Tehran Stock Price Modeling and Forecasting Using Support Vector Regression (SVR) and Its Comparison with the Classic Model ARIMA

Saeed hajibabaei[*]
Nematollah hajibabaei[**]
Syed mohammad hoseini[***]
Somaye hajibabaei[****]
sajad hajibabaei[*****]

* MSc; Department of Art and Architecture, Hamedan Branch, Islamic Azad University, Hamedan, Iran(**Corresponding Author**)
** MSc Candidate; Department of MANAGMENT, Buin zahra Branch, Islamic Azad University, Buin zahra, Iran
*** MSc Candidate; Department of Art and MANAGMENT, malayer Branch, Islamic Azad University, malayer, Iran
**** MSc; Department of Accounting, Hamedan Branch, Islamic Azad University, Hamedan, Iran
***** MSc Candidate ; Department of Art and Architecture, Hamedan Branch, Islamic Azad University, Hamedan, Iran

Abstract

Use of linear and non-linear models to predict the stock price has been paid attention to by investors, researchers and students of finance and investment companies, and organizations active in the field of stock. Timely forecasting stock price can help managers and investors to make better decisions. Nowadays, the use of non-linear methods in modeling and forecasting financial time series is quite common. In recent years, one of the new techniques of data mining with support vector regression (SVR) has had successful application in time series prediction. In this study, using support vector regression model, we examined the Tehran Stock prices and the predicted results were compared with ARIMA classic model. For this purpose, of the Tehran stock companies, 5 companies were selected during the years 2002 to 2012. Using benchmarks to evaluate the performance of MSE, MAE, NMSE these two methods were compared and the results (in the case of non-linear data) indicate the superiority of SVR model compared to the ARIMA model.

Keywords: stock investment, stock price forecasting, data mining, support vector regression, ARIMA models

1- Introduction

Achieving the long-term and continuous economic growth demands optimal allocation of resources at the national economics level which is impossible to obtain without the help of financial markets, especially extensive capital market .In a healthy economy, the presence of a financial system plays a role in proper distribution of funds and financial resources.

Usually, the financial markets are defined as a system composed of individuals and institutions, instruments and procedures that are collecting savers all in one place. One of the most attractive financial markets in which asset management and decision making quality are of extremely importance is stock market. In this market, the correct decision making has even greater importance than in traditional markets like commodities market and to do suitable activities and create wealth, the power of prediction ,direction and pace of change are very important. Stock price forecasting is a popular and important topic in financial and academic studies.(Yue xo, 2012) Yet, the number of individuals having experience and talent to guess the direction of stock market is few .Therefore , the systems which can help to make a proper decision is very valued.

Predicting stock index is a major activity of financial firms and private investors. However, stock index prediction is regarded as a challenging task of the prediction problem since the stock market is a complex, evolutionary, and nonlinear dynamic system (Lu et al, 2011). Highly accurate interval forecasting of a stock price index is fundamental to successfully making a profit when making investment decisions, by providing a range of values rather than a point estimate. (xiong, 2013)

Support vector regression (SVR) has successfully solved prediction problems in many domains, including the stock market. (Hung, tsai, 2009). Support vector regression (SVR) has become a popular alternative in stock index forecasting tasks due to its generalization capability in obtaining a unique solution. (dai et al, 2013)

The financial time series such as exchange rates, stock prices, inflation and oil prices, due to features such as no inertia, being non-linearity and noise are difficult to predict. There have been different studies in time prediction and several methods have been evaluated. (Engel, 1989) in recent years one of the machine learning algorithms named SM has had successful applications in time series prediction, especially financial time series. This

method, due to its outstanding properties such as simple geometric interpretation, reaching a general and unique solution is noticeable. Choosing best model in SVM model is equal to solving a quadratic programming problem, which leads to a unique and general solution. On the other hand, SVM can estimate the complex non-linear functions by applying a suitable core function and mapping the original data into a higher dimensional space. (Basak et al, 2007) due to these features, the use of genetic algorithm has become widespread as in predicting stock price in stock market. Also, one of the methods to predict stock price in market is mixed SVM-ARIMA model. (Ince and Trafalis, 2006)

2- Support Vector Machines for Regression Problems (SVR)

Based on Statistical Learning Theory, Support Vector Machine (SVM), was first developed by Vapnik(vapnik, 1998) Another version of SVM method for regression problems in year is called. support vector regression (SVR) (Vapnic, 1995) Application of support vector regression (SVR) could improve forecasting accuracy performance.(hong et al, 2013) The model produced by svc is only a subset of the training dependent sample .similarly , the produced model by the SVR is only a subset of training dependent samples because the loss function for model construction ignores the training data which are close to prediction model (at ε distance from the prediction model) This is corresponding to using loss function ε- insensitive suggested by Vepnic and produces one sparse set of support vectors .

Overall, SVR acts to estimate the regression function as the following

a. Estimate the regression function in a set of linear functions as $f(X)=(w,\Phi(X))+b$

b. Minimizes empirical risk with respect to loss function ε- insensitive

c. Performs structural risk minimization principle to minimize the real risk in that it considers the structure of the set of linear functions (defined by norm of coefficients of linear functions) as follows : $H_1 CH_2 C \ldots CH_k C \ldots$

Where H_k element of the structure is defined by the following inequality :

$$\| w \|^2 = (w,w) \leq C_k \tag{1}$$

And to find the model of optimal complexity, minimizes $\|w\|^2$.

2-1- Linear Mode

Suppose we have the training data S= {(x₁,y₁), (x₂,y₂) , … , (xₙ, yₙ)} $CR^d \times R$ in which n is the number of samples and R^d represents the input space . In SVR, our goal is to find a function f(x) which for all instances has ε values deiation from the real purpose of y_i and is as smooth as possible. In other words, the errors that are less than ε are overlooked, but it will not accept errors greater than e. linear regression function in the input space can be estimated as follows:

$$x, w \in R^d, b \in R \qquad f(x)=(w,x)+b \tag{2}$$

Smoothness of (2) means that we are looking for a model with minimal complexity. According to what was mentioned in the previous section, this is done with minimization of $\|w\|^2$. So the main problem can be written as follows: (smola, Scholkopf, 1998)

$$\min \frac{1}{2} \| \mathbf{w} \|^2 \tag{3}$$

$$\text{s.t.} \qquad y_i - \langle w, x_i \rangle - b \leq \varepsilon \quad , i = 1,2,....,n$$

$$\langle w, x_i \rangle + b - y_i \leq \varepsilon \quad , i = 1,2,...,n \tag{4}$$

Which is a quadratic optimization problem. The implicit assumption in (4) is that there really is a function f which can approximate all pairs (x_i, y_i) with accuracy ε. Or in other words, the above optimization problem can be solved. Sometimes it may happen that the relation (4) does not hold for all training samples. To deal with impossible constraints in the aboe optimization problem, we define the auxiliary variables of ξ_i^*, ξ_i so that they satisfy the following restrictions.

$$\min \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \tag{5}$$

$$\text{s.t.} \qquad y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i$$

$$\langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \tag{6}$$

$$\xi_i^{(*)} \geq 0, i = 1,2,...,n$$

In which $\xi_i^{(*)}$ refers to variables of ξ_i^*, ξ_i and constants C> 0 is a balance between smoothness of f and empirical risk. For example, if C=∞, then the aim is to minimize the empirical risk regardless of the complexity of the model.

(a)

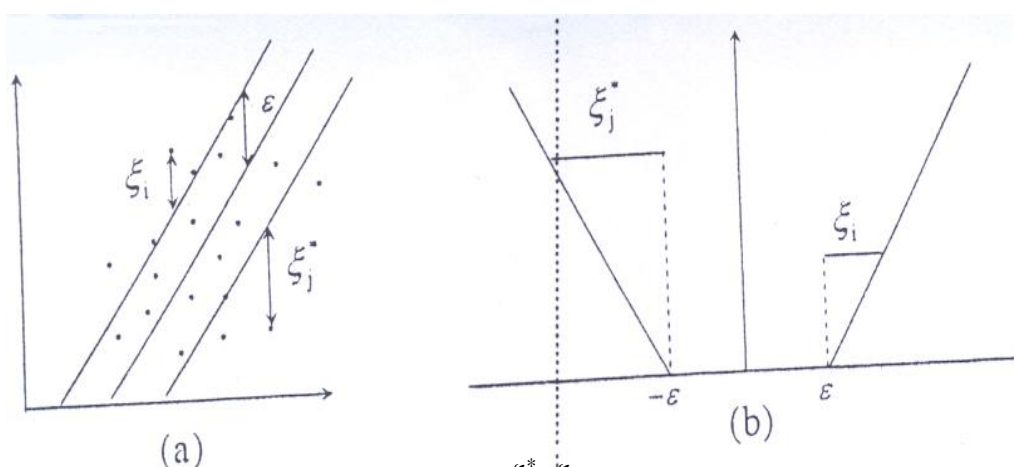Figure 1. (a) auxiliary variables (b) ξ_i^*, ξ_i are loss function ε- insensitive

(a) Auxiliary variables (b) ξ_i, ξ_i are loss function ε- insensitive which is defined as follows:

$$L_\varepsilon(y - f(X)) = \max(|\,y - f(x)\,| - \varepsilon, 0) = \begin{cases} 0 & |\,y - f(x)\,| \leq \varepsilon \\ |\,y - f(x)\,| - \varepsilon & 0.\omega \end{cases}$$

In most cases, the above optimization problem is solved easily in its dual formulation. Using coefficient method of Lagrange, we have:

$$L(w, b, \alpha, \beta) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) - \sum_{i=1}^{n}\alpha_i(\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b)$$

$$- \sum_{i=1}^{n}\alpha_i(\varepsilon + \xi_i^* - y_i + \langle w, x_i \rangle - b) - \sum_{i=1}^{n}(\beta_i\xi_i + \beta_i^*\xi_i^*)$$

$\beta_i^*, \beta_i, \alpha_i^*, \alpha_i$ Are Lagrange coefficients that satisfy the following conditions:

$$\alpha_i, a_i^*, \beta_i, \beta_i^* \geq 0$$

It is shown that this function can have a saddle point related to dual and main variables. The saddle point condition makes the partial derivations of L become zero interms of main variables to get optimization

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i) = 0 \tag{7}$$

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\mathbf{x_i} = 0 \;\;\Rightarrow\; \mathbf{w} = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\mathbf{x_i} \tag{8}$$

$$\frac{\partial L}{\partial \xi_i^{(*)}} = C - \alpha_i^{(*)} - \beta_i^{(*)} = 0 \Rightarrow \alpha_i^{(*)} = C - \beta_i^{(*)} \tag{9}$$

By putting (7), (8) and (9) in the Lagrange function, we have

Thus, the optimization problem is obtained under the dual form

$$\max_{\alpha} \; -\frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_h - \alpha_i^*) < \mathbf{x_i}, \mathbf{x_j} > \\ = -\frac{1}{2}(\mathbf{w},\mathbf{w}) - \varepsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{n}y_i(\alpha_i - \alpha_i^*) \\ -\varepsilon\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) + \sum_{i=1}^{n}y_i(\alpha_i - \alpha_i^*) \tag{10}$$

$$s.t \quad \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0 \tag{11}$$

$$\alpha_i, \alpha_i^* \in [0, C]$$

Calculating the Lagrange coefficients from equation (11) and using equation (8), we can write f(x) as the following:

$$f(X) = <\mathbf{w}, \mathbf{x}> + b \quad \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) < \mathbf{x_i}, \mathbf{x}> + b \tag{12}$$

As it can be seen, SVR algorithm is described in terms of the dot product between the data, and cwe not need to compute w explicitly to get f(x).

2-1-1 calculation b

To find f(x), w is required to be computed in addition to b. KKT conditions reflect the product of the dual variables and constraints must be zero.

$$\alpha_i(\varepsilon + \xi_i - y_i + <\mathbf{w}, \mathbf{x_i}> + b) = 0 \tag{13}$$

$$\alpha_i^*(\varepsilon + \xi_i^* + y_i - <\mathbf{w}, \mathbf{x_i}> - b) = 0 \tag{14}$$

$$(C - \alpha_i)\xi_i = 0 \tag{15}$$

$$(C - \alpha_i^*)\xi_i^* = 0 \tag{16}$$

The above relationships provide some useful conclusions. First, only the samples (x_i, y_i) corresponding to $\alpha_i^{(*)} = C$ are outside e- insensitive interval. Also, $\alpha_i, \alpha_i^* = 0$ means there is never a set of dual variables of α_i and α_i^* which become zero simultaneously .for $\alpha_i^{(*)} \in (0, C)$ we have $\xi_i^{(*)} = 0$ and therefore the second factor in 13 and 14 must be zero and b can be computed.

$$b = y_i - <\mathbf{w}, \mathbf{x_i}> - \varepsilon \quad a_i \in (0, C)$$
$$b = y_i - <\mathbf{w}, \mathbf{x_i}> + \varepsilon \quad a_i^* \in (0, C)$$

The relations (13) and (14) result that only samples with $|f(x_i) - y_i| \geq \varepsilon$ Lagrange coefficients may be nonzero. For samples with $|f(x_i) - y_i| < \varepsilon$ the second factor in (13) and (14) is non-zero, hence for KKT conditions to hold, α_i and α_i^* must be zero.

The samples which have nonzero Lagrange coefficients are called support vectors. Due to this and the relations (8) and (12), to compute f (x) and w , we do not need all of x_i, but only support vectors are involved in the calculation of w and f (x) . This property is called sparseness because the complexity of the model depends only on the number of support vectors.

2. Non -Linear mode

In SVR method when we cannot fit a linear regression function in input space to data, we can use a nonlinear mapping to transform data in to a space with greater dimension and the standard svr algorithm is applied in this new space. Thus, linear regression in the new space corresponds to nonlinear regression in the original input space.

As mentioned, the linear SVR algorithm is dependent on only dot product between data the nonlinear SVR algorithm only depends on the dot product. Replacing the kernel function for dot product of $<\Phi(x_i), \Phi(x_i)>$ in new space with great dimension, the linear regression function is written as follows.

$$f(x) = <s, \Phi(x)> + b = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) <\Phi(x_i), \Phi(x_j)> + b$$

$$= \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x) + b$$

Using kernel functions instead of a dot product of, calculations are done. Using data from the original input space, without having to know the mapping. There are a variety of kernel functions the common of which are polynomials of RBF Gaussian, RBF exponential and tangent hyperbolic.

3- Introduction to ARIMA models

ARIMA model is a linear - stochastic time series model in econometrics which is used for stationary and no stationary time series .one of the methods for transformation of no stationary series into a stationary series is subtraction method .for the Prediction of model ARIMA (p, d, q) the methodology of Jenkins box is used that has four stages: identification, estimation, diagnostic control and prediction. (Brillinger, 1981; Chattfield, 1975; box, Jenkins, 1975)

ARIMA model is one of the most common methods of forecasting time series. In an ARIMA model, the future value of a variable is expressed as a linear combination of past values and past errors as follows:

$$y_t = \theta_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \qquad (17)$$

Where y_t is the actual value, ε_t the random error at time t, φ_i and θ_j coefficients and p and q are integers which are the ARIMA model orders. In these models, with the appropriate pre-processing methods, the data are processed. ARMA model is then selected. To determine the model order, autocorrelation function and data partial autocorrelation are used.

4- Data mining

Data mining refers to extracting or mining knowledge from huge amount of data as an important step in the knowledge discovery process. (Han and Kamber, 2006), data mining techniques can be divided into two main categories: descriptive and predictive. In the descriptive approach, we aim to describe an event or a fact but in predictive methods, we aim to predict the unknown variables from future data (tan et al, 2006)
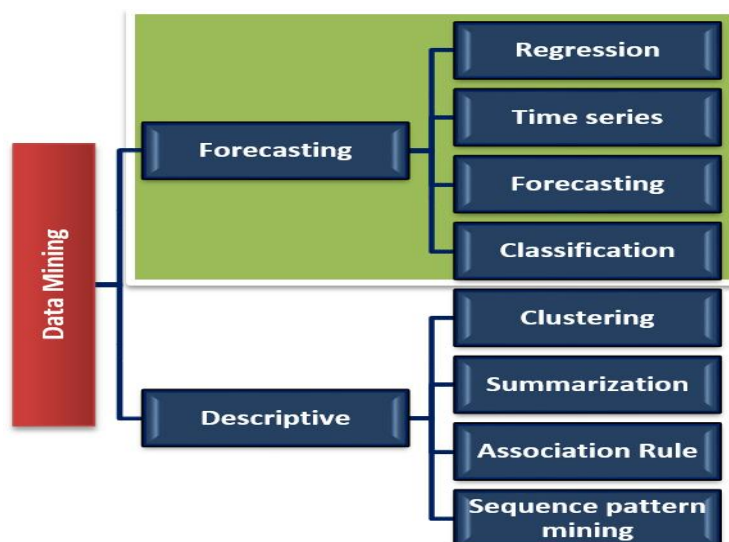
Figure 2: Data Mining Techniques

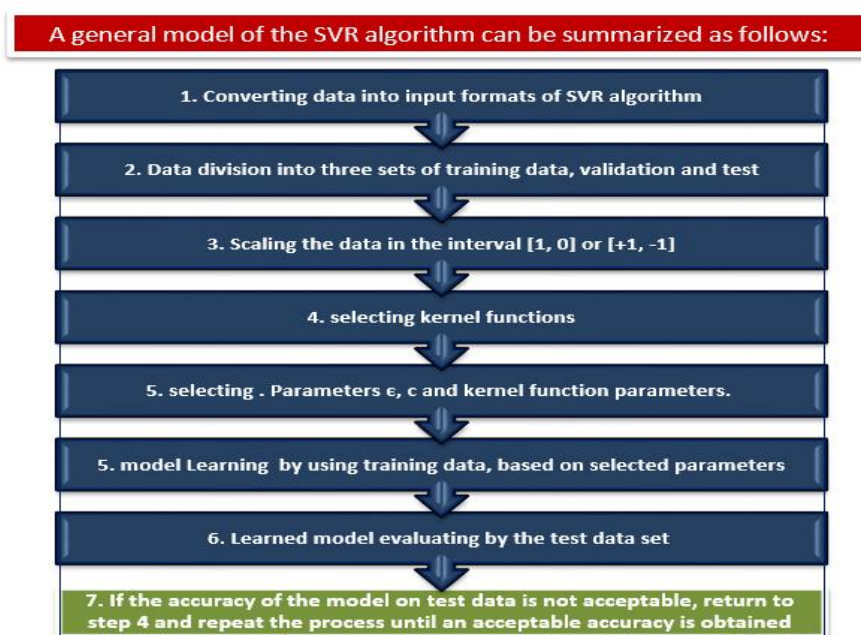## 5- Process Modeling and Implementation of SVR Algorithm

Figure 3: Process Modeling and Implementation of SVR Algorithm

5-1- Statistical Samples

According to SVM model, the value of the sample must be too high. Because of this, data were selected on a daily basis from the variables. In the SVM algorithm, every raw data sets divided into three sections. The first part, which comprises 60 percent of the data, is used as the training set to build the model. The second part comprising 30 percent of total data is called validation set used for selection of SVM model parameters and the remaining 10 percent is used as assessment of model accuracy. The table below shows how data are divided into three subsets of the training, validation and testing.

Table1: Separation of Data into Three Educational, Validation and Testing Subsets

COMPANY	TRAINING	VALIDATION	*TEST*	TOTAL
Ghadir	1150	550	95	1795
Khodro	1000	400	97	1497
Roy	870	450	152	1472
Saipa	750	350	98	1198
Setran	1200	600	163	1963

One of the steps of pre-processing time series data is using a moving window method. K window display is shown below.

$$x_1, x_2, x_3, ... x_k, x_{k+1}, x_{k+2}, ... x_n$$

The first input $\quad s.t \quad \sum (\alpha_i - \alpha_i^*) = 0$

The second input $\quad x_2, x_3, ..., x_{k+1}$

The third input $\quad x_3, x_4, ..., x_{k+2}$

The last entry $\quad x_{n-(k-1)}, ..., k_n$

To fix the remote point problem and prevent computational problems, it is better for data to be scaled in the interval [1, 0] or [+1, -1]. the following equation scales the data in the interval [0, 1]

$$Y_1 = \frac{X_i - X_{min}}{X_{ma} - X_{min}} \tag{18}$$

In this study, a moving window of size 5 is used to convert data into input format of SVR. Moreover, data are scaled in the interval [0, 1] to remove the outliers.

5-2- Choosing the Parameters

Before implementation of SVR algorithm, it is needed to specify the type of kernel function parameters and the parameters ε, and c. The accuracy of SVR models largely depends on the proper choice of kernel function parameters ε and c. Each of the parameters influences the model complexity and accuracy of predictions. ε Is defined as acceptable radius over deviations from the regression function such that an increase of ε reduces the number of support vectors. On the other hand, parameter c makes a balance between experimental risk and complexity such that increasing c increases the importance of learning error compared to model complexity. SVR algorithm, with the application of an appropriate kernel function and mapping the data into a higher dimensional space, is able to find nonlinear regression functions. Proper choice of kernel function and its parameters plays an important role in the complexity and accuracy of SVR models.

regarding the nonlinear data, we use a Gaussian RBF kernel function as applied in most applications of time series prediction (tay and cao, 1992) and a grid search method is used to select the model parameters in that taking two parameters constant, the values of the parameters will change based on the lowest MSE in data validation to select the best parameter value. (Cao and tay, 2001)

The following table shows selected parameters for the five data sets:

Table 2: Selected Parameters for 5 Sets of Data from Ghadir , Khodro, Roy, Saipa and Siman Tehran

	γ	c	ε
GHADIR	0.75	4900	0.1
KHODRO	5	700	0.00025
ROY	1	50000	1
SAIPA	0.25	7000	0.03
SETRAN	1.5	2000000	0.15

Parameters are selected based on the lowest MSE in validation data because data validation can be used to select the best model and to adjust the initial model.

5-3- Autocorrelation and Partial Autocorrelation Functions

One of the Interpretation tools for identification of model is the autocorrelation and partial autocorrelation functions of time series which is presented and discussed.

$$r_k = \frac{c_k}{c_0} = \frac{\sum_{t=1}^{n-k}(x_t - \overline{x})(x_{t+k} - \overline{x})}{\sum_{t=1}^{n}(x_t - \overline{x})} \quad \text{and} \quad (19)$$

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\text{cov}(x_t, x_{t+k})}{\text{cov}(x_t, x_t)}$$

In which p (k) is the autocorrelation function and r_k is autocorrelation coefficient

$$p_k = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{k-1} & \cdot & \cdot & \cdots & 1 \end{bmatrix} \qquad \phi_{kk} = \frac{\left| p_k^* \right|}{\left| p_k \right|} \qquad (20)$$

And ϕ_{kk} is the partial autocorrelation function, p_k is autocorrelation matrix and $[p_k]$ is the same matrix in which the last column is changed with vector (p₁, p₂,…, pₖ)
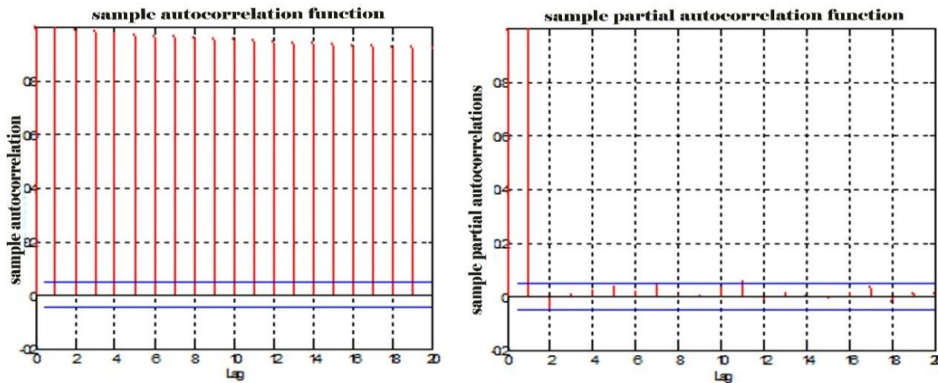
Figure 4: Autocorrelation and Partial Autocorrelation Functions Data of Ghadir Company
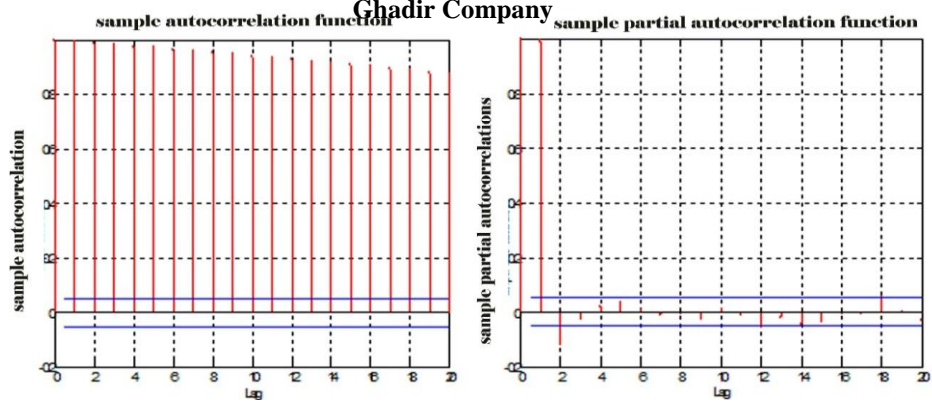
Figure 5: Autocorrelation and Partial Autocorrelation Functions Data of khodro Company
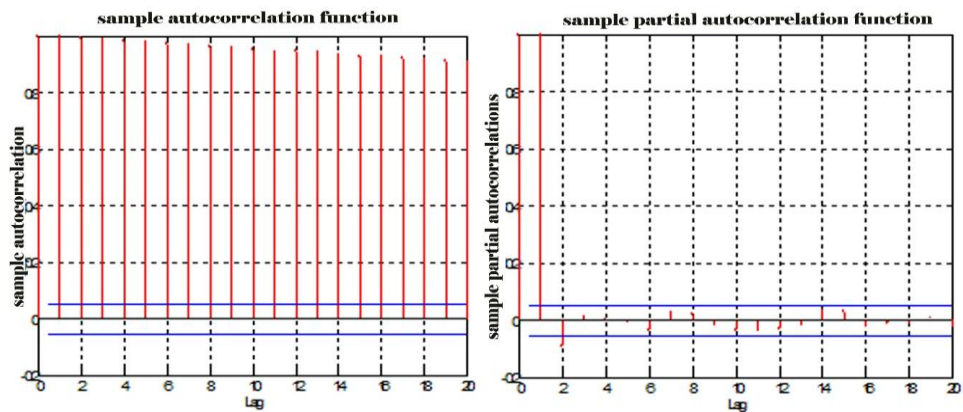
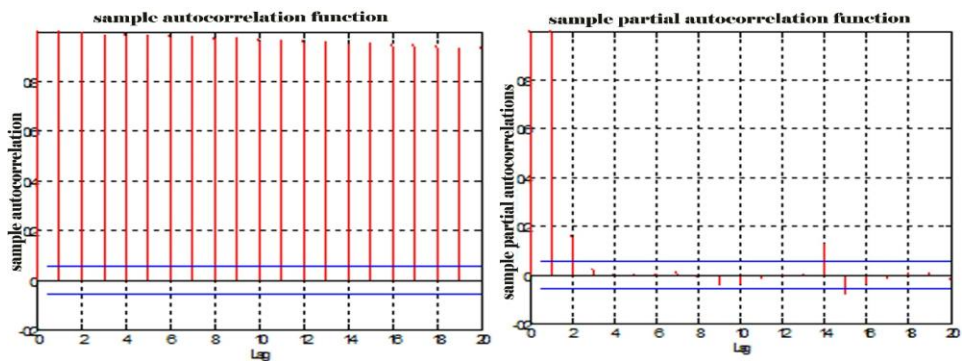**Figure 6: Autocorrelation and Partial autocorrelation functions data of roy Company**

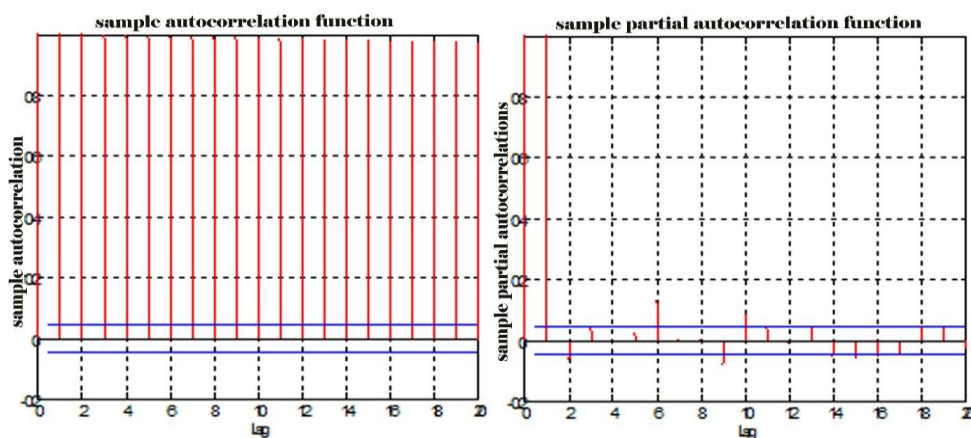**Figure 7: Autocorrelation and Partial Autocorrelation Functions Data of Saipa Company**

Figure 8: Autocorrelation and Partial Autocorrelation Functions Data of Siman Company

5-4- Pre- Processing Methods for Time Series Data

Financial time series demands preprocessing techniques due to features such as non-stationary, noise and nonlinearity .depending on data type, different methods are used to preprocess. One of the conditions of ARIMA model is the stationary mode of the data (either in terms of mean and variance). If the variance of data is non-stationary, by stability transformation, variance and the mean become stationary through differential.

5-5 sustainability Transformations of Variance

There are many time series which are stationary in the mean, but their variance is non-stationary. A sustainability transformation of Variance is necessary to solve these problems. Variance change of a non-stationary process is very common when the level changes. Important thing to mention here is that if the sustainability transformation of variance is necessary, it must come into force before any preprocessing such as differential.

Among five companies selected for this study, only data of Tehran Cement Company have unstable variance and requires sustainability transformations of Variance. Other companies have stationary variance.
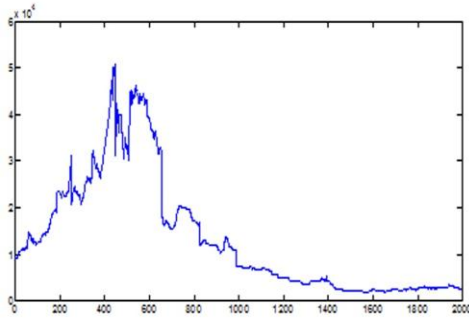
Figure 9: Main Data of Siman Tehran Company
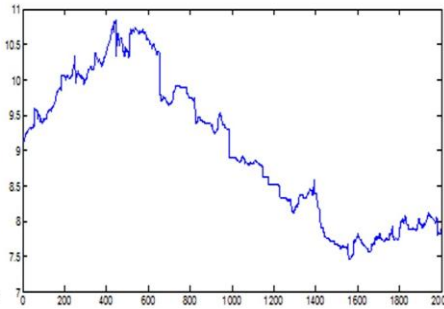
Figure 10: Data of Tehran Siman Company after Stabilization

In this company, regarding the fact that data variance is highly variable (over time), a logarithmic transformation was used to stabilize variance .after logarithmic transformation was used to stabilize the variance, it is observed that the data are non- stationary in the mean .to eliminate the problem , differential was used to make data stationary.

5-6- The Difference between Takings

One of the most common preprocessing techniques time series is differential used for making data stationary (on average). After taking the first difference of the time series, Original time series becomes:

$$Y_t = \nabla X_t = X_{t+1} - X_t \qquad (t = 1,2,....,n-1)$$

The following graphs show the subtracted data
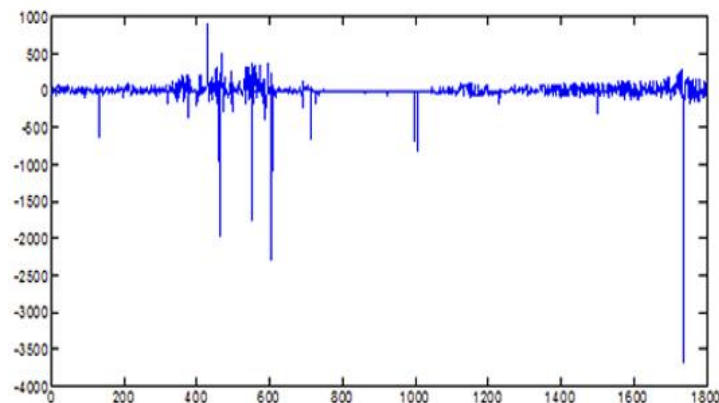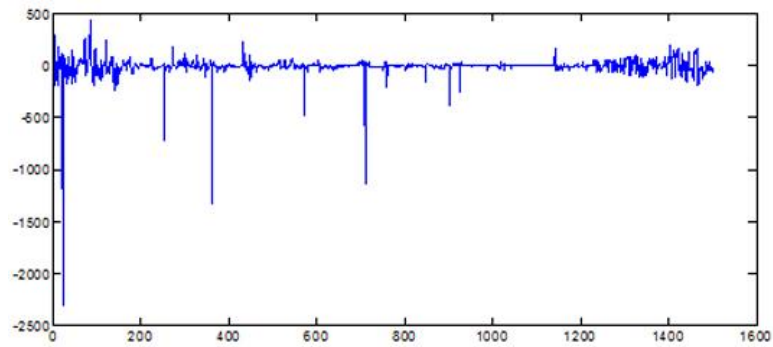
Figure 11: Subtracted data of Ghadir Company

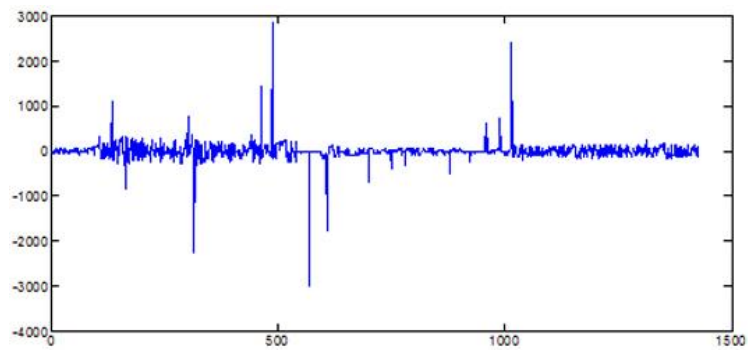**Figure 12: Subtracted Data of Khodro Company**

Figure 13: Subtracted data of Roy Company
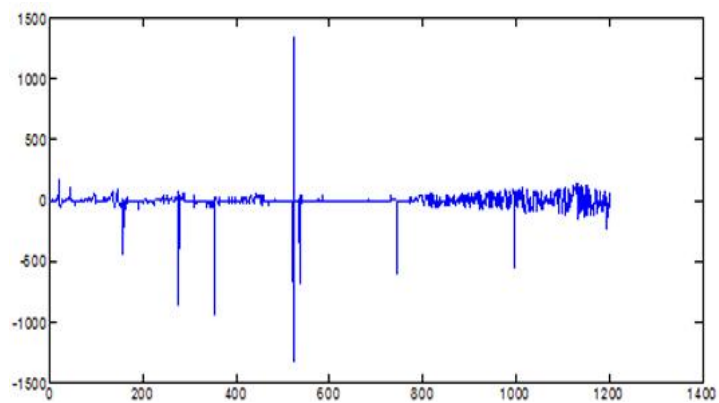

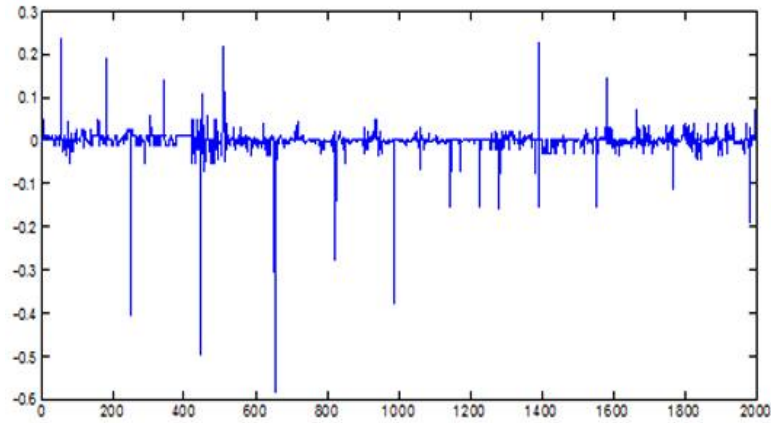**Figure 14: Subtracted Data of Saipa Company**

Figure 15: Subtracted Data of Siman Company

5-7- Autocorrelation and Partial Autocorrelation Functions after Stabilization and Subtracting

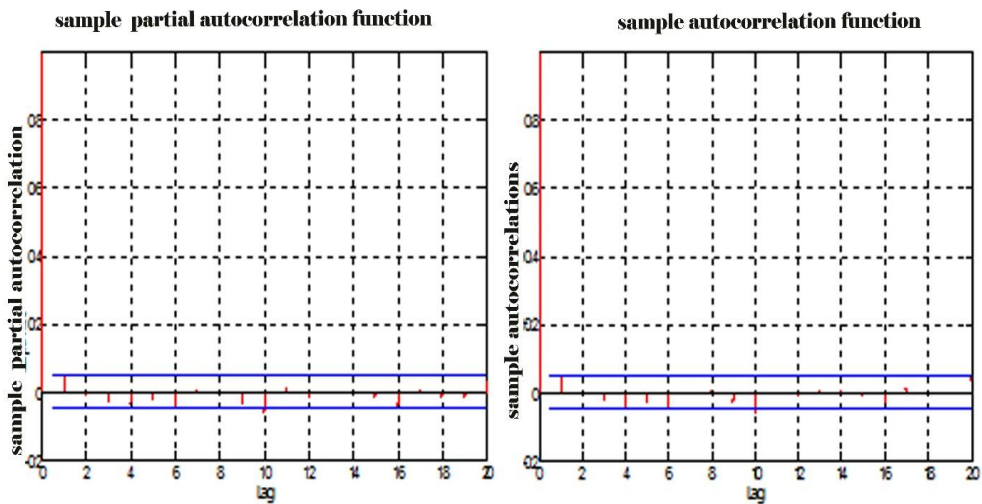After stabilization, partial autocorrelation and autocorrelation functions for stationary data are re-plotted:

Figure 16: Autocorrelation and Partial Autocorrelation Functions after Stabilization and Subtracting of Ghadir Company
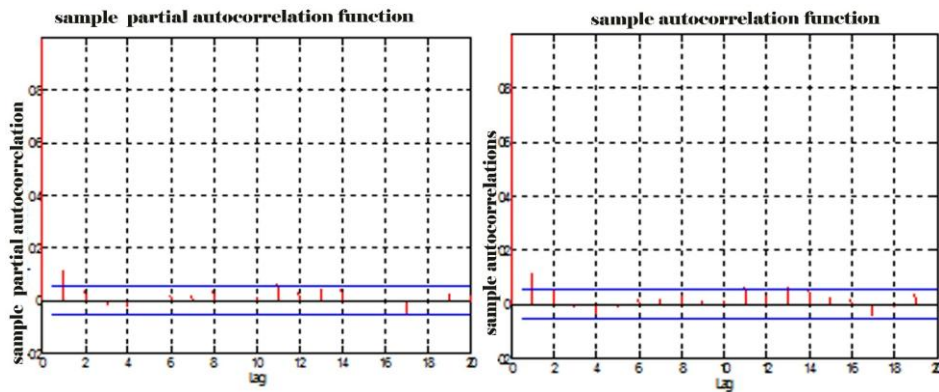
Figure 17: Autocorrelation and Partial Autocorrelation Functions after Stabilization and Subtracting of Khodro Company
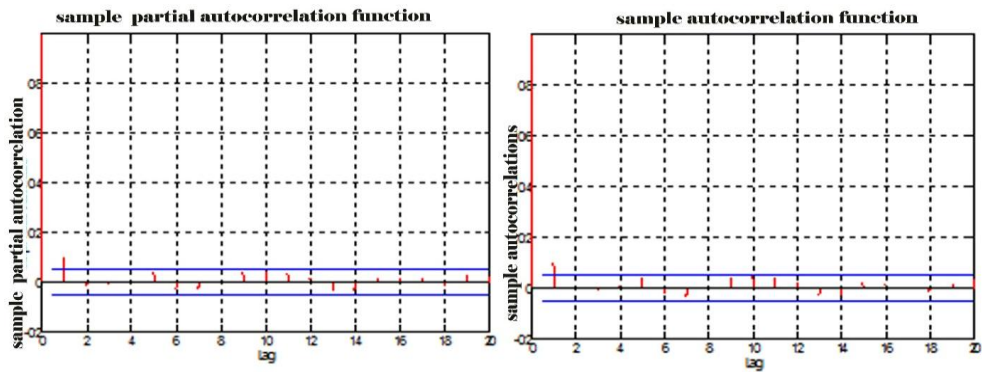

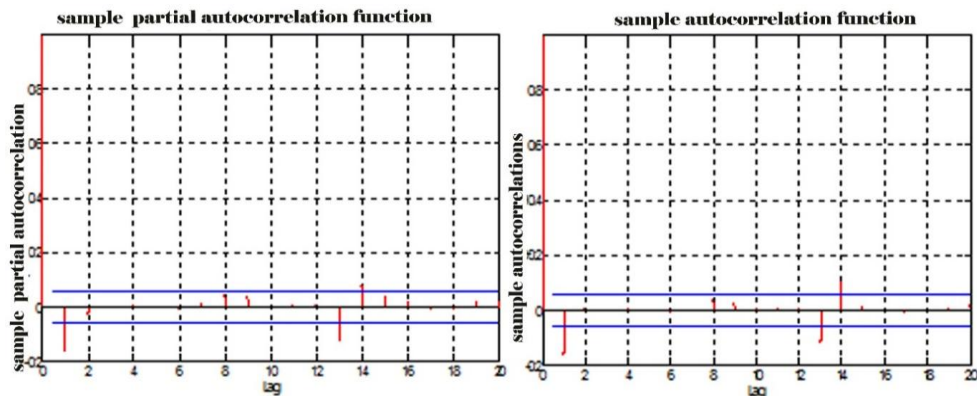**Figure 18: Autocorrelation and Partial Autocorrelation Functions after Stabilization and Subtracting of Roy Company**

Figure 19: Autocorrelation and Partial Autocorrelation Functions after Stabilization and Subtracting of Saipa Company
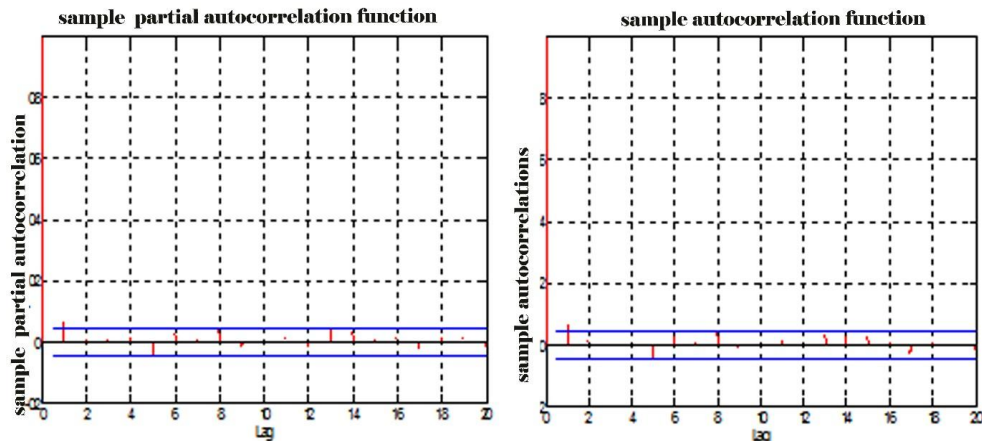
Figure 20: Autocorrelation and Partial Autocorrelation Functions after Stabilization and Subtracting of Siman Company

5-8- Detection of Model ARIMA (p, d, q)

Given the time-series graph and plotting the autocorrelation functions and partial autocorrelation, ARIMA (p, d, q) can be detected where d is the number of differentials and p and q are detectable with autocorrelation and partial autocorrelation. This means that the shape of the autocorrelation function plot after delay (q-p) tends to zero and the partial autocorrelation function of delay (p-q) tends towards zero.

In this study, because all data of company are stationary after one step of differential, for all five companies d = 1. After autocorrelation and partial autocorrelation functions of the observed data , it can be seen that with the exception of Ghadir Investment Company modelled ARIMA (0,1,0) model for other companies is ARIMA (1,1,1) .

5-9- Prediction of Stock Price of Companies Using SVR and Its Comparison with Arima
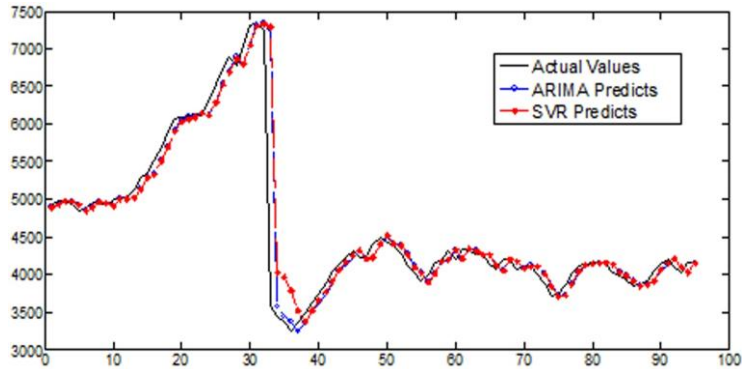

Figure 21: Prediction of Stock Price of Ghadir Company Using SVR and Its Comparison with Arima
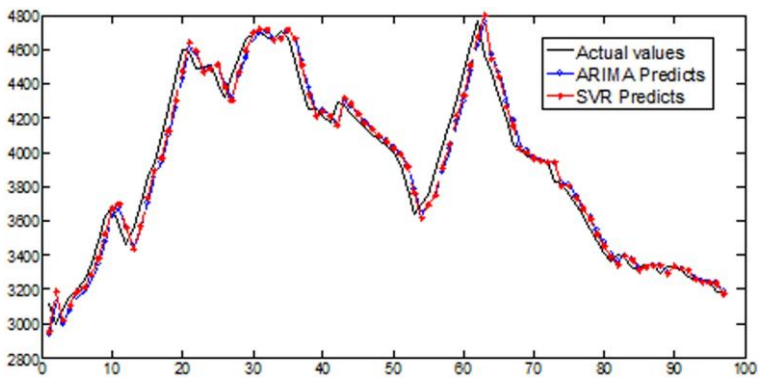

Figure 22: Prediction of Stock Price of KHODRO Company Using SVR and Its Comparison with Arima
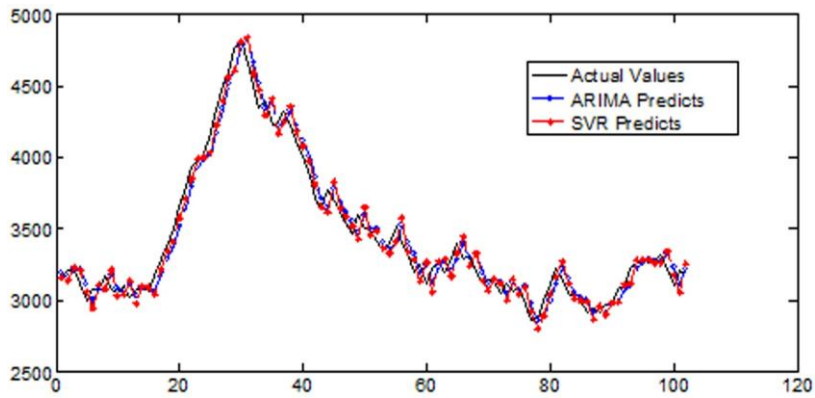
Figure 23: Prediction of Stock Price of Roy Company Using SVR and Its Comparison with Arima
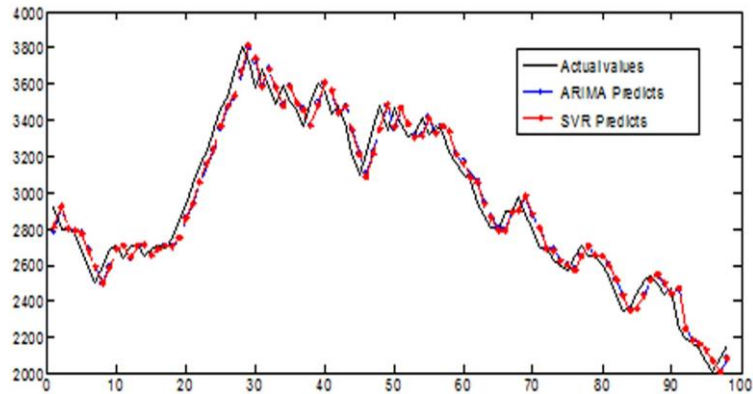
Figure 24: Prediction of Stock Price of Saipa Company Using SVR and Its Comparison with Arima
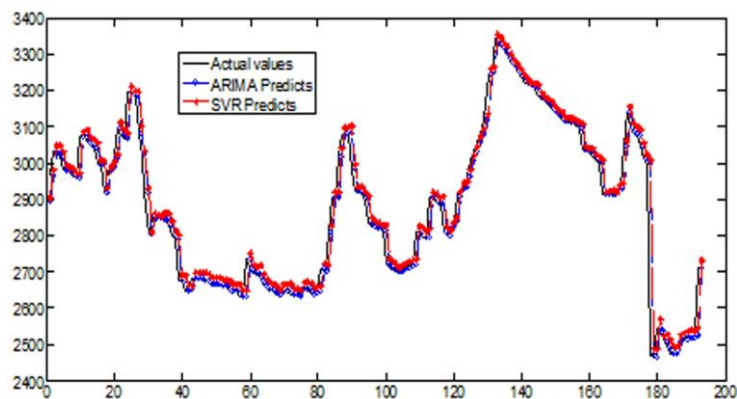
Figure 25: Prediction of Stock Price of Siman Company Using SVR and Its Comparison with Arima

5-10- performance evaluation criteria

Due to information limitations with a single criterion, the researchers use various criteria to evaluate network performance. For prediction problems, these criteria are related to differential between predicted output and real outputs. These criteria are used to investigate a prediction model or selection of the best model among different models by which proper decisions are made about acceptance or refusal of prediction model.

Overall, the closer actual amount is to the value predicted by the model, more accurate the prediction model is. Therefore, the model quality is assessed by the expected error rate. In this regard, three factors mean square error (MSE), mean absolute error (MAE) and the normalized mean square error (NMSE) will be looked into.

$$MSE = \frac{1}{l}\sum_{t=1}^{l}(a_t - p_t)^2$$

$$MAE = \frac{1}{l}\sum_{t=1}^{l}|(a_t - p_t)|$$

$$NMSE = \frac{1}{l\sigma^2}\Sigma(a_t - p_t)^2 \qquad \sigma^2 = \frac{1}{l-1}\sum_{t=1}^{l}(a_t - \overline{a})^2 \qquad \overline{a} = \frac{1}{l}\sum_{t=1}^{l}a_t$$

In which l is the number of testing data, a the real value in time t and p is the predicted value in time t.

Table 3: Prediction of Stock Price of Ghadir Company Using SVR and Its Comparison with Arima

GHADIR	MSE	MAE	NMSE
SVR	167284	142.2967	0.1717
ARIMA(1,1,1)	154.670	128.2211	0.1588

Table 4: Prediction of Stock Price of Khodro Company Using SVR and its comparison with Arima

KHODRO	MSE	MAE	NMSE
SVR	6959.18	66.6487	0.0278
ARIMA(1,1,1)	8659	74.9038	0.0345

Table 5: Prediction of Stock Price of Roy Company Using SVR and Its Comparison with Arima

ROY	MSE	MAE	NMSE
SVR	8415.1	77.916	0.0358
ARIMA(1,1,1)	8806.5	81.5858	0.0374

Table 6: Prediction of Stock Price of SAIPA Company Using SVR and Its Comparison with Arima

SAIPA	MSE	MAE	NMSE

SVR	7090.64	73.4283	0.0348
ARIMA(1,1,1)	7263.7	74.4982	0.0357

Table 7: Prediction of stock price of SIMAN company using SVR and its comparison with Arima

SIMAN	MSE	MAE	NMSE
SVM	3360.8	34.3946	0.0705
ARIMA(1,1,1)	3193.8	26.9102	0.067

Comparing the results predicted based on the criteria of MSE, MAE and NMSE in companies of Iran Khodro and Saipa indicates the superiority of the SVR method to ARIMA. But for Ghadir Investment Company and cement Tehran, ARIMA model forecasts better than SVR model. this is due to the fact that the data of these two companies are linear in some periods and for linear data , Arima model outperforms .Obviously, the model outperforms better in its own condition .For example ,in ARIMA models where the data are mostly linear, better results can be achieved.

6- Conclusions

Prediction of stock prices in the stock market is of utmost importance. Investors in stock market can gain profit if they can predict the trend of stock market regarding the changes in other markets such as exchange, gold and oil. In fact, when investors gain profit from the stock market, the market will provide life survival.

In recent years, the increase of computation power and data storage in different fields has brought about a necessary provision for new techniques and smart tools .data mining is a relatively young field originated from statistics, artificial intelligence and machine learning referring to extracting knowledge and information from big database.one of the most important applications of data mining is making prediction based on previous data done in different ways.

Support vector machine (SVM) is one of powerful new techniques for data mining, which has the successful application in classification problems, regression and time series prediction. These methods have attracted the attention of many researchers due to prominent features such as simple geometric interpretation, its general and unique solution to achieve and capability of modelling the complex nonlinear functions

In this study, we compare the classical regression model with SVR and high performance of computing and the results indicate that the SVR model outperforms the other. Of course, this does not mean ignoring other methods, and each model has its own efficiency. But Tehran Stock market should take advantage of the different methods for the implementation experience of the other markets, such as the New York and japan Stock Exchange.

All can be concluded is that using mathematical model in estimating stock price is of great importance.

Among all kinds of time series data, forecasting financial time series, due to characteristics such as non-stationary, noise and non-linearity, is more difficult. In this study, the performance of SVR for financial time series prediction has been investigated. To do so, five sets of data related to ghadir ,irankhodro, saipa , siman Tehran and rooy toseh zanjan were selected.to select model parameters ,we used network search method.

After implementation SVR model, the predicted results the classical were compared with ARIMA model.

Given Autocorrelation and partial autocorrelation functions, Non-stationary data were specified and the need to stabilize the variance and the difference was felt. a random walk model was used for ghadir data and for irankhodro, Saipa , Tehran cement and rooy toseh ,we used an ARIMA model. And the predicted results were compared. based on the criteria of MAE, MSE and NMSE The results of this study showed that in the data of companies of irankhodro,toseh rooy ,and saipa , and Saipa, SVR model is better than the ARIMA, based on criteria of MSE MAE and NMSE . But given Tehran cement and Ghadir, ARIMA model showed more accurate predictions. Then selection the appropriate model from existing models is one of the most important factors in understanding the stock market.

As in most developed countries such as japan and USA, SVR and other mathematical models model are used to predict stock market, some suggestions are presented below to establish a robust system in stock market and improve the accuracy of SVR model.

1) Education of the investors in stock market for the use of mathematical models to predict the stock price

2) Using experience in other stock markets in which non-linear models are used to predict stock prices.

3) Finding a more accurate and faster method for selecting model parameters of SVR.

4) Combining SVR approach with other methods and using appropriate pre-processing methods to increase the accuracy of the model.

5) Finding new kernel functions to be applied in various data.

References

1- Basak, D. et al. (2007), Suort Vector Regression, Neural Information Processing- Letters and Reviews, 11, No. 10, 203-223.

2- Boser, E, Guyon, I, Vapnik, V, A training algorithm for optimal margin classifiers. In Haussler, David (editor); 5th Annual ACM Workshop on COLT, pages 144–152

3- Box G. E. P. and Jenkins G. M. (1976), Time Series Analysis: Forecasting and Control. Revised Edition. Holden-Day.

4- Brillinger. (1981), Time series Data Analysis and Theory Expanded Edition. McGraw-Hill. Inc. New York.

5- Cao, L. J.and Tay, F. E.H. (2001), Financial Forecasting using Support Vector Machines, Neural Comput Applict, 10, 184-192.

6- Cao , L. J. and Tay, F. E.H. (2003), Support Vector Machine with Adaptive Parameters in Financial Time Series Forecasting, IEEE Transactions on Neural Networks, vol. 14, No. 6, PP.1506-1518

7- Chattfield C. (1975), the Analysis of Time Series. An introduction. New York, Chapman and Hall.

8- Chen, W, Shih, j-y. (2006), Comparison of Support Vector Machines and Back Propagation Neural Networks in Forecasting the Six Major Asian Stock Markets, International Journal of Electronic Finance,. 1, No. 1, 49, 49-67.

9- Dai, W, shao, Y, lu, C, (2013), Incorporating feature selection method into support vector regression for stock index forecasting, Neural Computing and Applications, Volume 23, Issue 6, pp 1551-1561

10-Engle, R. (1982), Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation, Economical, 50, Issue 4, 987-1008.

11-Engle, R. (2001), the Use of ARCH/GARCH Models in Allied Econometrics, Journal of Economic Perspective, 15, 4, 157-168.

12-Feng pai, P, sheng lin, C, (2005), A hybrid ARIMA and support vector machines model in stock price forecasting, Omega, Volume 33, Issue 6, Pages 497–505

13-Han, J. and kamber, M, (2006), data mining concepts and techniques, second editition, Elsevier, morgan koufmann peblishers.

14-Hong, W, dong, Y, zhang, W, chen, L, Panigrahi, B, (2013), Cyclic electric load forecasting by seasonal SVR with chaotic genetic algorithm, International Journal of Electrical Power & Energy Systems, Volume 44, Issue 1, Pages 604–614

15-Hung, C, tsai, C, (2009), A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting, Expert Systems with Applications, Volume 36, Issue 2, Part 1, Pages 1529–1539

16-Ince, H. and Trafalis, T. B. (2006), A Hybrid Model for Exchange Rate Prediction, Decision Suort Systems, 42, 1054-1062.

17-Lu, C, wu, J, chiu, C, tsai, Y, (2011), Predicting Stock Index Using an Integrated Model of NLICA, SVR and PSO, Advances in Neural Networks, Volume 6677, pp 228-237

18-Smola, A. and Scholkopf, B. (1998), A Tutorial on Support Vector Regression, Technical Report NeuroCOLT NC-TR-98-030.

19-Tan, P. N., Steinbach, M. and Kumar, V. (2006). Introduction to Datamining. Pearson Education, Inc., Boston, MA.

20-Tay, F. E. H. and Cao, L. J. (2001), Application of Support Vector Machines in Financial Time Series Forecasting, The International Journal of Management Science, Omega, 29, PP. 309-317.

21-Vapnik, V. N. (1995), the Nature of Statistical Learning Theory, New York: Springer-Verlag.

22-Vapnik, V.N. (1998), Statistical Learning Theory. Wiley, New York.

23-Xiong, T, bao, Y, hu, Z, (2013), Multiple-output support vector regression with a firefly algorithm for interval-valued stock price index forecasting, Knowledge-Based Systems, Volume 55, Pages 87–100

24-Yue xu, S, (2012), Stock Price Forecasting Using Information from Yahoo Finance and Google Trend, Department of Economics University of California, Berkeley, pages 1-22.