Credit Risk Modeling for Commercial Banks

Asrin Karimi

Full Terms & Conditions of access and use can be found at
http://hrmars.com/index.php/pages/detail/publication-ethics

Credit Risk Modeling for Commercial Banks

Asrin Karimi

Faculty of Economics and Management, University Putra Malaysia, 43400 Serdang, Selangor
Darul Ehsan, Malaysiam
Email: asrin.karimi@gmail.com

Abstract

The aim of this paper is to examine the efficiency of two credit risk modeling (CRM) to predict the credit risk of commercial Iranian banks: (1) Logistic regression model (LRM); (2) Artificial neural networks (ANNs). The calculations have been done by using SPSS and MATLAB software. Number of samples was 316 and 5 dependent variables. The results showed that, artificial neural network is more proper to identify bad customers in commercial bank. The major contribution of this paper is specifying the most important determinants for rating of customers in Iran's banking sector.
Keywords: Credit Risk, Modeling, Artificial Neural Network, Multiple Regression, Loan

Introduction

Collecting deposit of customers is one of the important roles of commercial banks. Banks as one of the financial institutions should allocate the collected deposit to borrowers and investors and should establish communication between lenders and borrowers.

Generally, officers' intrinsic assessment influence on loan applications' evaluations. Therefore, these evaluations do not have a unique form. Usually, they are inefficient. Accordingly, a proper model is needed to assist correctly in decision making regarding the application. Financial system efficiency and level of financial improvement play an important role in economic development of countries.

Banks face problems such as the probability of non-repayment of received loans at the due date or non-repayment that are called "credit risk"(Nazari & Alidadi, 2013). Previous studies have reported that credit risk is widely studied topic in bank lending decisions and profitability (Angelini et al., 2008). Generally, lenders do not have adequate data about the project to be finances, hence borrowers usually have proper information about those project (Matoussi & Abdelmoula, 2009). A good customer with power of lending loans bring high profit for commercial banks on the other hand bad customers who don't repay loans in due date, it will likely go bankrupt (Nazari & Alidadi, 2013). Regression models and neural networks are able to estimate credit risk in financial market (Vasconcelos et al., 1999). There are various ways to predict credit risk such as Probability & Deterministic Simulation, Legit Analysis, Prohibit Analysis, Arbitrage Pricing Theory, Option Pricing Theory and Linear Probability Model (Saunders and Allen, 2002).The main purpose of this paper is to identify the efficiency of two credit risk model in one of the commercial banks in Iran. This paper can estimate the credit risk of each customer and helps to make the right decision toward granting of loan to

customers. Artificial neural network (ANN) and logistic regression model (LRM) are used in this paper to achieve the above-mentioned.

The LRM model is a discrete-time event history model. This model is designed to predict the risk of an event occurring, as a function of specified variables measured before the event occurs. The logistic regression can be used to predict the risk of an event within a certain time period. This is equivalent and estimated by applying a logistic regression to issuer-year of data.

The ANN models use the same input and output parameters as in the logistic models. These models have three primary ingredients: the input data layers, the hidden layers and the output measure layer. Each of these layers contains nodes, and these nodes are connected to nodes at adjacent layers. The hidden layers contain two processes: the weighted summation functions and the transformation function. Both of these functions relate the values from the input data to the output measures. The weighted summation function is typically used in a feed forward/back propagation neural network model.

Literature Review

Credit rating is one of technical factor in credit risk evaluation (Khashman, 2010). Credit rating has two kinds of applicants which include good credit and bad credit (Ghodselahi & Amirmadhi, 2011). Multilayer feed forward networks are a class of universal approximation (Hornik et al., 1989). The ANN models have a high predictive power (Steiner et al., 2006). Networks are non-parameters model which is an arbitrarily specified degree of precision. One of the characteristics of networks are highly-sophisticated pattern recognition (Hall et al., 2009).

Pacelli and Azzollini (2011) illustrated that ANN in combination with linear methods have further supported. Salehi and Mansoury (2011) investigate the efficiency of neural network and logistic regression in forecasting customer credit risk. They state that both models have same efficiency. Credit rating is also investigated with other methods of artificial intelligence. Ghodselahi and Amirmadhi (2011) Use a hybrid method for credit rating. They use Support Vector Machine, Neural Networks and Decision Tree as base classifiers. They found that accuracy of this hybrid model is more than other credit rating methods. Credit rating is also investigated with three methods including Logistic Regression, Neural Networks and Genetic Algorithms (Gouvea & Goncalves, 2007). According of this research's result, logistic regression and neural networks are good and similar. Although neural network is slightly better and genetic algorithms take third place. The important roles of ANN in financial application are pattern recognition, classification and time series forecasting (Eletter & Yaseen, 2010).

Methodology and Data Analysis

Data has been collected through stratified random sampling method over the period 2010-2013, based on documents and records of applicants for an Iranian commercial bank. Sample estimation has been done by a pretest sample size of 90 cases and according to the sample size formula, 316 are selected as the number of samples, which are derived from legal customers' profiles. In this study dependent variables are good and bad legal customers; the aim of this study is to estimate the likelihood of good or bad customer being and also realize that how important are those factors. In this regard, good customer is a company which repays its loan plus the profit at the due date and in contrast, bad customer is a company which don't repay at the due date. To differentiate between good and bad customers in our neural network model and logistic regression model calculations, 0 is indicate good customers

and 1 is indicate bad customers. According to pervious literatures and interview with bank experts, five variables are introduced as independent variables. Independent variables include: Interest rate which it expressed as percentage and it determine amount of bank's profit, history of customer relationship with the bank, period of re-payment of loan, delay in repayment of loan, the company which is in industrial sector.

A Statistical Model: Logistic Regression Model

The logistic regression model can be used to predict the risk of an event within a certain time period. This is equivalent and estimated by applying a logistic regression to issuer-year of data. This model is designed to predict the risk of an event occurring, as a function of specified variables measured before the event occurs. The logistic regression takes the following form:

$$L = In\left(\frac{P}{1-P}\right) = B_0 + B_1 X_1 + \cdots + B_n X_n \tag{1}$$

Where p is the probability of the even occurring, and X independent variables, β, are each weighted by a coefficient.

This logic function preferable to the leaner regression function because it limits p, the probability of event, to be between 0 and 1 (or 0 and 100%), where applying the conventional leaner regression to dichotomous outcomes would instead allow nonsensical results like probabilities greater than 1 or less than 0.

$$P(Y=1) = P = \frac{e^{B'x}}{1+e^{B'x}} \quad P(Y=0) = (1-P) = \frac{1}{1+e^{B'x}} \tag{2}$$

This difference in structure between this model and more familiar linear regression models gives rise to differences in how the results of the model can be interpreted. Changing one independent variable by a fixed amount changes the level of the dependent variable by an amount that is identical, no matter what the levels of the other independent variables. In this paper to fit logistic regression mode used SPSS 18 software. The rate of repayment for legal customer at commerce bank is estimated around 0.641 by logistic regression (table 1).

Table 1
Regression test

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	-.089	.211	.178	1	0.673	0.641

The result of regression test is 0.000. It means that the regression test is significance (table 2).

Table 2

Significant Regression Test

		Chi-square	df	Sig.
Step 1	Step	78.651	10	0.000
	Block	78.651	10	0.000
	Model	78.651	10	0.000

Table 3 illustrates the regression coefficient. According to Cox & Snell, Coefficient of determination is 0.583. It means that around 60% of changes of predictive response variables explained by variations.

Table 3

The Regression Coefficient

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	45.938[a]	.583	0.777

The validity of the model to estimate the probability of good customer is 90.7 percent. As a result there is an error around 10 percent to estimate good customer (table 4). The validity of the model to estimate the probability of bad customer is 87.2 percent.

Table 4

Classification Table

		Kind of customer		Percentage Correct
	Observed	Bad customer	Good customer	
Step 1	Kind of Bad customer	41	6	87.2
customer	Good customer	4	39	90.7
	Total percentage			88.9

The estimated values of the coefficients of the independent variables in column B are observed. Also it can be seen the significance of each independent variables. Column Exp (B) demonstrates the odds ratio of independents variables. The values which are greater than one indicate a greater chance of success than failure. Also, the values which is less than one, means the chance of failure is less than success. The results of this study indicate credit risk has increasing trend when interest rate and the time delay in repayment are increase. Besides, according to results credit risk grew up if legal customers are relevant to the industry sector. Column Exp (B) demonstrates that history of customer relationship with the bank and time period of re-payment of loan have inversely related to credit risk because their Exp (B) amount is greater than 1. It means when the amount of history of customer relationship with the bank and time period of re-payment of loan are increase, the rate of credit risk is reduced.

Table 5
Coefficient of Independent Variable

	B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1ᵃ Interest rate	-.188	0.143	1.731	1	0.041	0.829
time period of re-payment of loan	1.044	0.355	8.640	1	0.003	2.841
History of customer relationship with the bank	.269	0.113	5.671	1	0.017	1.309
time delay in repayment of loan	-1.641	0.432	14.401	1	0.000	0.194
industry	-2.198	0.814	7.296	1	0.007	0.111
Constant	2.492	3.298	0.571	1	0.450	12.081

Logistic regression model is as follows:

$$\ln\left(\frac{p}{1-p}\right) = -0.188x_1 + 1.044x_2 + 0.269x_3 - 1.641x_4 - 2.198x_5$$

Where:

X_1 = Interest rate;
X_2 = Period of repayment of loan;
X_3 = History of legal customer relationship with the bank;
X_4 = Delay in repayment of loan;
X_5 = Company which is in industrial sector.

A Neural Network Approach

An artificial neural network consists of elementary computational units, known as Processing Elements (PE), proposed by MC Cullock and Pitts in 1943. Artificial Neural Networks (ANN), play an increasingly important role in financial applications for such tasks as pattern recognition, classification, and time series forecasting. The ANN models use the same input and output parameters as in the logistic regression model. Figure 1 demonstrates a simplified neural network.
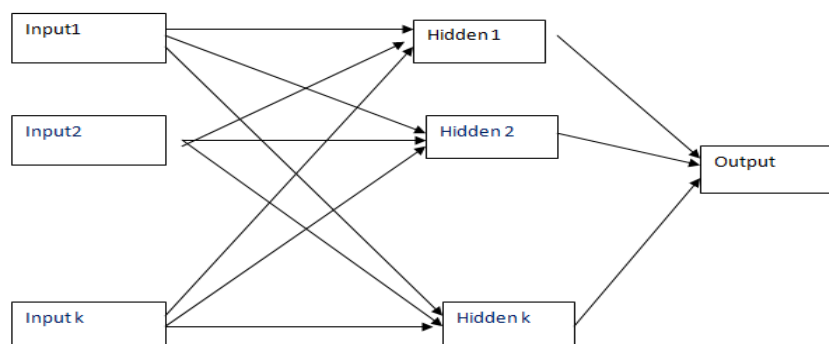

Figure 1. Simplified neural network

Each neuron is characterized by a transition function and a threshold value. The threshold is the minimum value that input must have to activate the neuron. Each neuron of this layer sums the inputs that are presented to its incoming connections. In mathematical terms, each neuron performs the summation of inputs, which are the product of output neurons of the first layer and weight of the connection. The result of this sum is again drawn on the basis of the transfer function of each neuron.

The simplest network consists of a single neuron with "*n*" inputs and one output. The basic learning algorithm of the perceptron analyzes the configuration (pattern) in- put and weighting variables through synapses, deciding which category of output is associated with the configuration.

However, this architecture presents the major limitation to solve only linearly separable problems (for each neuron output); the output values that activate the neuron must be clearly separate from the disabled through a hyperplane separation size 1 - *n*.

Artificial neural networks can be placed among those of dynamical systems with processing the experimental data, knowledge beyond transmits data to the network. The model is as follows:

1. All data are gathered in one place.
2. Data were divided into two groups: training data and test data.
3. Data entry form for entry into the system.
4. Select the network architecture.
5. Use the network which is obtained.

The proposed neural network model is a two-layer perceptron .In this paper, Matlab software is used to implement neural networks. The output of the neuron network includes a good customer of the bank account or the account of bad form. The algorithm is able to identify all six variables which are used in the model. The estimation results suggest that some variables are effective and others less effective. After training the neural network model (estimated coefficients) based on information from 90 customers, 40 customers in the performance prediction model for testing (outside of course) will be evaluated. The results are presented in table 6.

Table 6
Model Summery

	Good customers Class 0	Bad customers Class 1	Total
Good customers	12	2	14
Bad customers	7	29	36
Total	19	31	50
Percentage of correct predictions	63.16	93.54	-
Percentage of wrong predictions	36.84	6.45	-

The percentage of correct prediction for good customer is around 63 % and the percentage of correct prediction for bad customer is about 94%.

Conclusion

This study set out to determine the importance and necessity of customers' credit risk measurement and explained the logistic regression and ANN models. After determining of

required variables, the collected data were entered into the models. Results of this study show that logistic regression model has high accuracy in estimating good customers compared with ANN model and also ANN models in a comparison with LRM has high accuracy in estimating bad customers. It means that bank managers and policy makers to identity bad customers try to use ANN models. These models help banks to decline credit risk.

References

Angelini, E., di Tollo, G., & Roli, A. (2008). A neural network approach for credit risk evaluation. *The Quarterly Review of Economics and Finance*, 48(4), 733-755. http://dx.doi.org/10.1016/j.qref.2007.04.001

Eletter, S. F., & Yaseen, S. G. (2010). Applying Neural Networks for Loan Decisions in the Jordanian Commercial Banking System. *International Journal of Computer Science and Network Security*, 10(1), 209-214.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feed forward networks are universal approximators. *Neural Networks*, 2(5), 359-366, http://dx.doi.org/10.1016/0893-6080(89)90020-8

Ghodselahi, A., & Amirmadhi, A. (2011). Application of Artificial Intelligence Techniques for Credit Risk Evaluation. *International Journal of Modeling and Optimization*, 1(3), 243-249. http://dx.doi.org/10.7763/ IJMO.2011.V1.43

Gouvea, M. A., & Goncalves, E. B. (2007). Credit Risk Analysis Applying Logistic Regression, Neural Networks and Genetic Algorithms Models. *Paper presented at the Production and Operations Management Society* (POMS), Dallas, Texas, U.S.A.

Hall, M. J. B., Muljawan, D., Suprayogi, & Moorena, L. (2009). Using the artificial neural network to assess bank credit risk: a case study of Indonesia. *Applied Financial Economics*, 19(22), 1825-1846. http://dx.doi.org/10.1080/09603100903018760

Khashman, A. (2010). Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. *Expert Syst. Appl.*, 37(9), 6233-6239, http://dx.doi.org/10.1016/ j.eswa.2010.02.101

Matoussi, H., & Abdelmoula, A. (2009). Using a Neural Network-Based Methodology for Credit–Risk Evaluation of a Tunisian Bank. *Middle Eastern Finance and Economics*, 4, 117-140.

Pacelli, V., & Azzollini, M. (2011). An Artificial Neural Network Approach for Credit Risk Management. *Journal of Intelligent Learning Systems and Applications*, 3(2), 103-112. http://dx.doi.org/ 10.4236/jilsa.2011.32012

Salehi, M., & Mansoury, A. (2011). An evaluation of Iranian banking system credit risk: Neural network and logistic regression approach. *International Journal of the Physical Sciences*, 6(25), 6082-6090. http://dx.doi.org/10.5897/IJPS10.640

Steiner, M. T. A., Neto, P. J. S., Soma, N. Y., Shimizu, T., & Nievola, J. C. (2006). Using Neural Network Rule Extraction for Credit-Risk Evaluation. *International Journal of Computer Science and Network Security*, 6(5A), 6-16.

Vasconcelos, G. C., Adeodato, P. J. L., & Monteiro, D. S. M. P. (1999). *A Neural Network Based Solution for the Credit Risk Assessment Problem.* Paper presented at the IV Brazilian Conference on Neural Networks, São José dos Campos.