Comparing Single Vs. Hybrid models in Time Series Forecasting

By

Dr. Abdel Rahim Awad Bassiouni

PhD Statistics, Faculty of Commerce, Tanta University

dr-abdelreheembassuny@outlook.co

Dr. Maha Farouk Tawfik Ibrahim

Department of Statistics, Mathematics, and Insurance-Faculty of Commerce – Tanta University, Tanta-Egypt

Maha.ibrahim@commerce.tanta.edu.eg

Scientific Journal for Financial and Commercial Studies and Research (SJFCSR)

Faculty of Commerce – Damietta University

Vol.5, No.2, Part 1., July 2024

APA Citation:

Bassiouni, A. R. A. and **Ibrahim**, M. F. T. (2024). Comparing Single Vs. Hybrid models in Time Series Forecasting, *Scientific Journal for Financial and Commercial Studies and Research*, Faculty of Commerce, Damietta University, 5(2)1, 735- 760.

Website: https://cfdj.journals.ekb.eg/

Comparing Single Vs. Hybrid models in Time Series Forecasting

Dr. Abdel Rahim Awad Bassiouni and Dr. Maha Farouk

Abstract:

The research aims to forecast time series relying on individual models SVR, ARIMA, and the hybrid model "ARIMA-SVR" through different integration methods applied to global oil price data from January 2004 to December 2023, comprising monthly data with 240 observations and compare its results to identify the best model for forecasting global oil price. The integration methods include the additive hybrid model, the multiplicative hybrid, and the regression hybrid model as hybrid models comparing with single models SVR, and ARIMA models. The results showed that the additive hybrid model, ARIMA-SVR Additive is the best model among all models under studying, as it provides the lowest values of prediction accuracy metrics: MAE, MPE, MAPE, MSE. Using the Ljung-Box test for the resulting series it has the first ranking. The additive hybrid model, ARIMA-SVR Additive as the best model for modeling global oil price data is followed by the regression hybrid model, then the multiplicative hybrid model, SVR, and finally ARIMA.

Keywords: Hybrid Model (ARIMA–SVR), Additive Model, Multiplicative Model, Hybrid Regression Model.

1. Introduction:

Forecasting of specific phenomenon taking a time series shape can be is done using various statistical methods. Due to the fluctuations and the presence of both linear and non-linear patterns in time series, dealing with such series using either a linear or non-linear model alone may not accurately reflect the behavior of the series. This has led to the emergence of a modern technique known as "Hybrid Models" (Zhang, 2003) .ARIMA is dealing with a linear model, SVR is dealing with a non-linear model, but the hybrid models combine a linear model with a non-linear model to create a new model. By using this hybrid model, more accurate predictions can be obtained for time series than using either linear or nonlinear models.

Comparing forecasting with the single models (ARIMA and SVR) and the hybrid models were used in many studies:

Zhang, Zhou (2024) proposed a combining hybrid model, the autoregressive integrated moving average (ARIMA), support vector regression (SVR), which uses a combination of and peak over threshold (POT) method from the extreme value theory (ARIMA-SVR-POT), and compared its performance with three other models, namely ARIMA-EGARCH, ARIMA-SVR, and ARIMA-EGARCH-POT. The proposed model provides a more precise reflection of potential losses when estimating VaR. El Malkey, et al. (2023) compared ANN model and ARIMA model with the hybrid model "ARIMA-ANN" for predicting the EGX 30 stock index. The results concluded that the hybrid model performed better in terms of prediction accuracy measures. Nassar, et al (2023), compared three models for predicting monthly time series data of global gold prices, the auto regressive integrated moving average based on discrete wavelet transform (DWTARIMA), the support vector regression based on discrete wavelet transform (DWT-SVR) and hybrid model by combining the ARIMA model with the SVR model based on discrete wavelet transform (DWT-ARIMA-SVR). The results showed the superiority of DWT-ARIMA-SVR on other models with lower values of prediction accuracy measures, MAE, MSE, and Theil's U statistic.

Similarly, Sareminia (2023) proposed (SVM-ARIMA-3LFFNN hybrid model) focused on predicting time series data of spare parts consumption. Compared with a single models SVR and ARIMA, the proposed model improved the RMSE, MAPE, and sMAPE (up to 30% improvement).

Alsuwaylimi (2023) compared the hybrid models (ARIMA-ANN), ANN, and ARIMA for forecasting daily oil prices using real-world datasets. The results showed that the hybrid model performed better, and it was the more appropriate model. Also, Pannakkong et al. (2022) compared individual models multiple linear regression, SVM, ANN, and ARIMA with hybrid model (ARIMA-SVM) artificial neural network (ANN), support vector machine, hybrid models, and ensemble models, are implemented, using daily electricity consumption time series in Thailand. The results concluded that the hybrid model outperformed the others based on prediction quality criteria MAPE and MSE. Al Reweili and Fawzy (2022) compared individual models ANN and ARIMA with the hybrid model "ARIMA-ANN" for petroleum price forecasting. The study found that the hybrid model performed better. Lai et al. (2021) utilized SVR, ARIMA, and their hybrid model to predict unemployment rates in five developing and five advanced countries during the COVID-19 pandemic. The study showed the superiority of the hybrid model (ARIMA-SVR). Zheng et al (2021) focused on predicting daily coal prices using hybrid models ARIMA-SVR compared to individual models SVR and ARIMA in a Chinese port. The study found that the hybrid model performed better. Nawi et al. (2021) compared individual models SVM and ARIMA with their hybrid model for forecasting sea surface temperature (SST) data. The study concluded that the hybrid model was superior. Ayub & Jafri, (2020) compared individual models ANN and ARIMA with the hybrid models ANN -ARIMA applied on KSE viz data. results obtained show the excellence of Hybrid NN-ARIMA model over ANN and ARIMA. Xu et al. (2020) focused on predicting drought rates in water stations in Henan province, China, using hybrid models. The study found that the hybrid models outperformed individual models ARIMA and SVR.

All previous studies have focused on comparing individual models with hybrid models, which result from combining or integrating the linear and nonlinear parts of the time series. However, the current study aims to predict based on individual models, namely Autoregressive Integrated Moving Average (ARIMA) and Support Vector Regression (SVR), as well as the hybrid model combining them using different integration methods, the additive hybrid model, the multiplicative hybrid model, and the regression hybrid model. This will be applied to a monthly time series of global oil prices.

The remaining parts of the research will address the methodology in the second section, the applied study in the third section, and the results and recommendations in the fourth section.

2. Methodology

In this research, we focus on the individual models SVR and ARIMA, as well as the hybrid model "ARIMA - SVR" in Time Series Forecasting applied to a monthly time series of global oil prices and compare its results.

2.1 Autoregressive Integrated Moving Average (ARIMA) Model:

The ARIMA model is generated by combining the autoregressive model AR(p) and the moving average model MA(q) after taking the necessary differences to make the time series stationary. This model is denoted as ARIMA (p, d, q) (Box, et al (2016))

$$y_t = \emptyset_0 + \emptyset_1\, y_{t-1} + \cdots \ldots + \emptyset_p\, y_{t-p} + e_t - \theta_1 e_{t-1} \ldots - \theta_q e_{t-q} \quad (1)$$

Where;

$\emptyset_1, \emptyset_2 \dots \dots, \emptyset_p$: represent the parameters of the autoregressive model.

p: is the order of the autoregressive model.

$\theta_1, \theta_2 \dots\dots \theta_q$ .represent the parameters of the moving average model.

q :is the order of the moving average.

e_t .represents the random error.

When the time series is non-stationary, it can be transformed into a stationary series by taking differences of order (d). In this case, the ARIMA model is denoted as ARIMA (p, d, q), and its equation is written as:

$$\emptyset_p(B)\nabla^d y_t = \theta_q(B)e_t \qquad (2)$$

$$\emptyset_p(B)(1-B)^d y_t = \theta_q(B)e_t \quad (3)$$

Where.

d: is the differencing order for achieving stationarity in the time series.

B: is the backward shift operator.

Box-Jenkins stages.

The Box-Jenkins methodology consists of four main stages for obtaining the optimal model for forecasting the studied time series (Box, et al (2016) as follows:

1. **Identification**

The aim of this stage is to determine the model parameters p, d, q. The number of differences required for achieving stationarity in the time series, d, is determined through the Dickey-Fuller and Phillips-Perron tests. The values of p and q are determined using the autocorrelation function (ACF), and partial autocorrelation function (PACF).

To choose the best time series models, some criteria: Akaike information criterion (AIC), and the Bayesian information criterion (BIC), can be used and it can be calculated as follows:

$$AIC = n\, Ln\,(\hat{\sigma}^2 n) + 2m \qquad (4)$$

Where;

m: Number of model parameters,

n: Number of observations,

$\hat{\sigma}_n^2$: Variance of the residuals,

$$\text{Bic}(m) = n\,\text{Ln}\,(\hat{\sigma}_n^2) + m\,\text{Ln}\,(n) \qquad (5)$$

Therefore, the best model is the one with the lowest both BIC(m) and AIC(m) values.

2. **Estimation**

After identifying the proposed model which represent the time series data, the model parameters are estimated. There are several methods for parameter estimation, with the most important ones being the Maximum Likelihood method and the Ordinary Least Squares method.

3. **Diagnostic stage**

For Diagnostic the proposed model represents the time series data in the previous stages, checking the suitability of the model for the data is determined through analyzing the residuals obtained from applying the model. The residuals should be randomly distributed, and this is assessed based on the that assumptions $H_0: \rho = 0$, $H_1: \rho \neq 0$,where: ρ is the correlation coefficient between the residuals' units.

The goodness of fit of the model to the data is tested through two tests:

a) **Confidence Interval Test:**

If the autocorrelation coefficient of the residuals $r_k(e_t)$ falls within the confidence interval with a probability of 0. 95,

$$-1.96\frac{1}{\sqrt{n}} \leq r_k(e_t) \leq 1.96\frac{1}{\sqrt{n}},$$

Then the errors are normally distributed, the model is a good fit for the data.

b) **Ljung – Box (Q) statistic:**

4. To test if the proposed model is not suitable for the data, the test is done as follows:
5. H_0: the model fits the data well.
6. H_1: the model does not fit the data well.
7. Test statistic:

$$Q_m = n(n+2)\sum_{k=1}^{m}\frac{r_k^2(e)}{n-k} \qquad (6)$$

$r_k(e)$: The autocorrelation coefficient for the residuals at lag (k).

n: represent the number of errors.

m: denote the number of lag periods.

If the P-value < α, then the model is not suitable for the data, and thus another model should be chosen.

8. Forecasting

Forecasting is the final stage of Box-Jenkins models and represents the primary objective of model building.

2.2. Support Vector Regression model (SVR)

The Support Vector Regression (SVR) model is a powerful method for analyzing and forecasting data, distinguished by its high capability to handle non-linear data. The SVR model has been widely used for predicting time series data, which often exhibit non-linear patterns, such as financial and economic data. The idea behind SVR involves classifying the input data and separating it from each other through the optimal Hyperplane, regardless of whether the data is separable or not. The general formulation is as follows (Cao and Tay; 2003):

$$f(x,w) = w^t \varphi(x_i) + b \qquad (7)$$

Where;

w: is the regression coefficients vector.

$\varphi(x_i)$: plots in a high-dimensional space.

b: is the bias term, and prediction error is made using the loss function f(x). (Rosenbaum et al; 2013)

One of the main objectives of SVR is prediction and classification. Its concept revolves around separating two different sets of data with more than one line. The optimal line, or Hyperplane, is the one that maximizes the distance between the nearest point in the first set and the nearest point in the second set, and these points are called support vectors.

The decision boundaries expand, and the larger the distance between the nearest points in the first and second sets, the better the classification of new observations in one of the sets. The following equation represents the loss function:

$$L(y, f(w,x)) = \begin{cases} 0 & |f|y_i - f(w,x)| \le \epsilon \\ |y_i - f(w,x)| - \epsilon & other\ wise \end{cases}$$

The objective of using SVR is to find the function f (w, x) that agrees with the deviation ϵ for all training data in the prediction model. The problem is formulated as follows:

$$\text{Minimize: } \frac{1}{2}\ \|w\|^2$$

$$\text{Subject to } \begin{cases} y_i - f(x_i, w) - b \le \epsilon + \delta_i \\ f\ (x_i, w) + b - y_i \le \epsilon + \delta_i \end{cases}$$

Sometimes it's not possible to predict all training data within the deviation ϵ. Therefore, slack variables are introduced. Consequently, the problem is reformulated as follows:

$$\text{Minimize: } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\delta_i + \delta_i^*)$$

$$\text{Subject to } \begin{cases} y_i - f(x_i, w) - b \le \epsilon + \delta_i^* \\ f(x_i, w) + b - y_i \le \epsilon + \delta_i \end{cases}$$

$$\delta_i^*, \delta_i \ge 0$$

Where;

C: is the regularization parameter, where C > 0, also known as the regularization term.

δ_i^*, δ_i: are variables that measure deviations larger than ϵ.

The Lagrange equation is used to solve the problem as follows:

$$\text{f(x)} = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\varphi(x_i, x) + b \qquad (8)$$

Where;

$(\alpha_i - \alpha_i^*)$: Lagrange multipliers.

and by using the kernel function, the solution becomes.

$$\text{f(x)} = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)k(x_i, x) + b \qquad (9)$$

Where;

$k(x_i, x)$: This function is known as Kernel, and it is used when linear separation of data is challenging. The data is transformed from the original space to a high-dimensional space, allowing for better data separation. (Cosgun et al.; 2011).

2.3 Hybrid Model:

The time series in Hybrid Model is assumed to be as follows:

$$y_t = L_t + N_t \qquad (10)$$

Where;

L_t: The linear component of the time series.

N_t: The nonlinear component of the time series.

The merging process is carried out in various ways as the Additive Hybrid Model (ARIMA-SVR), the Multiplicative Hybrid Model (ARIMA – SVR) and the Hybrid regression model (ARIMA – SVR) as follows.

2.3.1 Additive Hybrid Model (ARIMA-SVR)

The construction of the Additive hybrid model follows these steps (Zhang, 2003):

1. Building the ARIMA model starting from the identification phase to prediction.
2. Obtaining the forecasted values from ARIMA to represent the linear part ($\hat{L}_t$).
3. Constructing the SVR model to model the residuals resulting from ARIMA.

Obtaining the forecasted values from the SVR model to represent the non-linear part ($\widehat{N}_t$), and the Additive hybrid model is formulated as and drown as shown in table (1)

2.3.2. Multiplicative Hybrid Model (ARIMA – SVR):

The multiplicative hybrid model (wang, et al; 2013) assumes that the series consists of two parts, linear and nonlinear, and the Additive hybrid model is formulated and drown as shown in table (1)

1. Obtaining the predicted values from ARIMA to represent the linear part ($\hat{L}_t$).
2. The nonlinear part is obtained as follows:
$$n_t = y_t / \hat{L}_t$$
3. It is estimated using an SVR model to obtain ($\widehat{N}_t$) to represent the nonlinear part.

The multiplicative Additive hybrid model is formulated and drown as shown in table (1)

2.3.3. Hybrid regression model (ARIMA – SVR)

The hybrid regression model is constructed relying on the predicted values from ARIMA and SVR as individual models to perform a multiple linear regression model to determine the fusion weights as follows: (Leenawong, Chaikajonwat ,(2022).

$$y_t = \beta_0 + \beta_1 F_{ARIMA} + \beta_2 F_{SVR}$$

Where;

y_t: Actual data (dependent variable).

F_{ARIMA}: Predicted values from ARIMA model.

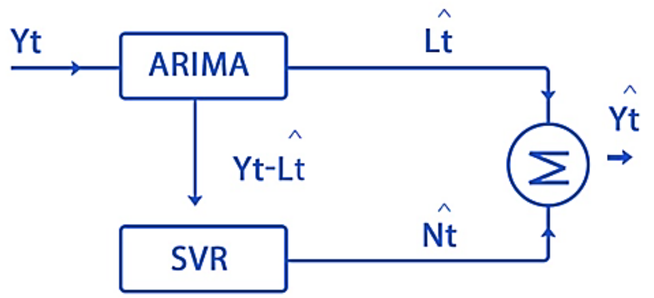
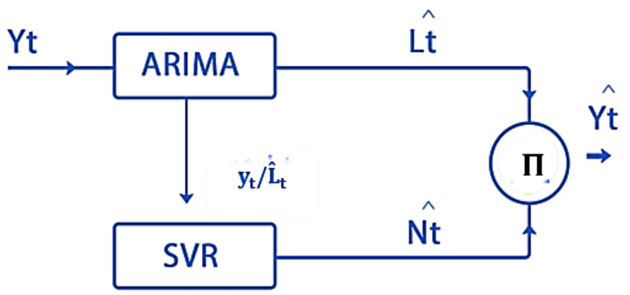
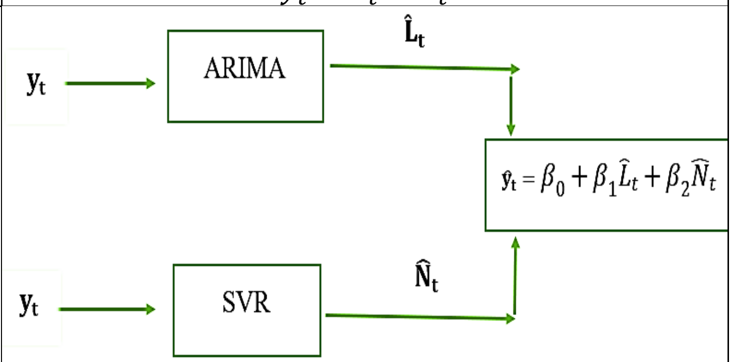
F_{SVR}: Predicted values from SVR model.

β_1: First linear regression coefficient (first weight).

β_2: Second linear regression coefficient (second weight).

Considering $\beta_0 = 0$, the regression model will consist of the dependent variable being the actual values of global oil prices and the independent variables being the predictions resulting from using SVR and ARIMA. the hybrid regression model is drawn as shown in table (1)

Table (1), The additive hybrid model, Multiplicative hybrid model and Hybrid regression model

The additive hybrid model	 source: Zhang; 2003
formula	$\hat{y}_t = \hat{L}_t + \hat{N}_t$
Multiplicative hybrid model	 source: Zhang; 2003
formula	$\hat{y}_t = \hat{L}_t * \hat{N}_t$
Hybrid regression model	 source: **Researcher's Preparation**
formula	$y_t = \beta_0 + \beta_1 F_{ARIMA} + \beta_2 F_{SVR}$

2.4. Model Accuracy and Efficiency Metrics

The efficiency of the performance of the models is measured to test the best model according to several statistical criteria, which are: (Wei, 2006):

1. **Mean square error (MSE):** $MSE = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2$
2. **Mean absolute percentage error:** $MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{y_t} \times 100$
3. **Mean Absolute error:** $MAE = \sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{n}$

3. applied study.

The research aims to forecast time series relying on individual models SVR, ARIMA, and the hybrid model "ARIMA-SVR" through different integration methods applied to global oil price data from January 2004 to December 2023, comprising monthly data with 240 observations. 216 observations were utilized as the training set, representing 90% of the total data, while the remaining 24 observations served as the test set, accounting for 10% of the data. This was done to compare actual and estimated values to test the model's predictive ability. The analysis was conducted using the software programs EViews 12 and Stat graphics 19 for analyzing the time series data of global oil prices.

3.1: Autoregressive Integrated Moving Average (ARIMA) Model:

To construct a better model for forecasting and estimating its parameters and to ensure the model's suitability for time series data of global oil prices using (ARIMA) Model, it is necessary to analyze the series according to the following steps:

1- Time Series Stationarity:

Stationarity is a fundamental concept in time series analysis. The Box-Jenkins methodology cannot be applied to analyze time series unless the series under study is stable. The first assumption to be tested is the stationarity of the series. Therefore, before applying the methodology, data must be prepared to verify the stationarity of the time series, either graphically by plotting the original series or through the graphical representation of the autocorrelation and partial autocorrelation functions, or through other statistical methods such as the Augmented Dickey-Fuller (ADF) test, Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test.

Figure (1) shows the plotting of the original series.

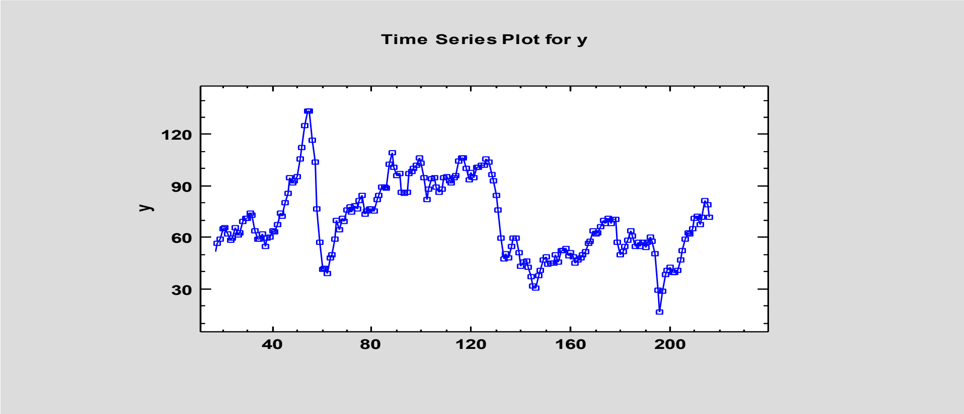
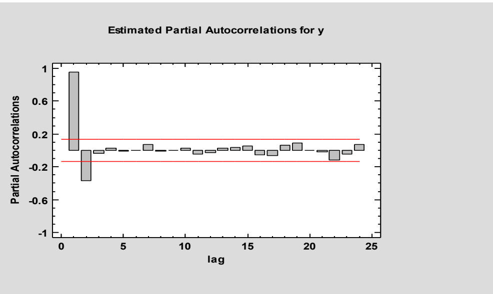


Figure (1): Time Series of Global Oil Prices

Figure (1) shows that there is no general trend in the time series of global oil prices, and therefore the series is stable over time. To identify the order of integration (d) through the shape of the autocorrelation function (ACF) and the autocorrelation coefficient (ACF)as shown below:



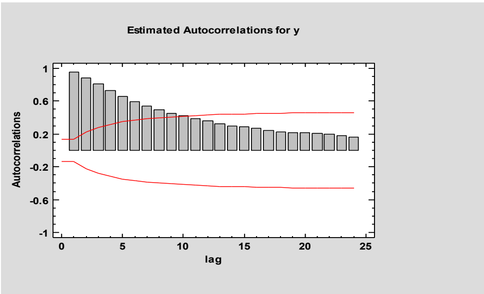


Figure (2) PACF **Figure (3) ACF**

Figures (2), and (3) showed that the ACF values are significant for many lags and the series does not decay slowly, indicating that the series is stationary at its original level. This means that it is integrated as zero order I (0). To further confirm the stationarity and non-stationarity of the time series, the following unit root tests were performed (ADF, PP, KPSS) as shown in table (2)

Table (2), unit root tests

Test	Value	P – value	I (d)
ADF	-3.2859	0.0167	I (0)
PP	-2.8690	0.0406	I (0)
KPSS	0.34425	0.34425	I (0)

From table (2), we observe that in both the ADF and PP tests, the p-value is less than the significance level of 0.05. Therefore, we reject the null hypothesis, which means that the global oil price time series is stationary at its original level. However, in the KPSS test, the p-value is greater than the significance level of 0.05, so we accept the null hypothesis that the series is stationary at its original level, I (0).

2- Model Estimation:

After confirming the stationarity of the time series for global oil prices at the original level and observing the autocorrelation function, the proposed model is ARIMA (2,0,0), and its equation is written as follows:

$$y_t = \mu + \emptyset_1 \, y_{t-1} + \emptyset_2 \, y_{t-2} + e_t$$

The model parameters are estimated as follows:

Table (3): Estimation of the Proposed ARIMA (2,0,0) Model

Parameter	Estimate	St. error	P – value
AR (1)	1.31571	0.06255	0.000
AR (2)	-0.368983	0.06246	0.000
Mean	68.67037	6.42035	0.000
Constant	3.65857		

Therefore, the **ARIMA (2,0,0)** model is formed as:

$$\boldsymbol{\hat{y}_t = 3.65857 + 1.31571 y_{t-1} - 0.368983 y_{t-2}}$$

3- Model Diagnosis tests.

After estimating the model parameters and determining its order, it is essential to verify the model's adequacy and efficiency. This is done by calculating the residuals' autocorrelation coefficients and conducting the following tests.

a) Ljung-Box Test:

The results of the test were as shown in table (4)

Table (4) Ljung – Box

Test	Statistics (Q)	P – value
Ljung – Box	15.7789	0.8267

Through table (4), the p-value for the test is greater than the significance level of 0.05. Therefore, we accept the null hypothesis $H_0: \rho_i(e_t) = 0$,indicating that the errors are random and uncorrelated. Consequently, the model is adequate and efficient for representing the time series of global oil prices.

b) Residual Test:

By plotting the partial autocorrelation and the autocorrelation functions of the residuals as shown in figure (4), and figure (5), we notice that all the autocorrelation coefficients of the residuals fall within the 95% confidence interval. This indicates that the residuals are uncorrelated random variables, and the model used is good and represents the data effectively.

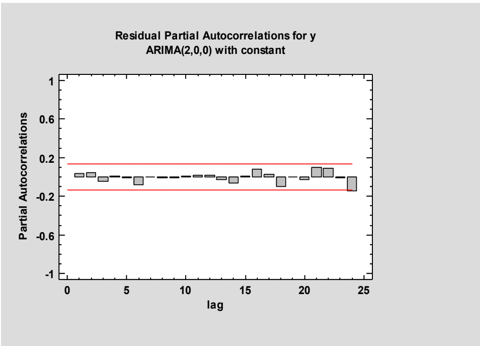


Figure (4): PACF of the Residuals

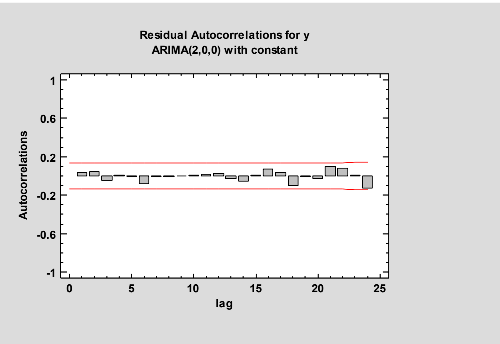


Figure (5): ACF of the Residuals

4- Forecasting:

After the model passed the diagnostic tests, indicating the model's suitability for forecasting global oil prices for 24 observations, which constitute the test set, the forecasting results were as shown in Table (9) and the following figure illustrates the estimated and actual values for the ARIMA (2,0,0) model:

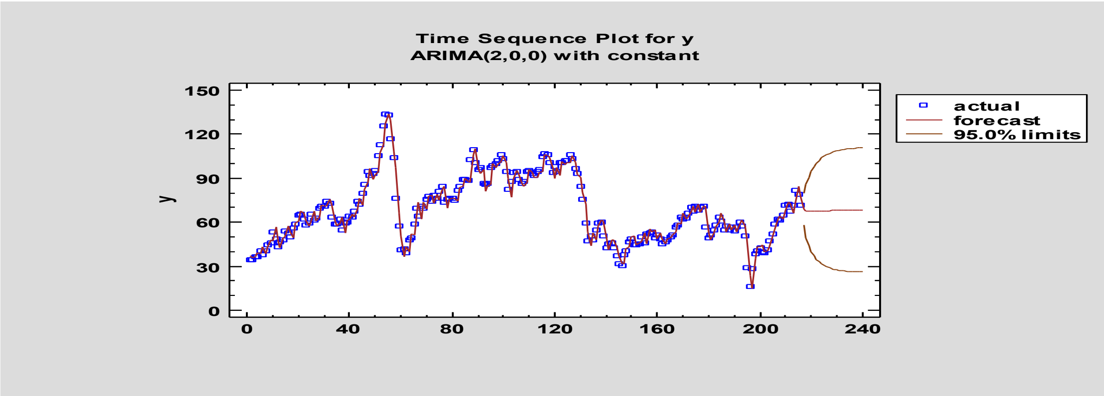


Figure (6): Plot of Estimated and Actual Values for ARIMA (2,0,0).

Figure (6) shows that the performance of the ARIMA model in the forecasting process is unsatisfactory for monthly global oil prices.

3.2. Support Vector Regression (SVR) model.

To estimate the SVR model for the global oil price series, the data was divided into two sets: the training set, consisting of 216 observations, which accounts for 90% of the total data, and the test set, consisting of 24 observations, accounting for 10% of the total data. descriptive statistics for both the training and test sets are presented as follows:

Table (5) Descriptive statistics

Statistics	Value (Training)	Value (Test)
n	٢١٦	٢٤
Mean	٨٨,٢٧٠	٨٦,٢٠٣٨
Median	٦٣,٨٣	٨٣,٧٤
Min	١٦,٥٥	٧٠,٢٥
Max	١٣٣,٨٨	١١٤,٨٤
Standard Division	٢٢,٤٧٧٥٢	١٢,٨٣١٥٩
Kurtosis	-٠,٤٤٣	-٠,١٩٤
Skewness	٠,٤٤٥	٠,٨٦٨

From table (5) and through the skewness coefficient value, we noticed that there is a slight positive skewness, and therefore the data is close to the normal distribution.

To estimate the model, it is necessary to determine the appropriate type of Kernel function for the model as shown in table (6).

Table (6): Types of Kernel Functions

Measure	Linear	Polynomial	RBF	Sigmoid
MSE	0.00000037	409.42	57.54	197.53
R^2(Training)	100%	18.588%	88.55%	60.63%

From table (٦), the linear Kernel function was selected for predicting global oil prices as it had the lowest mean squared error and the highest coefficient of determination (R-squared) for the training set. After conducting several trial tests on the training data, C=1.0 and ε=0.1 were chosen as the best parameters due to their lowest mean squared error values. These parameters were used to train the model, and predictions were made for the test set consisting of 24 observations from January 2022 to December 2023 using Stat graphics 19 software.

Using linear (SVR) model, the convergence between the estimated and actual values, is as illustrated in figure (7):

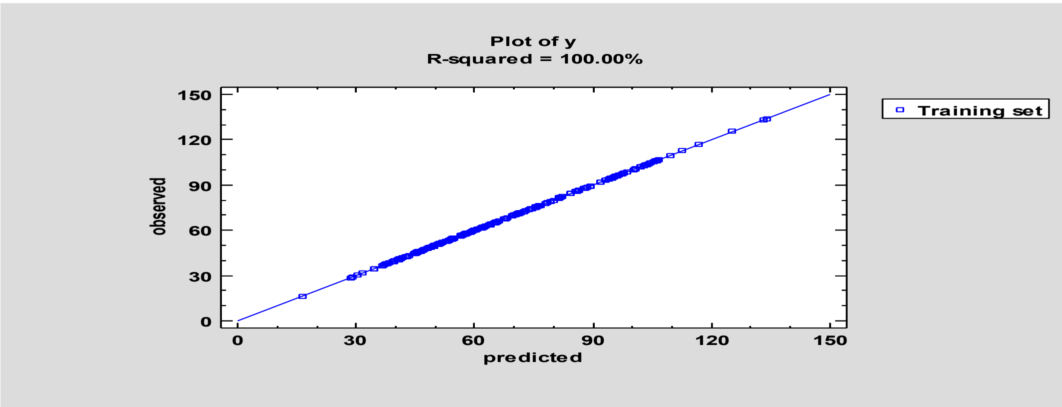


Figure (7): Estimated and Actual Values for the SVR Model

From figure (7), the performance of (SVR) model in forecasting is better than ARIMA model but it still unsatisfactory for data. To test the adequacy of the model for monthly time series data of global oil prices using the Ljung-Box test, the results are as follows:

Table (7) Ljung-Box Test

Test	Q	P – value
Ljung – Box	11.321	0.7125

Based on the results of the Ljung-Box test, the p-value equals 0.7125 >0.05. Therefore, we accept the null hypothesis that the errors are random and uncorrelated, indicating that the model is suitable and effective for representing the monthly global oil price series. Additionally, the autocorrelation coefficients fall within the 95% confidence interval, suggesting no significant autocorrelation among the residuals, further evidence of the quality of the SVR model. The predicted values of ARIMA and SVR model are shown in Table (11).

3.3. The Hybrid Model (ARIMA-SVR)

To estimate the hybrid model (ARIMA-SVR), The integration methods include the additive hybrid model, the multiplicative hybrid, and the regression hybrid model is used to identify the best model to fit the oil data.

3.3.1. The Additive Hybrid Model (ARIMA-SVR):

To estimate the Additive hybrid model (ARIMA-SVR**)**, first, estimated values from the ARIMA model representing the linear part of the series ($\widehat{L}_t$) are obtained. Then, an SVR model is constructed based on the residuals from ARIMA, and the predicted values from the SVR model represent the nonlinear part ($\widehat{N}_t$).

Then, the Additive hybrid model is used to forecast 24 observations, which constitute the test set, as illustrated in table (12). Through comparing the predicted values with the actual ones using the Additive hybrid model, we noticed a strong convergence between them. This is evident in Figure (8), indicating a significant improvement in the performance and predictive ability of the hybrid model for monthly global oil prices.

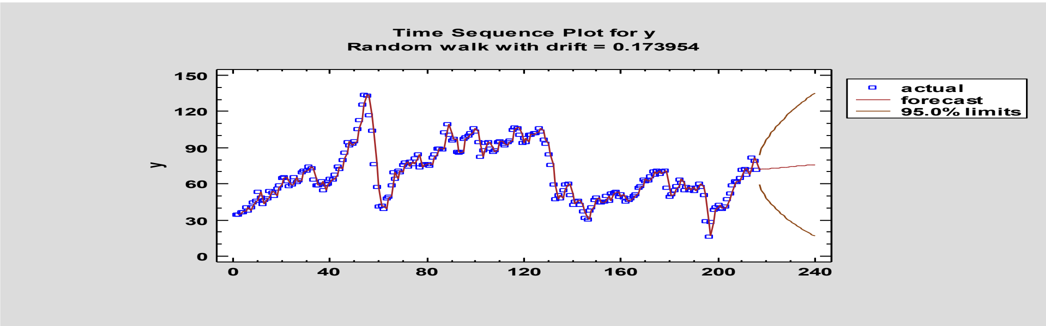


Figure (8): Actual and Estimated Values using the Additive Hybrid Model Additive (ARIMA-SVR)

From Figure (8), the performance of **Additive (ARIMA-SVR)** model in forecasting is better it still unsatisfactory for data. To ensure the suitability of the model for the global oil price series data, the Ljung-Box test was conducted, and the results are as shown in table (8):

Table (8): Ljung-Box Test

Test	Q	P – value
Ljung – Box	10.132	0.2976

The p-value was found to be (0.2976>0.05. Therefore, the null hypothesis of the **Additive** hybrid model's adequacy in representing the monthly global oil price series is accepted.

3.3.2. The Multiplicative Hybrid Model (ARIMA-SVR)

Using the multiplicative hybrid model to forecast a set of 24 test observations yielded results as shown in table (12). It is noticeable that there is an improvement in the performance of the multiplicative hybrid model in forecasting monthly global oil prices, as depicted in Figure (9).

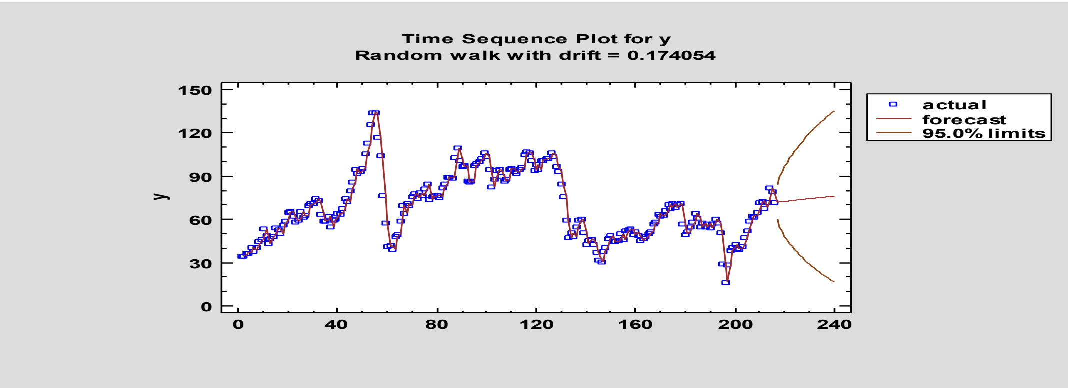


Figure (9): Actual and Estimated Values using the Multiplicative Hybrid Model Multiplicative (ARIMA-SVR)

From figure (9), the performance of Multiplicative (ARIMA-SVR) model in forecasting is closely to Additive (ARIMA-SVR).

Conducting the Ljung-Box test to verify the adequacy and validity of the model in representing the global oil price series, the results are shown in table (9).

Table (9): Ljung-Box Test

Test	Q	P – value
Ljung – Box	13.201	0.6515

Table (9) showed that p-value >0.05, then accepting the null hypothesis that the multiplicative hybrid model is valid and effective in representing the monthly global oil price series.

3.3.3. Hybrid Regression Model (ARIMA – SVR):

This model results from a multiple regression equation representing the actual values of the training set, which consists of 216 observations. The dependent variable is represented by the first independent variable, which is the predicted values by the ARIMA model, and the second independent variable is the predicted values by the SVR model. Therefore, the estimated regression equation without a constant is:

$$\hat{y}_t = 0.989\, F_{ARIMA} + 1.493\, F_{SVR}$$

Thus, the estimated and actual values by the multiple regression hybrid model are shown in Table (13), indicate high performance in the prediction process. This is evident from the convergence between the actual and predicted values, as illustrated in figure (10):

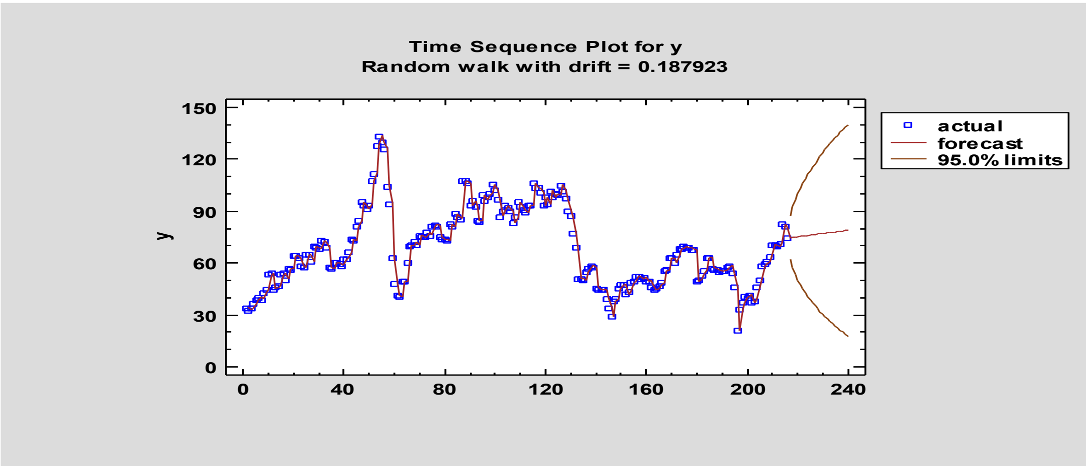


Figure (10): Actual and Predicted Values by the Hybrid Regression Model Regression (ARIMA – SVR)

From figure (10), the performance of **Regression** (ARIMA-SVR) model in forecasting is better than all previous models to Additive (ARIMA-SVR).

Using the Ljung-Box test for the resulting series, the results are as follows:

Table (10): Ljung-Box Test

Test	**Q**	**P – value**
Ljung – Box	7.1253	0.175401

The p-value >0.05, indicating acceptance of the null hypothesis that the hybrid regression model is suitable for the global oil price series data. Table (1٢) shows the Predicted and Actual Values for Hybrid Models.

Table (11), The predicted values of ARIMA and SVR model

date	ARIMA			SVR	
	Actual values	Predicted values	Residuals	Predicted values	Residuals
1/2022	83.32	80.11	3.11	83.15	0.07
2/2022	91.46	90.23	1.23	90.11	1.35
3/2022	108.5	106.71	1.79	104.55	3.95
4/2022	101.78	97.33	4.45	101.11	0.67
5/2022	109.55	102.33	7.22	100.75	8.8
6/2022	114.84	109.12	5.72	111.66	3.18
7/2022	101.62	100.69	0.93	102.01	-0.39
8/2022	93.67	90.53	3.14	90.6	3.07
9/2022	84.37	82.63	1.63	83.33	0.93
10/2022	87.55	89	-1.45	85.31	2.24
11/2022	84.37	80.25	4.12	80.17	4.2
12/2022	76.44	71.21	5.23	77.5	-1.06
1/2023	78.12	77.02	1.1	79.20	-1.08
2/2023	76.83	73.63	3.2	70.6	6.23
3/2023	73.28	70.58	2.7	72.18	1.1
4/2023	79.45	72.96	6.49	79.01	0.44
5/2023	71.58	70.65	0.93	68.88	2.7
6/2023	70.25	65.06	4.59	71.2	-0.95
7/2023	76.07	79.01	-2.94	73.71	2.36
8/2023	81.39	79.08	2.31	80.19	1.2
9/2023	89.43	88.22	1.21	85.34	4.09
10/2023	85.64	80.11	5.53	86.05	-0.41
11/2023	77.69	73.35	4.34	76.10	1.59
12/2023	71.9	73.69	-1.79	65.9	6

From table (11), the statistical predictions for the individual models ARIMA, SVR is close each other. Comparing the actual and predicted values using the SVR model, the performance of the SVR model in forecasting is better than the ARIMA model.

Table (12): Predicted and Actual Values for Single Models and Hybrid Model.

date	(ARIMA-SVR) Additive		ARIMA-SVR Multiplicative		ARIMA-SVR Regression	
	Predicted values	Residuals	Predicted values	Residuals	Predicted values	Residuals
1/2022	82.219	1.001	79.35	3.87	82.05	1.17
2/2022	90.32	1.14	89.32	2.14	89.001	2.45
3/2022	108.93	-0.43	105.36	3.14	99.61	8.89
4/2022	101.76	0.02	99.07	2.71	99.72	2.06
5/2022	105.51	4.04	110.01	-0.46	107.3	2.25
6/2022	112.71	2.13	107.32	7.52	115.9	-1.06
7/2022	101.61	0.01	99.15	2.47	101.69	-0.07
8/2022	91.62	2.05	93.01	0.66	89.01	4.66
9/2022	83.25	1.01	83.16	1.1	84.25	0.01
10/2022	85.54	2.01	86.17	1.38	87.19	0.36
11/2022	83.36	1.01	82.46	1.91	83.45	0.92
12/2022	76.45	-0.01	77.22	-0.78	75.49	0.95
1/2023	77.12	1	79.15	-1.03	76.15	1.97
2/2023	75.81	1.02	70.82	6.01	75.8	1.03
3/2023	70.18	3.1	72.13	1.15	72.91	0.37
4/2023	74.14	5.31	77.23	2.22	77.11	3.34
5/2023	71.50	0.08	70.69	0.89	68.32	3.26
6/2023	69.11	1.14	69.10	1.15	69.19	1.06
7/2023	72.05	4.02	78.15	-2.08	75.13	0.94
8/2023	80.13	1.26	80.23	1.16	79.65	1.74
9/2023	87.56	1.87	84.12	5.31	88.09	1.34
10/2023	82.46	3.18	82.38	3.26	82.31	3.33
11/2023	73.82	3.87	75.01	2.68	72.11	7.58
12/2023	71.35	0.55	69.02	2.88	72.13	-0.23

From table (12), a comparison between actual and predicted values for three hybrid Models is made. The statistical measures from various hybrid models were obtained, which will be also so closed.

3.4. Comparison Between ARIMA-SVR and Hybrid Model Using Different Integration Methods.

Now, the test set consisting of 24 observations is being utilized, and the results are presented in the following table.

Table (13): Comparison between Single Models and Hybrid Model

Model	MSE	MAPE	MPE	MAE	rank
ARIMA	13.7378	0.037545	0.030809	3.21458	5
SVR	10.48598	0.027933	0.02378	2.419167	4
"ARIMA-SVR" Additive	5.059067	0.020396	0.020055	1.719208	1
"ARIMA-SVR" Multiplicative	8.872858	0.027733	0.023156	2.415	3
"ARIMA-SVR" Regression	7.94678	0.022799	0.021706	2.001667	2

From table (13), the following points are evident:

1. The ARIMA model represents the least efficient and predictive capability among the models, displaying weak performance with significant errors.
2. The SVR model performs better in predicting monthly global oil prices compared to the ARIMA model, exhibiting lower errors.
3. The additive hybrid model (ARIMA-SVR) demonstrates the highest efficiency in predicting monthly global oil prices, possessing the lowest values for statistical measures such as MAE, MPE, MAPE, and MSE. It is followed by the regression hybrid model, then the multiplicative hybrid model, then the SVR model, and finally the ARIMA model.

4. Conclusion:

The research aims to forecast time series using single models SVR and ARIMA, as well as their hybrid model with different integration methods, applied to a monthly time series of global oil prices from January 2004 to December 2023. The study relies on several accuracy metrics to reach the optimal model and draws the following conclusions:

By comparing single models SVR, ARIMA, and the hybrid models including the additive hybrid model, the multiplicative hybrid model, and the multiple regression hybrid model, the additive hybrid model showed superiority due to having the lowest statistical criteria values (MAE, MPE, MAPE, MSE), followed by the regression hybrid model, then the multiplicative hybrid model, then SVR, and finally ARIMA.

Future Research:

From the research Conclusion, we recommended that.

1. Using simulation methods to generalize the results obtained.
2. Re-estimating the models with a larger time series than that used in the study.
3. Conducting new experiments to estimate integration weights and explore integration between different prediction methods to achieve the best hybrid prediction methods.

References:

1. Alrweili, H., & Fawzy, H. (2022). “Forecasting crude oil prices using an ARIMA-ANN hybrid model”. Journal of Statistics Applications & Probability, Vol.11(3), Pp.845-855.
2. Alsuwaylimi, A. A. (2023).” Comparison of ARIMA, ANN and Hybrid ARIMA-ANN models for time series forecasting”. Information Science Letters an International Journal, Vol.12(2), Pp. 1003-1016. http://dx.doi.org/10.18576/isl/120238
3. Ayub S., Jafri Y. Z., (2020) “Comparative Study of an ANN-ARIMA Hybrid Model for Predicting Karachi Stock Price”. American Journal of Mathematics and Statistics, Vol. 10(1), Pp. 1-9. Doi: 10.5923/j.ajms.20201001.01
4. Box G. E. P., Jenkins G. M., Reinsel G. C., Ljung G. M., (2016). *Time Series Analysis: Forecasting and Control*. Fifth Edition, John Wiley & Sons, Inc., Hoboken, New Jersey.
5. Cao, L. J., & Tay, F. E. H. (2003). “Support vector machine with adaptive parameters in financial time series forecasting”. IEEE Transactions on neural networks, Vol.14(6), Pp. 1506-1518. https://doi.org/10.1109/TNN.2003.820556
6. Cosgun, E., Limdi, N. A., & Duarte, C. W. (2011). “High-dimensional pharmacogenetic prediction of a continuous trait using machine learning techniques with application to warfarin dose prediction in

African Americans”. Bioinformatics, Vol.27(10), Pp. 1384-1389. https://doi.org/10.1093/bioinformatics/btr159

7. El-malky, M., Kamel, M., & Abdelnabi, W. (2023). “Hybrid Model "ARIMA -ANN" Using for forecasting Stock Index EGX30”. Commerce and Finance, VOL.43(3), Pp.352-379.
8. Lai, H., Khan, Y. A., Thaljaoui, A., Chammam, W., & Abbas, S. Z. (2021). “RETRACTED ARTICLE: COVID-19 pandemic and unemployment rate: A hybrid unemployment rate prediction approach for developed and developing countries of Asia”. Soft Computing.Vol.27, p.615. https://doi.org/10.1007/s00500-021_05871-6
9. Leenawong , C., Chaikajonwat T. ,(**2022**). “Event Forecasting for Thailand’s Car Sales during the COVID-19 Pandemic”. data. Vol.86 (7), https://doi.org/10.3390/data7070086
10. Nawi, W. I. A. W. M., Lola, M. S., Zakariya, R., Zainuddin, N. H., Abd Hamid, A. A. K., Aruchunan, E., & Nazzrol, N. S. A. (2021). “Improved forecasting sea surface temperature based on hybrid ARIMA and support vector machines models”. Malaysian Journal of Fundamental and Applied Sciences, Vol. 17(5), Pp. 609-620.
11. Nassar, H. M. M., Rashwan, N. I., Ismail, M. M., Talat, A. A., (2023). “Time Series Prediction Using a Hybrid Model of Autoregressive Integrated Moving Average and Support Vector Regression”. Commerce and Finance.VOL.43(3), PP.1043-1086.
12. Pannakkong, W., Harncharnchai, T., & Buddhakulsomsiri, J. (2022). “Forecasting daily electricity consumption in Thailand using regression, artificial neural network, support vector machine, and hybrid models”. Energies, Vol. 15(9), Pp. 3105. https://doi.org/10.3390/en15093105
13. Rosenbaum, L., Dörr, A., Bauer, M. R., Boeckler, F. M., & Zell, A. (2013).” Inferring multi-target QSAR models with taxonomy-based multi-task learning”. Journal of cheminformatics, Vol. 5(1), Pp .1-20.
14. Sareminia, S. (2023). “A support vector-based hybrid forecasting model for chaotic time series: Spare part consumption prediction”. Neural Processing Letters, Vol.55(3), pp.2825-2841.
15. Wang, L., Zou, H., Su, J., Li, L., & Chaudhry, S. (2013).” An ARIMA-ANN hybrid model for time series forecasting”. Systems Research and Behavioral Science, Vol. 30(3), Pp. 244-259.

https://doi.org/10.1002/sres.2179

16. Wei, W. W. (2006). *Time series analysis: Univariate and multivariate methods*. Pearson education inc.
17. Xu, D., Zhang, Q., Ding, Y., & Huang, H. (2020). "Application of a hybrid ARIMA–SVR model based on the SPI for the forecast of drought—a case study in Henan Province, China". Journal of applied meteorology and climatology, Vol. 59(7), Pp.1239-1259. Doi: 10.1175/jamc-d-19-0270.1
18. Zhang, G. P. (2003). "Time series forecasting using a hybrid ARIMA and neural network model". Neurocomputing, Vol. 50, Pp. 159-175. https://doi.org/10.1016/S0925-2312(01)00702-0
19. Zheng, Y., Yuan, K., & Dai, L. (2021). "Research on Coal Price Forecast based on ARIMA and SVM combination Model". In E3S Web of Conferences Vol. 257. https://doi.org/10.1051/e3sconf/202125702008
20. Zhang C., Zhou X., (2024)." Forecasting value-at-risk of crude oil futures using a hybrid ARIMA-SVR-POT model". Heliyon, Vol. 10.

مقارنة النماذج المفردة أو النماذج الهجينة في التنبؤ بالسلاسل الزمنية

المستخلص

يهدف البحث إلى التنبؤ بالسلاسل الزمنية بالاعتماد على النماذج الفردية SVR و ARIMA والنموذج الهجين " ARIMA-SVR"من خلال أساليب التكامل المختلفة المطبقة على بيانات أسعار النفط العالمية من يناير ٢٠٠٤ إلى ديسمبر ٢٠٢٣، متضمنة بيانات شهرية بـ ٢٤٠ ملاحظة ومقارنة نتائجها لتحديد أفضل نموذج للتنبؤ بأسعار النفط العالمية. تشمل طرق التكامل النموذج الهجين الإضافي، والهجين المضاعف، والنموذج الهجين الانحداري كنماذج هجينة مقارنة مع النماذج الفردية SVR، ونماذج ARIMA. أظهرت النتائج أن النموذج الهجين الإضافي ARIMA-SVR Additive هو النموذج الأفضل بين جميع النماذج قيد الدراسة، لأنه يوفر أقل قيم مقاييس دقة التنبؤ: MAE، MPE، MAPE، MSE. وباستخدام اختبار Ljung-Box للسلسلة الناتجة، حصل على الترتيب الأول. النموذج الهجين الإضافي ARIMA-SVR Additive كأفضل نموذج لنمذجة بيانات أسعار النفط العالمية يليه النموذج الهجين الانحداري ثم النموذج الهجين المضاعف SVR وأخيراً ARIMA