Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 231394, 7 pages http://dx.doi.org/10.1155/2015/231394



Research Article

A Hybrid Least Square Support Vector Machine Model with Parameters Optimization for Stock Forecasting

Jian Chai, Jiangze Du, Kin Keung Lai, 1,2 and Yan Pui Lee³

¹International Business School, Shaanxi Normal University, Xian 710062, China

Correspondence should be addressed to Kin Keung Lai; mskklai@cityu.edu.hk

Received 30 May 2014; Accepted 20 August 2014

Academic Editor: Shifei Ding

Copyright © 2015 Jian Chai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes an EMD-LSSVM (empirical mode decomposition least squares support vector machine) model to analyze the CSI 300 index. A WD-LSSVM (wavelet denoising least squares support machine) is also proposed as a benchmark to compare with the performance of EMD-LSSVM. Since parameters selection is vital to the performance of the model, different optimization methods are used, including simplex, GS (grid search), PSO (particle swarm optimization), and GA (genetic algorithm). Experimental results show that the EMD-LSSVM model with GS algorithm outperforms other methods in predicting stock market movement direction.

1. Introduction

Stock market is one of the most sophisticated and challenging financial markets since many factors affect its movement, including government policy, global economic situation, investors' expectations, and even correlations with other markets [1]. References [2, 3] described financial time series as essentially noisy, dynamic, and deterministically chaotic data sequences. Hence, a precise prediction of stock index movement can help investors make decisions to take or shed positions in the stock market at the right time and make profits. Many works have been published by researchers to maximize investment profits and minimize risk. Therefore, predicting stock market is quite important and significant.

Neural networks have been successfully applied in forecasting of financial time series during the past two decades [4–6]. Neural networks are general function approximations which can approximate many nonlinear functions regardless of the properties of time series data [7]. Besides, neural networks are able to learn dynamic systems which make them a more powerful tool for studying financial time series compared with traditional models [8–10]. However, there are a couple of weaknesses when neural networks are used in forecasting financial time series. For instance, when the typical back-propagation neural network is applied, a huge number of parameters are required to be controlled for. This makes the solution unstable and causes overfitting. The overfitting problem results in poor performance and becomes a critical issue for researchers.

Accordingly, [11] proposed a support vector machine (SVM) model. According to [12–14], there are two advantages of using SVM rather than neural networks. One is that SVM has a better performance in terms of generalization. Unlike the empirical risk minimization principle in traditional neural networks, SVM reduces generalization error bounds based on the structural risk minimization principle. SVM seeks to achieve an optimal structure through finding out a balance between generalization errors and Vapnik-Chervonenkis (VC) confidence interval. Another advantage is that SVM prevents the model from getting stuck into local minima.

Since the introduction of SVM, it has been developed rapidly in the real world. There are mainly two ways for applying SVM: one is classification and the other is regression. For classification, [15] constructed a SVM based model to accurately evaluate the consumers' credit score and solve

²Department of Management Sciences, City University of Hong Kong, Hong Kong

³School of Business, Tung Wah College, Hong Kong

classification problems. Also, SVM is widely used in the area of forecasting. Reference [16] used SVM to predict the direction of daily stock price in the Korea composite stock price index (KOSPI). More recently, [17] applies the Support Vector Regression to forecast the Nikki 225 opening index and TAIEX closing index after detecting and removing the noise by independent component analysis (ICA).

However, the performance of SVM mainly depends on the input data and is sensitive to parameters. Recent empirical studies have demonstrated that properties of the model performance are influenced by two aspects: low level of signal to noise ratio (SNR) and instability of model specification during the estimation process. For example, [18] investigates the hyperparameters selection for support vector machine with different noise distributions to compare the model performance. Moreover, [19] applied wavelet to denoise the bearing vibration signals by improving the SNR and then figure out the best model according to the performances of ANN and SVM.

To improve the classification and forecasting accuracy, several researchers including [20, 21] have proved that the combined classifying and forecasting models perform better than any individual model. Also, [22] showed that the ensemble empirical model decomposition (EEMD) can be integrated with extreme learning machine (ELM) to an effective forecasting model for computer products sales. In this paper, we propose a hybrid EMD-LSSVM (empirical mode decomposition least squares support vector machine) with different parameters optimization algorithms. The experimental results prove that the EMD-LSSVM model has a better performance than the WD-LSSVM (wavelet denoising least squares support vector machine) model. Firstly, we use the empirical mode decomposition and wavelet denoising algorithm to deal with the original input data. Secondly, parameters of SVM are optimized by different methods, including simplex, grid search (GS), particle swarm optimization (PSO), and genetic algorithm (GA). Results from empirical studies show that the hybrid model EMD-LSSVM with GS parameter optimization outperforms the other model.

2. EMD-LSSVM Model and WD-LSSVM

- 2.1. Empirical Mode Decomposition (EMD). References [23, 24] proposed empirical mode decomposition (EMD) which decomposes data series into a number of intrinsic mode functions (IMFs). It was designed for nonstationary and nonlinear data sets. In order to apply EMD, time series data set must satisfy the following two conditions.
 - (1) The sum of local maxima and local minima must equate to the total number of zero crossings or the difference between them is 1. In other words, for every local maxima and local minima, there must be one zero crossing following up.
 - (2) The local average is zero, which means that mean value of the upper envelope (defined by local maxima) and lower envelope (defined by local minima) must be zero.

Thus, if a function is an IMF, it represents a signal symmetric to local mean zero. An IMF is a simple oscillatory mode which is more general than the simple harmonic function and the frequency and amplitude of the IMF can be variable. Then, data series x(t) (t = 1, 2, ..., n) can be decomposed by the following sifting procedure.

- (1) Find all local maxima and minima in x(t). Then use the cubic spline line to connect all local maxima to generate upper envelope $x_{\rm up}(t)$ and connect all local minima to generate lower envelop $x_{\rm low}(t)$.
- (2) According to the upper and lower envelopes obtained in Step (1), calculate the envelope mean $m_1(t)$:

$$m_1(t) = \frac{\left(x_{\rm up}(t) + x_{\rm low}(t)\right)}{2}.$$
 (1)

(3) Data series x(t) minus envelope mean $m_1(t)$ gives the first component $d_1(t)$:

$$d_1(t) = x(t) - m_1(t).$$
 (2)

(4) Check if $d_1(t)$ satisfies the IMF requirements; if $d_1(t)$ does not satisfy them, go back to Step (1) and replace x(t) with $d_1(t)$ to conduct the second sifting procedure; that is, $d_2(t) = d_1(t) - m_2(t)$. Repeat the sifting procedure k times $d_k(t) = d_{k-1}(t) - m_k(t)$ until the following stop criterion is satisfied:

$$\sum_{t=1}^{T} \frac{\left[d_{j}(t) - d_{j+1}(t)\right]^{2}}{d_{j}^{2}(t)} < SC,$$
 (3)

where SC is the stopping condition. Normally, it is set between 0.2 and 0.3. Then, we get the first IMF component; that is, $c_1(t) = d_k(t)$.

- (5) Subtract first IMF component $c_1(t)$ from data sets x(t) and get the residual $r_1(t) = x(t) c_1(t)$.
- (6) Treat $r_1(t)$ as the new data series and repeat Steps (1) to (5). Then get the new residual $r_2(t)$. In this way, after repeating n times, we get

$$r_{2}(t) = r_{1}(t) - c_{2}(t),$$
 $r_{3}(t) = r_{2}(t) - c_{3}(t),$

$$\vdots$$

$$r_{n}(t) = r_{n-1}(t) - c_{n}(t).$$
(4)

When the residual $r_n(t)$ becomes a monotonic function, the data sets cannot be decomposed anymore. The whole EMD is completed. The original date series can be described as the combination of n IMF components and a mean trend $r_n(t)$; that is,

$$x(t) = \sum_{j=1}^{n} c_j(t) + r_n(t).$$
 (5)

In this way, the original data series x(t) can be decomposed into n IMFs and a mean trend function. Then, we use the IMFs for instantaneous frequency analysis.

The traditional Fourier transform decomposes a data series into a number of sine or cosine waves for the analysis. However, the EMD technique decomposes the data series into several sinusoid-like signals with variable frequencies and a mean trend function. The EMD has several advantages. First, this method is relatively easy to understand and is also widely applied since it avoids complex mathematical algorithms. Secondly, EMD is suitable to deal with nonlinear and nonstationary data series. Thirdly, EMD is more suitable for analysing data series with trends such as weather and economic data. Finally, EMD is able to find the residual which reveals the data series trends [25–27].

2.2. Wavelet Denoising Algorithm. While the traditional Fourier analysis can only remove noise of certain patterns over the entire time horizon, wavelet analysis can deal with multiscales and more detailed data and is more suitable for financial time series. Wavelets are continuous functions which satisfy the unit energy and admissibility condition in

$$C_{\varphi} = \int_{0}^{\infty} \frac{\left| \varphi(f) \right|}{f} df < \infty, \qquad \int_{-\infty}^{\infty} \left| \psi(t) \right|^{2} dt = 1, \quad (6)$$

where φ is the Fourier transform of frequency f. ψ is the wavelet transform.

The continuous wavelet function can orthogonally transform the original data into subdata series in the wavelet domain. Consider

$$W(u,s) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) dt, \tag{7}$$

where u is the dilation parameter and s is the translation parameter.

The wavelet synthesis rebuilds the original data series, guaranteed by the properties of orthogonal transformation in

$$x(t) = \frac{1}{C_{u}} \int_{0}^{\infty} \int_{-\infty}^{\infty} W(u, s) \psi_{u, s}(t) du \frac{ds}{s^{2}}.$$
 (8)

In wavelet analysis, the denoising technique separates the data and noise from the original data sets by selecting a threshold. The raw data series are first decomposed into some data subsets. Then, based on a certain strategy of selecting the threshold, the boundary between noises and data is set. Depending on the boundary, smaller data points are eliminated and the remaining data are handled by setting certain thresholds. Finally, these denoised data sets are rebuilt from the decomposed data points [28].

2.3. LSSVM in Function Estimation. This section shows the basic theory of the least squares support vector machine. The support vector methodology has been used mainly in two areas, that is, classification and function estimation. Considering regression in the set of function $f(x) = \omega^T \varphi(x) + b$ with given training data inputs $x_k \in \mathbb{R}^n$ and outputs $y_k \in \mathbb{R}$,

we apply $\varphi(x)$ to map x_k from R^n to R^{n_h} . Notice that $\varphi(x)$ can be of infinite dimensional and is defined only implicitly. Also, vector ω can also be infinite dimensional. Thus, the optimization problem becomes

min
$$J_P(\omega, \xi, \xi^*) = \frac{1}{2}\omega^T \omega + c \sum_{k=1}^N (\xi + \xi^*),$$

s.t. $y_k - \omega^T \varphi(x_k) - b \le \varepsilon + \xi_k, \quad k = 1, ..., N,$ (9)
 $\omega^T \varphi(x_k) + b - y_k \le \varepsilon + \xi_k^*, \quad k = 1, ..., N,$
 $\xi_k, \xi_k^* \ge 0, \quad k = 1, ..., N.$

The constant c>0 defines the tolerance of deviations from the desired ε accuracy. It defines the weight of the regularization term empirical risk. The larger the c is, the more important it is for the empirical risk to grow, compared with the regularization term. ε is called the tube size and represents the accuracy required in training data points.

By introducing Lagrange multipliers α , α^* , η , $\eta^* \ge 0$, we obtain the Lagrangian for this problem. Consider

$$L(\omega, b, \xi_{k}, \xi_{k}^{*}; \alpha, \alpha^{*}, \eta, \eta^{*})$$

$$= \frac{1}{2}\omega^{T}\omega + c\sum_{k=1}^{N} (\xi + \xi^{*})$$

$$- \sum_{k=1}^{N} \alpha_{k} (\varepsilon + \xi_{k} - y_{k} + \omega^{T} \varphi(x_{k}) + b)$$

$$- \sum_{k=1}^{N} \alpha_{k}^{*} (\varepsilon + \xi_{k}^{*} + y_{k} - \omega^{T} \varphi(x_{k}) - b)$$

$$- \sum_{k=1}^{N} (\eta_{k} \xi_{k} + \eta_{k}^{*} \xi_{k}^{*}).$$
(10)

The reason of introducing another Lagrange multiplier α_k^* is that there are other slack variables ξ_k, ξ_k^* . By maximizing the Lagrangian

$$\max_{\alpha,\alpha^*,\eta,\eta^*} \min_{\omega,b,\xi_k,\xi_k^*} L\left(\omega,b,\xi_k,\xi_k^*;\alpha,\alpha^*,\eta,\eta^*\right),\tag{11}$$

we obtain

$$\frac{\partial L}{\partial \omega} = 0 \longrightarrow \omega = \sum_{k=1}^{N} (\alpha_k - \alpha_k^*) \varphi(x_k),$$

$$\frac{\partial L}{\partial b} = 0 \longrightarrow \sum_{k=1}^{N} (\alpha_k - \alpha_k^*) = 0,$$

$$\frac{\partial L}{\partial \xi_k} = 0 \longrightarrow c - \alpha_k - \eta_k = 0,$$

$$\frac{\partial L}{\partial \xi_k} = 0 \longrightarrow c - \alpha_k^* - \eta_k^* = 0.$$
(12)

Category	Input variables selection	Output variable	Sample	Data number
Training and validation testing	CSI 300, USDX, SHIBOR, REPO, CDS, PE, M2/mkt cap, short-mid not/mkt cap, New Loan/mkt cap	CSI 300	05/01/2009-23/08/2011 24/08/2011-20/01/2012	643 100

TABLE 1: Input and output variables.

Then we obtain the following dual problem:

$$\max_{\alpha,\alpha^*} J_D(\alpha,\alpha^*) = -\frac{1}{2} \sum_{k=1}^N (\alpha_k - \alpha_k^*) (\alpha_l - \alpha_l^*) K(x_k, x_l)$$

$$-\varepsilon \sum_{k=1}^N (\alpha_k + \alpha_k^*) + \sum_{k=1}^N y_k (\alpha_k - \alpha_k^*),$$
s.t.
$$\sum_{k=1}^N (\alpha_k - \alpha_k^*) = 0, \quad \alpha_k, \alpha_k^* \in [0, c].$$
(13)

Here we use the kernel function $K(x_k, x_l) = \varphi(x_k)^T \varphi(x_l)$ for k, l = 1, ..., N. Then the function estimation becomes

$$f(x) = \sum_{k=1}^{N} \left(\alpha_k - \alpha_k^*\right) K\left(x_k, x_l\right) + b, \tag{14}$$

where α_k, α_k^* are solutions of the above quadratic programming problem and b is obtained from the complementarity of KKT conditions. It is obvious that the decision function is determined by the support vectors in which coefficients $(\alpha_k - \alpha_k^*)$ are not zero. In practice, a larger ε results in a smaller number of support vectors and thus the sparser of the solution. Also, the larger the ε is, the worse the accuracy of training points will be. Hence, ε can be applied to control the balance between closeness to training data and sparseness of the solution.

Kernel function can be obtained by seeking the function which satisfies Mercer's condition. Here are some popular kernel functions [14, 29, 30]:

linear:
$$K(x, x_k) = x^T x_k$$
;

polynomial: $K(x, x_k) = (x^T x_l + 1)^d$, where *d* is the degree of the polynomial kernel;

RBF kernel: $K(x, x_k) = \exp(-\|x - x_k\|^2/\sigma^2)$, where σ^2 is the bandwidth of the Gaussian kernel.

Parameters of the kernel function define the structure of the high dimensional feature space $\varphi(x)$ and also control the accuracy of the final solution. Thus, they should be selected carefully.

3. Empirical Study

3.1. Data Description. The CSI 300 is chosen for empirical analysis and to examine the performance of the proposed model. This index comprises 179 stocks from Shanghai stock

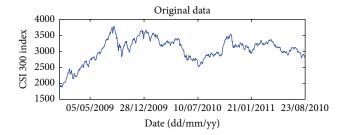
exchange and 121 stocks from Shenzhen stock exchange and is managed by the China Securities Index Company Ltd.

Most researchers have chosen international indices in the past, including S & P 500, NIKKEI 225, NASDAQ, DAX, and gold price as input variables. They have examined the cross relationship between stock market index and macroeconomic variables. The potential input variables that can be used for forecasting model mainly consist of the gross domestic product (GDP), gross national product (GNP), short-term interest rate (ST), long-term interest rate (LT), and term structure of interest rate (TS) [1, 31, 32].

Although China has overtaken Japan to become the world's second largest economy and the Chinese stock market has developed into one of the most important markets in the global economy, Chinese consumption capacity is limited in the domestic market. The movement of the stock market has a close relationship with the money available of the investors, which is determined by the money supply and the interest rate. Considering that the Chinese stock market is affected by the global economic situation as well as the domestic economic development, we choose US Dollar Index (USDX), Shanghai Interbank Offered Rate (SHIBOR), P/E ratio (PE), money supply (M2), repurchase agreement (REPO), China CNY Monthly New Loan, market capitalization of the 300 publicly traded companies (mkt cap), People's Bank 5-year CDS, and short-mid note as input variables.

The lag of input variables is 3 days. We use the daily data to predict the CSI 300 index by nonlinear SVM regression. Since M2, short-mid note, and New Loan are published once a month, we transform these variables into daily variables by dividing them by a daily variable. We divided all data sets into two sections and used the first section as the training part to find the optimal parameters for the LSSVM and avoid overfitting by training and validating the model. The other section is used for testing. As shown in Table 1, we choose nine variables as the input variables and one variable as the output variable including 643 daily data from May 1, 2009, to August 23, 2011, to train the parameters in the model. Once we obtain these parameters, we use the same input and output variables from August 24, 2011, to January 20, 2012, including 100 daily data to examine the performance of different model in the testing part.

In the hybrid wavelet denoising least squares support vector machine model (WD-LSSVM), we first denoise the CSI 300 index with wavelet denoising technique. As shown in Figure 1, the original data, which is depicted in the upper part of the figure, is packed with irrelevant noise. Then the wavelet denoising algorithm is applied to reduce the noise in the upper figure of Figure 1. The denoised data is depicted in the lower part of Figure 1 and it is clear that the denoised data



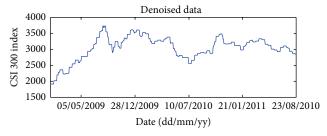


FIGURE 1: The original and denoised daily CSI 300 index.

TABLE 2: Parameters setting for simplex, GS, GA, and PSO.

Method	Parameter setting
Simplex	Chi: 2, Gamma: 0.5, Rho: 1, Sigma: 0.5
GS	TolX: 0.001, maxFunEvals: 70, grain: 7, zoomfactor: 5
GA	Sizepop: 20, maxgen: 200, <i>c</i> min: 0, <i>c</i> max: 100, <i>g</i> min: 0, <i>g</i> max: 1000
PSO	Sizepop: 20, maxgen: 200, <i>c</i> min: 0, <i>c</i> max: 100, <i>g</i> min: 0, <i>g</i> max: 1000, <i>k</i> : 0.6

can better reveal the trend of the index. Also, in both EMD-LSSVM and WD-LSSVM models, we preprocess the input data by scaling to the range of [0, 1] to prevent small numbers in the data sets from being overshadowed by large numbers, resulting in loss of information.

3.2. Optimization Methods and Parameters Setting. In both EMD-LSSVM models and WD-LSSVM, we try four kinds of search methods, that is, simplex, GS, GA, and PSO. In the simplex method, we define the parameters of expanding (Chi), contracting (Gamma), reflecting (Rho), and shrinking (Sigma) and get the optimal parameters for SVM through iteration until the stopping criteria is satisfied. Also, by calculating the objective function, we can get all points in the grids, which are related to the range and the unit grid search size.

The optimal parameters can be obtained from the point which has the lowest cost. Another effective method to solve optimization problem is the genetic algorithm. The first step of this method is to randomly select parents from the population. Then, parents produce children continuously. Step by step, the population eventually develops and optimal solution can be obtained when the stopping criteria are met. The PSO algorithm works by moving the candidate solution (particles) within the given search range. These particles are moved by the best known positions of particles and the entire

TABLE 3: Performance metrics and their calculations.

Metrics	Calculation			
NMSE	NMSE = $\frac{\sum_{i=1}^{n} (a_i - p_i)^2}{\delta^2 n}$ $\delta^2 = \frac{\sum_{i=1}^{n} (a_i - \overline{a})^2}{n-1}$			
MAPE	MAPE = $\frac{1}{n} \sum_{i=1}^{n} \left \frac{a_i - p_i}{p_i} \right \times 100\%$			
HR	$HR = \frac{\sum_{i=1}^{n} d_i}{n}$ $d_i = \begin{cases} 1 & \text{if } (a_i - a_{i-1}) (p_i - p_{i-1}) \ge 0\\ 0 & \text{otherwise} \end{cases}$			

TABLE 4: Results of eight different forecasting models.

Model	NMSE	MAPE	HR
EMD-LSSVM (simplex)	0.0253	0.79222%	77.7778%
EMD-LSSVM (GS)	0.0245	0.78834%	79.798%
EMD-LSSVM (GA)	2.5749	9.0641%	42.4242%
EMD-LSSVM (PSO)	9.4471	18.2733%	40.404%
WD-LSSVM (simplex)	0.0521	1.1772%	65.6566%
WD-LSSVM (GS)	0.0609	1.1357%	61.6162%
WD-LSSVM (GA)	0.0657	1.2997%	62.6263%
WD-LSSVM (PSO)	0.0910	1.5072%	63.6364%

swarm in the search space. When the particles arrive at a better position, they guide the swarm to move. The procedure is repeated until the stopping criteria are satisfied. In our experiment, Table 2 shows the setting of each optimization method.

3.3. Performance Criteria. We evaluate the performance of these models using three measurement methods, that is, normalized mean squared error (NMSE), mean absolute percentage error (MAPE), and the hitting ratio (HR) (Table 3). NMSE and MAPE are designed to measure the deviation of predicted value from the actual value; smaller values of NMSE and MAPE indicate better performance of the model. In the stock market, smaller values of MAPE and NMSE are able to control investment risk. We also introduce hitting rate to evaluate the model since the HR reveals accuracy of prediction of the CSI 300, which is valuable for individual and institutional traders.

3.4. Experiment Results. The experiments explore four parameter selection methods in both EMD-LSSVM and WD-LSSVM. Results of the experiments are as in Table 4. From the results, we can see that the hybrid model EMD-LSSVM with GS parameter optimization method not only has the smallest NMSE and MAPE but also gets the best hitting rate, which means it outperforms the other model with different parameter search methods.

From the experiment results, we can draw three conclusions.

- (1) For overall accuracy, the EMD-LSSVM (GS) is the best approach, followed by EMD-LSSVM (simplex), WD-LSSVM (simplex), WD-LSSVM (GA), and WD-LSSVM (GS). Hitting rates of the other approaches are below 60%. Prediction accuracy of all methods is also related to the chosen sample. So it is difficult to identify which model is the best and performs the best. However, tests based on the same sample may help us identify which is the best model.
- (2) According to the experiments, the PSO and GA need more computational time to obtain the best parameters for the model compared with simplex and GS optimization methods. Although the PSO and GA algorithm are relatively more complex than the other two methods, they do not perform better than GS and simplex.
- (3) Another interesting finding is that thresholds of the denoising algorithm also influence the performance of the model. When the threshold is too large, useful information in the data gets damaged. Besides, a small threshold makes the denoising process insignificant for handling noise. Therefore, we argue that the performance of the wavelet denoising algorithm is sensitive to the estimation method of the threshold level.

4. Conclusion

We have examined the use of the hybrid EMD-LSSVM and WD-LSSVM models to predict financial time series by four different parameters selection methods in this paper. The study shows that the hybrid EMD-LSSVM model provides a better way to forecast financial time series compared with WD-LSSVM. The key findings contain two aspects. First, empirical mode decomposition can serve as a potential tool for removing noise from original data during the modeling process and improving the prediction accuracy. Second, we compare four kinds of search methods for parameters in the experiments. The results show that the EMD-LSSVM with GS parameter optimization method provides the best performance. Use of the GS algorithm reduces the computation time and improves the prediction accuracy of the model for forecasting financial time series.

Future research in this direction mainly includes gaining better understanding of the relationship between optimal loss function, noise distribution, and the number of training samples. In this paper, we only consider applying different algorithm to denoise the original data without considering the distribution of the noise. The research on the density of noise which will be reduced for the SVM model will attract the effort of us. Moreover, another interesting research direction is to figure out the minimum number of samples based on which a theoretically optimal loss function will indeed have superior generalization performance.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] W. Huang, Y. Nakamori, and S.-Y. Wang, "Forecasting stock market movement direction with support vector machine," *Computers and Operations Research*, vol. 32, no. 10, pp. 2513–2522, 2005.
- [2] J. W. Hall, "Adaptive selection of U.S. stocks with neural nets," in *Trading on the Edge: Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets*, G. J. Deboeck, Ed., John Wiley & Sons, New York, NY, USA, 1994.
- [3] Y. S. Abu-Mostafa and A. F. Atiya, "Introduction to financial forecasting," *Applied Intelligence*, vol. 6, no. 3, pp. 205–213, 1996.
- [4] W. Cheng, L. Wagner, and C.-H. Lin, "Forecasting the 30-year US treasury bond with a system of neural networks," *Journal of Computational Intelligence in Finance*, vol. 4, pp. 10–16, 1996.
- [5] R. Sharda and R. B. Patil, "A connectionist approach to time series prediction: an empirical test," in *Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real World Performance*, R. R. Trippi and E. Turban, Eds., Irwin Professional Publishing, Chicago, Ill, USA, 1996.
- [6] J. R. van Eyden, The Application of Neural Networks in the Forecasting of Share Prices, Finance and Technology Publishing, Haymarket, Va, USA, 1996.
- [7] I. Kaastra and M. S. Boyd, "Forecasting futures trading volume using neural networks," *Journal of Futures Markets*, vol. 15, pp. 853–970, 1995.
- [8] G. Zhang and M. Y. Hu, "Neural network forecasting of the British pound/US dollar exchange rate," *Omega*, vol. 26, no. 4, pp. 495–506, 1998.
- [9] W.-C. Chiang, T. L. Urban, and G. W. Baldridge, "A neural network approach to mutual fund net asset value forecasting," *Omega*, vol. 24, no. 2, pp. 205–215, 1996.
- [10] F. E. H. Tay and L. Cao, "Application of support vector machines in financial time series forecasting," *Omega*, vol. 29, no. 4, pp. 309–317, 2001.
- [11] V. N. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, NY, USA, 2nd edition, 2000.
- [12] K.-R. Muller, A. J. Smola, G. Ratsch, B. Scholkopf, J. Kohlmorgen, and V. N. Vapnik, "Predicting time series with support vector machines," in *Proceedings of the International Conference on Artificial Neural Networks*, pp. 999–1004, Lausanne, Switzerland, 1997.
- [13] S. Mukherjee, E. Osuna, and F. Girosi, "Nonlinear prediction of chaotic time series using support vector machines," in Proceedings of the IEEE Workshop on Neural Networks for Signal Processing (NNSP '97), pp. 511–520, Amelia Island, Fla, USA, September 1997.
- [14] V. N. Vapnik, S. E. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," *Advances in Neural Information Processing* Systems, vol. 9, pp. 281–287, 1996.
- [15] C.-L. Huang, M.-C. Chen, and C.-J. Wang, "Credit scoring with a data mining approach based on support vector machines," *Expert Systems with Applications*, vol. 33, no. 4, pp. 847–856, 2007.

- [16] K.-J. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, no. 1-2, pp. 307–319, 2003.
- [17] C.-J. Lu, T.-S. Lee, and C.-C. Chiu, "Financial time series forecasting using independent component analysis and support vector regression," *Decision Support Systems*, vol. 47, no. 2, pp. 115–125, 2009.
- [18] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural Networks*, vol. 17, no. 1, pp. 113–126, 2004.
- [19] G. S. Vijay, H. S. Kumar, P. P. Srinivasa, N. S. Sriram, and R. B. K. N. Rao, "Evaluation of effectiveness of wavelet based denoising schemes using ANN and SVM for bearing condition classification," *Computational Intelligence and Neuroscience*, vol. 2012, Article ID 582453, 12 pages, 2012.
- [20] L. Zhou, K. K. Lai, and L. Yu, "Least squares support vector machines ensemble models for credit scoring," *Expert Systems* with Applications, vol. 37, no. 1, pp. 127–133, 2010.
- [21] Y. Bao, X. Zhang, L. Yu, K. K. Lai, and S. Wang, "An integrated model using wavelet decomposition and least squares support vector machines for monthly crude oil prices forecasting," New Mathematics and Natural Computation, vol. 7, no. 2, pp. 299–311, 2011.
- [22] C.-J. Lu and Y. E. Shao, "Forecasting computer products sales by integrating ensemble empirical mode decomposition and extreme learning machine," *Mathematical Problems in Engineer*ing, vol. 2012, Article ID 831201, 15 pages, 2012.
- [23] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *The Royal Society of London. Proceedings A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
- [24] N. E. Huang, Z. Shen, and S. R. Long, "A new view of nonlinear water waves: the Hilbert spectrum," *Annual Review of Fluid Mechanics*, vol. 31, pp. 417–457, 1999.
- [25] L. Yu, S. Wang, and K. K. Lai, "An EMD-based neural network ensemble learning model for world crude oil spot price forecasting," in Soft Computing Applications in Business, B. Prasad, Ed., vol. 230 of Studies in Fuzziness and Soft Computing, pp. 261–271, Springer, 2008.
- [26] L. Yu, K. K. Lai, S. Wang, and K. He, "Oil price forecasting with an EMD-based multiscale neural network learning paradigm," in *International Conference on Computational Science*, pp. 925–932, 2007.
- [27] S. Zhou and K. K. Lai, "An improved EMD online learning-based model for gold market forecasting," in *Proceedings of the 3rd International Conference on Intelligent Decision Technologies*, pp. 75–84, 2011.
- [28] K. He, C. Xie, and K. K. Lai, "Estimating real estate valueat-risk using wavelet denoising and time series model," in Computational Science—ICCS 2008, vol. 5102 of Lecture Notes in Computer Science, pp. 494–503, Springer, Berlin, Germany, 2008.
- [29] S. Zhou, K. K. Lai, and J. Yen, "A dynamic meta-learning rate-based model for gold market forecasting," *Expert Systems with Applications*, vol. 39, no. 6, pp. 6168–6173, 2012.
- [30] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, A Practical Guide to Support Vector Classification, Department of Computer Science and Information Engineering, University of National Taiwan, Taipei, Taiwan, 2003.

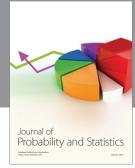
- [31] J. Lakonishok, A. Shleifer, and R. W. Vishny, "Contrarian investment, extrapolation, and risk," *Journal of Finance*, vol. 49, pp. 1541–1578, 1994.
- [32] M. T. Leung, H. Daouk, and A.-S. Chen, "Forecasting stock indices: a comparison of classification and level estimation models," *International Journal of Forecasting*, vol. 16, no. 2, pp. 173–190, 2000.



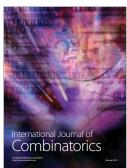










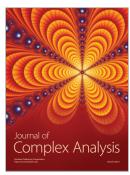


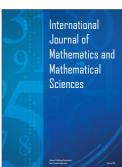


Submit your manuscripts at http://www.hindawi.com











Journal of Discrete Mathematics

