Scientific Research Publishing

The Corporate Financial Forecasting Based on Least Squares Support Vector Machines Methods

Shunquan Zhu

School of Finance, Guangdong University of Finance & Economics, Guangzhou, China
Email: hdxgzsq@163.com

Abstract

This paper analyzed the present domestic and foreign financial forecasting situation of listed companies and it is based on least squares support vector machines. According to our country's capital markets, 44 listed companies are modeling data samples, 10 listed companies are forecasting data samples, and building financial forecasting model of listed companies obtains satisfaction financial forecasting results. The empirical study results show that we may use entirely least squares support vector machines methods to build financial forecasting models, and to distinguish financial credit risks of listed companies; comparing to traditional statistical methods and neural network methods, financial forecasting method based on least squares support vector machines is an ideal listed company's financial forecasting method. It is used to extensive fields that have high extending value.

Keywords

Financial Data Samples, Support Vector Machines, Financial Forecasting, Modeling and Simulations

1. Introduction

Recently, financial forecasting has been widely draw attention by the academic, financial worker and government. As one of the most important areas of financial market, capital market has been developed for approximately 20 years. Investment rules such as bonds, stocks, funds as well as financial derivatives are well accepted by citizens, but there still exists many manipulate behaviors in capital market. Listed company's boards and managers change the company's financial data to make fraud, aiming to sell their shares with a high price to make a huge fortune; this could hurt other financial institutes and share holders. There

are many listed companies being special treated because of abnormal financial situation. Therefore, in order to protect the sustainable development of capital market, it becomes a key issue that how to apply advanced scientific approaches to forecast listed company's financial situation. This can help to create a fair contended market atmosphere and enhance the credit awareness of listed company as well as regulate the company's finance more correctly. Traditional financial forecast approaches such as statistic and BP neural net can't solve the problems because of the features such as small sample, part minima, high-dimension, function approach and poor classification ability, slow learning speed as well as instruction in need. A new financial forecasting approach which is based on Least Squares Support Vector Machines is in need to adapt to the complex capital market. This classification can overcome the problems listed above so as to improve the quality of financial forecasting. Therefore, the research in this article could not only enrich the theory and approaches of financial forecasting, but also has a significant influence to most investors, entrepreneurs, government regulators and the health of capital market. The quality of the financial forecast will have a direct impact on the company's risk management and cost control; it also generates a profound influence on the financial regulations, commercial banks, investment banks, fund companies, insurance companies as well as other listed companies.

2. Literature

In other countries, the research of financial forecasting went through some phases: qualitative analysis, statistic methods and market value based methods, qualitative analysis include 5C elements analysis (Character, Capacity, Capital, Collateral, Condition) and LAPP principles (Liquidity, Activity, Profitability, Potentialities), DuPont Financial analysis system and Walter Weight Method. All these methods have a common flaw that is there are too much subjectivity, it is easily influenced by people's subjectivity, in order to overcome the weakness of qualitative analysis methods such as poor comprehensive analysis ability, lacking integrated generalization, lacking quantitative analysis, foreign countries began to use statistic analysis method since 1960s, Beaver [1] first introduced the forecasting function of financial variables into empirical area, he established the single variable financial forecasting model; Altman [2] then introduced the dual variables statistic analysis methods into financial forecasting. However, the statistic analysis methods require strict data such as the data should obey multivariable normal distribution and the data shouldn't have Multicollinearity as well as the paired sample covariance matrix should be the same and so on. But in reality, the data is difficult to meet this requirement. Therefore, fellow-up scholars adopted the Probit or Logistic methods to establish model, Ohlson, J. A. [3] suggested using Logistic regression to build financial forecasting model, the different effect was more significant than the researches before. Collin R.A. and R.D. Green [4] proved that the effect of Logistic model was better than that of multivariate analysis model. In the 80s and 90s of 20th century, as the technolo-

gy developed, neural network was introduced into financial forecasting; it can overcome the non-normal, non-linear financial forecasting problem. But it can't solve the problems such as small sample, part minima, high-dimension, function approach and poor classification ability, slow learning speed etc. Many financial forecast models were build in the late 90s 20th century, the most representative is Credit Metrics established by JP Morgen Bank in 1997 which is based on VaR model as well as KMV model developed by KMV company. But during establishing the Credit Metrics model, we need to confirm several parameters such as rating migration matrix, the correlation coefficient between assets, long-term yield, etc. These parameters come from long time data statistic, so far there seldom exist similar statistic data in China, it is not easy to gather large amount of data in a short time. Therefore, at present the confirmed parameters problem can't be solved if we want to establish the Credit Metrics financial forecasting model.

Chinese scholars started to research financial forecasting since 90s 20th century, Suchas: Chen, X. [5], Cheng, P. [6], Lv, C. J. [7], Wang, C. F., Zhang, W. [8], Wu, S. N. [9], Yang, S. E., *et al.* [10], Yu, L. A., Wang, S. Y. [11], Zhang, M., *et al.* [12], Zhu, S. Q. [13] [14] *et al.* These scholars published some research papers. For instance, Wu, S. N., *et al.* [9] used Fisher's linear discriminate, linear regression and Logistic regression these three methods analysis listed company, it turned out: in terms of the same data set, Logistic model got the best result. Lv, C. J. [7] did some empirical research on company's financial status and result show that profitability, asset-liability ratio, firm size company have a significant influence on the financial crisis.

In conclusion, most of the research about financial forecasting is accomplished by classical statistic and neural network methods, these methods can't solve the problems such as small sample, part minima, high-dimension, function approach and poor classification ability, slow learning speed etc. In order to solve these problems listed above, we need to explore new financial forecasting methods to adjust the complicated capital market. Fortunately the classifying method Support Vector Machines based on the Statistical learning theory can solve these problems. Therefore, after inheriting and integrating the achievements made by domestic and abroad, this article which use the China listed company financial data set as the sample data try to apply the least square Support Vector Machines to analyze the listed companies' financial forecasting. We adopt the Least Squares Support Vector Machines to build the twice classification model and simulate analysis.

3. LSSVM Method

The traditional statistics and neural network classification method are effective with enough samples, but in the practical application, this premise is generally unable to be satisfied, therefore, some theorical learning approaches are matural yet not applicable in reality Back propagation neural network (BPNN), radial basis function neural network(RBFN) *et al.* have some difficult problems, such

as the problem that how to ascertain the network structure, the over-fitting problem, the local minima problem. These are essentially because the contradictions of infinite sample in theory and finite sample in practice. Unlike the traditional statistics and neural network, Vapnik *et al.* proposed the statistical learning theory in 1968, devoted to the small-sample statistics theory. In 1995, based on statistical learning, Vapnik *et al.* presented the support vector machine theory to study how to continue pattern recognition and regression prediction within limited learning samples. It can solve small sample, non-linear, high dimension and local minima problems, so as to overcome the inherent defects of traditional statistics and neural network.

LSSVM model is a modified SVM regression; it uses the least square linear system instead of the quadratic programming in the traditional support vector machine to solve the problem of pattern recognition. The LSSVM reclassified problem in $\min \Phi(\omega,b,e) = \frac{1}{2}\omega^{\mathrm{T}}\omega + \frac{1}{2}\gamma \sum_{k=1}^{n} e_k^2$ space is formulated (Bai, P. *et al.* [15]):

$$\min \Phi(\omega,b,e) = \frac{1}{2}\omega^{\mathrm{T}}\omega + \frac{1}{2}\gamma \sum_{k=1}^{n} e_k^2 \tag{1}$$

Subject to:

$$y_k\left[\omega^{\mathrm{T}}\varphi(x_k)+b\right] = 1 - e_k, \ k = 1,2,\cdots,n \tag{2}$$

where w = an adjustable weight vector

Define the Lagrange function as follows

$$L(\omega,b,e,\alpha) = \Phi(\omega,b,e) - \sum \alpha_k\left\{\omega^{\mathrm{T}}\varphi(x_k)+b\right\} - 1 + e_k \tag{3}$$

$\alpha_k \in R$; α_k = Lagrange multiplier. seek the extreme value point of the Equation(3), let the partial derivatives of $L(\omega,b,e,\alpha)$ with respect to ω, with respect to b, with respect to e_k, and with respect to α_k are zero, the solution is:

$$\begin{bmatrix} I & 0 & 0 & Z^{\mathrm{T}} \\ 0 & 0 & 0 & Y^{\mathrm{T}} \\ 0 & 0 & \gamma I & -I \\ Z & Y & I & 0 \end{bmatrix}\begin{bmatrix} \omega \\ b \\ e \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1_v \end{bmatrix} \tag{4}$$

Simplify the matrix equation by eliminating ω and e, we can get:

$$\begin{bmatrix} 0 & Y^{\mathrm{T}} \\ Y & \Omega + \gamma^{-1}I \end{bmatrix}\begin{bmatrix} b \\ e \end{bmatrix} = \begin{bmatrix} 0 \\ 1_v \end{bmatrix} \tag{5}$$

where $Z^{\mathrm{T}} = \left[y_1\varphi(x_1), y_2\varphi(x_2),\cdots, y_n\varphi(x_n)\right]$, $Y = \left[y_1, y_2,\cdots, y_n\right]^{\mathrm{T}}$, $1_v = \left[1,1,\cdots,1\right]^{\mathrm{T}}$, $e = \left[e_1,e_2,\cdots,e_n\right]^{\mathrm{T}}$, $\alpha = \left[\alpha_1,\alpha_2,\cdots,\alpha_n\right]$, $\Omega = y_k y_l\left[\varphi(x_k)\right]^{\mathrm{T}}$, $\varphi(x_l) = y_k y_l K(x_k,x_l)$. In this paper, put $K(x,x_k) = \exp\left(-\|x-x_k\|^2/\sigma^2\right)$. Then we get the classification function as follows:

$$y(x) = \mathrm{sgn}\left[\sum_{k=1}^{n} \alpha_k y_k K(x,x_k) + b\right] \tag{6}$$

4. Variable Selection and Modeling Samples

We mainly consider the earnings per share (EPS), return on equity (ROE), net asset value per share (NAVPS), Operating revenue per share, net cash flow per share and other financial factors as the financial indicators of listed companies,. The data acquisition in this paper comes from RESSET/DB, which is developed by Zhu Shiwu, professor at the school of economics and management, Tsinghua University.

The web address is: http://www.resset.cn/cn/homepage.jsp. We selected the 54 typical listing companies—mainly industrial manufacturing companies as the data sample in Table 1.

5. Least Squares Support Vector Machine (LSSVM) Experimental Modeling and Simulation Results

Download the advanced version of Matlab: LS-SVM lab1.5 for Windows from LS-SVM lab Toolbox at http://www.esat.kuleuven.ac.be/sista/lssvmlab/ for modeling and simulation, and run it under the matlab content environment.

We conduct the empirical study by 54 listing companies with normal financial condition, the front 44 companies are considered as training dataset, remaining set of 10 data are considered as the testing dataset. The whole modeling process and simulation results are as follows:

>>X= [−0.02 −0.0241 0.84 0.22 0.05; 0.01 0.01041 0.16 0.01 0.94 0.31; −0.03 −0.027 0.07;...]% 44 listed company's financial data are considered as training dataset to construct the model

>>Y= [1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1];% we use symbol 1 as normal in the 44 listed company according to their financial features.

> > gam = 10;% set penalty factor parameter

> > sig2 = 0.2;% set radial basis kernel function σ^2 parameter

> > type ='classification';% set classification type parameters

>> [alpha, b] = trainlssvm ({X, Y, type, GAM, sig2,'RBF_kernel'});% The design values of the σ will be determined during the training of LSSVM

>>Xtest= [0.01 0.00857 0.79 0.2 0.02 0.05 0.0262 1.8 0.28 0.09;... −0.02 −0.0508 0.38 −0.04]., 4.73;% 10 listed company's data to estimate the model performance.

Table 1. 54 listed companies' financial data.

Stock code	EPS	ROE	NAVPS	Operating revenue per share	net cash flow per share	financial situation	identifier
000004	−0.02	−0.0241	0.84	0.22	0.05	Normal	1
000008	0.01	0.0104	1	0.16	0.01	Norma	1
000018	−0.03	−0.027	0.94	0.31	0.07	Norma	1
…	…	…	…	…	…	…	...
900939	−0.03	−0.071	0.38	0.26	0.1	Norma	1
900946	−0.02	−0.0508	0.38	4.73	-0.04	Norma	1

>>Ytest = simlssvm ({X, Y, type, GAM, sig2,'RBF_kernel'}, {alpha, b}, Xtest);% 10 listed company's data are considered as testing dataset to estimate the model performance

>>Ytest =1 1 1 1 1 1 1 1 1 1 % the performance of the 10 listed company's data，and symbol 1 means normal

6. Results Analysis and Conclusion

In this paper, we use the listed companies' financial data from China's capital market in December 2008 to model and simulate support vector machine for 54 non-ST companies' financial situation. It indicated that: after applying the least squares support vector machines on 44 companies financial data to establish the model, we use another 10 companies with this model; the simulating results were exactly the same with what the financial was expect to be evaluated; simulation accuracy reached 100%; this verifies that this research method is high useful in capital market.

Notice that compared with the traditional statistical and neural network methods, least squares support vector machine has the following advantages: (1 This method is specialized for finite samples; the goal is to get the optimal value within the existing information rather than the optimal value when the number of samples tends to be infinite in classical statistic. (2 This approach turns to the quadratic program so as to obtain the integrate optimum and avoid the local extreme value of classical neural network method. 3) LSSVM transforms the non-linear to high dimension space, constructs linear discriminant function in the high dimension space to replace the nonlinear discriminant function of the initial space. In other words, SVM turns the non-linear separable in low dimension space to hyperplane linearly separable in high dimension space. The property can guarantee that the machine has good generalization ability, and solves "curse of dimensionality", then the complexity of this method is independent with the dimension of the sample. Therefore, the application of least squares (LSSVM) method in China capital market can effectively identify the listing corporation's financial risk, compared with the classical statistics and neural network method; this method has its unique advantages, and it will be widely used in the future.

Acknowledgements

References

[1] Beaver, W. (1966) Financial Ratios as Predictors of Failure. *Journal of Accounting Research*, **4**, 71-102. https://doi.org/10.2307/2490171

[2] Altman, E.I. (1968) Financial Ratios, Discriminated Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, **23**, 589-609.

https://doi.org/10.1111/j.1540-6261.1968.tb00843.x

[3]   Ohlson, J.A. (1980) Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, **18**, 109-131. https://doi.org/10.2307/2490395

[4]   Collin, R.A. and Green, R.D. (1982) Statistical Methods for Bankruptcy Forecasting. *Journal of Economics and Business*, **43**, 304-349. https://doi.org/10.1016/0148-6195(82)90040-6

[5]   Chen, X. (2000) The Theory and Method and Application of Financial Distress. *Investment Research*, No. 6, 23.

[6]   Cheng, P. and Wu, C.F. (2002) New Method of Credit of Listed Cooperate. *Systems Engineering Theory* & *Practice*, **6**, 89-93.

[7]   Lv, C.J. (2004) Comparative Analysis of Financial Distress and Bankruptcy of Listing Corporation. *Economic Research*, **8**, 46-55.

[8]   Wang, C.F. and Zhang, W. (1999) Credit Risk Assessment of Commercial Banks Based on Neural Network Technology. *Systems Engineering Theory* & *Practice*, **9**, 25-31.

[9]   Wu, S.N., *et al.* (2001) Research on Financial Distress Prediction Model of Listing Corporation in China. *Economic Research*, **6**, 46-55.

[10] Yang, S.E., *et al.* (2005) Financial Early-Warning Model of Listing Corporation Based on BP Neural Network. *Systems Engineering Theory* & *Practice*, **1**, 12-26.

[11] Yu, L.A. and Wang, S.Y. (2009) Fuzzy Least Squares Support Vector Machine Model and Its Application Based on Kernel Principal Component Analysis with Variable Penalty Factor. *System Science and Mathematics*, **10**, 1311-1326.

[12] Zhang, M., *et al.* (2005) The Dynamic View of an Empirical Study on Financial Early Warning of the Listing Corporation. *Finance and Economics Research*, **31**, 62-70.

[13] Zhu, S.Q. (2009) The Credit Classification Modelling and Applications of Listed Companies Based on Option Pricing Theory. *Statistics and Information Forum*, **7**, 23-38.

[14] Zhu, S.Q. (2009) Financial Modeling and Computation. Electronic Industry Press, Beijing.

[15] Bai, P., *et al.* (2008) The Theory of Support Vector Machine and Its Application in Engineering, Xidian University Press, Xi'an, 71-72.