current Liability Ratio	Quick Ratio	Inventory Turning Rate	Return of Equity
	Fixed Asset Depreciation	Interest Cover	Fixed asset Turning Rate	Return of Asset
	Net Capital Ratio		Capital Turning Rate	
	Fixed asset ratio			

Table 3. The result of Multivariate Discrimination Analysis

Set I					Set II				
			Estimation					Estimation	
			Non-default	Default				Non-default	Default
Reality	Non-default	Sum	2438	322	Reality	Non-default	Sum	99	4
		Proportion	88.33%	11.67%			Proportion	96.12%	3.88%
	Default	Sum	92	85		Default	Sum	9	21
		Proportion	51.98%	48.02%			Proportion	30.00%	70.00%

The result of Logistic Regression is as followed (Table 4):

Table 4. The result of Logistic Regression

Set I					Set II				
			Estimation					Estimation	
			Non-default	Default				Non-default	Default
Reality	Non-default	Sum	2706	54	Reality	Non-default	Sum	97	6
		Proportion	98.04%	1.96%			Proportion	94.17%	5.83%
	Default	Sum	147	30		Default	Sum	5	25
		Proportion	83.05%	16.95%			Proportion	16.67%	83.33%

The result of Decision Method is as followed (Table 5):

Table 5. The result of Decision Tree

Set I					Set II				
			Estimation					Estimation	
			Non-default	Default				Non-default	Default
Reality	Non-default	Sum	2732	28	Reality	Non-default	Sum	102	1
		Proportion	98.99%	1.01%			Proportion	99.03%	0.97%
	Default	Sum	105	72		Default	Sum	0	30
		Proportion	59.32%	40.68%			Proportion	0%	100%

The result of Neural Network Method is as followed (Table 6):

Table 6. The result of Neural Network

Set I					Set II				
			Estimation					Estimation	
			Non-default	Default				Non-default	Default
Reality	Non-default	Sum	2746	14	Reality	Non-default	Sum	102	1
		Proportion	99.49%	0.51%			Proportion	99.03%	0.97%
	Default	Sum	127	50		Default	Sum	1	29
		Proportion	71.75%	28.25%			Proportion	3.45%	96.55%

4. ROC Analysis and Results Analysis

4.1 ROC Analysis

The size of the area below ROC curve, which is called AUC, can be applied to measure the effect of one classification method. The larger the area is, the more powerful the method is.

Through the results before, obviously, the result of Multivariate Discrimination Analysis is much worst the other three methods, so this paper only gives out the ROC curve of the other three methods.

(1) Set I

1) Logistic Regression　　2）Decision Method　　3）Neural Network Method

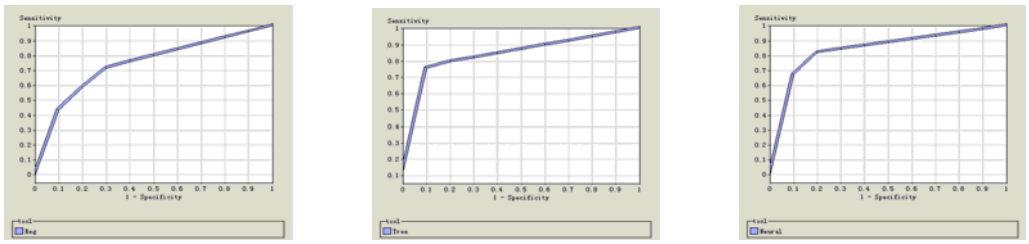

Figure1. ROC Curve (1)

(2) Set II

1) Logistic Regression　　2）Decision Method　　3）Neural Network Method

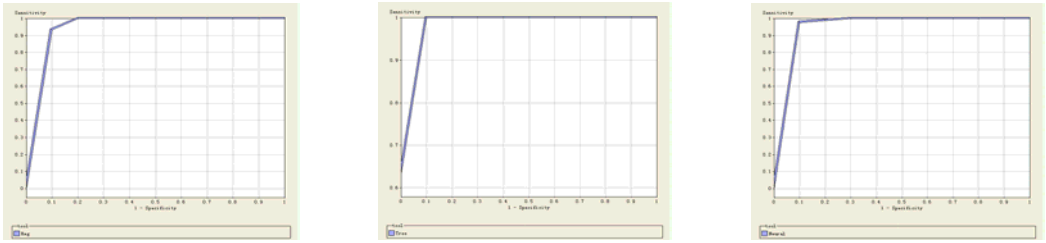

Figure2. ROC Curve (2)

Then the AUC Table is as followed (Table 7):

Table 7. The AUC results of 4 methods

	Logistic Regression	Decision Method	Neural Network
AUC (Set I)	75.03%	83.48%%	83.63%
AUC (Set II)	93.33%	98.67%%	93.77%

Through the table above, the AUCs of Decision Method & Neural Network Method are larger, which indicates that these two methods are more effective.

4.2. Results Analysis

Through the Empirical results, this paper can be concluded (Table 8):

Table 8. The total result

	Method	Accuracy to Non-default	Accuracy to Default	Second Fault Rate	First Fault Rate	Fault Rate	Total Accuracy
Set I	Multivariate Discrimination	88.33%	48.02%	3.13%	10.96%	14.09%	85.91%
	Logistic	98.04%	16.95%	5.01%	1.84%	6.85%	93.15%
	Decision	98.99%	40.68%	3.58%	0.95%	4.53%	95.47%
	Neural Network	99.49%	28.25%	4.32%	0.48%	4.80%	95.20%
Set II	Multivariate Discrimination	96.12%	70.00%	6.77%	3.01%	9.78%	90.22%
	Logistic	94.17%	83.33%	3.76%	4.51%	8.27%	91.73%
	Decision	99.03%	100%	0%	0.75%	0.75%	99.25%
	Neural Network	99.03%	96.55%	0.75%	0.75%	0.75%	98.50%

Through the total accuracy, Multivariate Discrimination Analysis Method is the worst, Logistic Regression Method also has the bad performance, while the other two methods have high accuracy.

For the estimating result of default performance, “Accuracy to Default” & “Second Fault Rate” are the concerned ratios. Through the Table above, Decision Tree Method is best one, and Neural Network Method is the second best one.

For the estimating result of non-default performance, “Accuracy to Non-default” & “First Fault Rate” are the concerned ratios. Through the Table above, Neural Network Method is best one, and Decision Tree Method is the second best one.

In General, Decision Method and Neural Network Method have the more estimating power, especially, Decision Tree Method have the more effective result; while the estimating power of Logistic Regression Method isn’t so good as the former two methods; Multivariate Discrimination Analysis is the worst one in the four.

5. Conclusion

This paper applied 4 different methodologies of classification technology to do empirical research on credit risk by two different data sets, during the research, “industry” & “year” variables were used to solve the influence of macroeconomics & industries to credit risk.

Within all the models, Multivariate Discrimination Analysis is the worst one, whose accuracy is low, and if it was used in the credit risk management of commercial banks, it would bring in quite a lot potential loss; Logistic Regression method has a low accuracy in the Second Statistical Fault, if it was used, it would drive the commercial banks in the adverse selection dilemma, which would increase the rate of non-performed loans; Decision Tree & Neural Network has better performance than the two before, especially, the former one, whose accuracy is the highest in all the models. As a result of it, commercial banks should apply Decision Tree Method to manage the credit risk themselves, which can decrease moral risk brought by adverse selection, also can decrease the proportion of non-performance loans to save more cost for commercial banks.

Correspondence to:
Duan Yun, Zhang Jianying, Cao Lin
School of Management
Harbin Institute of Technology
Harbin, Heilongjiang 150001, China
Telephone: 01186-451-8641-4011
01186-139-3609-9051(mobile)
Email:duanyun_hit@163.com

Received: 5/30/2007

References

1. Altman, Edward I., Caouette, J., Narayanan, P.. Managing credit risk: The ironic challenge in the next decade[M]. John Wiley and Sons, Inc. 1998.
2. Saunders. A.. Credit Risk Measurement: New Approaches to Value at Risk and other Paradigms[M]. John Wiley & Sons. Inc. 1999.
3. Michel, C., G., Dan and M., Robert. A Comparative Analysis of Current Credit Risk Models[J]. Journal of Banking and Finance. 2000.Vol(24):59-117.
4. L. Breiman, et al.. Classification and Regression trees[M]. Monterey, CA: Wadsworth and Brooks. 1984.
5. Huang, Z., et al.. Credit rating analysis with support vector machines and neural networks: a market comparative study. Decision Support Systems[M]. 2004.37(4):543-558.
6. Galindo, J and P. Tamayo. Credit risk assessment using statistical and machine learning: basic methodology and risk modeling applications[J].Computational Economics. 2000.15(1-2).
7. Johnson, R.A. and D.W. Wichern. Applied multivariate statistical analysis[M]. Upper Saddle River, N.J.: Prentice Hall. 2002.
8. Stephen A. Ross, Randolph W. Westerfield and Jeffery F. Jaffe. Corporate Finance[M]. Sixth edition. 2005.
9. Minhua Guo, Credit Rating[M]. China Renmin University Press.2004.(Chinese).