Advances in Mathematical Finance & Applications
www.amfa.iau-arak.ac.ir
Print ISSN: 2538-5569
Online ISSN: 2645-4610
Doi:10.22034/AMFA.2020.1883402.1332

Research Paper

Interval Forecasting of Stock Price Changes using the Hybrid of Holt's Exponential Smoothing and Multi-Output Support Vector Regression

Sayyed Mohammadreza Davoodi*, Mehdi Rabiei

Department of Management, Dehaghan Branch, Islamic Azad University, Dehaghan, Iran.

ARTICLE INFO

Article history:
Received 2019-11-16
Accepted 2020-06-28

Keywords:
Smoothing
Support vector machine
Multi-output least-squares vector
Regression interval forecasting.

ABSTRACT

Given the importance of investment in stock markets as a major source of income for many investors, there is a strong demand for models that estimate the future behavior of stock prices. Interval forecasting is the process of predicting an interval characterized by two random variables acting as its upper and lower bounds. In this study, a hybrid method consisting of Holt's exponential smoothing and multi-output least squares support vector regression is used to forecast the interval of the lowest and highest prices in a stock market. First, Holt's smoothing method is used to smooth the two bounds of the interval and then the residuals of the smoothing process are modeled with multi-output vector support regression. The output of the regression step is the error of the two bounds of the interval. The method is implemented on the weekly data of the overall index of the Tehran Stock Exchange from 1992 to 2016, with the interval defined as the distance between the lowest and highest overall index values. The results demonstrate the high accuracy of the hybrid method in producing in-sample and out-of-sample forecasts for the movement of the two bounds of the interval, that is, the weekly highs and lows of the overall index. Also, the hybrid method has achieved a lower mean squared error than the Holt's smoothing method, indicating that multi-output vector support regression has improved the performance of the smoothing method.

1 Introduction

The effective role of capital in every country flows through giving guidelines for capital and resources, generalizing companies and sharing development projects with public, and also adding accredited companies stock market requires appropriate decision making for shareholders and investors who are willing to buy shares based on price mechanism. [16] A stock exchange is a place for collecting uninvested savings and using them to finance long-term investment projects. People with uninvested savings compare the return on investment of stock exchange with other investment options (housing, banks, gold, direct production) and make the final decision after weighing the risk and return of each option. Stock trend forecasting is a one of the main factors in choosing the best investment, hence prediction and comparison

* Corresponding author. Tel.: *09132290367*
E-mail address: Smrdavoodi@ut.ac.ir

of different firms' stock trend is one method for improving investment process. [15] It is impossible to make proper investment decisions without information, and oftentimes, information is not fully available and needs to be complemented with sensible predictions [3]. The financial-economic literature contains several methods for predicting the return on investment: technical analysis, fundamental analysis, classical time series forecasting, and intelligent methods [2].

Fundamental Analysis: This method involves analyzing the general conditions of the economy and the condition of industries and companies to model prices and returns [5].

Technical Analysis: This method is to analyze share prices and exchange volumes with a series of oscillation indicators and patterns to look for suitable buy and sell positions [1]. Technical analysis is one of the financial market analysis tools. Technical analysis is a method of anticipating prices and markets through studying historical market data. [17]

In structural models, the relationships of variables are estimated based on the rational behavior of economic factors, including consumers, producers and economic policymakers using different econometric models. Despite the extensive use of technical and structural models in stock market forecasting, research has shown that these methods are sufficiently successful in predicting stock markets. Thus, researchers have used time series models and artificial neural networks to make better predictions [8]. Time series models assume that all factors and relationships that affect a variable are manifested in its values, and therefore it is possible to make forecasts only using the prior information of the variable itself. However, many studies have shown that these models are not very reliable. Further, all of these models are based on a concept called stationarity, which is not necessarily applicable to all price series. Given the complex and dynamic nature of stock prices, nonlinear models, including neural networks and support vector regression have found extensive use in stock price forecasting. Unlike linear models, these nonlinear models can discover the relationships between data with high flexibility and ultimately formulate an equation for these relationships. Support vector regression (SVR) is a method to process empirical data in order to transfer their inherent knowledge or rules into the network structure. Since this approach involves learning a series of general rules by processing numerical data or instances, the systems based on this approach are called intelligent systems [10]. Research has shown that stock price is a data series with nonlinear behavior, which is why artificial intelligence models are better suited to process these data than statistical models. The advantage of SVR over econometric models is that in this method, time series need not be stationary and there is no need to have linear relationships [9]. Another concept used in the present study is interval forecasting. While conventional forecasting methods tend to predict closing prices, interval forecasting can offer investors extra information such as the lowest and highest prices. In the present study, Holt's exponential smoothing, as a linear method, and multi-output support vector regression (MSVR), as a nonlinear method, are combined into a hybrid forecasting system in order to make use of both linear and nonlinear properties simultaneously. On the other hand, the hybrid model has the benefit of high process accuracy, suggesting that the evaluation of the similarity in the volatility of return at the level of market or industry constituent units is better than the simple technique of time seriess model for the entire market (instead of evaluation at unit levels) [18]. In this study, forecasting is of interval type, which means predicting the lowest and highest stock prices.

2 Theoretical Foundations and Research Background

Investment in stock markets is an important source of income for many people. Therefore, investors have developed various analytics and methods to predict and maximize the returns on their investments. The last few decades have seen the development and expansion of several predictive tools and methods, each with their own advantages and drawbacks. The combination of these methods into hybrid forecasting systems has always been of interest to researchers and those interested in price forecasting. The goal of interval forecasting is to predict an interval consisting of two values marking the lower and higher bounds of a random variable. A larger overlap between the interval and the real range of the target variable means more accurate forecasting.

2.1 Holt's Exponential Smoothing

Time series are sequences of data that are collected at equal intervals. A time series of data up to time t is denoted by $(x_1, x_2, \dots, x_{t-1}, x_t)$. One of the characteristics of time series is the interdependence of adjacent observations. In the analysis of time series, it is very important to discover the relationship between the entries and construct an appropriate model accordingly. The primary purpose of building such a model is to predict the future value of the analyzed time series based on past information [9]. In this study, entries of the time series are random two-dimensional vectors or intervals consisting of the lowest and highest prices at any time:

$$X_t = [x_t^H, x_t^L]^t = \begin{bmatrix} x_t^H \\ x_t^L \end{bmatrix} \quad t = 1, 2, \dots, n \tag{1}$$

Where x_t^H is the highest price and x_t^L is the lowest price at time t. Smoothing eliminates short-term noises, thus providing a better view of price range behavior and trends in the range changes. In Holt's smoothing, the time series is first approximated to series $\hat{x}_t$ as shown below:

$$\hat{X}_{t+1} = \hat{p}_t^X + \hat{Q}_t^X \tag{2}$$

where:

$$\begin{aligned} \hat{p}_t^X &= AX_t + (I - A)(\hat{p}_{t-1}^X - \hat{Q}_{t-1}^X) \\ \hat{Q}_t^X &= B(\hat{p}_t^X - \hat{p}_{t-1}^X) + (I - B)\hat{Q}_{t-1}^X \end{aligned} \tag{3}$$

such that:

$$\hat{p}_t^X = \begin{bmatrix} \hat{p}_t^U \\ \hat{p}_t^L \end{bmatrix} \quad \hat{Q}_t^X = \begin{bmatrix} \hat{Q}_t^U \\ \hat{Q}_t^L \end{bmatrix} \quad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{4}$$

These equations provide a linear approximation of a time series based on the linear combination of the prior entries of two-time series. In their open form (matrix), the above equations are expressed as follows:

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \quad B = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \tag{5}$$

$$\begin{aligned} \hat{p}_t^X &= \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} x_t^H \\ x_t^L \end{bmatrix} + \begin{bmatrix} 1-\alpha_{11} & -\alpha_{12} \\ -\alpha_{21} & 1-\alpha_{22} \end{bmatrix} \begin{bmatrix} \hat{p}_{t-1}^U + \hat{Q}_{t-1}^U \\ \hat{p}_{t-1}^L + \hat{Q}_{t-1}^L \end{bmatrix} \\ \hat{Q}_t^X &= \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} \hat{p}_t^U - \hat{p}_{t-1}^U \\ \hat{p}_t^L - \hat{p}_{t-1}^L \end{bmatrix} + \begin{bmatrix} 1-\beta_{11} & -\beta_{12} \\ -\beta_{21} & 1-\beta_{22} \end{bmatrix} \begin{bmatrix} \hat{Q}_{t-1}^U \\ \hat{Q}_{t-1}^L \end{bmatrix} \end{aligned} \tag{6}$$

Matrices A and B are obtained by the minimization of the sum of squared error (the error of the observed and estimated random vector) using Model (7).

$$\min_{\alpha_{ij},\beta_{ij}} \; error(A,B)$$
$$s.t: 0 \le \alpha_{ij}, \beta_{ij} \le 1$$
$$where$$
$$error(A,B) = \sum_{t=3}^{n} (x_t - \hat{x}_t)^t (x_t - \hat{x}_t)$$
$$= \sum_{t=3}^{n} (x_t^U - \hat{p}_{t-1}^U - \hat{Q}_{t-1}^U)^2 + \sum_{t=3}^{n} (x_t^L - \hat{p}_{t-1}^L - \hat{Q}_{t-1}^L)^2 \tag{7}$$

At the end of the smoothing process, the approximation residuals are calculated by Equation (8):

$$e_i = x_i - \hat{x}_i = \begin{bmatrix} e_t^H \\ e_t^L \end{bmatrix} \quad i = 1,2,...,n \tag{8}$$

2.2 Multi-output Support Vector Machine

Multi-output least squares support vector regression is used to find a nonlinear model for the residuals of Holt's exponential smoothing. The multi-output support vector regression (MSVR) model is trained with the following input and output with a suitable lag such as d:

$$input_t = [e_t^H, e_t^L, e_{t-1}^H, e_{t-1}^L, ..., e_{t-d+1}^H, e_{t-d+1}^L]^t \tag{9}$$

$$output_t = \begin{bmatrix} e_t^H \\ e_t^L \end{bmatrix} \tag{10}$$

Therefore, the error observed at the present time is assumed to be caused by the previously observed errors. The method of training of the MSVR model is described in the following. We first start with the single-output mode. Suppose there is one training dataset $X = \{(x_i, y_i)_{i=1}^{l}\}$, where the following is true for every x and y:

$$\forall i : x_i \in R^d, y_i \in R \tag{11}$$

For training, the vector W (slope of the line) and the scalar b (intercept) in the estimation model $Y = \varphi(x)^t w + b1_l$ (first the input x is transferred to the new space $\varphi(x)$, where φ is the kernel function, and then the output is determined by linear regression) are obtained from the optimization of Model (12).

$$\min_{w \in R^{n_h}, b \in R} \; J(w,\xi) = \frac{1}{2} w^t w + \gamma \frac{1}{2} \xi^t \xi$$
$$st: Y = Z^t w + b1_l + \xi \tag{12}$$

where:

$$Z = (\varphi(x_1), \varphi(x_2), ..., \varphi(x_l)) \in R^{n_h \times 1}$$
$$\varphi: R^d \to R^{n_h}$$
$$(\xi_1, ..., \xi_l) \in R^l$$
$$\gamma \in R_+ \tag{13}$$

As can be seen, the objective function is the sum of the least squares of errors ($\xi^t \xi$) and weights ($w^t w$). The sum of weights is minimized is to prevent over-training. For this optimization, the Lagrangian function is formulated as follows:

$$L(w,b,\xi,\alpha)=J(w,\xi)-\alpha^t (Z^t w+b1_l+\xi-Y) \qquad (14)$$
$$(\alpha_1,\alpha_2,...,\alpha_l)\in R^l$$

Computing the partial derivative of the Lagrangian function with respect to the variables gives:

$$\begin{cases} \dfrac{\partial L}{\partial w}=0 \Rightarrow & w=Z\alpha \\ \dfrac{\partial L}{\partial b}=0 \Rightarrow & \alpha^t 1_l=0 \\ \dfrac{\partial L}{\partial \xi}=0 \Rightarrow & \alpha=\gamma\xi \\ \dfrac{\partial L}{\partial \alpha}=0 \Rightarrow & Z^t w+b1_l+\xi-Y=0_l \end{cases} \qquad (15)$$

And by substituting w, ζ , we arrive at:

$$\begin{bmatrix} 0 & 1_l^t \\ 1_l & H \end{bmatrix}\begin{bmatrix} b \\ \alpha \end{bmatrix}=\begin{bmatrix} 0 \\ Y \end{bmatrix} \qquad (16)$$

where:

$$H=K+\gamma^{-1}I_d \in R^{l\times l}$$
$$K=Z^t Z \in R^{l\times l} \qquad (17)$$
$$K_{ij}=\varphi(x_i)^t \varphi(x_j)=\ker(x_i,x_j)$$

and $\ker(.,.)$ is the kernel function. To solve this system of equations, one can use its simplified version shown below:

$$\begin{bmatrix} s & 0_l^t \\ 0_l & H \end{bmatrix}\begin{bmatrix} b \\ \alpha+bH^{-1}1_l \end{bmatrix}=\begin{bmatrix} 1_l^t H^{-1}Y \\ Y \end{bmatrix} \qquad (18)$$
$$s=1_l^t H^{-1} 1_l$$

After finding the optimum values $\alpha^*=(\alpha_1^*,\alpha_2^*,...,\alpha_l^*)$ and b^* , the input-output relationship function of the support vector regression is obtained as follows:

$$f(X)=\varphi(X)^t w^* + b^* = \varphi(X)^t Z\alpha^* + b^* = \qquad (19)$$
$$\sum_{i=1}^{l} \alpha_i^* \varphi(X)^t \varphi(x_i)^t + b^* = \sum_{i=1}^{l} \alpha_i^* \ker(x, x_i) + b$$

In the multi-output mode, inputs and outputs are in the form of $\{(x_i, y^i)\}_{i=1}^{l}$ where the following is true for every x and y:

$$x_i \in R^d, y^i \in R^m \qquad (20)$$

Therefore, w, b are in the form of Equation (21).

$$W = (w_1, w_2, ..., w_m) \in R^{n_h \times m}$$
$$b = (b_1, b_2, ..., b_m)^t \in R^m \qquad (21)$$

To calculate w, b, Model (22) must be optimized:

$$\min_{W \in R^{n_h \times m}, b \in R^m} J(W, \Xi) = \frac{1}{2} trace(W^t W) + \gamma \frac{1}{2} trace(\Xi^t \Xi)$$
$$st.\ Y = Z^t W + repmat(b^t, l, 1) + \Xi \qquad (22)$$
$$\Xi = (\xi_1, \xi_2, ..., \xi_m) \in R_+^{l \times m}$$

To optimize Model (22), $w_i \in R^{n_h}$ is decomposed into Equation (23) and then Model (24) is optimized:

$$w_i = w_0 + v_i$$
$$w_0 \in R^{n_h} \qquad (23)$$
$$V = (v_1, v_2, ..., v_m) \in R^{n_h \times m}$$

$$\min_{W_0 \in R^{n_h}, V \in R^{n_h}, b \in R^m} J(w_0, V, \Xi) = \frac{1}{2} w_0^t w_0 + \frac{1}{2} \frac{\lambda}{m} trace(V^t V) +$$
$$\gamma \frac{1}{2} trace(\Xi^t \Xi)$$
$$st.\ Y = Z^t W + repmat(b^t, l, 1) + \Xi$$
$$\Xi = (\xi_1, \xi_2, ..., \xi_m) \in R^{l \times m} \qquad (24)$$
$$W = (w_0 + v_1, ..., w_0 + v_m) \in R^{n_h \times m}$$
$$Z = (\varphi(x_1), \varphi(x_2), ..., \varphi(x_l)) \in R^{n_h \times l}$$
$$\lambda, \gamma \in R_+$$

According to the Lagrangian multiplier method, the Lagrangian function of the above optimization is as follows:

$$L(w_0, V, b, \Xi, A) = J(w_0, V, \Xi) -$$
$$trace(A^t (Z^t W + repmat(b^t, l, 1) + \Xi - Y)) \qquad (25)$$
$$A = (\alpha_1, \alpha_2, .., \alpha_m) \in R^{l \times m}$$

By taking partial derivatives, we arrive at:

$$\begin{cases} \frac{\partial L}{\partial w_0} = 0 \quad \Rightarrow \quad w_0 = \sum_{i=1}^{m} Z \alpha_i \\ \frac{\partial L}{\partial V} = 0 \quad \Rightarrow \quad V = \frac{m}{\lambda} ZA \\ \frac{\partial L}{\partial b} = 0 \quad \Rightarrow \quad A^t 1_l = 0_l \\ \frac{\partial L}{\partial \Xi} = 0 \quad \Rightarrow \quad A = \gamma \Xi \\ \frac{\partial L}{\partial A} = 0 \quad \Rightarrow \quad Z^t W + repmat(b^t, l, 1) + \Xi - Y = 0_{l \times m} \end{cases} \qquad (26)$$

Therefore, $w_0=\frac{\lambda}{m}\sum_{i=1}^{m} v_i$, which means w_0 is a linear combination of v_i . Hence, the above model can be replaced with the following model:

$$\begin{aligned}&\min_{V\in R^{nh},\,b\in R^{m}} J(V,\Xi)=\frac{1}{2}\frac{\lambda^2}{m^2}V\,1_m 1_m^t V^t+\frac{1}{2}\frac{\lambda}{m}trace(V^tV)+\\&\gamma\frac{1}{2}trace(\Xi^t\Xi)\\&st:Y=Z^tW+repmat(\frac{\lambda}{m}Z^tV\,1_m,1,m)+repmat(b^t,l,1)+\Xi\end{aligned}\tag{27}$$

Thus, to find w, b , the following system of equations is formed:

$$\begin{bmatrix}0_{ml\times m} & P^t\\ P & H\end{bmatrix}\begin{bmatrix}b\\ \alpha\end{bmatrix}=\begin{bmatrix}0_m\\ Y\end{bmatrix}\tag{28}$$

where:

$$\begin{aligned}&P=blockdiag\,(\overbrace{1_l,\ldots,1_l}^{m})\in R^{ml\times m}\\&H=\Omega+\gamma^{-1}I_{ml}+\frac{m}{\lambda}Q\in R^{ml\times ml}\\&\Omega=repmat\,(k,m,m)\in R^{ml\times ml}\\&Q=blockdiag\,(\overbrace{k,\ldots,k}^{m})\in R^{ml\times ml}\\&K=Z^tZ\in R^{l\times l}\\&\alpha=(\alpha_1^t,\alpha_2^t,\ldots,\alpha_m^t)\in R^{ml}\end{aligned}\tag{29}$$

In the above equations:

$$\begin{aligned}&repmat\,(A,m,n)=\begin{bmatrix}A & A & \cdots & A\\ A & A & \cdots & A\\ \vdots & \vdots & \vdots & \vdots\\ A & A & \cdots & A\end{bmatrix}_{m\times n}\\&blockdiag\,(A,n)=\begin{bmatrix}A & 0 & \cdots & 0\\ 0 & A & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \cdots & A\end{bmatrix}_{m\times n}\end{aligned}\tag{30}$$

If $\alpha^*=(\alpha_1^{*t},\alpha_2^{*t},\ldots,\alpha_m^{*t})^t$ is the optimum solution, then the decision function is:

$$\begin{aligned}&f(x)=\varphi(x)^tW^*+b^{*t}=\varphi(x)^t\,repmat\,(W_0^*,1,m)+\varphi(x)^tV^*+b^{*t}=\\&\varphi(x)^t\,repmat\,(\sum_{i=1}^{m}Z\alpha_i^*,1,m)+\frac{m}{\lambda}\varphi(x)^tZA^*+b^{*t}\\&+repmat\,(\sum_{i=1}^{m}\sum_{j=1}^{l}\alpha_{i,j}^*\ker(x,x_j),1,m)+\frac{m}{\lambda}\sum_{j=1}^{l}\alpha^{j*}\ker(x,x_j)+b^{*t}\end{aligned}\tag{31}$$

Once the support vector regression model is fully trained and its output is obtained as $\hat{e}_{t+1}=\begin{bmatrix}\hat{e}_{t+1}^{H}\\ \hat{e}_{t+1}^{L}\end{bmatrix}$, the final forecast of the hybrid system is obtained by aggregating the forecast of MSVR with Holt's smoothing. In the end, we measure the accuracy of the hybrid model on test data by calculating the percentage deviation of forecasts from the actual trend of lowest and highest prices using the mean squared error.

2.3 Research Background

Although this is the first study on the combined use of interval forecasting and multi-output support vector regression (MSVR) for stock price forecasting in Iran, support vector regression and Holt's smoothing have been used separately in different fields. In a article by Davoodi et al. [12] titled "The Effectiveness of the Automatic System of Fuzzy Logic-Based Technical Patterns Recognition: Evidence from Tehran Stock Exchange", The present research proposes an automatic system based on moving average (MA) and fuzzy logic to recognize technical analysis patterns including head and shoulder patterns, triangle patterns and broadening patterns in the Tehran Stock Exchange. The automatic system was used on 38 indicators of Tehran Stock Exchange within the period 2014-2017 in order to evaluate the effectiveness of technical patterns. Having compared the conditional distribution of daily returns under the condition of the discovered patterns and the unconditional distribution of returns at various levels of confidence driven from fuzzy logic with the mean returns of all normalized market indicators, we observed that in the desired period, after recognizing the pattern, all patterns investigated at the confidence level 0.95 with a fuzzy point 0.5 contained useful information, practically leading to abnormal returns.

In a study by Ghazanfari et al. [18] titled "Corporate Bankruptcy Prediction Based On Hybrid Intelligent Systems", they developed an intelligent and coherent system based on artificial neural network, support vector machine, extreme learning machine with the help of imperialist competitive algorithm, cultural algorithm, and harmony search in order to add ress the deficiencies of previous models. In this study, the best performance without outlier elimination was achieved when the support vector machine was combined with the harmony search and the imperialist competitive algorithm. In a study titled "Multiple-Step-Ahead Forecasting of Value at Risk Based on Holt-Winters Exponential Smoothing Multiplicative Method", Mohammadian Amiri and Ebrahimi [4] used multiplicative Holt-Winters exponential smoothing to obtain multiple-step-ahead forecasts of value at risk for auto and bank indices. The results of this study showed that multiplicative Holt-Winters exponential smoothing provides reliable and accurate estimates of the value at risk of these stock indices. In a study by Shahriari et al. [3] titled "Financial Time series Forecasting using Holt-Winters in H-step Ahead", they used the Holt-Winters method to forecast the non-stationary time-series data of the sales profit of an intermediate product. This study reported that the proposed method outperforms the classical and S-filtered methods in forecasting future values. In another research, Mansourfar et al [5] used a support vector machine model to predict financial distress based on cash flow combinations. In a article by Xiong et al. [9] titled "Interval-valued time series forecasting using a novel hybrid Holt's and MSVR model", they developed a hybrid method consisting of Holt's smoothing and multi-output support vector regression (MSVR) to forecast daily stock price intervals, that is highest and lowest daily price values. The idea of this work was to use the useful features of the linear and nonlinear methods together within a hybrid system. The resulting hybrid system was implemented on three stock exchange and energy databases and the evaluations performed in terms of the returns obtained from a simple trading strategy based on the interval forecasts showed the good accuracy of the system for the test data.

In a study titled “Stock Price Trend Forecasting Using Modified Support Vector Machine with Hybrid Feature Selection," Bajlan et al. [19] developed a model for forecasting stock price trends by a support vector machine that was weighted by daily trading volumes and equipped with a hybrid feature selection process. To evaluate the prediction accuracy, the paired t-test was used to compare the results with the estimates of a simple support vector machine with hybrid feature selection and also other feature selection techniques including information gain, symmetric uncertainty and correlation-based feature selection. This study claimed that compared to the existing methods, the modified support vector machine model with hybrid feature selection has a superior performance in stock price prediction.
In another research by Parmezan et al. [13] titled “Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model”, In this paper, From 95 datasets, we evaluate eleven predictors, seven parametric and four non-parametric, employing two multi-step-ahead projection strategies and four performance evaluation measures. We report many lessons learned and recommendations concerning the advantages, drawbacks, and the best conditions for the use of each model. The results show that SARIMA is the only statistical method able to outperform, but without a statistical difference, the following machine learning algorithms: ANN, SVM, and *k* NN-TSPI. However, such forecasting accuracy comes at the expense of a larger number of parameters.
In a study by Shaolong Sun et al. [14] titlet “Interval decomposition ensemble approach for crude oil price forecasting”, This paper proposes an interval decomposition ensemble (IDE) learning approach to forecast interval-valued crude oil price by integrating bivariate empirical mode decomposition (BEMD), interval MLP (MLP^I) and interval exponential smoothing method (Holt’s). Firstly, the original interval-valued crude oil price is transformed into a complex-valued signal. Secondly, BEMD is used to decompose the constructed complex-valued signal into a finite number of complex-valued intrinsic mode functions (IMFs) components and one complex-valued residual component. Thirdly, MLP^I is used to simultaneously forecast the lower and the upper bounds of each IMF (non-linear patterns), and Holt’s is used for modeling the residual component (linear pattern). In a study by Chen et al. [11] titled “A Feature Weighted Support Vector Machine and K-Nearest Neighbor Algorithm for Stock Market Indices Prediction”, they tried several methods to find the most accurate solutions for forecasting stock price indices. Among the tested machine learning methods, support vector machine technique was found to have the best performance in this application. The method was further tested on two well-known indices of the Chinese stock market (Shanghai and Shenzhen stock exchange indices). The algorithm proposed in this study can also be modified to make forecasts for other stock market indices. As mentioned, many studies on the subject of stock price forecasting have used support vector regression. Holt’s method has also been used in a few studied in this area. However, hybrid models have shown to be very efficient in stock price forecasting as they can use both linear and nonlinear features in their predictions. Such hybrid methods have been the subject of a few articles, the most important of which is the one published by Xiong [7]. The main difference of the present study from that work is the use of a different time interval and also the use of a different type of support vector regression known as the least squares support vector regression. The proposed method of Holt-MSVR interval forecasting is introduced in the following section.

3 Research Methodology

In this descriptive-analytical study, the goal was to predict the interval of the overall index of the Tehran Stock Exchange, that is, the interval of the lowest and highest stock price index. The method used for this purpose is a hybrid method consisting of the combination of Holt's exponential smoothing and multi-output support vector regression (MSVR). As explained in Section 2, first, the highest and lowest price indices were approximated and smoothed by Holt's smoothing method, and then the residuals of approximation were fed to MSVR as a time series. Finally, the forecasts were obtained by summing the outputs of MSVR and Holt's smoothing. In the end, we tested the accuracy of the model on a test dataset. The method was implemented in Matlab and Eviews. The algorithm of the forecast method is illustrated in Fig. 1.

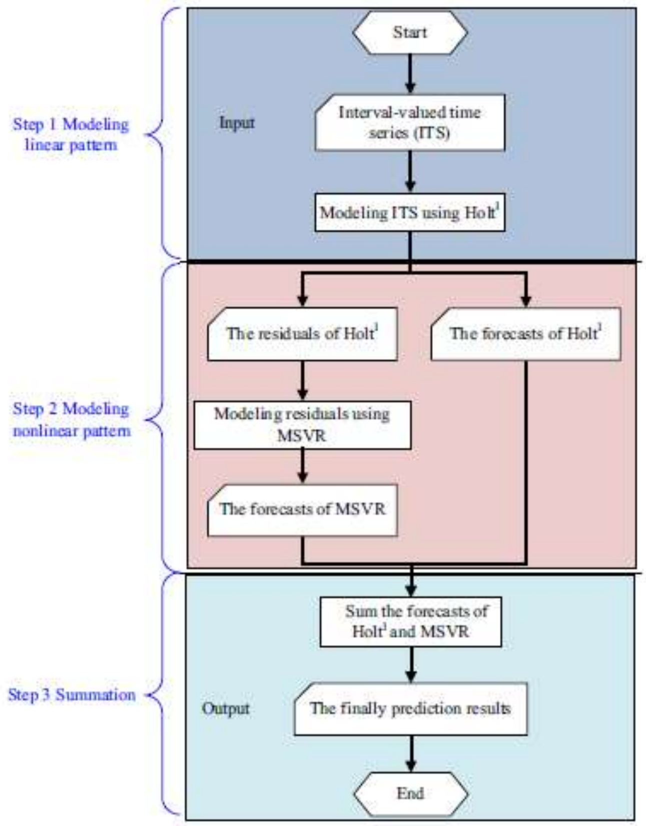

Fig. 1: Algorithm of the Forecasting Model

3.1 Research Questions

1-How to predict the intervals of stock price changes including the lowest and highest prices by combining Holt's smoothing with multi-output support vector regression?
2 – How accurate the aforementioned hybrid method would be in forecasting the overall index of the stock market?

3.2 Statistical Population and Sample

The statistical population of this study was the records of the overall index of the Tehran Stock Exchange in the period of 1982-2016, which was collected from Rahavardnovin software. The daily data of the overall index, which contains 5980 data points, was divided into weekly data (five days per week) and the lowest and highest index values in each week were determined. Of the 1196 weekly intervals available (including the vector of minimum and maximum index values), 1000 were used to train Holt's smoothing and MSVR and the last 196 intervals were used as test data. (approximately 20% of the data was used for testing data. In most studies this ratio is 15-20%).

4 Data Analysis

The daily values of the overall index over the period 1982-2016, which include 5980 data points, are plotted in Chart 1.

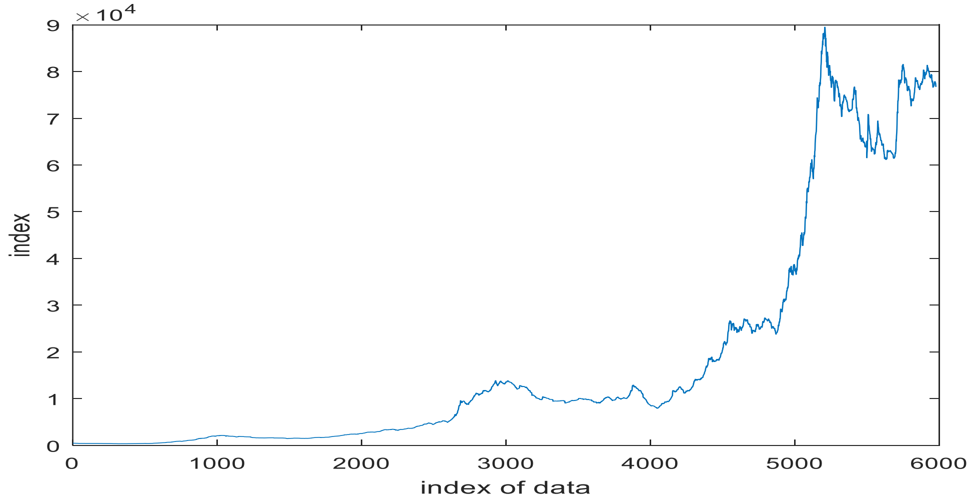

Chart 1: Overall Index of Tehran Stock Exchange

Some of these interval-valued data are displayed in Chart 2:

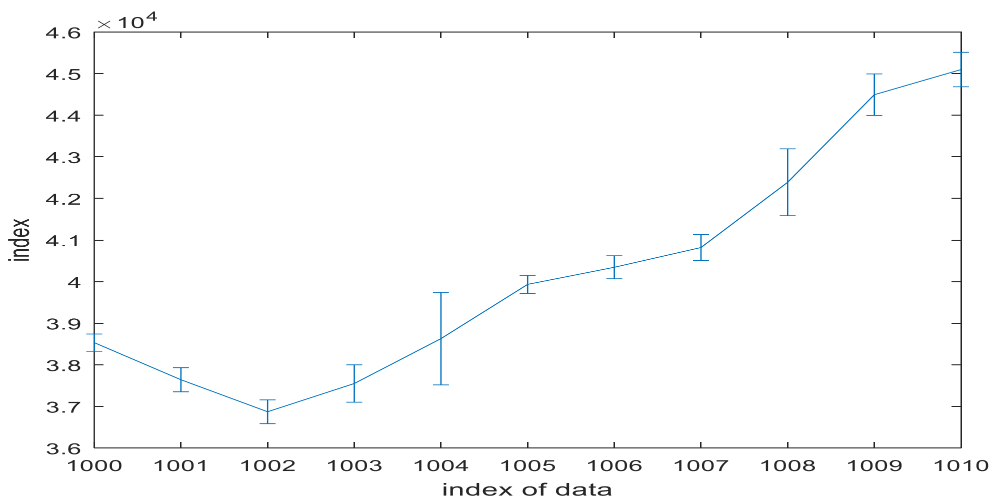

Chart 2: Examples of the Interval-Valued Data Used in the Research

4.1 Descriptive Statistics

The Descriptive Statistics of the Data are presented in Table 1.

Table 1: Descriptive Statistics of the Research Data

Statistical index	Value
Mean	0.000864
Median	0.000392
Maximum	0.054017
Minimum	-0.05513
Standard deviation	0.005669
Skewness	0.470124
Kurtosis	14.52453
Jarque-Bera statistic	33307.76
Probability value	0

The Jarque-Bera statistic was used to check the normality of the data distribution, and as statistic value shown in Table 1 indicates, it was found that data does not follow a normal distribution.

4.2 Results of the Holt Model

The interval-valued series of the overall index were smoothed by the Holt's exponential smoothing method. For smoothing, matrices A and B were calculated by optimizing the sum of squared error. This optimization was done by a Genetic Algorithm with specifications given in Table 2 in Matlab.

Table 2: Parameter Setting of the Genetic Algorithm

Parameter	Value
Population size	500
Number of iterations	100
Crossover probability	0.90
Mutation probability	0.05

The result of the optimization of the two matrices is provided below:

$$A = \begin{bmatrix} 0.9978 & 0.2551 \\ 0.6872 & 0.5270 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.2043 & 0.0132 \\ 0.1822 & 0.0132 \end{bmatrix} \qquad (32)$$

As an example, Chart 3 compares a portion of the interval-valued data smoothed by the Holt method with the corresponding original data.

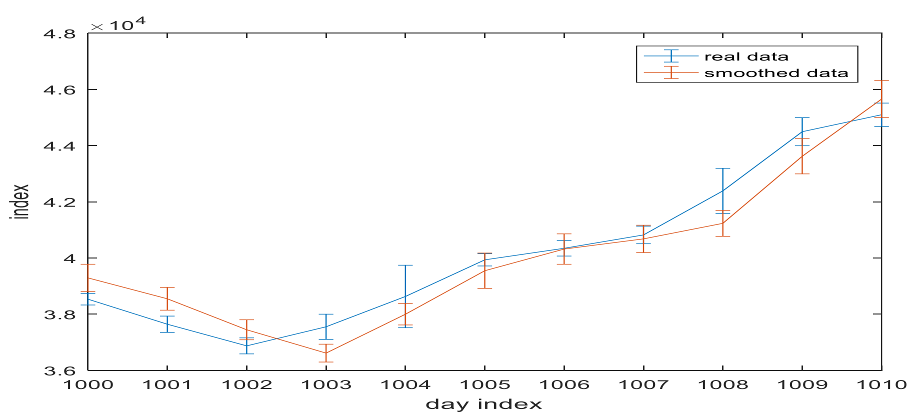

Chart 3: Comparison of the Smoothed Series and Original Interval-valued Data

4.3 Performance of Holt's smoothing in training and test data

The summarized results of performance evaluation of Holt's smoothing in training and test data are presented in Table 3.

Table 3: Performance of Holt's Smoothing

Index	in training data	in test data
Correct forecasting of the trend of the highest price	0.85	0.68
Correct forecasting of the trend of the lowest price	0.83	0.65
Correct prediction of the trends of both highest and lowest prices	0.71	0.64
mean squared error in the forecasting of the highest price	10.21	13.45
mean squared error in the forecasting of the lowest price	8.10	17.28
Total mean squared error in the forecasting of highest and lowest prices	18.31	30.78

4.4 Descriptive Statistics of Approximation Residuals

After applying the Holt's smoothing to the training data, the approximation residuals, that is, the result of subtracting Holt's approximations from the initial interval-valued data was calculated. The results obtained for some of these residuals and their descriptive statistics are presented in Table 4 and Figure 4.

Table 4: Descriptive Statistics of the Holt's Smoothing Residuals

Index	Residuals for the highest price	Residuals for the lowest price
Mean	-3.857771	12.94811
Median	-1.035678	2.89036
Maximum	5955.25	3916.826
Minimum	-3957.991	3977.957
Standard deviation	585.3408	481.0047
Skewness	1.11555	-0.480046
Kurtosis	24.33853	24.53066
Jarque-Bera statistic	22938.81	23147.15
Probability value	0	0

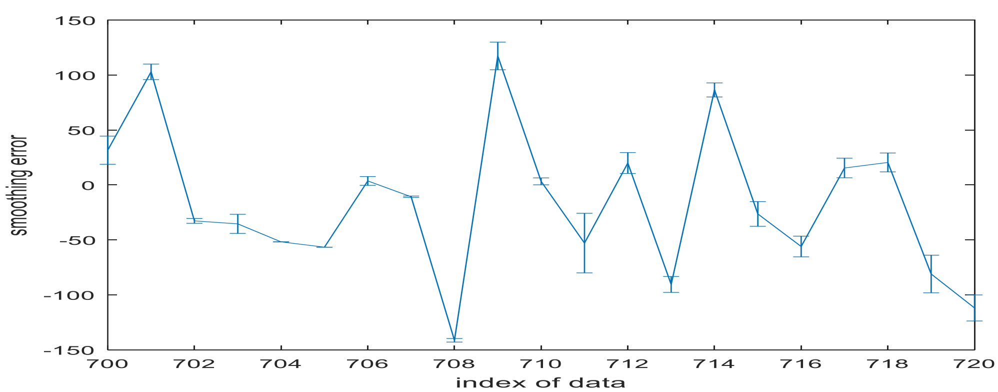

Chart 4: Part of the Holt's Smoothing Residuals

4.5 Hybrid System (modeling of residuals)

Residuals were modeled using multi-output least squares support vector regression. Using the Akaike statistic, the optimal lag length for the residual series was calculated to 4. The support vector regression was trained with γ=0.50, λ=4, and ρ=2 and Gaussian kernel. The optimum values of regression parameters β and α (a matrix with 2 rows and 1000 columns) are provided below:

$$b=\begin{bmatrix}-4.12591845616201\\12.9537919184076\end{bmatrix} \tag{33}$$

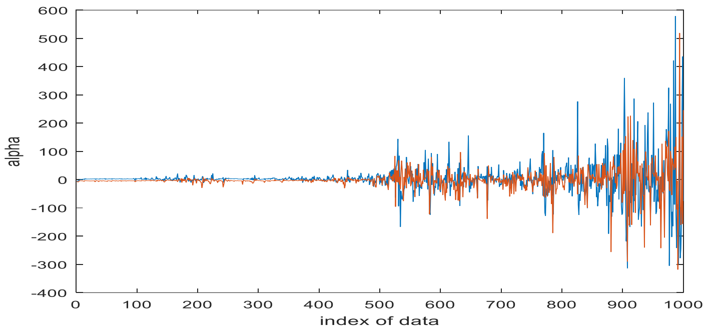

Chart 5: Optimum α

4.6 Forecasting of Test Data

The trained support vector regression model was used to calculate the vectors of residuals are for test data. The resulting values were then added to Holt approximations to obtain the final forecasts. Some of the forecasts obtained for these data are shown in Figure 6:

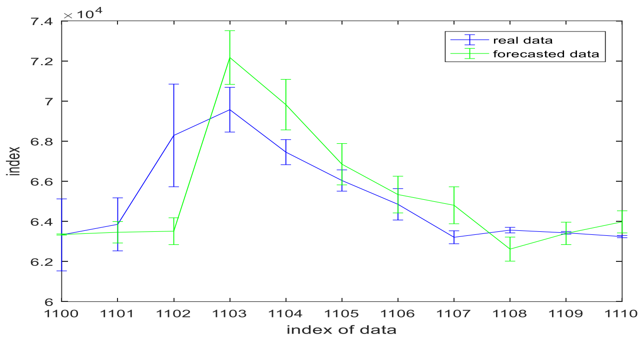

Chart 6: Part of the Forecasts Obtained from the Hybrid Model

The summary of the results of the performance evaluation of the hybrid in training and test data are presented in Table 5.

Table 5: Performance of the Hybrid Method

Index	in training data	in test data
Correct forecasting of the trend of the highest price	0.88	0.72
Correct forecasting of the trend of the lowest price	0.87	0.70
Correct prediction of the trends of both highest and lowest prices	0.76	0.68
mean squared error in the forecasting of the highest price	8.25	7.05
mean squared error in the forecasting of the lowest price	7.00	10.20
Total mean squared error in the forecasting of highest and lowest prices	15.31	17.25

Comparison of the results presented in Tables 3 and 5 shows that the modeling of residuals with multi-output vector regression has significantly improved the performance of Holt's model, ultimately leading to satisfactory results.

5 Discussion and Conclusion

In this study, a hybrid system of forecasting was implemented for the overall index of the Tehran Stock Exchange. First, data were partitioned into weekly (5 days) groups and the lowest and highest prices in each week were used to form the interval vector of prices. Of the resulting interval vectors, 1000 were used to train the smoothing model and the multi-output support vector regression model and 196 were used as test data. The results indicate that the modeling of the smoothing residuals can increase the accuracy of forecasting for both in-sample and out-of-sample data. Also, this approach is capable of predicting future trends of the lowest and highest prices. As shown in Table 5, the accuracy of the method in predicting the trend of the highest price was 0.88 for the training data and 0.72 for the test data. For the trend of lowest price, the accuracy was 0.87 for the training data and 0.70 for the test data. In the simultaneous prediction of the trends of both lowest and highest prices, the method showed an accuracy of 0.76 for the training data and 0.68 for the test data.

Xiong et al. [9] have also developed a hybrid model based on Holt's smoothing and multi-output vector regression for forecasting the daily intervals of stock prices. After implementing this model on three stock exchange and energy databases, evaluations performed in terms of the returns obtained from a simple trading strategy based on the interval forecasts showed that the developed model enjoys good accuracy on the test data. It should be noted that the hybrid system of the present study uses a different type of support vector regression called the least-squares support vector regression, and also uses weekly (5 days) intervals. As also indicated by Xiong et al., by incorporating the nonlinear capabilities of multi-output vector regression, the hybrid approach can improve the linear capabilities of the Holt's smoothing model.

There has been a great number of studies on stock price forecasting, many of which have reported the superiority of nonlinear methods over linear methods. As mentioned, the present study has attempted to combine the linear and nonlinear methods of stock price forecasting. In a study by Falahpour et al. [6], where they presented a hybrid model of support vector machine based on genetic algorithm for predicting stock price trends, it was showed that hybrid approach performs much better and is much more accurate than the simple support vector machine. Considering the difficulty of accurate identification of linear and nonlinear patterns in economic and financial time series, Raee et al. [20] used a combination of autoregressive integrated moving average models and support vector machine to predict weekly prices of Brent crude oil as an oil indicator. This study reported that in most cases the hybrid model had a lower error in predicting crude oil prices than other models. In the studies of Mohammadian Amiri and Ebrahimi [4] and Shahriari et al. [3], forecasts made by Holt's exponential smoothing method demonstrated the high accuracy of this method. Chen et al. [7], who used the support vector machine for stock forecasting also reported the high accuracy of this method.

Suggestions and limitations of this research can be stated as follows:

1. Forecasting means making optimal use of available information to estimate future values of a random variable. There are many ways to combine the available data for this purpose. These include statistical methods such as ARIMA and GARCH and artificial intelligence methods such as neural networks and

support vector regression. To build an accurate model, it is necessary to run the models with different types of training data and evaluate their accuracy in forecasting test data.

2. The present study presented a new hybrid approach for forecasting stock price intervals. Using the results of the hybrid method on the test data, it is possible to evaluate its accuracy in forecasting the trend of price changes (increasing/decreasing) based on criteria such as mean squared error. Since the hybrid approach managed to improve the performance of Holt's smoothing model, investors and shareholders interested in forecasting stock prices, especially the lowest and highest prices are encouraged to train this model on their own data and use it for forecasting provided that it offers the desired accuracy for the test data. Researchers are also recommended to embrace the new paradigm of interval forecasting as a high potential field of research.

As a result, our suggestions for future research can be summarized as follows:

1. Researchers are recommended to also test other variants of multi-output support vector regression such as marginal likelihood and compare the results.
2. Researchers are recommended to apply the proposed hybrid method to other indicators of the stock market or other financial assets and report the resulting accuracy.
3. Researchers can test other optimization algorithms for Holt's smoothing and compare the results.
4. Researchers are recommended to define interval forecasts based on new intervals such as opening and closing prices and report the accuracy of the hybrid method.

References

[1] Raee, R., Nikahd, A., Habibi, M., *The Index Prediction of Tehran Stock Exchange by Combining the Principal Components Analysis, Support Vector Regression and Particle Swarm Optimization*, Financial Management Strategy, 2017, **4**(4), P.1-23. Doi: 10.22051/jfm.2017.12410.1175

[2] Raei, R., Saeidi, A., *Principles of financial engineering and risk management*, Tehran University Business School publications and Samt publications, 2004. (in Persian).

[3] Shahriari, H., Aghaie, A., Nezhad Afrasiabi, M., *Financial Time series Forecasting using Holt-Winters in H-step Ahead*, Financial Research Journal, 2016, **18**(3), P. 505-518. Doi: 10.22059/jfr.2016.62453

[4] Mohammadian Amiri, E., Ebrahimi, S., *Multiple-Step-Ahead Forecasting of Value at Risk Based on Holt-Winters Exponential Smoothing Multiplicative Method*, Financial Management Strategy, 2018, **6**(1), P.93-114. Doi: 10.22051/jfm.2018.15099.1355.

[5] Mansourfar, G., Ghayour, F., Lotfi, B., *The Ability of Support Vector Machine (SVM) in Financial Distress Prediction*, Empirical Research in Accounting, 2015, **5**(1), P.177-195. Doi: 10.22051/jera.2015.646

[6] Fallahpour, S., Golarzi, G., Fatourechian, N., *Predicting Stock Price Movement Using Support Vector Machine Based on Genetic Algorithm in Tehran Stock Exchange Market*, Financial Research Journal, 2013, **15**(2), P.269-288. Doi: 10.22059/jfr.2013.51081.

[7] Chen, Y., Hao, YA., *feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction*, Expert Systems with Applications, 2017, **80**, P.340 355. Doi:10.1016/j.eswa.2017.02.04.

[8] Falah Shams, M., Rashnoo, M., *Investigating the Prices Manipulation in the Tehran Stock Exchange by Using the SVM Model*, Financial Research Journal, 2012, **14**(1), P. 69-84. Doi: 10.22059/jfr.2012.36633

[9] Xiong, T., Li, C., Bao, Y., *Interval-valued time series forecasting using a novel hybrid Holt'I and MSVR model*, Economic Modelling, 2017, **60**, P.11-23. Doi: 10.1016/j.econmod.2016.08.019

[10] Huang, S. C., Chuang, P. J., Wu, C. F., Lai, H. J, *Chaos-based support vector regressions for exchange rate forecasting*, Expert Systems with Applications, 2010, **37**(12), P.8590-8598. Doi: 10.1016/j.eswa.2010.06.001

[11] Gavrishchaka, V., Banerjee, S., *Support vector machine as an efficient framework for stock market volatility forecasting*, Computational Management Science, 2006, **3**(2), P.147-160. Doi:10.1007/s10287-005-0005-5

[12] Abdolbaghi Ataabadi, A., Davoodi, S., Salimi Bani, M., *The Effectiveness of the Automatic System of Fuzzy Logic-Based Technical Patterns Recognition: Evidence from Tehran Stock Exchange*, Advances in Mathematical Finance and Applications, 2019, **4**(3), P.107-125. Doi: 10.22034/amfa.2019.585179.1185

[13] Parmezan, A. R. S., Souza, V. M., Batista, G. E., *Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model*, Information Sciences, 2019, **484**, P.302-337. Doi:10.1016/j.ins.2019.01.076

[14] Sun, S., Sun, Y., Wang, S., Wei, Y., *Interval decomposition ensemble approach for crude oil price forecasting*, Energy Economics, 2018, **76**, P.274-287. Doi: 10.1016/j.eneco.2018.10.015.

[15] Ehteshami, S., Hamidian, M., Hajiha, Z., Shokrollahi, S., *Forecasting Stock Trend by Data Mining Algorithm*, Advances in Mathematical Finance and Applications, 2018, **3**(1), P.97-105.
Doi: 0.22034/amfa.2018.539138

[16] Aghaeefar, N., Mohammad Pourzarandi, M., Afshar Kazemi, M., Minoie, M., *Applying Optimized Mathematical Algorithms to Forecast Stock Price Average Accredited Banks in Tehran Stock Exchange and Iran Fara Bourse*, Advances in Mathematical Finance and Applications,2019, **4**(3), P.77-94.
Doi: 0.22034/amfa.2019.580802.1150

[17] Abbasi, E., Samavi, M., Koosha, E., *Performance Evaluation of the Technical Analysis Indicators in Comparison whit the Buy and Hold Strategy in Tehran Stock Exchange Indices*, Advances in Mathematical Finance and Applications,2020, **5**(3), P.1-19. Doi:10.22034/amfa.2020.1893194.1376

[18] Panahian, H., Ghazifari, S., *Modelling and Investigating the Differences and Similarities in the Volatility of the Stocks Return in Tehran Stock Exchange Using the Hybrid Model PANEL-GARCH*, Advances in Mathematical Finance and Applications,2018, **3**(2), P.13-26. Doi:10.22034/amfa.2018.540828

[19] Bajalan, S., Fallahpour, S., Dana, N., *The Stock Trend Prediction Using Volume Weighted Support Vector Machine with a Hybrid Feature Selection Method to Predict the Stock Price Trend in Tehran Stock Exchange*, Financial Management Strategy,2016, **4**(3), P. 121-148. Doi:10.22051/jfm.2016.2575

[20] Rai, R., Habibi, M., *Prediction of Tehran Stock Exchange Index by combining methods of analysis of main components, support vector regression and cumulative particle motion*, Financial Management Strategy, 2017, **4**(4), P.1-23. Doi:10.22051 / JFM.2017.12410.1175