# Support Vector Machine based Forecasting of the Contract Prices of Stock Index Futures

Yang Yuanyuan[1], Su Mingche [2],Zhou Yang [3],He Chucong [4]

[1] School of Economics, Jilin University, Changchun China

[2]School of Computer Science and Technology, Jilin University Changchun China

[3] School of Philosophy and Social Sciences, Jilin University, [4] School of Business, Jilin University

sumingche@gmail.com

**Abstract-*Finance is the core of modern economy. Financial derivatives first appeared in the 80s of the 20th century. As one of the important financial derivatives, the stock index futures have developed only recently but have become one of the most successful derivatives. Its impact can be seen in many corners of the financial markets. With the continuous development of financial markets, it is necessary to forecast the trends of stock index futures. This article first discusses the importance of forecasting stock index futures and surveys the stock index futures forecasting methods including the artificial neural networks and support vector machines. Then it describes the support vector machine-based forecasting model for stock index futures' contract price and finally validates its scientific soundness using a case study.***

**Keywords:** ***Stock Index Futures, Forecasting of Stock Index Futures, Neural Network, Support Vector Machine***

## I. THE IMPLICATIONS OF FORECASTING STOCK INDEX FUTURES

Stock index futures refer to the futures contracts whose commodity is the stock index. The stock index is a reference number created by the stock exchange agency or the financial services provider to indicate the stock market's trend. Thus, stock index futures have the following characteristics: the stock index as the commodity and index points as the quotation unit. The contract price of the stock index futures is the product of the quotation of stock index futures and a certain kind of currency. The stock index futures' position is settled with the calculated difference in cash. Predicting the stock index futures has the following implications: (1) you can circumvent the risks in the stock market. The risks of stock market primarily consist of systematic and non-systematic risks. The nonsystematic risk can be prevented through a portfolio of investments which is not applicable to the prevention of systematic risks. The systematic risk is especially high in China and there is an urgent need to avoid suck kind of risk ;( 2) it is conducive to fostering institutional investors and to promoting the standardization of the stock market. Take China as an example, the institutional investors account for only a small portion of all investors which is not beneficial to the standardization of the stock market. Forecasting the stock index futures can provide institutional investors with an effective governance tool, diversify the investment products, promote the long term development of the investment portfolio and rational trading, increase the market liquidity, reduce transaction costs of institutional investors and improve efficiency of funds use; (3) the prediction of stock index futures can promote the reasonable fluctuation of the stock price and function as the economic barometer. The lack of risk prevention mechanism forces the institutional investors to rely on only the insider information to carry out the short-term investments and subject to the fluctuation of the stock price.

Finally, the prediction of stock index futures

978-0-7695-4527-1/11 $26.00 © 2011 IEEE
DOI 10.1109/BIFE.2011.113

can improve the stock market's function and enhance China's capital market's competitiveness in the international competition. As can be seen clearly in the international market, the stock index futures can effectively improve the function of capital market system and this functionality has been widely recognized. The transition from the traditional transaction methods to a whole new perspective of economic development is conducive to the integration of China's capital market with the international financial markets and can provide the international investment capital with a place to prevent risk. China's capital market is still imperfect which will affect the entrance of foreign capital. Therefore, the development of stock index futures and its prediction mechanism are the top priorities in the market development as it has great strategic significance in promoting the development of China's capital market.

## II. Forecasting Methods for Stock Index Futures

Although the stock market is quite complex as a nonlinear dynamic system, especially, its operation is still amenable to some rules. Meanwhile, the increasingly close relation between the social and economic development ties the growth trend of stock index futures to the political, economic and social factors. The traditional statistical analysis methods are no longer superior in handling the constantly changing stock market. Therefore, some new forecasting methods are introduced in the financial sector this year.

### *A. Using Neural Network to Forecast the Stock Index Futures*

Since the 1980s, the concepts of non-linearism and non-equilibrium began to gradually enter the mindset of the people and there were more and more approaches to deal with this nonlinear system. As the scientific nature of such approaches was gradually recognized and the computational means progressed, many economic analysts are beginning to introduce the nonlinear system as a new tool into the economy.

In recent years, artificial neural network is used as an extension of the nonlinear system and played a significant role in the prediction of the stock index futures. This method can directly use the input and output variables of the neural network for training and correction which helps determine the mutual impacts and connections between these economic variables. The neural network has the characteristics of smooth interpolation which makes it more suitable to fit the data than other means.

The current domestic research on the artificial neural network can be divided into error generalizing neural networks, fuzzy neural networks, wavelet neural networks, radial basis networks and recurrent neural networks. Most of the neural network models require different premises and lack the scientific basis for choosing data variables which reduces the model's accuracy, produces many irrelevant variables and neglects many important variables. In addition, with the popularity of the neural network forecasting methods in recent years, many neural networks based prediction models have appeared. However, none of them has a clear idea on what type of artificial neural network is better for the prediction of stock index futures. Finally, some advanced technical methods have been proven to be effective in successfully establishing new hybrid systems whose initial values can be predetermined to improve the prediction speed.

### *B. Using Support Vector Machine to Forecast the Stock Index Futures*

Vapnik introduced the concept of Support Vector Machine in the 1990s. It is a new machine learning method that is built on the VC dimension and structural risk minimization principle and based on the statistical learning. Its sound theoretical foundation, good learning performance and prediction performance have made it receive wide attention from the research community. In addition, SVM can transform the nonlinear problems into higher dimensional special problem which can be solved by a summation function and avoid the dimensionality and local minima problems.

However, the SVM's prediction accuracy still needs improvement and the sequence's selection and handling needs standardization and scientific justification. One way to improve them is to build a dynamic model and constantly add new training samples to the model. In addition, the choice of summation function is not easy as different summation functions will produce completely different results. The use of different summation functions in the regression prediction can have huge impact on the data fitting results. Thirdly, the huge amount of data pose a computational challenge on SVM's learning speed.

## III. Support Vector Machine based Forecasting of the Contract Prices of Stock Index Futures

### *A. Overview of Support Vector Machine*

Support Vector Machine (SVM) is an emerging learning method in recent years. Compared with the general neural network, the support vector machine algorithm transforms the problem into a quadratic optimization problem and infers the global optimum which can avoid the local maxima problem in the neural network. The topology of support vector machine is determined by the support vector which avoids the experiment and trial method that is required in the neural network training. In addition, SVM's optimal solution is based on structural risk minimization idea and can generalize to more problems than other non-linear function approximation methods.

### *B.principles of SVM*

Given a data set $H = \{(x_i, y_i), i = 1,2,\ldots,n\}$ where $x_i$ is the input variable, $y_i$ is the expected value and $n$ the total number of data points as the training samples. The supervised learning will seek a function $f(x)$ such that $y_i = f(x_i)$ and there is a *y* for any $x_i$ that is not in the data set *H*. The estimated function is

$$f(x) = \omega\varphi(x) + b \quad \omega \in R^{nh}, b \in R$$

Where: $\varphi(x)$is the nonlinear mapping function from the input space $R^n$ to the high dimensional feature space $R^{nh}$ and $b$ is the offset. According to the structure risk minimization principle, the function estimation problem is to find the function $f(x)$ that minimizes the risk

$$\min Q(\omega, b) = \frac{1}{2}\|\omega\|^2 + CR_{emp}$$

Where $\|\omega\|^2$ reflects the generalization ability of the regression function $f(x)$ and is the normalized part; $C$ is the penalty factor and $R_{emp}$ is the empirical risk (*i.e.*, the accumulation of the sample loss function). The commonly used sample loss functions include the quadratic loss function, Huber function, Laplace function and $\varepsilon$ non-sensitive function. The $\varepsilon$ non-sensitive function can tolerate the regression error within $\varepsilon$ range so the sample loss is represented by the $\varepsilon$ non-sensitive function

$$|y - f(x)|_\varepsilon = \max\{0, |y - f(x)| - \varepsilon\}$$

This means the error term $\varepsilon$ that is less than is not penalized and r the risk is

$$R_{emp} = \frac{1}{n}\sum_{i=1}^{n}|y - f(x)|_\varepsilon$$

The search for the risk minimization problem becomes

$$\min Q(\omega, b) = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$

$$\text{s.t.} \begin{cases} y_i - f(x_i) \le \varepsilon + \xi_i \\ f(x_i) - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases}$$

By the dual principle, Lagrange multiplier method and nuclear technology, the dual form of the optimization problem becomes

$$\max Q(a_i, a_i^*) = -\frac{1}{2}\sum_{i,j=1}^{n}(a_i - a_i^*)(a_j - a_j^*)k(x_i, x_j) - \varepsilon\sum_{i=1}^{n}(a_i + a_i^*) + \sum_{i=1}^{n}y_i(a_i - a_i^*)$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{n}(a_i - a_i^*) = 0 \\ 0 \le a_i, a_i^* \le C \end{cases}$$

From the above optimization equation, we can obtain $a_i$, $a_i^*$ and the regression SVM model

$$f(x) = \sum_{i=1}^{n}(a_i - a_i^*)k(x_i, x) + b$$

We use the commonly used kernel function which is the radial basis function

$$k(x_i, x) = exp\left\{-\frac{\|x - x_i\|^2}{\gamma^2}\right\}$$

where $\gamma$、 $d$ are the nuclear parameters. The bias *b* can be calculated by the following conditions

$$b = \begin{cases} y_i - \sum_{i=1}^{n}(a_i - a_i^*)k(x_i, x) - \varepsilon, & a_i \in (0, C) \\ y_i - \sum_{i=1}^{n}(a_i - a_i^*)k(x_i, x) + \varepsilon, & a_i^* \in (0, C) \end{cases}$$

Therefore, the analytical expression of this sample set's the estimated function can be calculated.

### *C.factors Affecting the Contract Price of Stock Index Futures*

Before considering the training of support vector machine model, we need to consider the selection of the input and output factors. The contract prices of stock index futures are subject to the following factors from the macroscopic perspective: macroeconomic conditions, macroeconomic policy changes, and the factoring affecting the underlying stock index, international financial markets trend, stock index futures contract expiration date and changes in investor's psychology. The data indicators are mainly from daily data from Shanghai and Shenzhen 300 stock index future including: (1) the highest quote; (2) lowest quote; (3) opening quote; (4) closing quote; (5) volume; (6) the total amount of futures contracts traded and (7) average price. This article requires the next day's closing price of futures contracts in order to train the SVM model and predict the stock index futures' contract price. During the training of the support vector machines, a too small training sample size will be detrimental to the prediction accuracy of model prediction and cause poor generalization. Therefore, the sample data used in this simulation is the one year contract transaction history data from Shanghai and Shenzhen stock index futures in the first season of IFSC3. The data were from December 2008 to December 18, 2009 and consists of a total of 238 transaction records.

## IV. Data Preprocessing and Error Analysis

Since the sample data units range from 10k

Yuan, points to times, we preprocess the data to make them consistent and easy to be accepted by the SVM model. For the data with the same units, we normalize them by deducting the minimum value in the group from all data and dividing all the data with the difference between the maximum value and the minimum value. For example, given a group of data

$V = (v_1, v_2, v_3 \ldots v_n), v_i \geq 1$, the

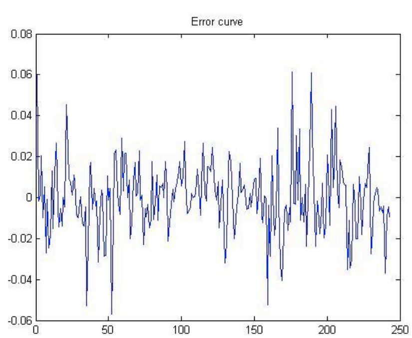
normalization function is $v_i = \frac{v_i - \min(V)}{\max(V) - \min(V)}$

where $i = 1, 2, \ldots \ldots n$. To test the accuracy of the model, we define the error margin function as follows:

$$\frac{1}{2}\sum_{i=1}^{n}(W_{真} - W_{预测})^2$$

## V. Concluding Remarks

With continuous training, we build an accurate support vector machine model. The above model is implemented in Mat lab 7.0 to validate the model's accuracy. Six variables of the training data are input to our model and the model will produce a predicted value. To measure the model's accuracy, we compare the predicted next day's closing price against the actual next day's closing price. The error margin results are shown in Figure 3.



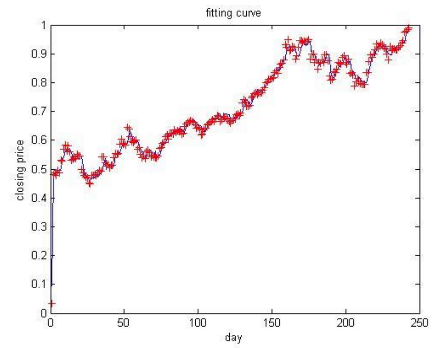




**Figure 3**

In our experiment, the SVM based prediction model is applied on the open, high, low, average price, trading volume, open interest and the closing price on Nov. 20 to produce a next day closing price of 3969.6 which is very close to the actually next day closing price 3999. As we can see, the error margin is only 0.74%.

## VI. References

[1]. Wang Baosen. Stock Index Futures Trading Strategy and Risk Management [D]. Tianjin University. November. 2003

[2]. Wu Xiaowei. Artificial Neural Network Based Prediction Model of the Stock Prices[D]. Dalian University of Technology. June .2004

[3]. Hu Ying. Support Vector Machine Based Securities Investment and Risk Management [D]. Xi'an University of Electronic Science and Technology in January. 2010

[4]. Li Haiqing. The Application of Support Vector Machines in the Prediction of Financial Markets [D]. Liaoning Normal University in May .2010

[5]. Shang Junsong, Mao Dingxiang: Improving the Application of BP Neural Network Prediction in the Stock Market [J] Value Engineering, 2004, (7) :26-30

[6]. Chen Shiyi. The application of Non-parametric Support Vector Regression and Classification Theory in the Financial Market Forecasting. Beijing University Press ,2008,5-21.