Grey Vector Model Based on Residual Error Correction and Its Application in Financial Risk Prediction

Hailei Zhao
School of Business, Jiangnan University, Wuxi 214122, Jiangsu, China

Dehuan Jin*
School of Finance, Shanghai University of Finance and Economics, Shanghai 200433, China
*Corresponding author

Abstract
Financial risk prediction is a kind of complex prediction problem of lacking the information, small sample and uncertainty. To improve the accuracy of financial risk evaluation, this paper puts forward a kind of financial risk evaluation model of support vector machine. Firstly preprocess the risk evaluation index via the fuzzy system, and then input the data into the support vector machines to proceed with the learning and optimize the support vector machines with residual error correction algorithm to build the intellectualized evaluation model of the financial evaluation and finally verify the performance of model through the simulation experiments. The simulation experiments show that, the support vector machine improves the accuracy of financial risk evaluation, overcomes the defects of conventional risk evaluation model and is a kind of risk evaluation model of information safety with excellent perform.

Key words: Evaluation System, Financial Risk, Risk Prediction, Bond, Shares, Financing.

1. INTRODUCTION

In recent years, with the constant development of economic globalization and information technology, financial internationalization increasingly speeds up and the effect of financial risk management is more and more important. Various transnational corporations and regulatory authorities of various countries all input a large number of energy, manpower and financial resources to develop the risk management technology. So, the measurement and analysis technique of the risk progress in rather rapid speed and can soon be transformed into the economic benefits. As financial risk has the peculiarity and its risk evaluation is different from the risk evaluation of the conventional industrials, domestic and foreign scholars have put many kinds of financial risk evaluation methods (Peng, et al., 2011; Eibe, et al., 2005; Li and Shi, 2012).

There mainly are qualitative evaluation methods, quantitative evaluation methods, comprehensive evaluation methods with combination of quantification and qualification. Qualitative evaluation methods, mainly containing the ALE-based method, OCTAVE method, CORAS method, relies on the knowledge and experiences of the experts, has strong subjectivity in the evaluation results and is hard to precisely measure the size of risk; quantitative methods adopts the traditional model description to describe the financial risks while the financial risk possesses the nonlinearity, complexity and other characteristics, cannot be exactly described by the conventional mathematic models and its implementation is too complex and tedious and has strong limitations. Comprehensive evaluation methods fuse the advantages of quantification and qualification and become the method of main current in the present financial risk evaluation, mainly including neutral network, wavelet neutral network, Markov chain and support vector machines (Liu and Dai, 2007). As the financial risk evaluation actually is a kind of identification problems small samples (data) and nonlinear models, Markov chain lacks the learning and adaptive ability and cannot describe the nonlinear variation features; in the case of large samples, neutral network has high performance, but for the small samples, it has the defects of over-fitting and local optimum, etc.; support vector machine is a kind of model recognition algorithm professionally aiming at the limited samples while its performance is superior to the neutral network and becomes the important research direction (Pauly and Fricke, 2009) of financial risk evaluation.

In the practical application, performance of support vector machine is closely related to its parameters while the genetic algorithm (GA), particle swarm optimization (PSO) are adopted to select the SVM parameters, but as the defect of local minimum is easily to occur in the GA, PAO, it is hard to find the optimized parameters (Giudici and Polasek, 2001). To further improve the accuracy of financial risk evaluation, starting from the nonlinear and small-samples characteristics of financial risk evaluation, and combining with the fuzzy theory, ant colony optimization and support vector machine, this paper puts forward a kind of evaluation models of financial risk with advanced support vector machine and verify the performance of models via specific tests.
2. CONSTRUCTION OF FINANCIAL RISK MODEL

2.1. Data Pre-processing and Slope Calculation

Financial pre-processing contains two parts of financial data screening and financial data sorting. Because of the complex and diversity forms of financial data gathered and the big interfering noise which severely influences the word frequency statistics, the financial data shall be screened. After screened, the financial data shall be sorted and the sorting ways are: (1) extract the time information of financial data and be sorted in time order; (2) divide the financial data into several groups in certain quantity interval and each group can be called a window (for example, a group of 300).

\[
c_n = \{C_n(1), C_n(2), \cdots, C_n(k)\}
\]  

(1)

\[
V_w = \left[\{(C_n(1), t_{d1}), (C_n(2), t_{d2}), \cdots, (C_n(k), t_{d_k})\}\right]
\]  

(2)

For which each word forms a time series, such as the word \( w \), we represent it as:

\[
c_w = \{C_w(1), C_w(2), \cdots, C_w(k)\}
\]  

(3)

In Formula (1), \( c_n(i) = TF_n(i) \) namely, is the frequency of the word \( w \) in the \( i \) window. Because the time information of financial data has been extracted while the financial data sorting, count the earliest and latest times of every widow and calculate the half of the difference between the both times to record the amount as \( t_{d} \), so the time series of every word can be represented with two-tuples sequence shown as follows:

\[
V_w = \left[\{(C_w(1), t_{d1}), (C_w(2), t_{d2}), \cdots, (C_w(k), t_{d_k})\}\right]
\]  

(4)

That is, word frequency of the word \( w \) in every window and the middle moment of the widow form a two-tuples.

To master the development trend of the network incidents in time, the slope is utilized to analyze the time series of two-tuples and find the evolution trend of suspected words of hot-spot themes to check out the hot spots. Give a time series: \( V_w = \left[\{(C_w(1), t_{d1}), (C_w(2), t_{d2}), \cdots, (C_w(k), t_{d_k})\}\right] \), and utilize the slope to calculate the growth trend of the word \( w \).

\[
SLO_w = \frac{C_w(i) - C_w(i-1)}{t_{d_i} - t_{d_{i-1}}}
\]  

(5)

In Formula (5), unit of \( t_{d} \) is hour.

That is to calculate the slope of every two windows (next to each other) of a word, if the slope is bigger and bigger or the growth trend appears by and large in the slope, we can think that the word may become the word of hot-spot theme. Via the mentioned ways, as it can be found that some words increase in high speed and some words grow very slow while the growth slope levels off to zero, we sort the growth speed of each word and select the word of which the growth speed is bigger than the threshold value \( T \) to be the theme word.

2.2. Risk Evaluation System

What this paper researches is involved from the conventional banks financing plans to the three parties of banks, medium and small enterprises and guarantor enterprises. There into, the guarantor enterprises are the enterprises in the same industry of the medium and small enterprises that have the supervised obligations for the medium and small enterprises to make the private profit of the medium and small enterprises to be \( b_i \) after shirked the duty while the external supervision cost \( \lambda_i \) shall be paid. If the project successes, the banks, medium and small enterprises and guarantor enterprises shall divide \( R \) in a certain principle; if the project fails, the scrap value \( C \) shall be taken out to be the guarantee, \( \delta_i C \) \( (0 < \delta_i < 1) \) will be gotten after the guarantor enterprises realized the scrap value and the guarantor enterprises are responsible for compensating for a certain amount of \( \beta_i \delta_i C \) \( (0 < \beta_i < 1) \) to the banks when the project fails. Expectant basic rate of income of banks in the loaning period is \( \gamma_i \). Incomes of the banks, medium and small enterprises and guarantor enterprises respectively are: \( R_{b_i}, R_{m_i}, R_{g_i} \). Relation among the three incomes is Table 1:
Table 1. Income Statement of Three Parties of Banks, Medium and Small Enterprises and Guarantor Enterprises

<table>
<thead>
<tr>
<th>Total amount</th>
<th>Banks</th>
<th>Medium and small enterprises</th>
<th>Guarantor enterprises</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success</td>
<td>$R$</td>
<td>$R_b^i$</td>
<td>$R_s^i$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$x_1$</td>
<td></td>
</tr>
<tr>
<td>Shirk</td>
<td>$R$</td>
<td>$R_b^i$</td>
<td>$R_s^i + b_i$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$x_1$</td>
</tr>
<tr>
<td>Fail</td>
<td>$C$</td>
<td>$x_1$</td>
<td></td>
</tr>
<tr>
<td>Shirk</td>
<td>$C$</td>
<td>$x_1$</td>
<td></td>
</tr>
</tbody>
</table>

There into: $R = R_b^i + R_s^i + R_i^i (i = G, B)$

To avoid the moral risk of the medium and small enterprises, meet the condition of the banks consenting the loan and satisfy the participation and supervision constraints, in the case of $A < I$, the model 1 is built:

$$
\text{max } P_{\mu}^i \cdot R^i_s - A
$$

(6)

s.t.

$$
\begin{align*}
&P_{\mu}^i \cdot R^i_s \geq P_{\mu}^i \cdot R^i_b + b_i \quad \cdots \\
&P_{\mu}^i (R - R_b^i - R_s^i) + (1 - P_{\mu}^i) \beta_i \delta C \geq (I - A)(1 + r_i) \cdots \\
&P_{\mu}^i \cdot R_s^i + (1 - P_{\mu}^i)(1 - \beta_j) \delta C - x_i \geq 0 \quad \cdots \\
&P_{\mu}^i \cdot R_s^i + (1 - P_{\mu}^i)(1 - \beta_j) \delta C - x_i \geq P_{\mu}^i \cdot R^i_s + (1 - P_{\mu}^i)(1 - \beta_j) \delta C \cdots
\end{align*}
$$

(7)

Analysis to the model 1:

Getting the Inequation (1) as the equality sign, the minimum income of the medium and small enterprises is: $R_s^i = \frac{b_i}{\Delta P^i}$. Obviously, in the cases of no supervision, as the medium and small enterprises can obtain the private income $B \ (B > b_i)$ while shirking the duty, to make the medium and small enterprises to fulfill its duty, the minimum income of the medium and small enterprises shall be: $\frac{B}{\Delta P^i}$. Because $\frac{b_i}{\Delta P^i} < \frac{B}{\Delta P^i}$, when the income of medium and small enterprises is, the medium and small enterprises will fulfill their duties without the supervision and will shirk their duties without the supervision and thus the In equation (3) and (4) produced. In equation (4) guarantees that the income of guarantor enterprises obtained during the supervision is bigger than that of the case without supervision and the guarantor enterprises will be certain to supervise the medium and small enterprises. Inequation (3) meets to the participation constraint, that is, the guarantor enterprise will get the positive income while the supervision. Inequation (2) guaranteed the participation constraint of banks when the medium and small enterprises fulfill their duties under the supervision.

$$
A \geq \bar{A}_i = I - \frac{P_{\mu}^i (R - \frac{b_i}{\Delta P^i} - \delta_i C) + \beta_i \delta_i C - P_{\mu}^i x_i}{1 + r_i} (i = G, B)
$$

(8)

In conclusion, when the amount of own funds is bigger than $\bar{A}_i$, the medium and small enterprises can loan from the banks and the three parties of banks, medium and small enterprises and guarantor enterprises all can get the incomes.

3. RESIDUAL ERROR CORRECTION SVM FOR FINANCIAL DATA RISK PREDICTION

3.1. Residual error correction SVM mode

The financial data hotspot by many factors with complex characteristics of multi scale, business, mutability and so on, which makes classical statistical theory and method have not been applied to the research of financial data hotspot prediction. Residual error correction SVM model has obtained good results in the field of linear and nonlinear prediction respectively. However, they all have their own shortcomings. It is not suitable to use the residual error correction model for nonlinear problems. Similarly it will make people confused to use SVM to solve the result for the linear problem. SVM should not blindly and mechanically apply for any type of data. However, in practical application, it is difficult to fully know the characteristics of financial data hotspot. Simply using the residual error correction SVM for prediction all likely leads to large error and low prediction accuracy. Therefore, it is a good option to combine the linear prediction model ARIMA and nonlinear prediction model SVM to forecast the financial data hotspot.

The working principle of combination model of residual error correction SVM financial data hotspot prediction is: first, the residual error correction SVM financial data hotspot prediction model is established to
obtain the preliminary prediction results of hotspot prediction of network financial data. Then according to the residual error correction SVM hotspot prediction of financial data, SVM is established. Finally, SVM prediction results revise the financial data hotspot prediction of residual error correction SVM to obtain the final prediction results of financial data hotspot prediction. The respective advantages of the residual error correction SVM are given full play to revise the residual error correction SVM results by adopting the strong points while overcoming weak points.

![SVM Financial Data Hotspot Prediction Model Diagram](image)

### 3.2. Financial data risk prediction process

Assumed that there are n learning samples in total \( \{x_i, y_i\}, i=1,2,\ldots,N; x_i \) is sample input; \( y_i \) model is the output expected value. SVM estimation function is: \( f(x) = w \cdot \varphi(x) + b \), where, \( w \) is weight vector; \( b \) is bias vector quantity.

Optimization function is used to optimize the target value. That is:

\[
\min J = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} (\xi_i^+ + \xi_i^-) \tag{9}
\]

Constraint conditions are:

\[
\begin{align*}
  y_i - w \cdot \varphi(x) - b & \leq \varepsilon + \xi_i^+ \\
  w \cdot \varphi(x) + b - y_i & \leq \varepsilon + \xi_i^- \\
  \xi_i^+, \xi_i^- & \geq 0, i = 1,2,\ldots,n
\end{align*}
\]

Where, \( \xi_i, \xi_i' \) is relaxing factor; \( C \) is penalty factor.

By introducing lagrangian multiplier, the above optimization problem is changed into typical convex quadratic optimization problem. That is:

\[
\begin{align*}
  L(w, b, \xi, \varepsilon, \alpha, \alpha', \gamma, \gamma') &= \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} (\xi_i^+ + \xi_i^-) \\
  &\quad - \sum_{i=1}^{n} \alpha_i (\xi_i^+ + \varepsilon - y_i + f(x)) - \sum_{i=1}^{n} \alpha'_i (\xi_i^- + \varepsilon - y_i + f(x)) \\
  &\quad - \sum_{i=1}^{n} (\xi_i \gamma - \xi_i' \gamma')
\end{align*}
\]

Where, \( \alpha_i \) and \( \alpha'_i \) are lagrangian multiplier.

In order to speed up the solution speed, convert the formula (11) turns into dual form. That is:

\[
W(\alpha, \alpha') = -\frac{1}{2} \sum_{i,j=1}^{n} (\alpha_i - \alpha_j')(\alpha_j - \alpha_j')(\varphi(x_i), \varphi(x_j))
\]

\[
\varphi(x_i) + \sum_{j=1}^{n} (\alpha_j - \alpha_j') y_j - \sum_{j=1}^{n} (\alpha_j - \alpha_j') \varepsilon
\]

For linear regression problem, support vector machine function is:

\[
f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i')(\varphi(x_i), \varphi(x)) + b
\]

375
For nonlinear prediction problem, \( k(x_i, x) \) is used to replace \( \phi(x_i), \phi(x) \) to operate, which can avoid curse of dimensionality. That is:

\[
f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha'_i) k(x_i, x) + b
\] (14)

From the SVM model process, there is a great correlation between the performance of SVM prediction and its parameters (Fu, J.J, 2015). Residual error correction algorithm is a kind of intelligent bionic algorithm, which absorbs the behavioral characteristics of intelligent algorithms and achieves remarkable results (Jing, S., Li, X., Jie, N. J., 2015) in combinatorial optimization problem solving through its internal search mechanism. For that reason, in this study, the residual error correction is used to optimize the SVM parameters and will be used in financial risk assessment to improve the accuracy of assessment.

4. EXPERIMENT AND ANALYSIS

4.1. Data preparation

Sequence chart of return series \( r_t \) of Shanghai Stock Exchange corporate bond is shown in Fig. 2. As can be seen from the figure 2, there is obvious fluctuation cluster phenomenon for return series. Namely a big fluctuation is closely followed by another big fluctuation; a small fluctuation is closely followed by another small fluctuation. From the basic statistical analysis of Table 1, return series coefficient of skew corporation bond index is less than 0; kurtosis coefficient is more than 0, which shows the distribution of corporate bond market return of our country is left advertence distribution with the obvious spike feature. From value P of JB statistical magnitude, \( r_t \) is not normal distribution. From QQ normal distribution diagram of \( r_t \) of Fig. 3, upper right of \( r_t \) deflects from the straight line and tilts down; lower left deflects from the straight line and upwards upward, which shows that the upper tail and lower tail of \( r_t \) have obvious thick shadow. It is not reasonable to use normal distribution to simulate the change of enterprise bond market return.

![Figure 2. Return Series Sequence Chart](image)

![Figure 3. Return Series Q-Q Chart](image)

<table>
<thead>
<tr>
<th>Table 2. Basic Statistical Characteristics of Bond Index Market Price Daily Return Rate Series</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean value</td>
</tr>
<tr>
<td>-------------------------------------------------------------</td>
</tr>
<tr>
<td>9.67E-05</td>
</tr>
</tbody>
</table>

Roots of unity test methods are often used for stationary test of return series. Test results are shown in Table 2. At the 1% significant level, return series \( r_t \) is stable.

<table>
<thead>
<tr>
<th>Table 3. ADF Roots of Unity Test Results of ( r_t ) Stationary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable quantity</td>
</tr>
<tr>
<td>-------------------------------------------------------------</td>
</tr>
<tr>
<td>( r_t )</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Notes: C, T, L of (C,T,L) respectively represent whether there is intercept term, trend term and lagging number. The determination of lagging number is carried out with automatic selection according to Schwarz information criteria.
4.2. Interpretation of result

Bank of Communications (601328) of Shanghai Stock Exchange is used for experimental subject. Opening price, the top price, the bottom price, amount of increase, amplitude, total hands, transaction amount, changed hands, number of transactions of stock and other index are regarded as input index of IELM. Second day closing price of stock is regarded as output index. 200 data are collected in total. Changing curve of closing price is shown in Fig. 1. The first 150 samples are selected as training set, which is used for training and improvement of extreme learning machine so as to build stock price forecasting model. The remaining 50 samples are regarded as test set to test the generalization ability of stock price model.

![Graph](image1.png)

**Figure 4.** Closing Price of 601328

![Graph](image2.png)

(a) Delay Time (τ) and Changing Curve of Mutual Information Function

![Graph](image3.png)

(b) Changing Curve of Embedded Dimension (m) and Correlation Dimension

**Figure 5.** Determination of Delay Time and Embedded Dimension of Closing Price of Stock
The stock closing price is affected by many factors with chaos characteristic. Therefore, the delay time (τ) and embedded dimension (m) are needed to be determined to reconstruct its phase space and explore information in time series of the stock closing price so as to establish learning sampling. Specific as follows: (1) mutual information method is used to calculate the delay time of stock closing price (τ). The changing curve of the delay time and mutual information value are shown in Figure 5(a). From Figure 5(a), when τ=8, mutual information function obtains the first minimum value without too great fluctuation, then the optimal stock closing price τ=8 is determined. (2) saturated correlation dimension is used to calculate the embedded dimension of stock closing price. The changing curve between correlation dimension and embedded dimension is shown in Figure 5(b). From Figure 5(b), when m=6, the correlation dimension is in a steady state basically without change, which shows that the optimal stock closing price m=6. Adopting τ=4, m=6 normalized stock closing price is reconstructed. Learning samples of extreme learning machine is obtained.

4.3. Performance comparison

Residual error correction SVM algorithm and DO-ELM are used to study the training set of stock closing price. Prediction model of corresponding stock closing price line is established, of which the prediction results of prediction set is shown in Figure 6. From Figure 6, prediction accuracy of IELM stock closing price is improved compared with DO-ELM, which better reflects the changing trend of the stock closing price.

![Figure 6. Residual Error Correction SVM and Closing Price Prediction Results of DO-ELM](image_url)

Computation time of residual error correction SVM and DO-ELM stock closing price in prediction of step from 1 to 40 is shown in Figure 7. From Figure 7, compared to DO-ELM, the computation time of residual error correction SVM is greatly reduced. The greater the prediction step, the more obvious of advantages of residual error correction SVM, which is mainly because residual error correction SVM introduces Cholesky decomposition method to make the network output weight value βi carry out recursion solving updating on the basis of historical training data so as to simplify the modeling process and improve the learning efficiency and better overcome the shortcomings that each time when a sample is added to DO-ELM, the output weight value is needed to be solved again. Compared to DO-ELM, residual error correction SVM is more suitable for real time and online prediction of financial analysis.

![Figure 7. Comparison of Calculation Time between IELM and DO-ELM](image_url)
5. SUMMARY

Specific to small sample, nonlinear characteristics of financial risk, and financial risk assessment model with improved support vector machine is proposed and is carried out with specific simulation experiment. The minimum self-owned amount of money needed when loaning of bank, Internet banking, logistics finance is compared. Under the circumstances of different personal income, interest rate, residual rate, the cost of supervision, financial analysis prediction evaluation system and method are determined. Simulation results show that residual error correction with support vector machine has a good adaptability with more reliable evaluation results, which has a very good application prospect in the financial security risk intelligence evaluation.

REFERENCES


