# A novel KMV-based commercial bank credit risk assessment model

**Xian Sheng Chen** *

*School of Management, HeFei University of Technology, Anhui PR China*

*Received 1 October 2014, www.cmnt.lv*

**Abstract**

Commercial banks has great risks. Commercial banks credit risk evaluation, as an effective way of financial regulation, is an integral part of banking regulation system. Given that scholars applied the same credit risk evaluation method to all kinds of banks, error are unavoidable. This paper targets at Z BANK, a commercial bank in China and designs a suitable credit risk evaluation model to evaluate its customer credit and debt credit. This paper draws KMV method when calculating the possibility of default PD. With some adjustments, this method can be applied to other commercial banks, providing an effective approach to financial regulation in China.

*Keywords:* commercial banks; credit risk; evaluation model; possibility of default

## 1 Introduction

While promoting the development of the banking industry, financial globalization also brings huge potential risk. As special companies that can finance asset, commercial banks are risky in many ways owing to its special business object, extensive social connections and powerful influence. Currently, the world financial regulation institution attaches great importance to credit risk evaluation of banks. In January 2001, Basel Committee on Banking Supervision announced the New Basel Accord draft that made specific recommendations on applying credit rating to financial regulation and encouraging countries to adopt credit rating in financial regulation [1] (Han, 2001).

In Western countries, credit risk management of commercial banks is mature. Many quantitative techniques and supporting tools and software have been widely used. In China, modern commercial banking system is still young. It lacks complete information and credit risk analysis is still in the traditional stage, in which only a single financial indicator is evaluated, weighted and averaged to determine the risk. The biggest flaw of this method is that determining the weights of indicators has much subjectivity, so that the results are diverged from real situation [2]. Here calls for scientific methods to determine the effective indicator, combine qualitative and quantitative analysis and establish an accurate comprehensive evaluation model to address the issue. On the other hand, Chinese commercial banks lag behind foreign ones in terms of organizational form, economic strength and the main business, etc. Chinese commercial banks should draw merits from foreign counterparts and learn about internationally accepted regulatory standards and advanced credit risk management approaches tailored to Chinese commercial banks [3]. Meanwhile, comprehensive credit risk evaluation index system is necessary for scientific credit risk evaluation.

This paper attempts to draw domestic and foreign commercial banks credit risk to build an evaluation index system for Z BANK, and uses KMV-based PD possibility of default model to do comprehensive evaluation. It intends to provide a scientific and feasible method for China's credit risk evaluation, which will be significant to perfect the financial regulation system and improve macro-control ability [4].

## 2 Literature review

At present, traditional evaluation methods of enterprise credit risk include classical analysis and multivariate statistical method. Classical analysis refers to that commercial banks rely on subjective judgment of highly trained experts. Corporate credit analysis plays upon their common sense and their own individual judgment [5-6]. Multivariate statistical method is based on historical accumulation of samples to establish the mathematical model and predicting the likelihood of the occurrence of certain events in new samples, including linear probability model, LOGIT method, PROBIT method and Multivariate Discriminant Analysis (MDA).

But the two most popular credit risk econometric models are JP Morgan's Credit-Metric model and EDF model of KMV. The main difference between the two lies in that Credit-Metric model uses historical data of corporate credit rating for analysis while KMV Model

* ***Corresponding author's*** e-mail: cxc5566@sohu.com

takes the advantage of the stock market information to measure credit risk. KMV Model plays upon market data that are changing all the time and thus forward-looking [7]. In comparison, Credit-Metric model focuses on historical data from rating agencies and thus are backward-looking.

These studies did not distinguish samples. But all are subject to the same method. This paper believes that this way of processing samples may generate bias and errors. Therefore, this paper designs credit risk evaluation index system for Z BANK and uses specific evaluation method to measure the risk, hoping that it can optimize the credit risk evaluation model for commercial banks.

## 3 Credit risk evaluation model

This section will introduce common credit risk evaluation models, including MDA and LOGIT regression model.

### 3.1 MULTIPLIED DISCRIMINANT ANALYSIS (MDA)

Discriminant model is divided into three categories in statistics: First, get the discriminant function with the smallest average probability of miscarriage of justice in the context of knowing the general distribution, which is called Bayes discriminant function; Second, without knowing the general distribution function, get the optimal linear discriminant function according to Fisher criterion; Third, without knowing the general distribution function, get the distance discriminant function between individual to the whole [8-10].

First of all, we discuss two general categories, namely $G_1$ and $G_2$. X stands for a sample, composing of p financial indicators. We define:

$$d^2(X,G_1)=(X-u^{(1)})E_1^{-1}(X-u^{(1)})^T \tag{1}$$

$$d^2(X,G_2)=(X-u^{(2)})E_2^{-1}(X-u^{(2)})^T \tag{2}$$

Expression (1) and (2) refers to the distance of X to $G_1$ and $G_2$. $u^{(1)}$, $U^{(2)}$, $E_1$ and $E_2$ are the average and covariance matrix of $G_1$ and $G_2$. Use function W(X) to express the distance between a sample and a general category. There is:

$$\begin{aligned} W(X)&=d^2(X,G_1)-d^2(X,G_2)\\ &=(X-u^{(1)})E_1^{-1}(X-u^{(1)})^T\\ &\quad-(X-u^{(2)})E_2^{-1}(X-u^{(2)})^T \end{aligned} \tag{3}$$

Suppose $E_1=E_2=E$, refer to (Fang, 1989), it is easy to prove

$$W(X)=-2[X-(u^{(1)}+u^{(2)})/2]^T E^{-1}(u^{(1)}-u^{(2)}) \tag{4}$$

Make $u=(u^{(1)}+u^{(2)})/2, a=E^{-1}(u^{(1)}-u^{(2)})$, then the discriminant function is:

$$W(X)=(X-\bar{u})^T E^{-1}(u^{(1)}-u^{(2)})=a^T(X-\bar{u}) \tag{5}$$

The discriminant rule is defined as:

$$\begin{cases} X\in G_1 & \text{when } \mathrm{w}(x)<0\\ X\in G_2 & \text{when } \mathrm{w}(x)>0\\ \quad\text{Wating to be discriminated} & \text{when } \mathrm{w}(x)=0 \end{cases}$$

As u(1), u(2), E1, E2 are usually unknown, we can estimate their value through evolvement. Suppose $X_1^{(1)},...,X_{n_1}^{(1)}$ are samples from $G_1$, $X_1^{(2)},...,X_{n_2}^{(2)}$ are samples from $G_2$. One of the unbiased estimations for $u^{(1)}$ and one for $u^{(2)}$, namely $\overline{X^{(1)}}$、$\overline{X^{(2)}}$ , are:

$$\overline{X^{(1)}}=\frac{1}{n_1}\sum_{i=1}^{n_1}X_i^{(1)} \tag{6}$$

$$\overline{X^{(2)}}=\frac{1}{n_2}\sum_{i=1}^{n_2}X_i^{(2)} \tag{7}$$

The unite unbiased estimation $S_p$of E is:

$$S_p=\frac{1}{n_1+n_2-2}(A_1+A_2)\,, \tag{8}$$

in which, $A_i=\sum_{j=1}^{n_i}(X_j^{(i)}-\overline{X^{(i)}})(X_j^{(i)}-\overline{X^{(i)}})^T, i=1,2$

However, in real situation, there are more than two categories. So, we need to deduce the multiple discriminant function from that of two general categories.

Under multiple categories, the discriminant function is:

$$\begin{aligned} W_{ij}(X)&=(X-u^{(i)})E_i^{-1}(X-u^{(i)})^T\\ &\quad-(X-u^{(j)})E_j^{-1}(X-u^{j})^T \end{aligned} \tag{9}$$

Suppose $E_i=E_j=E$, then the discriminant function is:

$$W_{ij}(X)=[X-(u^{(i)}+u^{(j)})/2]^T E^{-1}(u^{(i)}-u^{(j)}) \tag{10}$$

The unbiased estimation of u(i), u(j) and E, namely, $\overline{X^{(i)}}$、$\overline{X^{(j)}}$、 $S_p$ are calculated through expression (11):

$$\begin{cases} \overline{X^{(n)}}=\dfrac{1}{n_a}\sum_{j=1}^{n_a}X_j^{(a)} & a=1,...,k\\ S_p=\dfrac{1}{n-k}\sum_{j=1}^{k}A_j & a=1,...k\\ A_a=\sum_{j=1}^{n_a}(X_j^{(a)}-\overline{X^{(a)}})(X_j^{(a)}-\overline{X^{(a)}})^T & a=1,...,k \end{cases} \tag{11}$$

In which, $n=n_1+n_2+\ldots+n_k$ and i, j are positive integers.

Corresponding discriminant rule is defined as:

$$\begin{cases} X\in G_i & \text{when } \mathrm{W}_{ij}(X)<0,\forall j\neq i\\ \quad\text{Wating to be discriminated} & \text{when } \mathrm{W}_{ij}(X)=0 \end{cases},$$

in which, I and j are positive integers.

### 3.2 LOGIT REGRESSION MODEL

We may be concerned about the probability p of some event (such as default). It is difficult to directly put the probability p as the dependent variable to establish the regression model, because: 1) it is hard to describe the relationship between p and the independent variable in a linear way or hard to make sure the range of the dependent variable is between 0 to 1. 2) when p is close to 0 or 1, it is not sensitive to the variation of the independent variable. So we aim to find a function f(p) about p to address the issue. Hoping that the function would not be too complicated, monotonic function of p is the best choice [11]. So there is:

Use $\log\mathrm{it}(p)$ to replace $p$ and get the Logistic regression model :

$$\log\mathrm{it}(p) = \ln\frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \ldots\ldots + \beta_k x_k \tag{12}$$

or $$p = \frac{\exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)}{1+\exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)} = \frac{e^{\theta}}{1+e^{\theta}} \tag{13}$$

In which, $\frac{p}{1-p}$ is the ratio between the event that occurred and the event that don't occur. It is also called the relative risk. $\beta_i$ is the LOGIT regression coefficient that affects the sub independent variable $x_i$. Thus, LOGIT REGRESSION MODEL is a popularized form of the multiple linear regression model [12-13]. The prediction of the occurrence of the event or identifying unkonwn samples depends on whetehr p is above 0.5.

Suppose y is the dependent variable for 0-1 type. Relevant independent variables are $x_1, x_2, \ldots, x_k$. Samples are $(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i)(i = 1, 2, ..., n)$. For any sample i, its possibility $p_i$ fits:

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik})}{1+\exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik})} \tag{14}$$

$$P(y_i) = p_i^{y_i}(1-p_i)^{y_i}, i = 1, 2, ..., n, y_i = \{0|1\} \tag{15}$$

Then, we can get the Log Likelihood (LL) function:

$$L = \prod_{i=1}^{n} P(y_i) = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{y_i} \tag{16}$$

Get the logarithm for expression (16). Simplify it and there is:

$$\ln L = \sum_{i=1}^{n} [y_i(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik})] - \ln(1+\exp\{\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}\})] \tag{17}$$

Expression (17) shows the estimation value of regression coefficients $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ through gradually iterative calculation.

## 4 Enterprise credit comprehensive evaluation of Z BANK

Z BANK is one of the top five commercial banks in China, with business covering commercial banking, investment banking, insurance and air leasing. Z BANK provides financial services for individuals and companies over the globe. Its credit rating system includes customer rating and debt rating.

### 4.1 CUSTOMER RATING

The model is at the core of Z BANK customer rating under the support of the management systems and IT system. The result is production and human resources run throughout. Z BANK customer rating module is shown in Figure 1 [14].

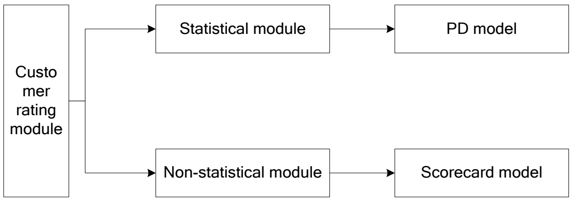


FIGURE 1 Z BANK customer rating module

PD model is categorized by size and industry segments. Financial factors and some qualitative indicators decide the output value. This model is suitable for customers who have comprehensive financial data. Scorecard model is segmentation model. It is qualitative indicators and some financial factors that decide the output value. This model is suitable for customers who lack comprehensive financial data.

PD model ratings are divided into 15 grades according to the actual PD distribution and default rate from low to high. Based on back-testing result, sections can be adjusted. Default client is directly identified as D (the lowest) level. Scorecard model ratings are divided into 10 levels according to the actual final score from high to low. Default client is directly identified as D (the lowest) level.

Customer credit rating is Z BANK authorized credit business management and an important basis for customer access and withdrawal. It is also a reference for credit approval decisions, credit pricing and credit risk classification of assets. Specifically speaking, Z BANK adopts KMV MODEL to predict the PD. Merton (1974) based on the model of Black and Scholes and proposed a model to estimate the company's debt default of credit risk according to the option theory, which came to be the famous KTV model.

It analyzes the changes in credit risk from a microscopic point of view and holds that some of the assets held by the company and its capital structure determine the company's credit risk level. When corporate liabilities such as market value or the assets value happen, there could be a default. This method is based on Merton's option pricing theory and calculates the Expected Default Frequency (EDF) through a series of processes. EDF refers to the expected probability of the assets value borrower's being less than the face value of debt, as is shown in the shadow below.

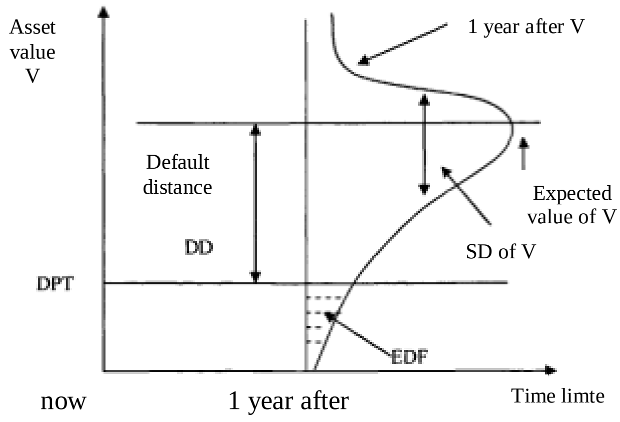


FIGURE 2 expected default frequency EDF

Merton model has two hypotheses: the option has a fixed risk free rate r before maturity; the underlying securities are traded continuously; each part can be traded; investors can trade any proportion of the securities owned; price change of the underlying securities follows geometric Brownian motion. If VE stands for the company equity value, VA for the company assets value, D for underlying exercise price of the option and r is the averaged risk free rate within the risk cycle, then according to Merton's option completely pricing model, the relationship between VE and VA is:

$$V_e = V_A \cdot N(d_1) - De^{-rt} \cdot N(d_2) \tag{18}$$

In which, $N(d_1)$ and $N(d_2)$ are standard normal distribution function.

According to Ito lemma, the relationship between company's equity volatility $\sigma_E$and assets volatility σA is [32]:

$$\sigma_E = \frac{V_A}{V_E} \cdot N(d_1) \cdot \sigma_A \tag{19}$$

Combine (18) and (19), we can get $V_A$ and $\sigma_A$.

DuanJie (2008) did an empirical study of KMV MODEL. This paper establishes a model by citing the stock information of 60 health companies in 2005 and forecasts "the possibility of deterioration" (ie given delisting warning in 2007, dubbed "ST (Special Treatment) by the stock exchange so as to remind investor of the risk of such companies). When selecting 60 stocks samples, representative industry is taken into account. 30 cases are ST shares, and the rest 30 are for the industry with a branch ST share corresponding to normal stocks.

We draw merits from DuanJie's empirical study to design the PD for Z BANK (2008):

Step 1: calculate equity value $V_E$

$V_E$ = value of tradable shares +value of non-tradable shares

= number of tradable shares × closing price at the end of year + number of non-tradable share×(-0.475+1.038×net asset value per share)

Step 2: calculate annual stock price volatility (standard deviation $\sigma_E$)

In KMV MODEL, suppose the stock price is subject to lognormal distribution. Relative yield per day is $u_i = \mathrm{Ln}(s_i / s_{i-1})$, in which, $s_i$ refers to the closing price for week *i*, $s_{i-1}$ refers to the closing price for week i-1. Thus we can get the standard deviation of weekly earning:

$$\sigma_0 = \sqrt{\frac{1}{n-1}\sum_{I=1}^{n}(u_i - \bar{u})^2} \tag{20}$$

$\bar{u} = \frac{1}{n}\sum_{I=1}^{n} u_i$. Suppose there are 250 trading days, 5 for each week. Then the annual standard deviation of stock price is $\sigma_E = \sigma_0\sqrt{250/5}$.

Step3: Based on Step 1 and 2, solve the non-linear equations of (18) and (19) and calculate the assets value $V_A$ and assets volatility $\sigma_A$ of lited companies.

Suppose the risk cycle is one year and take the one-year deposit interest rate x% issued by PBOC in some year as the risk free rate, namely, r=x%. Solve the non-linear equations.

Step 4: Do statistics of default point DPT. Make the definition according to KMV's DPT: Short Term Debt (STD) plus half of the Long Term Debt(LTD), namely, DPT=STD+0.5LTD. These data are available from annual reports of listed companies.

Step 5: Use the following expression to calculate the default distance DD

$$DD = \frac{V_A - DPT}{V_A \cdot \sigma_A} \tag{21}$$

Step 6: using historical data, to calculate the experience corresponding possibility of default value of default distance.

$$\textit{experience default possibility} = \frac{\begin{array}{c}\textit{real default times of the company whose default distance}\\ \textit{at the beginning of the year is DD within the risk cycle}\end{array}}{\textit{the number of default companies with the default distance of DD}} \tag{22}$$

For the first step, it is more difficult to calculate the value of non-tradable shares. Among Chinese listed companies, there are a large proportion of non-tradable shares [15-16]. Theoretically, the main factor that affects the value of non-tradable share is the net asset value per share. Non-tradable shares cannot be traded until its lift of ban (it is just as tradable shares which belong to the same stake in listed companies). If there is no flow, there is no price. Its equity value is manifested through net asset value per share. Since the relevant data cannot be disclosed, the common practice is to take the net asset value per share as the value of non-tradable share. This method of calculating PD has a good theoretical basis and is more objective Therefore Z BANK can use this method to predict the possibility of default PD.

### 4.2 DEBT RATING

Debt rating is a reflection of the specific risk factors during transaction, such as mortgage, priority of claim, product categories. The New Basel Accord requires banks to establish a two-dimensional internal rating system of credit business customer rating and debt rating.

Debt rating is based on Loss Given Default (LGD). Sorted by size, the debt is divided into 21 grades.

LGD is used to measure the ratio of actual losses occurred after the customer default [17]. It equals to the ratio of economic loss to default risk exposure (EAD). The economic loss is the part that failed to be recovered after default risk exposure.

Default debt recovery process typically includes collateral recovery, margin recovery and credit recovery. It should also consider direct and indirect cost of recovery.

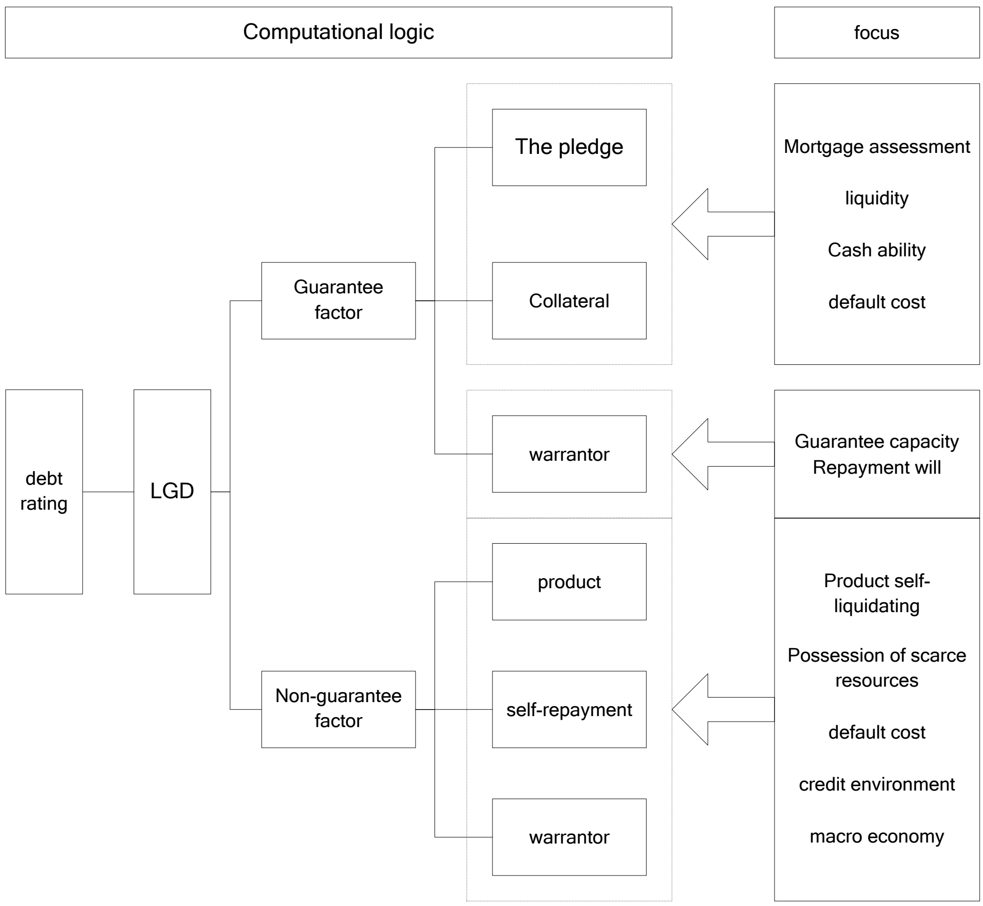


FIGURE 3 factors that influence Z BANK debt rating

Basic factors that affect Z BANK's debt rating is shown in Figure 3. After the evaluation process in Figure 3, through Figure 3, we can clearly define its customer registration and debt level. As the method is tailored for Z BANK, therefore customer credit and debt credit are real. Thus, such evaluation will be helpful to reduce the credit risk.

## 5 Conclusion

To address the poor credit assets quality and lack of credit risk prediction of China's commercial banks, and also to make up for the deficiencies of previous scholars, this paper proposes a set of customized credit risk evaluation model for Z BANK, a commercial bank in China. The model evaluates customer credit and debt credit of Z bank. When calculating the possibility of default PD during customer credit evaluation, KMV method is used. The proposed enterprise credit evaluation system only applies to Z BANK. With some adjustments, it may also apply to other commercial banks in China, providing an accurate evaluation method of enterprise credit risk for commercial banks, to improve the financial regulation system and macro-regulation capacity.

## References


[1] Han G W. 2001 On the establishment of personal credit system in China. *CHINA SOFT SCIENCE*,(06): 122-124.(*in Chinese*)

[2] Wang A J, Meng H. 2001 Solutions to China's personal credit system's establishment. *Economic Perspectives*, (02): 15-19. (*in Chinese*)

[3] Zhao H Y. 2002 Present, problems and policy suggestions to the establishment of China's personal credit system. *Review of Economic Research*, (17): 26-35. (*in Chinese*)

[4] Li W J, Li J B. 2003 Equity, board governance and enterprise credit of listed companies in China. *Management World*, (09): 103-109+156. (*in Chinese*)

[5] Yao Y L. 2004 Enterprise credit game during transition period. Economic Theory and *Business Management*, (02): 47-50. (*in Chinese*)

[6] Zhu W D, Wang J. 2013 The relative relationship between employees income rate and enterprise value based on value added *Metalurgia International*, 18(11): 49-52.

[7] Yan J. 2008 Credit risk assessment of commercial banks in China. *Shanghai: Shanghai University*. (*in Chinese*)

[8] Han L Y, Gu Z H, Ding X P. 2012 Credit risk metric under the condition of mutative variance. *Journal of Systems Engineering*, 27(5) : 633-640. (*in Chinese*)

[9] Wang J, Zhu W D. 2014 The impact of capital structure on corporate performance on panel threshold model. *Computer Modeling & New Technologies*, 18(5): 162-167.

[10] Chau A L, Li X O, Yu W. 2012 Data Selection Using Decision Tree for SVM Classification. *IEEE 24th International Conference on Tools with Artificial Intelligence*:742-749.

[11] Xu Q H, Geng S. 2012 A Fast SVM Classification Learning Algorithm Used to Large *Training Set Proc of the IEEE International Conference on Intelligent System Design and Engineering Application*. Hainan, China: 15-19.

[12] Fan R G, Chang K W, Hsieh Chojui, et al. 2008 LIBLINEAR: A library for large linear classification. Journal of Machine Learning Research, (9):1871-1874.

[13] Zhu X L. 2014 The Effect of Education on Economic Growth – An Empirical Research Based on the EBA Model. *Journal of Interdisciplinary Mathematics*, 17(1): 67-79.

[14] Chang F, Guo C Y, Lin X R. 2010 Tree Decomposition for Large-Scale SVM Problems Proc of IEEE International Conference on Technologies and Applications of Artificial Intelligence, *National Chiao Tung University*: 233-240.

[15] Menon A K. 2009 Large-scale support vector machines: algorithms and theory, *Research Exam, University of California, San Diego*:1-17.

[16] Zhu X L, Zhu W D. 2013 Utility and Collaborative Filtering-Based Evaluation Method. *International Journal of u- and e- Service, Science and Technology*, 6(4):81-90.

[17] Cui Q, Li Q, Liu Y, et al. 2011 Fast SVM classification method based on the decision tree. *Systems Engineering and Electronics*, 33(11):2558-2563. (*in Chinese*)


## Authors


**Chen Xian Sheng**

**He received his B.S. degree from Anhui Normal Universityof Technology, Wuhu, China in 1995. Now he is a doctoral candidate in School of Management, Hefei University of Technology. Currently, XianSheng Chen is serving as Vice President of the Hefei commercial Banks ,Anhui, China.**