Int. J. Nonlinear Anal. Appl. 14 (2023) 3, 87–102
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.7168

The use of ARIMA, ANN and SVR models in time series hybridization with practical application

Noor Saleem Mohammed Ali, Firas Ahmed Mohammed*

College of Administration and Economics, Department of Statistics, University of Baghdad, Iraq

(Communicated by Asadollah Aghajani)

Abstract

Forecasting is one of the important topics in the analysis of time series, as the importance of forecasting in the economic field has emerged in order to achieve economic growth. Therefore, accurate forecasting of time series is one of the most important challenges that we seek to make the best decision, the aim of the research is to suggest employing hybrid models to predict daily crude oil prices. The hybrid model consists of integrating the linear component, which represents Box Jenkins models, and the non-linear component, which represents one of the methods of artificial intelligence, which is the artificial neural network (ANN), support vector regression (SVR) algorithm and it was shown that the proposed hybrid models in the prediction process when conducting simulations for the time series and for different sample sizes and when applying them on the daily crude oil price data, it was more efficient than the single models, as the comparison between the single models and the proposed hybrid models was done by means of the comparison scale, the mean square error (MSE), the results showed that the proposed hybrid models gave more accurate and efficient results, in addition to its ability to predict crude oil prices well.

Keywords: ARIMA, Artificial Neural Network, Time Series Forecasting, Support Vector Regression, Hybrid Model
2020 MSC: 34A38

1 Introduction

Accurate forecasting of the time series of crude oil prices is one of the important topics in the economic field and one of the most important issues facing energy economists towards better forecasting decisions. Through forecasting, the necessary plans are made to avoid losses resulting from oil price fluctuations.

Hybrid models have been used, which results from merging a linear model with a non-linear model in order to address the existing deficiency if each model is used separately, as most of the time series data contain a linear pattern and another non-linear pattern, and the hybridization process aims to reduce errors resulting from using an inappropriate model to improve future predictions. The first model represents the linear component, which is the (ARIMA) model represents the Box Jenkins model, which was developed to represent stationary and non-stationary time series to predict time series values, and the second model represents the non-linear component, which is the use of some methods in artificial intelligence are machine learning models that represent artificial neural networks (ANN) and support vector regression (SVR) algorithm capable of modeling complex properties such as nonlinearities and fluctuations.

*Corresponding author
Email addresses: noor.ali1101@coadec.uobaghdad.edu.iq (Noor Saleem Mohammed Ali), firasmohana@coadec.uobaghdad.edu.iq (Firas Ahmed Mohammed)

Received: August 2022 *Accepted:* November 2022

2 The First Proposed Hybrid Model

The first proposed hybrid model is built according to the time series consisting of two components: the first component, which is the linear component, which represents Box-Jenkins models, i.e., the (ARIMA) model (L_t), and the second component, which represents non-linear models, which represents one of the methods of artificial intelligence (N_t) in (t) time. These two components are applied to the regression model, where the first independent variable represents the linear component and the second independent variable represents the nonlinear component.

$$\hat{y}_t = b_0 + b_1\hat{x}_{1(\textit{ Linear Component } [L_t])} + b_2\hat{x}_{2(\textit{ Non-Linear Component } [N_t])} \tag{2.1}$$

The parameters estimation of the regression model can be obtained by the method of the Ordinary Least Squares (OLS) according to the following mathematical formula [4]:

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{2.2}$$

The first proposed hybrid model [a] is built according to the following steps:

1. The (ARIMA) time series model is built and the predictive values are obtained by the Box Jenkins construction.
2. After obtaining the predictive values, which represent the linear component of the hybrid model, the residuals of the (ARIMA) model are obtained.
3. The non-linear component of the (ANN), (SVR), where the residuals of the (ARIMA) model are considered as inputs to the non-linear model, then the predictive values of the non-linear component are calculated.
4. The predictive values of the (ARIMA) model (whose inputs represent the observations of time series) are combined with the predictive values of the non-linear model (whose inputs represent the residuals of the ARIMA model) by means of the regression model in order to obtain the predictive values of hybrid model where the linear component represents the first independent variable and the non-linear component represents the second independent variable, the parameters of the regression model are estimated by the (OLS) method.

The equation for the first proposed hybrid model [a] can be written as follows:

$$\hat{y}_t = b_0 + b_1\hat{y}_{1(\textit{Linear Component } [L_t])} + b_2\hat{e}_{2(\textit{Non-Linear Component } [N_t])} \tag{2.3}$$

The first proposed hybrid model [b] is built according to the following steps:

1. The non-linear component of the (ANN), (SVR), are built, and the predicted values are obtained.
2. After obtaining the predictive values, which represent the non-linear component of the hybrid model, the residuals of the non-linear model are obtained.
3. The linear component of (ARIMA), where the residuals of the non-linear model are considered as inputs to the linear model, then the predicted values of the linear component are calculated.
4. The predicted values of the non-linear model (whose input are represent the observations of time series) are combined with the predictive values of the linear model (whose inputs represent the residuals of the non-linear model) by means of the regression model in order to obtain the predictive values of hybrid model where the non-linear component represents the first independent variable and the linear component represents the second independent variable, the parameters of the regression model are estimated by the (OLS) method.

The equation for the first proposed hybrid model [b] can be written as follows:

$$\hat{y}_t = b_0 + b_1\hat{y}_{1(\text{Non-Linear Component } [N_t])} + b_2\hat{e}_{2(\text{Linear Component } [L_t])}. \tag{2.4}$$

3 The Second Proposed Hybrid Model

The second proposed hybrid model is built according to the following steps:

1. The (ARIMA) time series model is built and the predictive values are obtained by the Box Jenkins construction.
2. After obtaining the predictive values, which represent the linear component of the hybrid model.
3. The non-linear component of the (ANN), (SVR), where the time series (y_t) are considered as inputs to the non-linear model, and the predictive values of the non-linear component are calculated.
4. The predictive values of the (ARIMA) model (whose inputs represent the observations of time series (y_t)) are combined with the predictive values of the non-linear model (whose inputs represent the observations of time series (y_t)) by means of the regression model in order to obtain the predictive values of hybrid model where the linear component represents the first independent variable and the non-linear component represents the second independent variable, the parameters of the regression model are estimated by the generalized ridge regression (GRR) method.

The equation for the second proposed hybrid model can be written as follows:

$$\hat{y}_t = b_0 + b_1 \hat{y}_{1(Linear\ Component\ [L_t])} + b_2 \hat{y}_{2(Non\text{-}Linear\ Component\ [N_t])} \tag{3.1}$$

The Generalized Ridge Regression (GRR) estimates for the linear regression model can be calculated according to the following equation [4]:

$$b_{GRR} = (X^{*'} X + K)^{-1} X^{*'} Y \tag{3.2}$$

Choice of the Estimator ridge (K) by (Hoerl), (Kennard) and (Baldwin) suggested that a choice for (K) is [5]:

$$\hat{K}_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}} \tag{3.3}$$

4 Autoregressive Integrated Moving Average (ARIMA) Model

The ARIMA model was introduced by Box and Jenkins in (1976), the ARIMA models are generated from the auto regressive models and the moving averages models, the differences for the time series are taken to convert the non-stationary series into a stationary series which is called the degree of integrated and is denoted by (d) and which it symbolizes by ARIMA (p, d, q).

The ARIMA models can be expressed using the backshift operator (B) as follows [6]:

$$\phi(B)(1 - B)^d y_t = \theta\,(B)\,\varepsilon_t \tag{4.1}$$

5 Artificial Neural Network (ANN)

Artificial neural networks are considered one of the most important applications of artificial intelligence in the field of machine learning. The idea of (ANN) is to simulate data in order to reach a model for this data for the purpose of classification, analysis, prediction or any other processing without the need for any prior model describing the relationship between the data or a proposed model for this data [9].

The architecture of the (ANN) consists of three levels: input level, hidden level, and output level, which can be explained as follows [1]:

The input level represents the first level in the (ANN), which contains the number of nodes that represent the number of independent variables. The hidden level represents the second level in the (ANN), this level contains one or more hidden layers and the hidden layer contains a number of hidden nodes, and each node has a weight that connects it with the input level and another weight that connects it with the next level that represents the output level.

The processing units in an (ANN) consist of the following basic components:

1. **The Sum Function:** It is a linear combination that calculates the input multiplied by the weight in the neural network. It can be expressed mathematically as follows [10]:

$$z = \sum_{i=1}^{n} w_{ji} x_i + b \quad , \quad i = 1, 2, \ldots, n \ , \quad j = 1, 2, \ldots, s \tag{5.1}$$

Where:
z : The sum of the multiplication between the input level variables and the weights.
w_{ji}: The weights between the input level (i) and the hidden level (j).
x_i: The input level variables.
n : The number of the input nodes.
s : The number of the hidden nodes.
b : The bias that improves the neural network and is represented by a constant value ($x_0 = 1$).

2. **The Activation Function:**

It is the function that converts the product of the weighted summation of the input level variables with the addition of bias into the product representing the output of the hidden node. The sigmoid function is used in the hidden nodes, this function takes output values between (0,1), the formula for the sigmoid function is [10]:

$$f(z) = \frac{1}{1 + \exp(-z)}. \tag{5.2}$$

The linear function is used as an activation function in the output node of the output level, the formula of the linear function is:

$$f(z) = z \tag{5.3}$$

The data should be normalized which is a process of initializing the data before it is processed in order to be used in the training process of the network, the normalized formula will be used and the formula is [2]:

$$X_{new} = \frac{X - X_{\min}}{X_{man} - X_{\min}} \tag{5.4}$$

The steps of the backpropagation algorithm are represented as follows [1] & [8]:

1. The initial values for the weights are generated in the neural network.
2. The number of nodes is initialized at the input level, hidden level, and output level, as well as the number of bias nodes at input level and hidden level.
3. Each node in the input level receives the data, and this data sent to each node in the hidden level.
4. The weighted input values at each node in the hidden level are summed according to the equation 5.1.
5. The activation function is applied to get an estimate of the output of the hidden level, these values are sent to each node in the output level according to the equation 5.2.
6. Each node in the output level receives the summation input signals weighted from the hidden level.
7. The loss function represented by (MSE) will be used as follows [8]:

$$J(w) = \frac{1}{2} \sum_{k=1}^{m} (y_k - a_k)^2$$

y_k: The actual values.
a_k: The predicted value of the output layer in the network.

8. Calculate the error factor in the output level is:

$$\delta_k = (y_k - a_k) g'(z_k)$$

9. The weights are updated which connects the output level and the hidden level by the gradient descent:

$$w_{kj(new)} = w_{kj(old)} + \Delta w_{kj}$$

10. The error rate at the hidden level is calculated as follows:

$$\delta_j = (\sum_k \delta_k w_{kj}) g'(z_j)$$

11. The weights are adjusted, which connects the hidden level and the input level:

$$w_{ji(new)} = w_{ji(old)} + \Delta w_{ji}$$

This process continues until the optimal weights are obtained that make the loss function as small as possible.

6 Support Vector Regression (SVR)

The Support Vector Regression (SVR) algorithm was introduced by (Vapnik & Cortes) in (1995). The goal of using (SVR) is to find the regression curve so that all data are closer to the curve and reduce the error rate in the results, the quality of the estimate is measured using the ϵ-insensitive loss function. The regression of any non-linear function or curve can be handled by applying kernel in the (SVR) algorithm, where the non-linear data is set in a space that makes the data linear.

Suppose that the data set:

$$D = \{(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_n, y_n)\} \quad , x_i \in R^n \ , y_i \in R \ , \ i = 1, 2, \ldots\ldots, n$$

where, D is the data set, $x_1, x_2, \ldots\ldots, x_n$ are the numbers of input variables ans $y_1, y_2, \ldots\ldots, y_n$ are the output values associated of (x_i)'s.

The mathematical equation for the non-linear (SVR) function is expressed as [3]:

$$f(x) = < w, \varphi(x) > +b \tag{6.1}$$

where: $\varphi(x)$ is the feature which is the non-linear mapping of inputs (x), $f(x)$ is the (SVR) function, w is the parameters or weights, b is the intercept term and called the bias which is a numerical value and $< w, \varphi(x) >$ is the inner product of the vectors (w) and (x_i).

The main objective of the (SVR) model is to find the values of the function and this is done according to the following steps [7]:

Step1: The values for weights $(w_1, w_2, \ldots\ldots, w_n)$ are calculated so that the value of the Euclidean criterion for weights is as low as possible:

$$\frac{1}{2} \parallel w \parallel^2$$

where, $\parallel w \parallel^2$ is the weights vector norm which is used to limit the amplitude of the model and this is in order to obtain the best generalize performance.

Step2: The coefficients (b) and (w) are estimated by reducing the regularization risk function as follows:

$$R(C) = C \frac{1}{N} \sum_{i=1}^{n} |y_i - f(x_i)|_\varepsilon + \frac{1}{2} \parallel w \parallel^2 \tag{6.2}$$

Subject to the constraints:

$$\begin{cases} y_i - w, x_i > -b \leq \varepsilon \\ < w, x_i > +b - y_i \leq \varepsilon \end{cases} \tag{6.3}$$

where, C is the regularization factor that balances between the estimation error and the amount of distance that is away about the main axis and $|y_i - f(x_i)|_\varepsilon$ represents the ϵ − insensitive loss function.

The (SVR) algorithm is trained on the basis of the loss function **(ϵ − insensitive) Loss Function** and can be expressed in the form:

$$|y_i - f(x_i)|_\varepsilon = \begin{cases} 0 & if \quad |y - f(x_i)| \le \varepsilon \\ |y_i - f(x_i)| - \varepsilon & o.w \end{cases} \tag{6.4}$$

Step3: The slack variables $(\xi_i, \xi_i{}^*)$ are entered because some points still fall outside the margin, so errors that are greater than epsilon must be calculated, the presence of these deviations is reasonable, but must reduce it as much as possible.

The (SVR) formula is given as the following minimization function:

$$\frac{1}{2} \parallel w \parallel^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i{}^*) . \tag{6.5}$$

Subject to the constraints:

$$\begin{cases} y_i - < w, \phi(x) > -b \le \varepsilon + \xi_i \\ < w, \phi(x) > +b - y_i \le \varepsilon + \xi_i{}^* \\ \xi_i, \xi_i{}^* \ge 0 \end{cases} \tag{6.6}$$

The optimization problem is the solution of primal problem.

Step4: After getting the solution to the primal problem, the dual problem is derived from the primal problem by introducing a Lagrange Multiplier to transform the quadratic programming function, where the dual problem transforms the problem to be solved in a convex state to be in a state that can reach the optimal solution:

$$\begin{aligned} L(\eta_i, \eta_i{}^*, \alpha_i, \alpha_i^*) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi^*{}_i) - \sum_{i=1}^{n}(\eta_i\xi_i + \eta_i{}^*\xi^*{}_i) - \sum_{i=1}^{n}\alpha_i(\varepsilon + \xi_i - \\ y_i + < w, \varphi(x_i) > +b) - \sum_{i=1}^{n}\alpha_i^*(\varepsilon + \xi_i^* - y_i + < w, \varphi(x_i) > +b) \end{aligned} \tag{6.7}$$

where, L is the Lagrange function, $(\eta_i, \eta_i{}^*, \alpha_i, \alpha_i{}^*)$ is the Lagrange multiplier, $\alpha^{(*)}{}_i = (\alpha_1, \alpha^*{}_1, \ldots\ldots, \alpha_n, \alpha^*{}_n)^T$ and $\eta^{(*)}{}_i = (\eta_1, \eta^*{}_1, \ldots\ldots, \eta_n, \eta^*{}_n)^T$ are the Lagrange Multiplier vectors. Take the partial derivatives with respect to the initial variables and set the resulting derivatives to zero, the solution is given as follows:

$$-\frac{\mathbf{1}}{\mathbf{2}}\sum_{i=\mathbf{1}}^{n}\sum_{j=\mathbf{1}}^{n}(\alpha^*{}_i - \alpha_i)(\alpha^*{}_j - \alpha_j)\, k < x_i, x_j > -\varepsilon\sum_{i=\mathbf{1}}^{n}(\alpha^*{}_i + \alpha_i) + \sum_{i=\mathbf{1}}^{n}(\alpha^*{}_i - \alpha_i)\, y_i \tag{6.8}$$

with the constraints:

$$0 \le \alpha_i, \alpha_i{}^* \le c$$

$$\sum_{i=1}^{n}(\alpha^*{}_i - \alpha_i) = 0$$

The optimization problem is the solution of dual problem.

Step5: The dual problem can be simplified by deleting the variable $\eta^{(*)}$ so that it is a problem for only one variable which is $\alpha^{(*)}$, it can be rewritten as a minimization problem as follow:

Minimize

$$\frac{\mathbf{1}}{\mathbf{2}}\sum_{i=\mathbf{1}}^{n}\sum_{j=\mathbf{1}}^{n}(\alpha^*{}_i - \alpha_i)(\alpha^*{}_j - \alpha_j)\, k < x_i, x_j > -\varepsilon\sum_{i=\mathbf{1}}^{n}(\alpha^*{}_i + \alpha_i) + \sum_{i=\mathbf{1}}^{n}(\alpha^*{}_i - \alpha_i) y_i \tag{6.9}$$

with constraints:

$$\sum_{i=1}^{n}(\alpha^*{}_i - \alpha_i) = 0$$

$$0 \le \alpha_i{}^{(*)} \le c \qquad , \; i = 1, 2, \ldots\ldots, n$$

This is the dual problem of the primal problem.

$$k < x_i, x_j >=< \phi(x_i), \phi(x_j) > \tag{6.10}$$

$$= \exp(-\gamma \parallel x_i - x_j \parallel^2)$$

where:
k: The inner product of a nonlinear mapping $\phi(x_i)$ and $\phi(x_j)$.
$k < x_i, x_j >$: The kernel function.
$\parallel x_i - x_j \parallel^2$: The squared Euclidean distance between feature vectors.
γ: The width parameter of RBF kernel.

Step7: The optimal values for weights can be obtained from the equation:

$$w = \sum_{i=1}^{n} (\alpha^*{}_i - \alpha_i) y_i \tag{6.11}$$

The output values for the (SVR) model can be found from the equation:

$$f(x) = \sum_{i=1}^{n} (\alpha^*{}_i - \alpha_i) k < x_i, x_j > +b \tag{6.12}$$

7 Prediction Accuracy Criterion

The prediction accuracy criterion aims to measure the degree of accuracy of forecasting, and the comparison criterion was relied on the mean square error (MSE) to compare the single models and the proposed hybrid models to find out what is the most accurate model in forecasting, the general form is:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_t - \hat{y}_t)^2 \tag{7.1}$$

where:
y_t: The actual values.
$\hat{y}_t$: The predicted values.
n: The number of the observations.

8 The Simulation

The simulation is one of the most important scientific methods that are used as a method of experimental proof to find out which model is the best among a group of models in case it is unable to prove it theoretically, as the simulation builds a mathematical model of a model or method and reconfigures the model in an experimental manner that allows us to evaluate the performance of this method or model and study its properties or solve problems that may be related to it.

The simulation experiments included generating data for a group of the single models and the proposed hybrid models with different sample sizes (small, medium and large), which are (250, 500, 1500 & 3000) where the sample size is of great importance in the accuracy and the efficiency of the extracted results, these models using the comparison criterion (MSE), the programs (R) were used in the simulation and the experiment was repeated (250) times.

a. The simulation result for the single models for the sample size (250):

Table 1: Calculating the (MSE) of the individual models at a sample size (250)

Model	ARIMA	ANN [1]	ANN [2]	SVR
MSE	1.234411	0.980979	0.996053	1.457884

The table (1) shows the results of the (MSE) criterion for each of the models individually and the superiority of the (ANN [2]) model with a one hidden layer at small sample sizes (250), it has the lowest (MSE) value equal to (0.980979), followed by the (ANN [1]) model with two hidden layers which has value equal to (0.996053). The model that gave the highest value for the (MSE) is the (SVR) model, where it has a value equal to (1.457884).

b. The simulation result for the single models for the sample size (500):

Table 2: Calculating the (MSE) of the individual models at a sample size (500)

Model	ARIMA	ANN [1]	ANN [2]	SVR
MSE	1.186498	1.002452	1.004662	1.283591

The table (2) shows the case of sample sizes (500) also the stability of the results for the other of the models and the superiority of the (ANN [2]) with a one hidden layer, which has the lowest value of the (MSE) criterion equal to (1.002452). The (SVR) model also that gave the highest value for the (MSE) is, where it has a value equal to (1.28359).

c. The simulation result for the single models for the sample size (1500):

Table 3: Calculating the (MSE) of the individual models at a sample size (1500)

Model	ARIMA	ANN [1]	ANN [2]	SVR
MSE	1.095499	1.024594	1.016957	0.997173

The table (3) the superiority of the (SVR) model at sample sizes (1500), as it has the least value of the (MSE) criterion equal to (0.997173), then (ANN [1]) with two hidden layers which have value equal to (1.016957). The (ARIMA) model also that gave the highest value for the (MSE) is, where it has a value equal to (1.095499).

d. The simulation result for the single models for the sample size (3000):

Table 4: Calculating the (MSE) of the individual models at a sample size (3000)

Model	ARIMA	ANN [1]	ANN [2]	SVR
MSE	1.06265	1.031744	1.013328	0.992531

The table (4) shows at large sample sizes also the stability of the results and the superiority of the (SVR), which has the lowest value of the (MSE) criterion equal to (0.992531). Then the (ANN [1]) with two hidden layers which have value equal to (1.013328), the (ARIMA) model that gave the highest value for the (MSE) is, where it has a value equal to (1.06265).

e. The simulation result for the first proposed hybrid model [a] for the sample size (250):

Table 5: Calculating the (MSE) of the first proposed hybrid models [a] at a sample size (250)

Model	ARIMA-ANN [1]	ARIMA-ANN [2]	ARIMA-SVR
MSE	0.525373	0.530408	0.132256

The table (5) shows that the (ARIMA-SVR) hybrid model has outperformed the other of the hybrid models as it has the lowest value for the (MSE) criterion equal to (0.132256) at small sample sizes (250), then followed by the hybrid model (ARIMA-ANN [2]) which have value equal to (0.525373).

f. The simulation result for the first proposed hybrid model [b] for the sample size (250):

Table 6: Calculating the (MSE) of the first proposed hybrid models [b] at a sample size (250)

Model	ANN [1]-ARIMA	ANN [2]-ARIMA	SVR-ARIMA
MSE	0.737142	0.733967	0.169762

The table (6) shows that the best hybrid model is (SVR-ARIMA), which has the lowest (MSE) criterion equal to (0.169762), then followed by the hybrid model (ANN [1]-ARIMA) which have value equal to (0.733967).

g. The simulation result for the first proposed hybrid model [a] for the sample size (500):

Table 7: Calculating the (MSE) of the first proposed hybrid models [a] at a sample size (500)

Model	ARIMA-ANN [1]	ARIMA-ANN [2]	ARIMA-SVR
MSE	0.5273258	0.5280812	0.1804434

The table (7) also shows that the (ARIMA-SVR) hybrid model is superior to other hybrid models equal to (0.1804434) then followed by the hybrid model (ARIMA-ANN [2]) which have value equal to (0.5273258).

h. The simulation result for the first proposed hybrid model [b] for the sample size (500):

Table 8: Calculating the (MSE) of the first proposed hybrid models [b] at a sample size (500)

Model	ANN [1]-ARIMA	ANN [2]-ARIMA	SVR-ARIMA
MSE	0.769074	0.7657282	0.4138947

The table (8) also shows that the (SVR-ARIMA) hybrid model at sample sizes (500) is the best among the hybrid models as it has the lowest value for the (MSE) criterion equal to (0.4138947), followed by the (ANN [1]-ARIMA) hybrid model which have value equal to (0.7657282).

i. The simulation result for the first proposed hybrid model [a] for the sample size (1500):

Table 9: Calculating the (MSE) of the first proposed hybrid models [a] at a sample size (1500)

Model	ARIMA-ANN [1]	ARIMA-ANN [2]	ARIMA-SVR
MSE	0.513005	0.515781	0.208167

The table (9) shows that the (ARIMA-SVR) hybrid model outperformed of the other hybrid models which have the lowest value for the (MSE) criterion equal to (0.208167) when applied to the sample sizes (1500), followed by the (ARIMA-ANN [2]) hybrid model which have value equal to equal to (0.513005).

j. The simulation result for the first proposed hybrid model [b] for the sample size (1500):

Table 10: Calculating the (MSE) of the first proposed hybrid models [b] at a sample size (1500)

Model	ANN [1]-ARIMA	ANN [2]-ARIMA	SVR-ARIMA
MSE	0.781807	0.777783	0.728664

The table (10) shows that the best hybrid model is (SVR-ARIMA) that has the lowest value for the (MSE) equal to (0.728664) when applied to the sample sizes (1500), followed by the (ANN [1]-ARIMA) hybrid model which have value equal to (0.777783).

k. The simulation result for the first proposed hybrid model [a] for the sample size (3000):

Table 11: Calculating the (MSE) of the first proposed hybrid models [a] at a sample size (3000)

Model	ARIMA-ANN [1]	ARIMA-ANN [2]	ARIMA-SVR
MSE	0.510309	0.510143	0.232618

The table (11) shows that the best hybrid model is (ARIMA-SVR), where it outperformed the other hybrid models at large sample sizes (3000) and has the lowest value for the (MSE) criterion equal to (0.232618), followed by the (ARIMA- ANN [1]) hybrid model which have value equal to (0.510143).

l. The simulation result for the first proposed hybrid model [b] for the sample size (3000):

Table 12: Calculating the (MSE) of the first proposed hybrid models [b] at a sample size (3000)

Model	ANN [1]-ARIMA	ANN [2]-ARIMA	SVR-ARIMA
MSE	0.777676	0.774375	0.963179

The table (12) shows that the best hybrid model is (ANN [1]-ARIMA) which has the lowest value for the (MSE) criterion equal to (0.774375) at a sample size (3000), followed by the (ANN [2]-ARIMA) hybrid model which has value equal to (0.777676).

m. The simulation result for the second proposed hybrid model for the sample size (250):

Table 13: Calculating the (MSE) of the second proposed hybrid models at a sample size (250)

Model	ARIMA-ANN [1]	ARIMA-ANN [2]	ARIMA-SVR
MSE	0.527608	0.526521	0.170033

The table (13) shows that the (ARIMA-SVR) hybrid model is the best and has the least value for the (MSE) criterion equal to (0.170033) at sample sizes (250), followed by the (ARIMA-ANN [1]) hybrid model which has value equal to (0.526521).

n. The simulation result for the second proposed hybrid model for the sample size (500):

Table 14: Calculating the (MSE) of the second proposed hybrid models at a sample size (500)

Model	ARIMA-ANN [1]	ARIMA-ANN [2]	ARIMA-SVR
MSE	0.5304257	0.5262908	0.412577

The table (14) also shows that the (ARIMA-SVR) hybrid model is the best and has the least value for the (MSE) criterion equal to (0.412577) at sample sizes (500), followed by the (ARIMA-ANN [1]) hybrid model which has value equal to (0.5262908).

o. The simulation result for the second proposed hybrid model for the sample size (1500):

Table 15: Calculating the (MSE) of the second proposed hybrid models at a sample size (1500)

Model	ARIMA-ANN [1]	ARIMA-ANN [2]	ARIMA-SVR
MSE	0.519788	0.519466	0.720183

The table (15) shows that the)ARIMA-ANN [1]) hybrid model is the best and has the least value for (MSE) criterion equal to (0.519466) at sample sizes (1500), followed by the (ARIMA-ANN [2]) hybrid model which have value equal to (0.519788).

p. The simulation result for the second proposed hybrid model for the sample size (3000):

Table 16: Calculating the (MSE) of the second proposed hybrid models at a sample size (3000)

Model	ARIMA-ANN [1]	ARIMA-ANN [2]	ARIMA-SVR
MSE	0.516402	0.513097	0.886358

The table (16) also shows that the (ARIMA-ANN [1]) hybrid model is the best and has the least value for the (MSE) criterion equal to (0.513097) at large sample sizes (3000), followed by the (ARIMA-ANN [2]) hybrid model which have equal to (0.516402).

9 The Application

The real data for daily crude oil prices were applied for the time period from (01/01/2010) to (30/06/2021) with (2999) observations on the best hybrid model that was obtained from the simulation. This data gets it from the SOMO (State Organization for Marketing of Oil) company in Iraq.

Time series should be stationary in terms of mean and variance. Therefore, the time series that shows the development of daily crude oil prices must first be drawn and this is done by applying some necessary transformations to make the time series stationary.

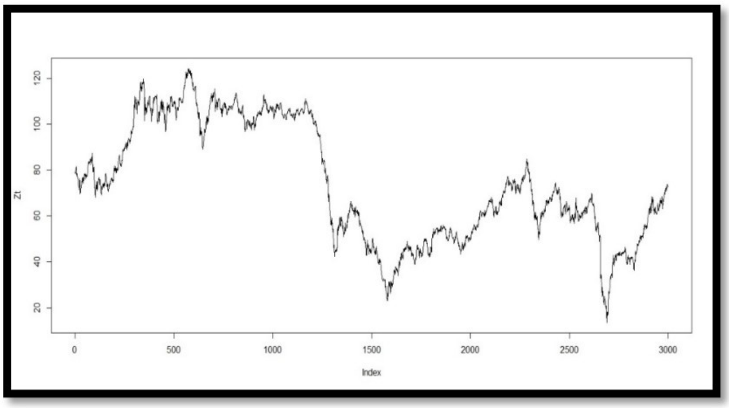

Figure 1: The time series of daily crude oil price

After plotting the time series, it was found that it is non stationary on the mean and variance with time. The logarithmic transformation of the original time series data represented by the oil prices was taken then the first differences (d=1) of the series were taken to the logarithm of the original series and then the new series was drawn as follows:

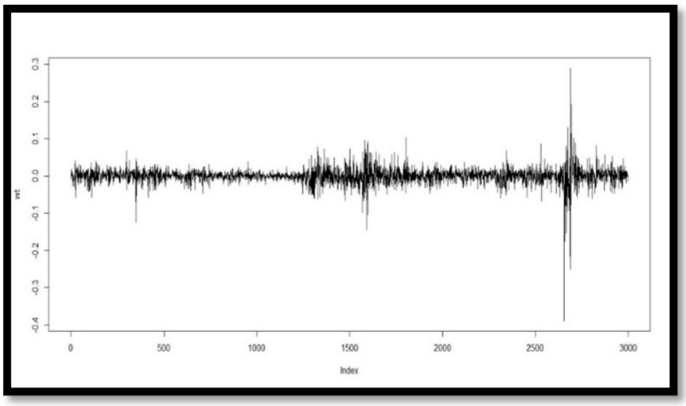

Figure 2: The time series after taking the first differences of the logarithm of the original series

To ensure the validity of the stationary, the unit root tests were carried out, namely the Augmented Dickey Fuller (ADF) test and the KPSS test. The following table shows the results of the two tests:

Table 17: The test results of (ADF) and (KPSS) for the time series after taking the first differences of the logarithm of the original series

The Test	p-value	The Decision
ADF	0.01 <0.05	Stationary
KPSS	0.1 >0.05	Stationary

The results of the table show the stationary of the time series after taking difference of the logarithm of the time series, as the p-value was less than (0.05) in the (ADF) test, which mean the null hypothesis was rejected, and that the p-value in the (KPSS) test was greater than (0.05), which means the null hypothesis was not rejected and this indicates that the time series is stationary.

To get the best model, a set of the models was nominated according to the lowest values of the criteria (AIC, HQC & AICc) then they are compared to get the appropriate model that has the lowest values of these criteria as in the following table:

Table 18: The ARIMA (p, d, q) candidate models and corresponding the statistical criteria for each model

Model	AIC	HQC	AICc
ARIMA(1,1,2)	-14136.74	-14135.1	-14136.74
ARIMA(2,1,1)	-14136.69	-14135.4	-14136.67
ARIMA(2,1,2)	-14160.35	-14167.09	-14160.32
ARIMA(2,1,3)	-14159.7	-14164.1	-14159.67
ARIMA(3,1,2)	-14165.82	-14164.2	-14165.78
ARIMA(3,1,3)	-14166.31	-14161.04	-14166.26
ARIMA(4,1,2)	-14153.49	-14149.01	-14153.44
ARIMA(5,1,2)	-14166.59	-14157.21	-14166.54
ARIMA(5,1,3)	-14191.58	-14180.48	-14191.52
ARIMA(5,1,4)	-14190.32	-14177.19	-14190.24

The table shows that the best model that has the lowest values for the statistical criteria (AIC, HQC & AICc) is the ARIMA (5,1,3) model, where the value of the (AIC) criterion equal to (-14191.58), the (HQC) criterion equal to (-14180.48) and the (AICc) criterion equal to (-14191.52), the best model ARIMA (5,1,3) that can be used in the estimation and prediction.

The parameters of the obtained model ARIMA (5,1,3) are estimated by maximum likelihood estimation (MLE) method and all the estimated parameters of the ARIMA (5,1,3) model are significant, because the p-value is less than the level of significance for the t-test ($\alpha = 0.05$), this indicates the importance of the parameters of the model for prediction.

The model is subjected to a number of tests which are the Box-Pierce and Ljung Box tests was done for the residuals of the ARIMA (5,1,3) model, and the following results were obtained:

Table 19: The Box-Pierce test for the ARIMA (5,1,3) model

Lag	1	5	10
Chi – Square	0.0198	3.0763	9.7567
DF	1	5	10
p-value	0.888	0.6882	0.4621

Table 20: The Ljung Box test for the ARIMA (5,1,3) model

Lag	1	5	10
Chi – Square	0.0198	3.0833	9.7898
DF	1	5	10
p-value	0.888	0.6871	0.4591

The tables (19) & (20), it is clear that the Box-Pierce and Ljung Box tests for the residuals of the model is random, this means that the variables are independent and there is no correlation between them.

By obtaining the predictive values of the ARIMA (5,1,3) model, the amount of prediction error between the actual and predictive values of the model was determined using the (MSE), the results were as follows:

Table 21: The MSE for prediction accuracy of (ARIMA) model

The Model	MSE
ARIMA	0.0005111879

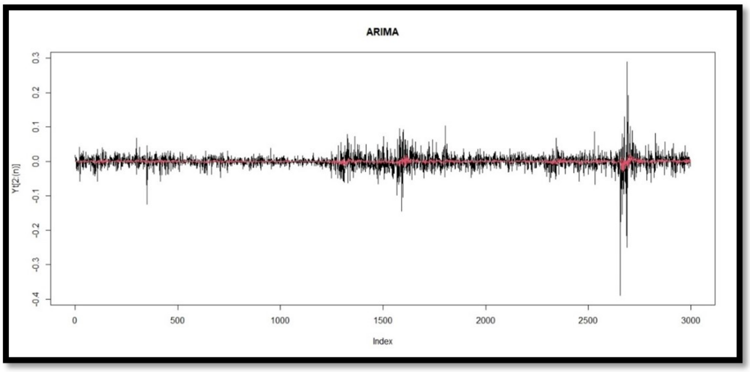

Figure 3: The real and predictive values of the (ARIMA) model

The Result of the First Proposed Hybrid Model for (ARIMA-SVR)

After using the ARIMA model and calculate the predicted values. Building the (SVR) model to model the residuals of the ARIMA (5,1,3). The input variables of the (SVR) model were determined through the trial and error of lags, the SVR-kernel is (RBF), the (C) has been set (1000) and the value of Gamma in the (RBF) as (10) and the Epsilon is (0.1).

After estimating the (SVR) model and obtaining its predictive values, the predictive values of the (ARIMA) model are combined with the predictive values of the (SVR) model, to determine the amount of error in the prediction between the actual values and the predictive values for the hybrid model, it is done using the (MSE) comparison measure, and the results are as follows:

Table 22: The MSE for prediction accuracy of the first proposed hybrid model (ARIMA-SVR)

The Model	MSE
ARIMA-SVR	0.0002522791

The predicted values of the first proposed hybrid model (ARIMA-SVR) are obtained by the estimates of the first independent variable, which represents the predicted values of the (ARIMA) model and obtaining the estimates of the second independent variable, which represents the predicted values of the (SVR), estimate the parameters by the (OLS) method and the first proposed hybrid model from which the final predicted values are obtained is:

$$\hat{y}_t = b_0 + b_1\hat{x}_{1(ARIMA)} + b_2\hat{x}_{2(SVR)}$$

The equation of the (ARIMA-SVR) can be written as follows:

$$\hat{y}_t = -0 \cdot 0002772 + 0 \cdot 7025015\ \hat{x}_{1(ARIMA)} + 0 \cdot 9278395\ \hat{x}_{2(SVR)}$$

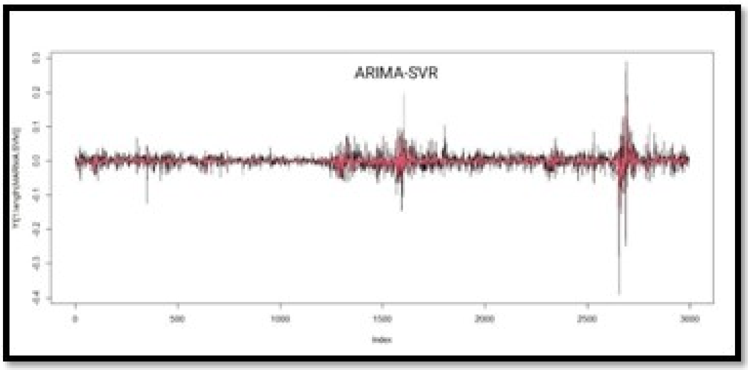

Figure 4: The real and predictive values of the first proposed hybrid model (ARIMA-SVR)

The Box-Pierce and Ljung Box tests was done for the residuals of the (ARIMA-SVR) hybrid model, and the following results were obtained:

Table 23: The Box-Pierce test for the first proposed hybrid model (ARIMA-SVR)

Lag	1	5	10
Chi – Square	2.2579	9.873	15.612
DF	1	5	10
p-value	0.1329	0.07891	0.1113

Table 24: The Ljung Box test for the first proposed hybrid model (ARIMA-SVR)

Lag	1	5	10
Chi – Square	2.2602	9.8899	15.649
DF	1	5	10
p-value	0.1327	0.07842	0.1101

In the above tables (23) & (24), it is clear that the Box-Pierce and Ljung Box tests for the residuals of the hybrid model is random, this means that the variables are independent and there is no correlation between them.

The Result of the Second Proposed Hybrid Model for (ARIMA-ANN [1])

The best second proposed hybrid model was obtained from the simulation is (ARIMA-ANN [1]) which will be applied to daily oil price data and after using the ARIMA (5,1,3) model, and calculate the predicted values. Building the (ANN) model where the network inputs are the previous observations and then they are modeled and the predictive values are obtained.

The data in the neural network is processed using the Min Max Normalized Method, the number of input nodes at the input level is determined by the rank of Box Jenkins model, the trial-and-error method was used to determine the number of hidden nodes in the first and the second hidden layers in the hidden level. One output node has been identified. The number of epochs was set at)500 (epochs in the process of training the neural network, the learning rate and the momentum factor value are (0.5) and (0.8), respectively.

After obtaining the optimal model of the (ANN [1]), it was used to obtain the predictive values, the amount of prediction error between the actual values and the output of the (ANN [1]) was determined using the (MSE), the results were as follows:

Table 25: The MSE for prediction accuracy of the (ANN [2]) model

The Model	MSE
ANN [2]	0.000523413

To determine the amount of error between the actual and predictive values of the second proposed hybrid model (ARIMA-ANN [1]), the (MSE) was used, the results were as follows:

Table 26: The MSE for prediction accuracy of the second proposed hybrid model (ARIMA-ANN [2])

The Model	MSE
ARIMA-ANN [2]	0.0002540483

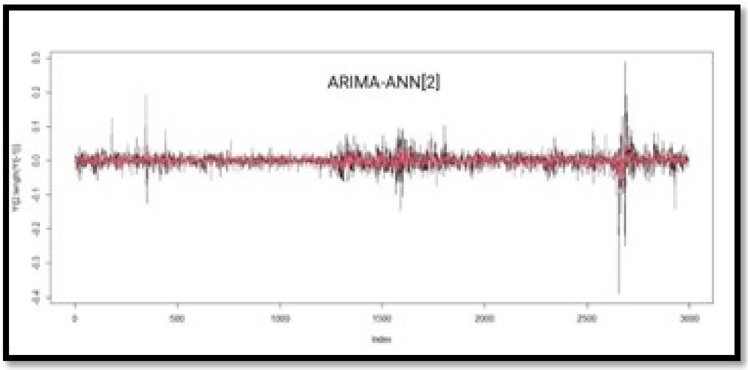

Figure 5: The real and predictive values of the second proposed hybrid model (ARIMA-ANN [1])

The predicted values of the second proposed hybrid model (ARIMA-ANN [1]) are obtained by the estimates of the first independent variable, which represents the predictive values of the (ARIMA) model and obtaining the estimates of the second independent variable, which represents the predicted values of (ANN [1]) and estimate the parameters by the (GRR) method, the second proposed hybrid model from which the final predicted values are obtained is:

$$\hat{y}_t = b_0 + b_1\hat{x}_{1(ARIMA)} + b_2\hat{x}_{2(ANN[2])}$$

The ridge parameter (K_{HKB}) has been obtained (0.001043597), the equation of the (ARIMA-ANN [1]) can be written as follows:

$$\hat{y}_t = -0 \bullet 0000061 + 0 \bullet 4847805\ \hat{x}_{1(ARIMA)} + 0 \bullet 4941542\ x2$$

The Box-Pierce and Ljung Box tests was done for the residuals of the (ARIMA-ANN [1]) hybrid model, and the following results were obtained:

Table 27: The Box-Pierce test for the second proposed hybrid model (ARIMA-ANN [2])

Lag	1	5	10
Chi − Square	1.0694	3.0072	13.27
DF	1	5	10
p-value	0.3011	0.6989	0.209

Table 28: The Ljung Box test for the second proposed hybrid model (ARIMA-ANN [2])

Lag	1	5	10
Chi − Square	1.0704	3.0125	13.312
DF	1	5	10
p-value	0.3008	0.6981	0.2067

In the above tables (27), (28), it is clear that the Box-Pierce and Ljung Box tests for the residuals of the hybrid model is random, this means that the variables are independent and there is no correlation between them.

10 The Conclusions

The most important conclusions reached through the results in the simulation are:

1. When applying sample sizes (250) and (500) to the hybrid models, the best model for the first proposed hybrid model is (ARIMA-SVR) & (SVR-ARIMA), the best second proposed hybrid model is (ARIMA-SVR).
2. When applying the sample sizes (1500) to the hybrid models, the best model for the first proposed hybrid model is (ARIMA-SVR) & (SVR-ARIMA) and for the second proposed hybrid model is (ARIMA-ANN [1]).

3. When applying the sample sizes (3000) the best hybrid models for the first proposed hybrid model is (ARIMA-SVR) & (ARIMA- ANN [1]) and for the second proposed hybrid model is (ARIMA-ANN [1]).

The most important conclusions reached through the results in the application are:

1. The best model (ARIMA) was obtained as it obtained the lowest value of the evaluation criteria for the candidate models (AIC, HQC and AICc), which is ARIMA (5,1,3) and it was found that the residuals of the model were random.
2. The residuals for first proposed hybrid model for (ARIMA-SVR) and the second proposed hybrid model for (ARIMA-ANN [1]) were analyzed and it was found the residuals are random according to the Box Pierce and Ljung Box tests.
3. When comparing the proposed hybrid models, it was found that the best hybrid model is the first proposed hybrid model (ARIMA-SVR) according to the (MSE), so it is preferred when using it to forecast daily crude oil prices.

References

[1] M.A. Ashour, *Artificial neural networks and time series prediction methods*, 1st ed., Memory for Publishing and Distribution, 2018.

[2] A. Gökhan, C.O. Güzeller, and M.T. Eser, *The effect of the normalization method used in different sample sizes on the success of artificial neural network model*, Int. J. Assess. Tools Educ. **6** (2019), no. 2, 170–192.

[3] X. Hu, *Support vector machine and its application to regression and classification*, MSU Graduate Theses, Missouri State University, 2017.

[4] A.H. Kadim, *Introduction to economic measurement*, 2009.

[5] G. Khalaf and M. Iguernane, *Ridge Regression and Ill-Conditioning*, J. Mod. Appl. Statist. Meth. **13** (2014), no. 2, 355–363.

[6] D.C. Montgomery, C.L. Jennings, and M. Kulahci, *Introduction to time series analysis and forecasting*, John Wiley & Sons, 2015.

[7] H. Mori and D. Kanaoka, *Application of support vector regression to temperature forecasting for short-term load forecasting*, Int. Joint Conf. Neural Networks, IEEE, 2007, pp. 1085–1090.

[8] A. Sadiq, *Intelligent systems and machine learning*, Al-thakera, Baghdad, 2016.

[9] A.T. Sadiq, *Machine learning methods and algorithms*, Al-thakera, Baghdad, 2021.

[10] S. Sivanandam, *Introduction to artificial neural networks*, Vikas Publishing House, 2009.