**Int. J. Nonlinear Anal. Appl. 14 (2023) 1, 1371–1383**
**ISSN: 2008-6822 (electronic)**
http://dx.doi.org/10.22075/ijnaa.2022.7110

# The use of ARIMA, LSTM and GRU models in time series hybridization with practical application

Noor Saleem Mohammed Ali, Firas Ahmmed Mohammed*

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

*(Communicated by Javad Vahidi)*

**Abstract**

The importance of forecasting has emerged in the economic field in order to achieve economic growth, as forecasting is one of the important topics in the analysis of time series, and accurate forecasting of time series is one of the most important challenges in which we seek to make the best decision. The aim of the research is to suggest the use of hybrid models for forecasting the daily crude oil prices as 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 long short term memory (LSTM) and the gated recurrent unit (GRU) which represents deep learning models. It was found that the proposed hybrid models in the prediction process when conducting simulation for different sample sizes and when applied to the daily crude oil price time series data, were more efficient than the single models, and the comparison between the single models and the proposed hybrid models was made by comparison scale, mean square error (MSE), and the results showed that the proposed hybrid models have the ability to predict crude oil prices, as they gave more accurate and efficient results.

Keywords: ARIMA, Long Short Term Memory, Time Series Forecasting, Gated Recurrent Unit, Hybrid Model
2020 MSC: 91B84

## 1 Introduction

Crude oil is the most traded commodity in the world as it is considered the main source of energy, and thus the movement of oil has a significant impact on the global economy and due to the presence of constant, irregular and non-linear fluctuations that affect the economic development of the country, i.e. any increase or decrease in crude oil prices It has a volatile impact on global markets, so oil is one of the most complex and difficult prices in the modeling process, and therefore it is difficult for the investor to make a decision. Therefore, the search for oil price fluctuations has become more complex due to the interaction between factors such as demand and supply in the market, dollar exchange rates and natural disasters, wars and therefore crude oil prices show a complex volatility.

Accurate forecasting of the time series that represents crude oil prices is one of the important issues in the economic field and is one of the most important issues facing energy economists towards decisions for better forecasting.

Various models for forecasting crude oil prices will be developed by conducting a study using hybrid models that combine a linear model and another non-linear model, as they have the ability to combine the strength of linear models

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

*Received:* August 2022 *Accepted:* October 2022

and the flexibility of non-linear models, and due to the presence of fluctuations in the crude oil price series, they it contains both linear and non-linear patterns, and thus it will have difficulty in determining the time series data if it is linear or non-linear, as most time series data contain linear and non-linear patterns, so it is not enough to use one model when predicting the time series and it is not enough to represent the relationship non-linearity. In addition, the errors extracted from the linear model contain information or a non-linear pattern, and this leads to a weakness in the prediction accuracy. Therefore, ARIMA models alone are not sufficient for modeling, so two models were combined instead of using one model to predict where the model cannot linear or non-linear processing of the two components together. The first model, which represents the linear component, which is the (ARIMA) model, which represents the Box Jenkins model, was developed to represent the stationary and non-stationary time series for prediction, and the second model represents the non-linear component, which represents the use of some methods used in artificial intelligence are deep learning models that represent a long short term memory (LSTM) network and gated recurrent unit (GRU) network capable of modeling complex properties such as nonlinearities and fluctuations.

## 2 Autoregressive Integrated Moving Average (ARIMA) Model

The ARIMA models are generated from the auto regressive models and the moving averages models, it was introduced by (Box and Jenkins) in (1976), 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 denoted by $(d)$. The ARIMA $(p, d, q)$ models can be expressed using the backshift operator (B) as follows [6]:

$$\phi(B)(1-B)^d y_t = \theta\,(B)\,\varepsilon_t \tag{2.1}$$

## 3 Long Short Term Memory (LSTM)

The long short term memory network was suggested by (Hochreiter) and (Schmidhuber) in (1997), this network is one of the recurrent neural network (RNN) designs, this network improves the gradient fading problem that appears when training the (RNN) as it provides a longer range without exploding or vanishing gradient by changing the (RNN) cell structure [1].

The network consists of a memory block, it is called memory cell represents the main component that enhances the ability of the (LSTM) network to model and predict the time series. The memory cell consists of the cell state, which is located at the top of the memory cell, it is symbolized by $(C_t)$, the cell state is characterized by that it does not contain activation functions and transmits the processed information, the cell state in LSTM can be expressed as follows:

$$C_t = \ f_t \ \odot \ C_{t-1} + i_t \ \odot \ \widetilde{C}_t \tag{3.1}$$

The memory cell contains three gates which are the forget gate, input gate and the output gate, where the outputs of three gates are, represented as $f_t, i_t$ and $o_t$ which are weighted functions as they control the flow of information in the memory cell, the forget gate determines the information that is allowed to pass or that is ignored from the previous case [1], the output of the forget gate variable produces a vector whose element values range between (0) and (1), the forget gate in LSTM can be expressed as follows [2]:

$$f_t = \sigma\left(W_{hf} \cdot h_{t-1} + W_{xf} \cdot x_t + b_f\right) \tag{3.2}$$

Where $(\sigma)$ is referred to as the sigmoid function and is an activation function that enables the nonlinear capabilities of the model. The input gate defines the information that is allowed to pass from the new state that was calculated from the current input and uses the sigmoid function as an activation function [1], it generates the variable $(i_t)$ and the candidate $(\widetilde{C}_t)$, the input gate in LSTM can be expressed as follows:

$$i_t = \sigma\left(W_{hi} \cdot h_{t-1} + W_{xi} \cdot x_t + b_i\right) \tag{3.3}$$

The candidate input can be obtained by inserting the same concatenation into the activation function which is uses the hyperbolic tangent function [2] , denoted (tanh function) whose outputs are between (-1) and (+1), the candidate in LSTM can be expressed as follows:

$$\widetilde{C}_t = tanh\left(W_{hc} \cdot h_{t-1} + W_{xc} \cdot x_t + b_c\right) \tag{3.4}$$

The output gate determines the information that is output based on the cell's memory and the input, it generates the state $(h_t)$ and the variable $(o_t)$, the output gate in LSTM can be expressed as follows:

$$o_t = \sigma\left(W_{ho} \cdot h_{t-1} + W_{xo} \cdot x_t + b_o\right) \tag{3.5}$$

The state $(h_t)$ delivers information to the next cell using a different path, the tangent function is applied as an activation function to the state of the cell as follows:

$$h_t = o_t \ \odot \tanh(C_t) \tag{3.6}$$

The activation vectors for each state and memory block are, denoted as $C_t$ and $h_t$. The weight matrices (W) and bias vectors (b) are utilized to build connections between the input layer, output layer and the memory block [8].

## 4 Gated Recurrent Units (GRU)

This network is a type of recurrent neural network (RNN) that was proposed by (K. Cho) in (2014), this network contains two gates, the update and the reset gates, which are vectors that specify the information that must be passed to the outputs. The forget gate and the input gate are combined into the update gate as this gate controls how much each hidden unit in the network can remember or forget [7].

The update gate helps the model determine how much previous information should be discarded and new information that should be stored and helps determine the information that should be passed to the future [5]:

$$u_t = \sigma\left(W_{hu} \cdot h_{t-1} + W_{xu} \cdot x_t + b_u\right) \tag{4.1}$$

The update gate generates a variable $(u_t)$ and the candidate $(\widetilde{h}_t)$, the mathematical formula of the candidate is:

$$\widetilde{h}_t = tanh\left(W_{hh} \cdot \ (r_t \ \odot \ \ h_{t-1}) + W_{xh} \cdot x_t + b_h\right) \tag{4.2}$$

Where: $\odot$ is the point-wise multiplication.

The reset gate determines how the new input is combined with the previous memory and it used to determine how much of the previous information should be forgotten:

$$r_t = \sigma\left(W_{hr} \cdot h_{t-1} + W_{xr} \cdot x_t + b_r\right) \tag{4.3}$$

The state of GRU stores the current information and transfers this processed information to the next cell and this is done by a linear combination, as in the following equation:

$$h_t = (1 - u_t) \ \odot \ h_{t-1} + u_t \ \odot \ \widetilde{h}_t \tag{4.4}$$

## 5 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 model, 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.

$$\hat{y}_t = b_{\mathbf{0}} + b_{\mathbf{1}}x1(Linear\ Component[Lt]) + b2x2(Non - Linear\ Component[Nt]) \tag{5.1}$$

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 [3]:

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{5.2}$$

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 (LSTM), (GRU), 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 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 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 hybrid model can be written as follows:

$$\hat{y}_t = b_{\mathbf{0}} + b_{\mathbf{1}}y1(Linear\ Component[Lt]) + b2e2(Non - Linear\ Component[Nt]) \tag{5.3}$$

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

1. The non-linear component of the (LSTM), (GRU), 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 represent the 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 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 hybrid model can be written as follows:

$$\hat{y}_t = b_{\mathbf{0}} + b_{\mathbf{1}}y1(Non - Linear\ Component\ [Nt]) + b2e2(Linear\ Component\ [Lt]) \tag{5.4}$$

## 6 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 (LSTM), (GRU), 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 time series ($y_t$)) are combined with the predictive values of the non-linear model (whose inputs represent the time series ($y_t$)) by means of the regression model in order to obtain the predictive values 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 hybrid model can be written as follows:

$$\hat{y}_t = b_{\mathbf{0}} + b_{\mathbf{1}} y1(Linear\ Component\ [Lt]) + b2y2(Non - Linear\ Component\ [Nt]) \tag{6.1}$$

The Generalized Ridge Regression (GRR) estimates for the linear regression model can be calculated according to the following equation [3]:

$$b_{GRR} = (X^{*'} X + K)^{-1} X^{*'} Y \tag{6.2}$$

Choice of the Estimator ridge by (Hoerl), (Kennard) and (Baldwin) suggested that a choice for (k) is [4]:

$$\hat{k}_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}} \tag{6.3}$$

## 7 The Simulation

Simulation is one of the most important scientific methods that are used as an empirical proof method in order to find out the best model among a group of models in the event that they are not able to be proven theoretically and it can be used to find the degree of efficiency of the models used and to perform a comparison between these models.

The simulation experiments included generating data for a group of single models and the proposed hybrid models, and different sample sizes (small, medium and large) were used (250, 500, 1500 and 3000). The models were compared using the (MSE) comparison scale and programs (R and Python) 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 | LSTM | GRU |
|---|---|---|---|
| MSE | 1.234411 | 1.101469 | 1.094299 |

Table (1) shows the results of the (MSE) criterion for each of the models individually and the superiority of the (GRU) model at small sample sizes (250) which equal to (1.094299), it has the lowest (MSE) value, followed by the (LSTM) model equal to (1.101469) and the model that gave the highest value for the (MSE) is (ARIMA) model, where it has a value equal to (1.234411).

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 | LSTM | GRU |
|---|---|---|---|
| MSE | 1.186498 | 1.081580 | 1.075327 |

Table (2) shows the case of sample sizes (500) also, the stability results for the other models, and the superiority of the (GRU) model which has the lowest value of the (MSE) criterion equal to (1.075327), followed by the (LSTM) model equal to (1.081580), the (ARIMA) model also that gave the highest value for the (MSE) is, where it has a value equal to (1.186498).

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 | LSTM | GRU |
|---|---|---|---|
| MSE | 1.095499 | 1.061319 | 1.057630 |

Table (3) shows the superiority of the (GRU) model at sample sizes (1500), as it has the least value of the (MSE) criterion equal to (1.057630), then followed the (LSTM) model which has value equal to (1.061319), 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 | LSTM | GRU |
|---|---|---|---|
| MSE | 1.06265 | 1.086326 | 1.054925 |

Table (4) shows the case of the large sample sizes also the stability of the results for the other models, and the superiority of the (GRU) model, which has the lowest value of the (MSE) criterion equal to (1.054925) and the (LSTM) model that gave the highest value for the (MSE) is, where it has a value equal to (1.086326).

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-LSTM | ARIMA-GRU |
|---|---|---|
| MSE | 0.963024 | 0.980916 |

Table (5) shows that the (ARIMA-LSTM) hybrid model has outperformed the (ARIMA-GRU) hybrid model as it has the lowest value equal to (0.963024) for the (MSE) than the other models in the small sample sizes (250), then followed by the hybrid model (ARIMA-GRU) equal to (0.980916).

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 | LSTM-ARIMA | GRU-ARIMA |
|---|---|---|
| MSE | 0.983074 | 0.972165 |

Table (6) shows that the best hybrid model is (GRU-ARIMA), which has the lowest (MSE) criterion of the other hybrid models and is equal to (0.972165), then followed by the hybrid model (LSTM-ARIMA) which have (0.983074) value.

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-LSTM | ARIMA-GRU |
|---|---|---|
| MSE | 0.993617 | 0.990649 |

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 | LSTM-ARIMA | GRU-ARIMA |
|---|---|---|
| MSE | 0.985432 | 0.987286 |

Table (8) also shows that the (LSTM-ARIMA) hybrid model is better than the (GRU-ARIMA) hybrid model at sample sizes (500) it has the lowest value for the (MSE) criterion equal to (0.985432), followed by the (GRU-ARIMA) hybrid model which have (0.987286) value.

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-LSTM | ARIMA-GRU |
|---|---|---|
| MSE | 0.989513 | 0.997967 |

Table (9) shows that the (ARIMA-LSTM) hybrid model outperformed the (ARIMA-GRU) hybrid model which has the lowest value for the (MSE) criterion equal to (0.989513) when applied to sample sizes (1500), followed by the (ARIMA-GRU) hybrid model which has (0.997967).

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 | LSTM-ARIMA | GRU-ARIMA |
|---|---|---|
| MSE | 0.995476 | 0.997785 |

Table (10) shows that the best hybrid model is (LSTM-ARIMA) that has the lowest value for the (MSE) equal to (0.995476) when applied to sample sizes (1500), followed by the (GRU-ARIMA) hybrid model which have (0.997785) value.

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-LSTM | ARIMA-GRU |
|---|---|---|
| MSE | 0.996682 | 0.996209 |

Table (11) shows that the best hybrid model is (ARIMA-GRU), where it outperformed the (ARIMA-LSTM) hybrid model equal to (0.996209) at large sample sizes (3000) and has the lowest value for the (MSE) criterion, followed by the (ARIMA-LSTM) hybrid model which has (0.996682).

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 | LSTM-ARIMA | GRU-ARIMA |
|---|---|---|
| MSE | 0.997612 | 0.995674 |

Table (12) shows that the hybrid model is (GRU-ARIMA) which has the lowest value for the (MSE) criterion equal to (0.995674). Followed by the (LSTM-ARIMA) hybrid model which has (0.997612).

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-LSTM | ARIMA-GRU |
|---|---|---|
| MSE | 0.831578 | 0.844807 |

Table (13) shows that the (ARIMA-LSTM) hybrid model is the best and has the least value for the (MSE) criterion equal to (0.831578) at small sample sizes (250), followed by the (ARIMA-GRU) hybrid model which have (0.844807).

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-LSTM | ARIMA-GRU |
|---|---|---|
| MSE | 0.964289 | 0.863138 |

Table (14) also shows that the (ARIMA-GRU) hybrid model is the best and has the least value for the (MSE) criterion equal to (0.863138) at sample sizes (500), followed by the (ARIMA-LSTM) hybrid model which have (0.964289) value.

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-LSTM | ARIMA-GRU |
|---|---|---|
| MSE | 0.966569 | 0.965505 |

Table (15) shows that the (ARIMA-GRU) has the least value for the (MSE) criterion equal to (0.965505) at sample sizes (1500), followed by the (ARIMA-LSTM) hybrid model has gave the value for the (MSE) equal to (0.966569).

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-LSTM | ARIMA-GRU |
|---|---|---|
| MSE | 1.036272 | 1.009991 |

Table (16) also shows that the (ARIMA-GRU) hybrid model is has the least value for the (MSE) criterion equal to (1.009991) at large sample sizes (3000), followed by the (ARIMA-LSTM) hybrid model which have (1.036272) value.

## 8 The Application

The real data for the 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.

The data is prepared before applying the Box Jenkins methodology, as the time series must be fixed in terms of mean and variance. The time series is drawn to know its basic features in terms of the presence of the general trend or not, and the stationary of variance or not. The time series graph shows the pattern that develops It has the daily crude oil prices and this is done by applying some necessary transformations to make time series stationary.

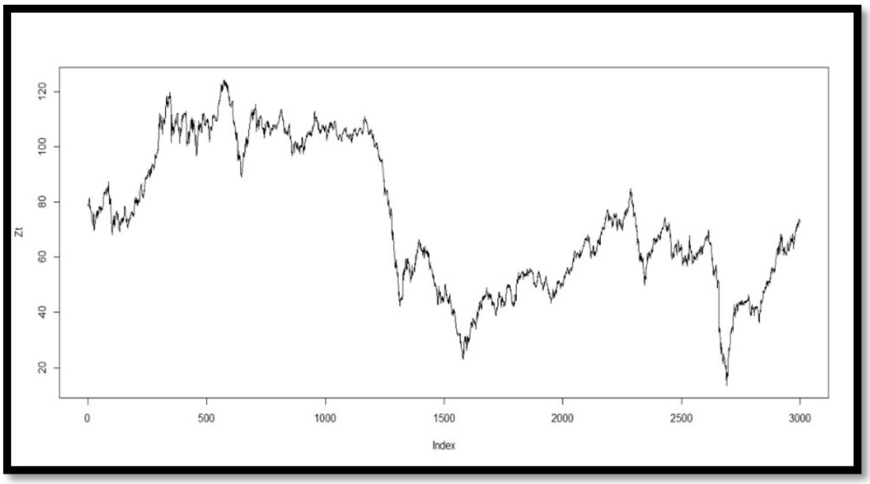


Figure 1: The time series of daily crude oil price

After plotting the time series, it was found that it is non stationary on the mean and variance with time where dispersion change around the level of the series can be observed. 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:

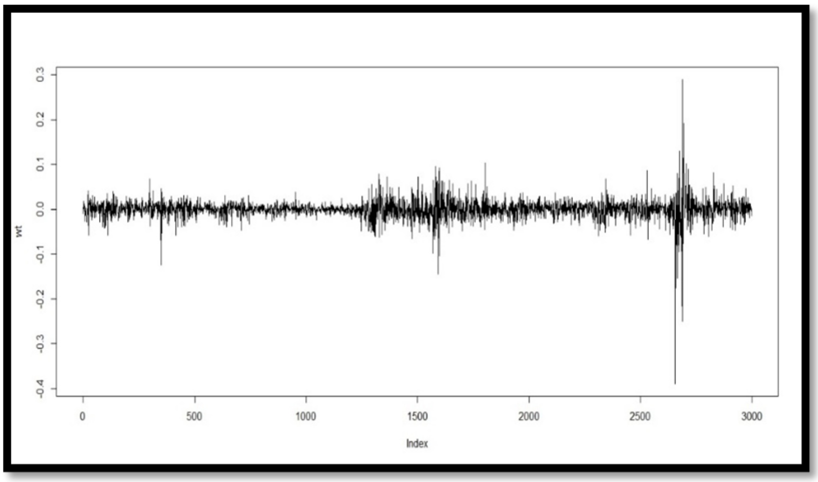


Figure 2: The time series after taking the first differences of the logarithm of the original series

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) and the best model with the lowest value for these criteria was obtained, which is the ARIMA (5,1,3) model. 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. It was found that the residuals of the ARIMA model are random and there is no correlation, that is, the variables are independent according to the Box-Pierce and the Ljung Box tests.

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

Table 18: The MSE for prediction accuracy of (ARIMA) model

| The Model | MSE |
|---|---|
| ARIMA | 0.0007719612 |

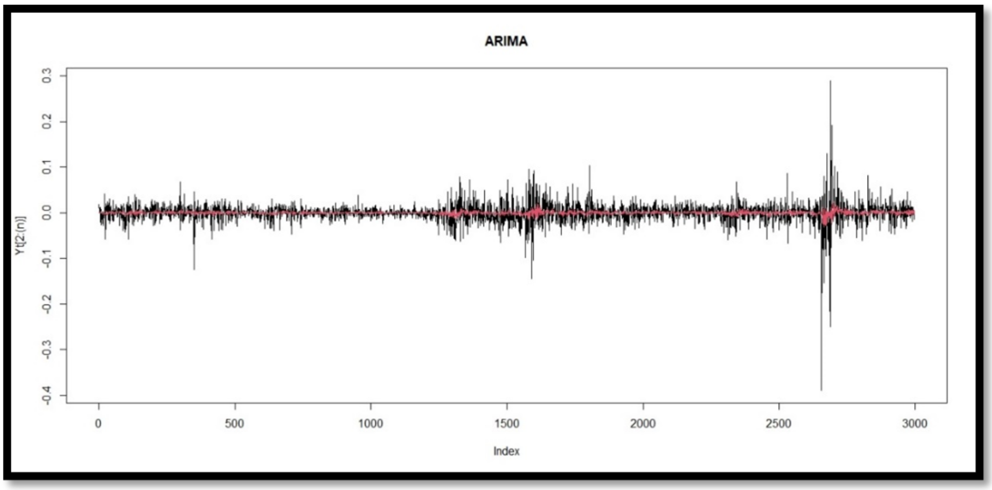


Figure 3: The real and predictive values of the ARIMA model

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

The best proposed hybrid model was obtained from the simulation for (ARIMA-GRU), which will be applied to daily oil price data. The application of the (ARIMA-GRU) hybrid model will be explained as follows:

After using the ARIMA model and calculate the predicted values. Building the (GRU) model to model the residuals of the ARIMA (5,1,3). The data in the (GRU) network is processed using the Min Max Normalized Method. The input variables of the (GRU) model were determined by Box Jenkins model, assume five neurons were selected. Determining the batch size equal to one and the number of epochs equal to 100. The optimizer Adam was determined, and the loss function is the mean squared error (MSE).

After estimating the (GRU) model and obtaining its predictive values, the predictive values of the (ARIMA) model are combined with the predictive values of the (GRU) 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), and the results are as follows:

Table 19: The MSE for prediction accuracy of the first proposed hybrid model (ARIMA-GRU)

| The Model | MSE |
|---|---|
| ARIMA-GRU | 0.0007718573 |

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

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

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

$$\hat{y}_t = -0.0000526 + 110.90\ \hat{x}_{1(ARIMA)} +\ 0.0000003\hat{x}_{2(GRU)}$$

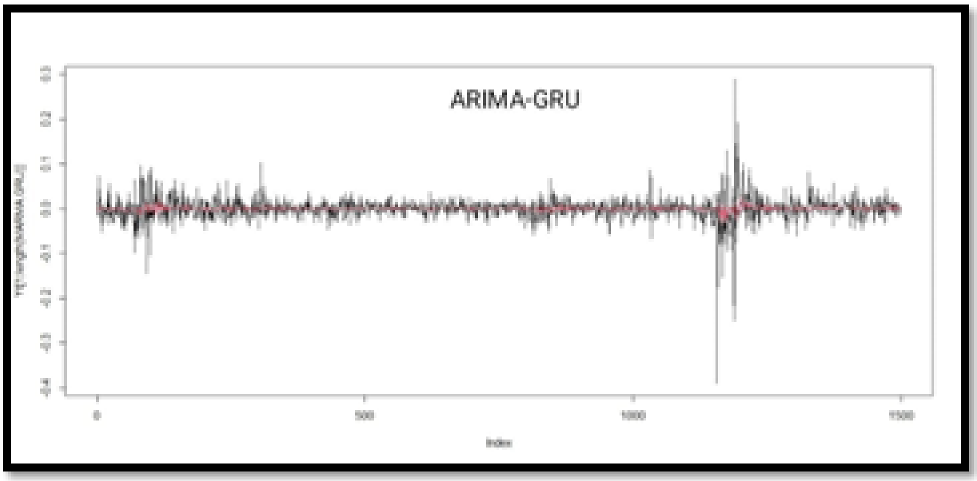


Figure 4: The real and predictive values of the first proposed hybrid model (ARIMA-GRU)

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

Table 20: The Box-Pierce test for the first proposed hybrid model (ARIMA-GRU)

| Lag | 1 | 5 | 10 |
|---|---|---|---|
| Chi – Square | 0.0627 | 0.7748 | 8.7242 |
| DF | 1 | 5 | 10 |
| p-value | 0.8022 | 2.5111 | 0.5585 |

Table 21: The Ljung Box test for the first proposed hybrid model (ARIMA-GRU)

| Lag | 1 | 5 | 10 |
|---|---|---|---|
| Chi – Square | 0.0628 | 2.5226 | 8.7823 |
| DF | 1 | 5 | 10 |
| p-value | 0.802 | 0.7731 | 0.5529 |

In the above tables (20) & (21), it is clear that the Box-Pierce and Ljung Box tests for the residuals of the (ARIMA-GRU) hybrid model is random and there is no correlation and the variables are independent.

## The Result of the Second Proposed Hybrid Model for (ARIMA-GRU)

After using the ARIMA (5,1,3) model and calculate the predicted values. Building the (GRU) model where the network inputs are the observations then they are modeled and the predictive values are obtained.

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

Table 22: The MSE for prediction accuracy of the (GRU) model

| The Model | MSE |
|---|---|
| GRU | 0.00079799663 |

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

Table 23: The MSE for prediction accuracy of the second proposed hybrid model (ARIMA-GRU)

| The Model | MSE |
|---|---|
| ARIMA-GRU | 0.000772001 |

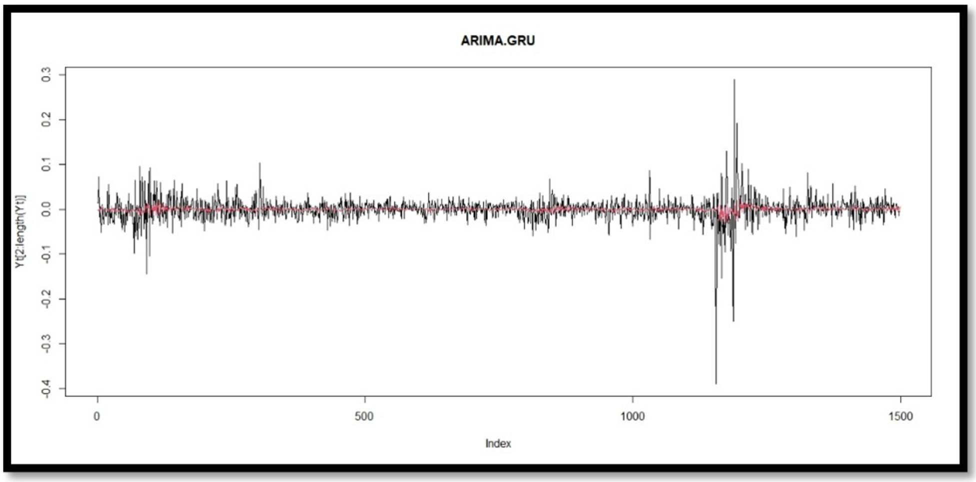


Figure 5: The real and predictive values of the second proposed hybrid model (ARIMA-GRU)

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

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

The ridge parameter ($K_{HKB}$) has been obtained (0.001256249), the equation of the (ARIMA-GRU) can be written as follows:

$$\hat{y}_t = -0.000052599 + \ 106.0998 \hat{x}_{1(ARIMA)} + 0.000000332 \ \hat{x}_{2(GRU)}$$

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

Table 24: The Box-Pierce test for the second proposed hybrid model (ARIMA-GRU)

| Lag | 1 | 5 | 10 |
|---|---|---|---|
| Chi – Square | 0.0615 | 2.9144 | 9.6799 |
| DF | 1 | 5 | 10 |
| p-value | 0.804 | 0.7132 | 0.469 |

Table 25: The Ljung Box test for the second proposed hybrid model (ARIMA-GRU)

| Lag | 1 | 5 | 10 |
|---|---|---|---|
| Chi – Square | 0.06171 | 2.9278 | 9.7442 |
| DF | 1 | 5 | 10 |
| p-value | 0.8038 | 0.7111 | 0.4632 |

In the above tables (24) & (25), it is clear that the Box-Pierce and Ljung Box tests for the residuals of the (ARIMA-GRU) hybrid model is random and there is no correlation and the variables are independent.

## 9 The Conclusions

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

1. When applying at small sample sizes (250) to the hybrid models, the best model for the first proposed hybrid model is (ARIMA-LSTM), the best second proposed hybrid model also is (ARIMA-LSTM).
2. When applying sample sizes (500) to the hybrid models, the best model for the first proposed hybrid model is (ARIMA-GRU), the best second proposed hybrid model also is (ARIMA-GRU).
3. When applying the sample sizes (1500) to the hybrid models, the best model for the first proposed hybrid model is (ARIMA-LSTM) and for the second proposed hybrid model is (ARIMA-GRU).
4. When applying at large sample sizes (3000) to the hybrid models for the first proposed hybrid model is (ARIMA-GRU) and for the second proposed hybrid model is (ARIMA-GRU).

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-GRU) and the second proposed hybrid model for (ARIMA-GRU) 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-GRU) according to the (MSE), so it is preferred when using it to forecast daily crude oil prices.

## References

[1] J. Brownlee, *Long short-term memory networks with python*, Machine Learn. Mastery **1** (2017), 1–229.

[2] H.K. Choi, *Stock price correlation coefficient prediction with ARIMA-LSTM hybrid model*, arXiv preprint arXiv:1808.01560 (2018).

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

[4] G. Khalaf and M. Iguernane, *Ridge regression and ill-conditioning*, J. Mod. Appl. Statist. Meth. **13** (2014), no. 2, 18.

[5] R. Mangalampalli, V. Pandey, P. Khetre, and V. Malviya, *Stock prediction using hybrid ARIMA and GRU models*, Int. J. Engin. Res. Technol. **9** (2020), no. 5, 2278–0181.

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

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

[8] B.B. Sahoo, R. Jha, A. Singh, and D. Kumar, *Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting*, Acta Geophysica **67** (2019), no. 5, 1471–1481.