Professor Cătălina-Lucia COCIANU, PhD
E-mail: Catalina.Cocianu@ie.ase.ro
The Bucharest University of Economic Studies
Hakob GRIGORYAN, PhD Student
E-mail: Grigoryanhakob90@yahoo.com
The Bucharest University of Economic Studies

MACHINE LEARNING TECHNIQUES FOR STOCK MARKET PREDICTION.ACASE STUDY OF OMV PETROM

Abstract. *The research reported in the paper focuses on the stock market prediction problem, the main aim being the development of a methodology to forecast the OMV Petrom stock closing price. The methodology is based on some novel variable selection methods and an analysis of neural network and support vector machines based prediction models. Also, a hybrid approach which combines the use of the variables derived from technical and fundamental analysis of stock market indicators in order to improve prediction results of the proposed approaches is reported in this paper. Two novel variable selection methods are used to optimize the performance of prediction models. In order to identify the most informative time series to predict a stock price, both methods are essentially based on the general forecasting error minimization when a certain stock price is expressed exclusively in terms of other indicators. After the variable selection is over, the forecasting is performed in terms of the historical values of the given stock price and selected variables respectively. The performance of the proposed methodology is evaluated by a long series of tests, the results being very encouraging as compared to similar developments.*

Keywords: *Machine learning; Artificial neural network;Nonlinear autoregressive with exogenous input;Support vector regression; Financial data forecasting; Clustering.*

JEL classification: C02, C14, C19, C45, C49, C61

1. Introduction

Financial markets are complex, non-stationary, and volatile systems conditioned by chaotic nature of stock data [1]. Stock market forecasting is one of the most complicated issues of time series analysis. The development of an accurate prediction models plays an important role in the design of effective trading strategies.

Recently, machine learning techniques have been successfully introduced into the field of financial time series analysis in order to help investors make qualitative decision in stock market forecasting. Artificial neural networks (ANN) have become one of the most popular and useful machine learning techniques for time series prediction due to their ability to deal with noisy and unstable data. In the area of financial data predictions, White (1988)was among the first scholars who introduced an artificial neural network based model for economic data prediction [2]. Furthermore, in [3] the potential of neural networks in forecasting of stock market prices was examined, showing promising experimental results. A combined method based on backpropagation neural network (BPNN) for Japanese stock market prediction was proposed in [4]. In addition, a series of studies carrying out comparative analyses experimentally proved the advantages of ANN based forecasting models against traditional statistical models. However, neural networks have some limitations, including overfitting problem, selection of many controlling parameters, beside ANNs require more training data and time. [5,6]

In the late 1990's, a novel technique called Support Vector Machines (SVM) has been introduced to solve non-linear regression problem in time series analysis. Unlike neural networks, for achieving generalized performance, the noted techniques attempt to minimize the generalized error bound instead of minimizing the training error [7]. Because of its good generalization capability, the Support Vector Regression (SVR) has been extensively applied in time series analysis and showed promising results. A SVM-based approach for financial data forecasting was proposed in [8]. The experimental analysis proved that SVM outperforms the BPNN from the point of view of prediction evaluation criteria. Also, studies aiming the forecasting of the stock price indices and their movement directions using SVM experimentally established that SVMs generally perform better than other forecasting methods [9,10].

In the past decade, long series of research had been carried out aiming to hybridize machine learning methods with feature selection techniques and econometric analysis tools in order to improve prediction accuracy have been reported. There is a tremendous amount of work done in the field of financial time series analysis that demonstrates the effectiveness of the use of combined artificial intelligence techniques for prediction task [11-14]. Nonetheless, there are still opportunities to improve the existing models and increase performances.

This research focuses mainly on the stock market prediction problem. The main goal is to develop a methodology to forecast the OMV Petrom stock closing price. Our proposed methodology is based on some novel variable selection methods and analysis of neural network and support vector machines based models for prediction of stock market prices. This paper presents a hybrid approach combining the use of the variables derived from technical and fundamental analysis of stock market indicators aiming to improve prediction results of the proposed approaches. Two novel variable selection methods are used to optimize the performance of prediction models. In order to identify the most informative time series to predict a stock price, both methods are essentially based on the general forecasting error minimization when a certain stock price, *Y*, is expressed exclusively in terms of other indicators (variables and/or stock prices). Also, the former proposed method includes cross-correlation analysis whereas the second one is derived from the cluster analysis and cross-correlation coefficients. After the variable selection is over, the forecasting of *Y* is performed in terms of the historical values of *Y* and selected variables respectively. Also, we had to analyze the performances of the forecasting model using both stock data and sets of variables obtained from technical and fundamental analysis against the forecasting model that uses only the stock closing price historical values to predict the closing price at the next moment of time. The last model is obviously the simplest one and yet very useful in many real world applications. In many cases it proved better results that much sophisticated models [15]. In our case the former model proved better results.

The rest of the paper is organized as follows. In Section 2 the methodology used in the research is discussed briefly. Data acquisition and preprocessing are presented in Section 3, while the general prediction model and two of the most intensively used machine learning-based techniques for data forecasting are exposed in the fourth section of the paper. In Section 5 the experimental results of the proposed methodology together with a comparative analysis are presented. Finally, the concluding remarks are given in Section 6.

2. Proposed methodology

This paper presents a three-stages architecture to develop a hybrid prediction model for stock exchange market.

The first stage is data preprocessing. Data preprocessing consists of the following steps. The technical indicators based on historical data are computed in the first step. Then we apply a data normalization technique to normalize data into one scale. The aim of the final step of is to choose the key variables for the input data in the model. We develop two different variable selection techniques based on the general forecasting error minimization when a certain stock price, *Y*, is expressed exclusively in terms of other indicators. The former technique includes also the cross-correlation method and the second one is based on a clustering process.

In the next stage, we present the general prediction model and two powerful forecasting techniques to deal with it, namely Nonlinear Autoregressive with eXogenous input (NARX) neural networks and support vector machines.

In the third stage, we evaluate the results achieved from the experiments based on different models and as well as comparing the obtained results.

Also, the analysis of the performances of the forecasting model using both OMV stock data and sets of variables obtained from technical and fundamental analysis against the forecasting model that uses only the stock closing price historical values to predict the closing price at the next moment of time is presented in the final part of the paper. Each stage is detailed in the following sections.

3. Data acquisition and preprocessing

3.1 Research data

The data used in this study is the historical stock prices taken from the Bucharest stock exchange (BVB). The whole data set covers the period from March 9, 2008 to November 30, 2014, a total of 350 weekly observations. The stock data consists of weekly observations of stock opening, closing, lowest, highest prices, and traded volume of OMV Petrom shares (symbol OMV). In addition, technical and fundamental analyses were used to accurately choose indicators that influence stock price's behavior. In this study 5 variables from fundamental analysis and 30 variables from technical analysis of the stock market were selected. OMV Petrom stock closing price was used as a forecasting variable.

3.2 Fundamental and Technical analysis

Fundamental analysis attempts to examine a variety of factors that could alter security's value, including global and market based macroeconomic and industry specific factors. This research includes 5 indicators obtained from fundamental analysis of market data. In contrast to fundamental analysis which examines market related economic, financial and other qualitative and quantitative factors, technical analysis examines only statistics from past market data, such as price and volume to determine market trends. Technical analysis is a set of techniques for prediction the stock price movements by analyzing the past sequence of stock prices. The main goal of this technique is to identify regularities in the stock data by extracting nonlinear patterns from noisy data. In this research we used 30 technical indicators. The complete list of technical indicators, as well as stock-based variables and fundamental analysis-based indicators are given in Table 1.

In the following, we assume that $\mathrm{Y_t}$ is the OMV stock closing value at the moment of time *t*. We consider $1 \leq t \leq T$. For each *t*, we denote by $\mathrm{XT_t} = \left(\mathrm{XT_t}(1), \mathrm{XT_t}(2), \dots, \mathrm{XT_t}(N)\right)^T$ the vector whose entries are the values of the following indicators: OMV stock opening, lowest, highest prices, and traded volume (stock-based indicators), 5 variables resulted from fundamental analysis and 30 variables obtained from technical analysis of the stock market. Note that $\mathrm{XT} = \left(\mathrm{XT}(1),\ \mathrm{XT}(2), \dots,\ \mathrm{XT}(N)\right)^T$ is a vector, each entry $XT(i)$ being the time series corresponding to the *i*th variable. In our case $N = 39$.

3.3. Data normalization

As the collected data samples have different scales with different values, it is necessary to normalize the time series prior to training step. The most commonly used approach for data normalization purposes is min-max method that normalizes the values of an attribute according to its minimum and maximum values. In our research, the data normalization range is chosen to be [0,1], and the equation for data normalization is given by,

$$V = \frac{v - v_{min}}{v_{max} - v_{min}} \tag{1}$$

where V is the normalized data, v is the original data, $v_{max} and v_{min}$ are maximum and minimum values of v.

3.4. Variable selection

Variable selection is the process of selecting input variables for the use in model construction in order to simplify training step. This method identifies a small subset of variables that provide the most important information about the given data. The selected variables minimize the generalization error and shorten training time.

The most commonly used selection process is exclusively based on the cross-correlation values and is described as follows. Let Tr be a given threshold value. The variable $XT(i)$ is selected as an input variable if the cross-correlation coefficient between $XT(i)$ and Y is bigger than Tr.

$$r_{XT(i),Y} = \frac{\sum_{t=1}^{T}(XT_t(i) - \overline{XT(i)})(Y_t - \bar{Y})}{\sqrt{\sum_{t=1}^{T}(XT_t(i) - \overline{XT(i)})^2 \sum_{t=1}^{T}(Y_t - \bar{Y})^2}} \tag{2}$$

$$r_{XT(i),Y} > Tr$$

Where:

$$\overline{XT(i)} = \frac{1}{T}\sum_{t=1}^{T} XT_t(i), \bar{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t$$

In order to improve the prediction accuracy, we proposed two novel approaches in variable selection process.

Selection method V1. The set $\mathcal{S}_f$ of variables $X(i),\ 1 \leq i \leq n, n \leq N$, components of XT, are selected as inputs if each cross-correlation coefficient between $X(i)$ and Y is bigger than a threshold $Tr, Tr_{min} \leq Tr \leq Tr_{max} < 1$,

$$r_{XT(i),Y} > Tr \tag{3}$$

and the general forecasting error of the model

$$\widehat{\mathrm{Y}}_{\mathrm{t}} = g\big(X_t(1), \dots, X_t(n)\big) \tag{4}$$

is minimized with respect to *Tr* on a certain set of possible values $TrS \subset [Tr_{min}, Tr_{max}]$.

The general forecasting error is expressed in terms of root mean squared error (RMSE), defined by:

$$RMSE(T,P) = \sqrt{\frac{1}{nr}\sum_{i=1}^{nr}\big(T(i) - P(i)\big)^2} \tag{5}$$

where $T = \big(T(1), T(2), \dots, T(nr)\big)$ is the vector of target values, $P = \big(P(1), P(2), \dots, P(nr)\big)$ is the vector of predicted values and nr is the number of data samples.

The idea behind this selection process is to determine $\mathcal{S}_f$, the most suitable subset of variables strongly enough correlated to Y, from the point of view of its closing price prediction capacity in terms of (4). In this case, the prediction capacity of (4) when $\mathcal{S}_f$ is the subset of selected variables is a measure of the influence of each element belonging to $\mathcal{S}_f$on the closing price *Y*.

Note that the set TrS can be established based on the computed cross-correlation values (in our case those displayed in Table 1).

Selection method V2. Another way to select a subset of key input variables is based on the following procedure.

- Apply a clustering method (for instance *k*-means) to the variable set and select the cluster $\mathcal{C} = \big\{X(1), \dots, X\big(n(\mathcal{C})\big)\big\}$such that the RMSE prediction error of the model (6) is minimized

$$\widehat{\mathrm{Y}}_{\mathrm{t}} = g\left(X_t(1), \dots, X_t\big(n(\mathcal{C})\big)\right) \tag{6}$$

- For further optimization, select a subset $\mathcal{S}_\mathcal{C}$ of $\mathcal{C}$ having the following properties: $x\epsilon\mathcal{S}_\mathcal{C}$ if and only if the cross-correlation between $\mathcal{C} \setminus \{x\}$ and Y and cross-correlation between $\mathcal{C}$ and Y are significantly different. The cross-correlation between a set of random variables $\mathcal{S} = \{V_1, \dots, V_n\}$ and a random variable Y is given by

$$r_{\mathcal{S},Y} = d^T(R_\mathcal{S})^+ d \tag{7}$$

Where:

$$d(i) = \frac{E\big((V_i - \overline{V_i})(Y - \overline{Y})\big)}{\sqrt{E((V_i - \overline{V_i})^2)E((Y - \overline{Y})^2)}}$$

$$R_{\mathcal{S}}(i,j) = \frac{E\left((V_i - \overline{V_i})(V_j - \overline{V_j})\right)}{\sqrt{E((V_i - \overline{V_i})^2)E\left((V_j - \overline{V_j})^2\right)}}$$

In terms of (7), $x \epsilon \mathcal{S}_{\mathcal{C}}$ if and only if

$$\left|r_{\mathcal{C}\setminus\{x\},Y} - r_{\mathcal{C},Y}\right| > \varepsilon \tag{8}$$

Note that the cluster $\mathcal{C}$ includes the variables strongly correlated to Y and its corresponding general error with respect to (6) is minimal on the cluster sets.

4. Machine learning-based models for data forecasting

4.1. The general forecasting model

In order to develop the general forecasting model, we used Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) to establish whether the time series are stationary or not.

In case ACF decays very slowly, the considered time series are non-stationary. In the following we consider the general model, when the time series are non-stationary. Let *d* be such that, for all considered variables, PACF function drops immediately after the *d*th lag. This means that the delay should be set to *d*.

We considered the general forecasting model, where both OMV stock data and sets of variables obtained from technical and fundamental analysis are taking into account. The model is described as follows:

$$\hat{\mathrm{Y}}_{(\mathrm{t+p})} = \mathrm{f}\left(Y_t^{(d)}, X_t^{(d)}\right) \tag{9}$$

$$Y_t^{(d)} = \{\mathrm{Y_t}, \mathrm{Y_{t-1}}, \mathrm{Y_{t-2}}, \dots, \mathrm{Y_{t-d+1}}\} \tag{10}$$

where:

$$X_t^{(d)} = \{\mathrm{X_t}, \mathrm{X_{t-1}}, \mathrm{X_{t-2}}, \dots, \mathrm{X_{t-d+1}}\} \tag{11}$$

$$\mathrm{X_t} = \left(\mathrm{X_t}(1), \mathrm{X_t}(2), \dots, \mathrm{X_t}(n)\right)^T \tag{12}$$

and $\mathrm{X_t}(1), \mathrm{X_t}(2), \dots, \mathrm{X_t}(n)$ are the selected components of $\mathrm{XT_t}$.

4.2. NARX neural networks

Artificial neural networks (ANNs) are non-parametric methods which are able to approximate a large class of functions with a high degree of accuracy. An important class of dynamic recurrent neural networks is the Nonlinear Autoregressive with eXogenous input (NARX) model.

The NARX network is a dynamical neural architecture used for input-output modeling of nonlinear dynamical systems. When applied to time series forecasting, the NARX network is designed as a feedforward Time Delay Neural Network and consists of a Multilayer Perceptron which takes as input a window of past input and output values and computes a prediction of the current output value. [16]

NARX networks are well suited for modelling non-linear time series as well as for multi-step ahead prediction. One of the most commonly used NARX architecture consists of an input layer,F_X, a hidden layer, F_H and an output layer, F_Y, each of which is connected to the other. The architecture of the three-layered NARX model is illustrated in Figure 1.

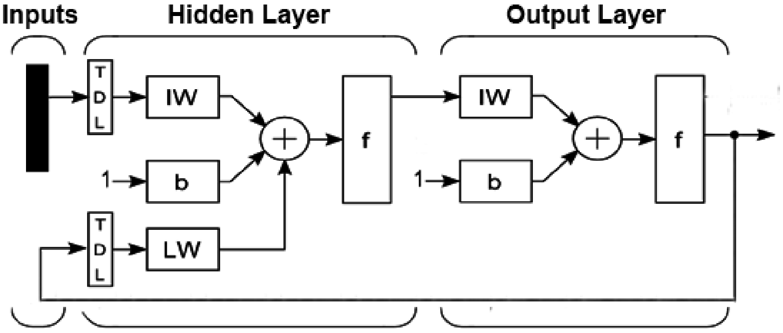

Figure 1. NARX network architecture

In the NARX model, the previous values of an independent input signal are used to predict the future value of output signal which can be mathematically represented as

$$y(n+1) = f\big[y(n), y(n-1), \dots, y\big(n-t_y+1\big), u(n-k), \\ u(n-k+1), \dots, u(n-t_u-k+1)\big] \quad (13)$$

where$y(n+1)$ is the next value of the dependent output variable y, and u is externally determined vector of variables that influence y at time n. The function$f$ is the mapping performed by a Multilayer Perceptron, that estimates the next value of $\{y(t)\}$, while $t_u \geq 1, t_y \geq 1, t_u \leq t_y$ are the input-memory and output-memory orders, respectively, and parameter k is a delay term, where $k \geq 0$.

By assuming that $k=0$, we get the standard version of NARX model

$$y(n+1) = f\big[y(n), y(n-1), \dots, y\big(n-t_y+1\big), u(n), \dots, u(n-t_u+1)\big](14)$$

The terms $u(n), u(n-1), \dots, u(n-t_u-1)$ are the exogenous inputs produced with an input delay line with memory of order t_u and $y(n), y(n-1), \dots,$ $y\big(n-t_y+1\big)$ are the endogenous inputs produced with a delay memory line of order t_y.

The training phase in NARX model can be carried out either in series-parallel mode or in parallel mode.

In series-parallel mode, the output's regressor consists only of the actual values of the system's output and can be computed by:

$$\hat{y}_{SP}(n+1) = \hat{f}\left[y(n), \dots, y\left(n - t_y + 1\right); u(n), \dots, u(n - t_u + 1)\right] \qquad (15)$$

In parallel mode, estimated outputs are fed back and included in the output's regressor.

$$\hat{y}_P(n+1) = \hat{f}\left[\hat{y}(n), \dots, \hat{y}\left(n - t_y + 1\right); u(n), \dots, u(n - t_u + 1)\right] (16)$$

The three-layered NARX network we used in forecasting model (9) for $p = 1$ is defined in terms of (15), where $d = t_y = t_u$.

The size of the hidden layer of NARX can be computed in many ways, some of the most frequently expressions being [17]

$$|F_H| = 2\left[\sqrt{(|F_Y| + 2)|F_X|}\right] \qquad (17)$$

where$|F_X|$ and $|F_Y|$ are the sizes of the input layer and the output layer respectively.

The activation functions of the neurons belonging to the hidden and output layers can be selected from a very large family. In our work, we considered the logistic type to model the activation functions of the neurons belonging to the hidden layers, and the unit functions to model the outputs of the neurons belonging to the output layers.

The training of the NARX architecture is of supervised type using a gradient descent approach. In our work the local memories of F_H and F_Y are determined using theLevenberg-Marquardt (LM) variant of the backpropagation learning algorithm. The LM algorithm is one of the most widely used and efficient optimization algorithm which provides a numerical solution of nonlinear least squares minimization problem. It is based on Newton optimization method, where the iterative process is of hill-climbing type and the computation of the updating values is computed based on Singular Value Decomposition technique. [18]

4.3. Support Vector Regression

Support Vector Machines (SVMs) are a relatively new supervised learning techniques based on the structured risk minimization (SRM) principle [19,20]. In contrast with neural networks, **S**VMs seek to minimize an upper bound of the generalization error instead of minimizing the observed training error.

Let $G = \{(x_i, y_i),\ i = 1, \dots, l\} \subset \mathcal{S} \times \mathbb{R}$be a given training data, where $\mathcal{S}$ denotes the space of the input patterns. In the following we consider $\mathcal{S} \subset \mathbb{R}^n$. The objective of the support vector regression (SVR) is to find a function *f* that, on one hand, has at most ε deviation from the target y_i and, on the other hand, *f* is as flat as possible.

In case of linear functions, *f* is given by [20]

$$f(x) = \langle w, x\rangle + b = w^T x + b \tag{18}$$

where,$w \in X$ is a weight vector,$b \in \mathbb{R}$ is a bias. In this case flatness means small values of the Euclidian norm of w, $\|w\|^2$. Consequently, the problem can be formulated as a convex optimization problem,

$$minimize \frac{1}{2}\|w\|^2$$
$$subjectto \begin{cases} y_i - w^T x_i - b \le \varepsilon \\ w^T x_i + b - y_i \le \varepsilon \end{cases} \tag{19}$$

If there exists f such that $|f(x_i) - y_i| \le \varepsilon$ for all $(x_i, y_i) \in G$, the convex optimization problem (19) is feasible. Otherwise, slack variables ξ_i, ξ_i^* should be considered to deal with unfeasible constraints of the optimization problem (19). Hence we obtain the following optimization problem [20],

$$minimize \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)$$
$$subjectto \begin{cases} y_i - w^T x_i - b \le \varepsilon + \xi_i \\ w^T x_i + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{20}$$

where $C > 0$ is a regularization constant expressing the trade-off between flatness of f and the amount of deviation larger than ε that is tolerated.

In order to solve the optimization problem (20), one considers its corresponding dual problem, using the standard Lagrange multipliers method. Let L be the Lagrange function

$$L = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) - \sum_{i=1}^{l}\alpha_i(\varepsilon + \xi_i - y_i + w^T x_i + b)$$
$$-\sum_{i=1}^{l}\alpha_i^*(\varepsilon + \xi_i^* + y_i - w^T x_i - b) - \sum_{i=1}^{l}(\eta_i\xi_i + \eta_i^*\xi_i^*) \tag{21}$$

where $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \ge 0$ are the dual variables. Using the saddle point condition, if (w, b) is a solution of (20) then

$$\partial_b L = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i) = 0$$

$$\partial_w L = w - \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) x_i = 0 \qquad (22)$$

$$\partial_{\xi_i^{(*)}} L = C - \alpha_i^{(*)} - \eta_i^{(*)} = 0$$

and the dual optimization problem corresponding to (20) is given by

$$maximize -\frac{1}{2} \sum_{i,j=1}^{l} (a_i - \alpha_i^*)(a_j - \alpha_j^*) {x_i}^T x_j - \varepsilon \sum_{i=1}^{l} (a_i + \alpha_i^*) + \sum_{i=1}^{l} y_i (a_i - \alpha_i^*) \qquad (23)$$

$$subjectto \begin{cases} \sum_{i=1}^{l} (a_i - \alpha_i^*) = 0 \\ a_i, \alpha_i^* \in [0, C] \end{cases}$$

Obviously, from (22) we get

$$w = \sum_{i=1}^{l} (a_i - \alpha_i^*) x_i \qquad (24)$$

and consequently,

$$f(x) = \left(\sum_{i=1}^{l} (a_i - \alpha_i^*) x_i \right)^T x + b = \sum_{i=1}^{l} (a_i - \alpha_i^*) {x_i}^T x + b \qquad (25)$$

The expression of b is given by [19]

$$\begin{aligned} b &= y_i - w^T x_i - \varepsilon \text{for } a_i \in [0, C] \\ b &= y_i - w^T x_i + \varepsilon \text{for } \alpha_i^* \in [0, C] \end{aligned} \qquad (26)$$

In the following we consider the general case, when the SV algorithm is nonlinear. The non-linear transform is a map $g: \mathbb{R}^n \to \mathcal{F}$, the image of $\mathcal{S}$ in the space $\mathcal{F}$ being given by $\mathcal{S}_g = \{g(x_i),\ x_i \in \mathcal{S}, i = 1, \dots, l\}$. For each $i = 1, \dots, l$, $g(x_i)$ is the new representation of x_i in the considered space $\mathcal{F}$. The transform g is called a feature extractor, and $\mathcal{F}$ is the feature space.

The main problem in designing the feature extractor g is to select a particular functional expression of g, such that, on one hand, the results of forecasting process are accurate enough, and on the other hand the problem of computing the parameter(w,b) is tractable. The "kernel trick" provides a solution to these problems [20]. It consists in selecting a function K that "covers" the explicit functional expression of g, therefore the evaluation of the regression function expression $f(x) = w^T g(x) + b$ is performed exclusively in terms of K. The main result in kernel-based approaches is the celebrated Mercer theorem. According to these results, if $K: \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0,\infty)$is a continuous symmetric function, the existence of a function g such that for any $x, x^{'} \in \mathbb{R}^n, K\left(x,x^{'}\right) = g(x)^T g(x')$holds, is guaranteed in case K satisfies a set of quite general conditions [21].

Using the "kernel trick", the non-linear Support Vector algorithm is expressed in terms of the following optimization problem

$$maximize -\frac{1}{2}\sum_{i,j=1}^{l}(a_i-\alpha_i^*)\left(a_j-\alpha_j^*\right)K(x_i x_j)-\varepsilon\sum_{i=1}^{l}(a_i+\alpha_i^*) + \sum_{i=1}^{l}y_i(a_i-\alpha_i^*)$$

$$subjectto\begin{cases}\sum_{i=1}^{l}(a_i-\alpha_i^*)=0\\ a_i,\alpha_i^*\in[0,C]\end{cases} \tag{27}$$

Similarly, we get

$$w=\sum_{i=1}^{l}(a_i-\alpha_i^*)g(x_i) \tag{28}$$

and therefore

$$f(x)=\sum_{i=1}^{l}(a_i-\alpha_i^*)K(x_i\,x)+b \tag{29}$$

A series of particular expressions of kernels satisfying the Mercer's conditions have been extensively used in the published literature [19]. In our tests we use the Gauss Radial Basis Function (GRBF), $K(x,x') = exp(-\gamma\|x - x'\|^2), \gamma > 0$ because of its proved efficiency in feature extraction process.

5. Case study of OMV PETROM stock closing price. Experimental analysis

In our test, the prediction performance is evaluated in terms of RMSE measure.

In case of the time series described in §2.2, ACF decays very slowly, therefore the considered time series are non-stationary. We analyzed the delay considered in order to establish a forecasting model. Using the correlogram-based analysis, we obtained that for all considered variables, PACF function drops immediately after the 2nd lag. This means that the delay for all variables should be set to 2 ($d = 2$).

Step 1. First, we analyzed the most commonly used model, which considered only one time series, namely $\mathrm{Y_t}$. Since $d = 2$, the following relation describes the forecasting model

$$\widehat{\mathrm{Y}}_{\mathrm{t+1}} = f1(Y_t, Y_{t-1}) \qquad (30)$$

We used the SVR and ANN methods to determine the function $f1$ in (30). We evaluated the forecasting capacity of (30) by splitting the available data into train data (70% of all data) and test data (30% of the all data, considered as new, unseen yet samples) and computing the RMSE between the real data and the forecasted ones. The best result in terms of RMSE when new data were predicted was obtained when SVR method was used.

The predicted closing price values versus the real ones are presented in Figure 2.

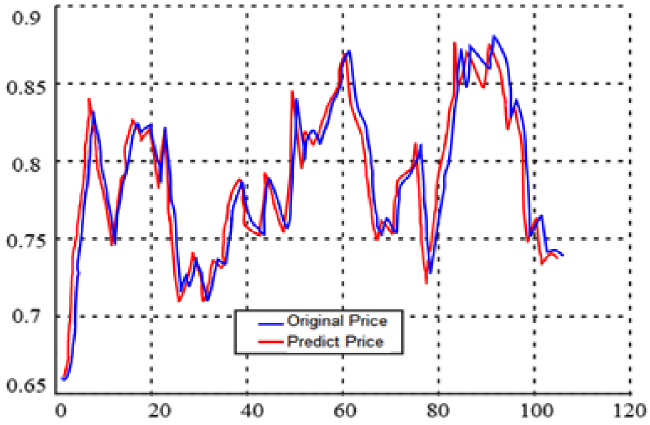

Figure 2. The predicted closing price values versus the real ones. The normalized case RMSE=0.025514

In case of un-normalized closing price, the obtained RMSE is 0.010511.

Step 2. In the following, we consider the model (9), where the maximum value of n is 39 and p=1.

$$\hat{\mathrm{Y}}_{(\mathrm{t+1})} = \mathrm{f}\left(Y_t^{(2)}, X_t^{(2)}\right)$$

$$Y_t^{(2)} = \{\mathrm{Y_t}, \mathrm{Y_{t-1}}\}, X_t^{(2)} = \{\mathrm{X_t}, \mathrm{X_{t-1}}\} \qquad (31)$$

$$\mathrm{X_t} = \left(\mathrm{X_t}(1), \mathrm{X_t}(2), \dots, \mathrm{X_t}(n)\right)^T$$

and $\mathrm{X_t}(1), \mathrm{X_t}(2), \dots, \mathrm{X_t}(n)$ are the selected components of $\mathrm{XT_t}$.

Step 2.1 Variable selection

The function g in (30) was computed based on SV regression. Note that the cross correlation between closing price and each component of XT is presented in Table 1.

V1. In case of using variant 1 for variable selection, taking into account the values of the cross-correlation coefficients in Table 1, the tests were conducted in case $Tr \in TrS = \{0.6, 0.8, 0.85, 0.9, 0.95\}$.

Note that, if we select all available variables as key variables, the generalized error of (4) is RMSE:0.07249.

The cross-correlation threshold value $Tr \in TrS$ for which the general forecasting error of the model (4) is minimized is 0.9 (the time series should be strongly correlated to the closing price time series), $n = 6$ and RMSE of (4) when new samples are forecasted is 0.0229912. The results are presented in Figure 3.

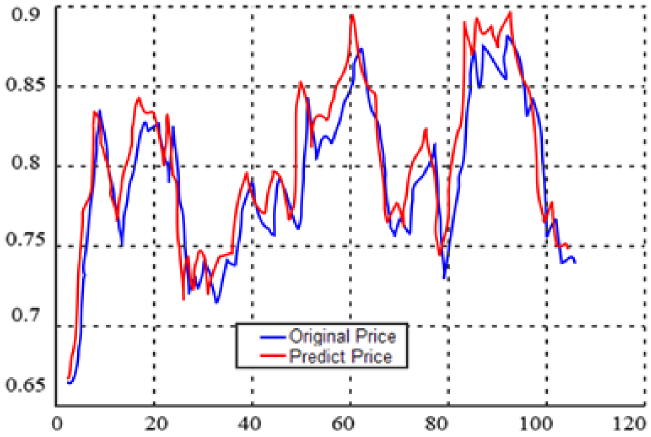

Figure 3. The performances of (4) for Tr=0.9. RMSE=0.0229912

V2. If we apply the variable selection process described in Variant2, we obtained the following results. In case of using4-means algorithm, the computed cluster $\mathcal{C}$ included 10 variables (1,2,3,5-10,37 in Table 1)

The error of (4) when new samples are forecasted is 0.051403.

In order to select the most informative variables from the cluster $\mathcal{C}$, we used $\varepsilon = 1.5 * 10^{-6}$. The computed subset $\mathcal{S}_{\mathcal{C}}$consists of 4 variables (high price, low price, KAMA and MA). The error of (4) when new samples are forecasted is 020422. The results are presented in Figure 4.

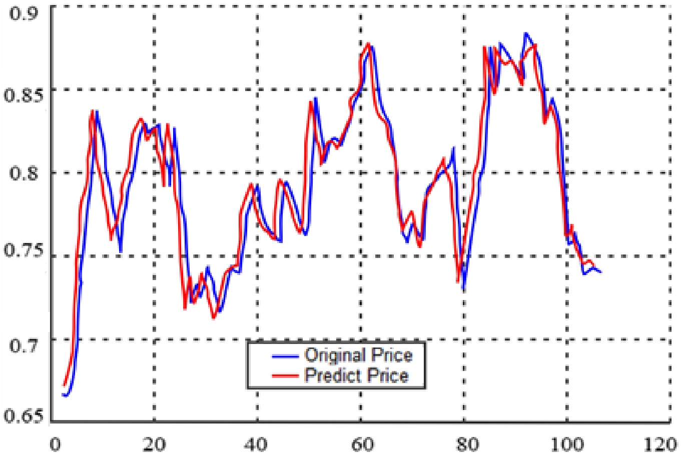

Figure 4. The performances of (4) when we used variant 2 of proposed selection process, k=4 and $= \mathbf{1.5 * 10^{-6}}$. RMSE=0.020422

Step 3. NARX-based forecasting of model (31)

We apply a neural network-based method for closing price forecasting in terms of the model (31). We implemented the NARX model using the following computation scheme:

1. The training step. Train the network using first 50% examples; the process is over when the neural network is capable to forecast the trained data such that $RMSE < \varepsilon'$. In our tests we considered $\varepsilon^{'} = 0.04$

2. The validation step uses the next 20% examples. If the network forecasts sufficiently well the examples belonging to the validation set, i.e. $RMSE < \varepsilon"$, then GOTO3. Otherwise GOTO 1. The parameter $\varepsilon"$ is set in the interval $[0.04{,}0.05]$.

3. The resulted network is tested on new data (the remaining 30% data).

In case of using variable selection method V1, the error of (31) measured in terms of RMSE when new samples are forecasted is around 0.0300. The results are shown in figure 5. In case of using variable selection process based on the proposed selection method V2, the RMSE error of (31) when new samples are forecasted is around 0.0275. The graphic representation of forecasted values versus the true ones is presented in figure 6.

Step 4. SVR-based forecasting using model (31)

We apply a SVR-based method for closing price forecasting in terms of the model (31). We evaluated the forecasting capacity of (31) by splitting the available data into train data (70% of all data) and test data (30% of the all data, considered as new, unseen yet samples) and computing the RMSE between the real data and the forecasted ones.

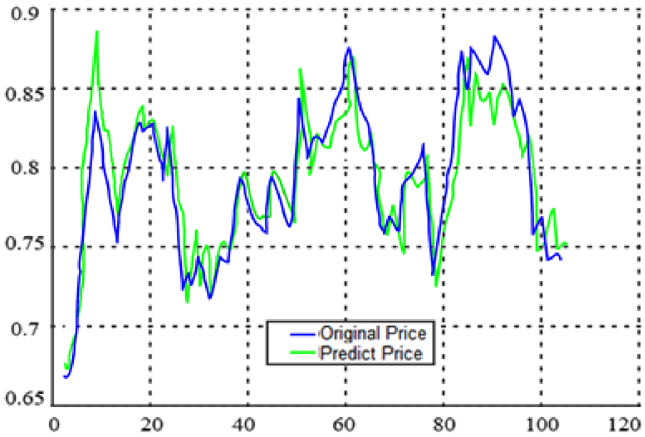

Figure 5.The performances of (31) when we used NARX-based forecasting method and variant 1 of proposed selection process. RMSE=0.03

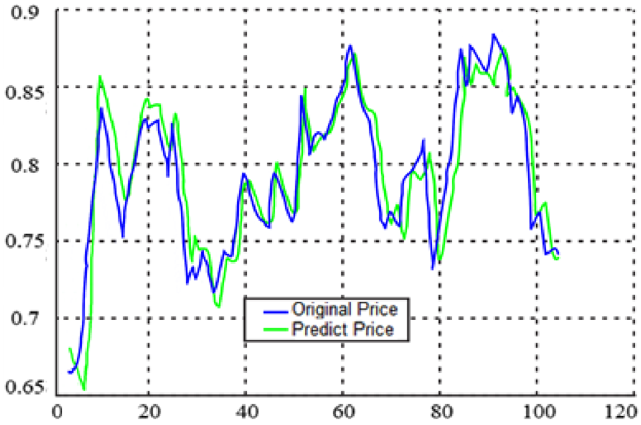

Figure 6.The performances of (31) when we used NARX-based forecasting method and variant 2 of proposed selection process. RMSE=0.0275

If we apply the selection method V1, the error of prediction model (31) when new samples are forecasted is 0.0206. The corresponding results are displayed in Figure 7.

If we use the variable selection method V2, the error of (31) when new samples are forecasted is 0.0200. The obtained results are presented in figure 10.In case of un-normalized closing price (real values), we obtained RMSE=0.0079319.

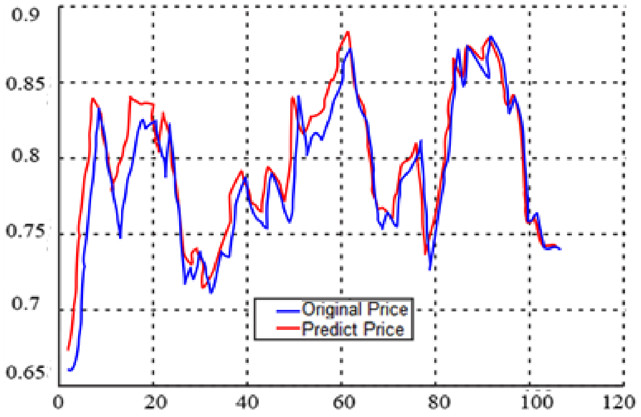

Figure 7.The performances of (31) when we used SVR-based forecasting method and variant 1 of proposed selection process. RMSE=0.0206

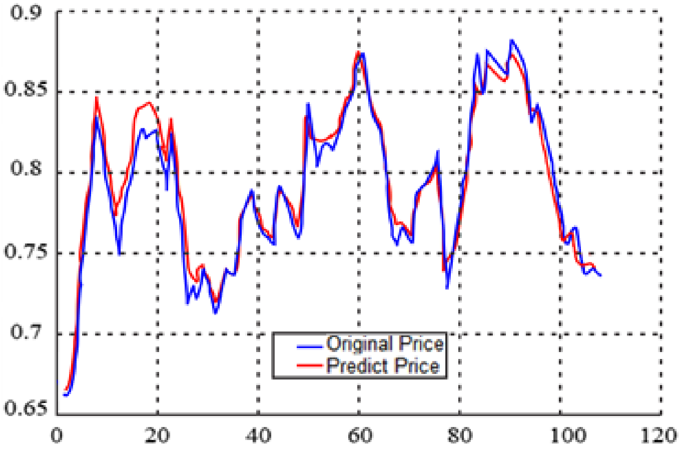

Figure 8.The performances of (31) when we used SVR-based forecasting method and variant 2 of proposed selection process. RMSE=0.0200

6. Experimentally derived conclusion and suggestions for further work

The research work reported in this paper aims the development of a hybrid approach which combines the use of the variables derived from technical and fundamental analysis of stock market indicators in order to improve prediction results of the proposed approaches. Two novel variable selection methods are used to optimize the performance of prediction models. In order to identify the most informative time series to predict a stock price, both methods are essentially based on the general forecasting error minimization when a certain stock price, *Y*, is expressed exclusively in terms of other indicators (variables and/or stock prices). Also, the former proposed method includes cross-correlation analysis and the second one is derived based on cluster analysis and cross-correlation coefficients. After the variable selection is over, the forecasting of *Y* is performed in terms of the historical values of *Y* and selected variables respectively.

First, we analyze the performances of the complex forecasting model (31) against the forecasting model (30). The last model is obviously the simplest one and yet very useful in many real world applications. In our case the former model proved better results.

In our work, the most suited forecasting model is given by (31) and the best prediction method resides in using the novel selection method V2 and using a SVR model for learning the regression function of (31)

Note that in model (30) the generalized error (RMSE) is 0.010511, while the performance of the proposed forecasting model (31) given in terms of RMSE is far better, the RMSE value being 0.0079319.

The long series of tests proved good results of the above described methodology entailing the hope that further and possibly more sophisticated extensions can be expected to improve it. Among several possible extensions, some work is still in progress concerning the use of different output functions for the hidden and output neurons, and the use of more hidden layers in the NARX neural architectures. Also, some other selection techniques combined to new techniques for feature extraction as well as the use of fuzzy SVR based learning schemes to forecast the data are expected to allow the removal of a larger amount of noise.

Table 1.Cross correlation between closing price and other variables

Stock Based Analysis		**Technical Analysis**			
Indicator	**Coefficient**	**Indicator**	**Coefficient**	**Indicator**	**Coefficient**
1. Opening	**0.99821946**	**5. BB**	**0.5676**	**20. MFI**	**0.6673**
2. High	**0.999401**	**6. EMA**	**0.9093**	**21. Momentum**	**0.3680**
3. Low	**0.99896**	**7. KAMA**	**0.9213**	**22. PPO**	**0.4010**
4. Volume	**-0.25555**	**8. MA**	**0.8694**	**23. ROC**	**-0.0349**
		9. WMA	**0.9306**	**24. RSI**	**0.5018**
		10. TRIMA	**0.8546**	**25. %K**	**-0.0863**
		11. ATR	**-0.0267**	**26. %D**	**-0.0803**
		12. ADX	**-0.2409**	**27. Ultimate Osc**	**0.0662**
Fundamental Analysis		**13. APO**	**0.4750**	**28.Williams %R**	**0.2169**
Indicator	**Coefficient**	**14. AROON**	**-0.2309**	**29. Minus DI**	**-0.577**
35. BET Index	**0.882602**	**15. BOP**	**0.1138**	**30. Plus DI**	**0.3653**
36. ROTX Index	**0.8423804**	**16. CCI**	**0.2713**	**31. Minus DM**	**-0.4402**
37. Crude Oil Price	**0.880853**	**17. CMO**	**0.5018**	**32. Plus DM**	**0.4482**
38.DowJones Index	**0.854413**	**18. DX**	**-0.1317**	**33. ChaikinOsc**	**-0.0283**
39. USD-RON	**-0.28212**	**19. MACD**	**0.5802**	**34. OBV**	**0.5896**

Acknowledgment

A major contribution to the research work reported in this paper belongs to Professor Luminita State, PhD, and distinguished member of the Romanian Academic Community. Professor Luminita State passed on January 2016. The authors will always remember her amazing spirit, as well as her brilliant mind. May God rest her soul in peace!

REFERENCES

[1] **Deboeck, G. (1994)**, ***Tradingon the Edge: Neural, Genetic and Fuzzy Systems for Chaotic Financial Markets*** (Vol. 39).*John Wiley & Sons*;

[2] **White, H. (1988, July)*, Economic Prediction Using Neural Networks: The Case of IBM Daily Stock Returns***. In *Neural Networks, 1988. IEEE International Conference on* (pp. 451-458). IEEE;

[3] **Yoon, Y. & Swales, G. (1991, January)*, Predicting Stock Price Performance: A Neural Network Approach***. In: *System Sciences, 1991.Proceedings of the Twenty-Fourth Annual Hawaii International Conference on* (Vol. 4, pp. 156-162).IEEE;

[4] **Baba, N. &Kozaki, M. (1992, June)*,An Intelligent Forecasting System of Stock Price Using Neural Networks*** .In *Neural Networks, 1992.IJCNN.International Joint Conference on* (Vol. 1, pp. 371-377). IEEE;

[5] **Moshiri, S. & Cameron, N. E. (1999)*,Neural Network versus Econometric Models In Forecasting Inflation***. *Journal of forecasting*, *19*;

[6] **De Faria, E. L., Albuquerque, M. P., Gonzalez, J. L., Cavalcante, J. T. P. & Albuquerque, M. P. (2009)*,Predicting the Brazilian Stock Market through Neural Networks and Adaptive Exponential Smoothing Methods***. *Expert Systems with Applications*, *36*(10), 12506-12509;

[7] **Vapnik, V., Golowich, S. E. & Smola, A. (1996)*,Support Vector Method for Function Approximation, Regression Estimation and Signal Processing***. *In: Advances in neural information processing systems 9*;

[8] **Tay, F. E.& Cao, L. (2001)*,Application of Support Vector Machines in Financial Time Series Forecasting***. *Omega*, *29*(4), 309-317;

[9**] Kim, K. J. (2003)*,Financial Time Series Forecasting Using Support Vector Machines***. *Neurocomputing*, *55*(1), 307-319;

[10] **Hui Qu and Yu Zhang (2016)*,A New Kernel of Support Vector Regression for Forecasting High-Frequency Stock Returns*;** *Mathematical Problems in Engineering, vol. 2016,* Article ID 4907654, 9 pages, doi:10.1155/2016/4907654;

[11] **Lam, M. (2004),*Neural Network Techniques for Financial Performance Prediction: Integrating Fundamental and Technical Analysis***. *Decision Support Systems*, *37*(4), 567-581;

[12] **de Oliveira, F. A., Nobre, C. N. & Zarate, L. E. (2013),*Applying Artificial Neural Networks to Prediction of Stock Price and Improvement of the Directional Prediction Index-Case study of PETR4***; Petrobras, Brazil. *Expert Systems with Applications*, *40*(18),7596-7606;

[13] **Wang, Y. & Choi, I. C. (2013),*Market Index and Stock Price Direction Prediction using Machine Learning Techniques***: An empirical study on the KOSPI and HIS;

[14] **Okasha, M. K. (2014),*Using Support Vector Machines in Financial Time Series Forecasting***. *International Journal of Statistics and Applications*, 28-39;

[15] **Duan, W. Q. & Stanley, H. E. (2011), *Cross-correlation and the Predictability of Financial Return Series***. *Physica A:Statistical Mechanics and its Applications*, *390*(2), 290-296;

[16**] Menezes, J. M. P. &Barreto, G. A. (2008),*Long-Term Time Series Prediction with TheNARX Network: An Empirical Evaluation***. *Neurocomputing*, *71*(16), 3335-3343;

[17]**Huang, G. B. (2003),*Learning Capability and Storage Capacity of Two-Hidden-Layer Feedforward Networks***. *Neural Networks, IEEE Transactions on*, *14*(2), 274-281;

[18]**Marquardt, D. W. (1963),*An Algorithm for Least-Squares Estimation of Nonlinear Parameters***. *Journal of the society for Industrial and Applied Mathematics*, *11*(2), 431-441;

[19] **Vapnik, V. (1998),*Statistical Learning Theory***. *John Wiley & Sons. Inc., New York*;

[20] **Abe, S. (2005)*, Support Vector Machines for Pattern Classification*** (Vol. 2). *London: Springer*;

[21]**State, L., Cocianu, C.& Mircea, M. (2014). *Improvements of the Recognition and Generalization Capacities of the Nonlinear Soft Margin Support Vector Machines***. *Economic Computation and Economic Cybernetics Studies and Research*;*ASE Publishing;* *48*(4), 17-38.