

Analyzing the Relationship between the Dow Jones Index and Oil Prices Using the ARIMAX Model

Haifa Taha Abd^{a,*}, Ameena Kareem Essa^a, Firas M. Jassim^a

^a Collage of Management and Economics, University of Mustansiriyah. Iraq

Corresponding author: *haefaa_adm@uomustansiriyah.edu.iq

Abstract—The values prediction of the Dow Jones index is essential in the global financial markets systems. The index provides a clear vision of what is happening in the market as a whole. Hence, it offers integrated information on that index to stipulate forecasts characterized as efficient for investors and shareholders. In this study, the ARIMAX model was used to predict the daily Dow Jones index values from 1/1/2020 to 1/ 5/2020 (the spread of COVID 19), considering Brent crude's effect daily prices as an external factor. The Dow Jones Index daily price prediction process went through several stages. The first stage is the time series stationary test phase through the Augmented Dickey-Fuller test. The second stage is achieving stationary by taking the first difference, passing through the stage of identifying the model, and determining the rank based on criteria (AIC), (BIC), (RMSE). The preference of the model was shown in ARIMA (0,1,2) for the Dow Jones index series. The ARIMA (1,1,0) model was shown for crude price Brent series and determining the order of the transfer function of the ARIMAX model. The comparison stage between the models ARIMA (0,1,2) and ARIMAX(3,1,1)(0,0,1) by residuals scatter plot and (Ljung-Box) test for each model. The results demonstrated the superiority of the ARIMAX model over the ARIMA model. The daily Dow Jones Index values were predicted based on corresponding Brent crude prices according to the ARIMAX (3,1,1) (0,0,1) model. The researchers did not find substantial differences in the index's behavior, except for a slight decrease in the index's value.

Keywords—ARIMA; ARIMAX; Dow Jones; Brent Crude; COVID (19).

Manuscript received 9 Jul. 2020; revised 20 Jan. 2021; accepted 12 Feb. 2021. Date of publication 30 Apr. 2021.
IJASEIT is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



I. INTRODUCTION

Time series analysis witnessed globally in the second half of the twentieth century a significant development, especially in the last three decades. This development is attributed to the modern methods presented by researchers in this field. ARIMA models under the Box-Jenkins methodology are considered the most common in the theoretical scientific circles. This methodology has proven high efficiency in Time series modeling and forecasting. The basic concept of ARIMA models is based on employing one-time series without using the information available in other time series that may be related to it or may include more than a variable to describe behavior. In many applications, the studied time series is subject to impact regular from another time series. In this case, we need to use a model that includes more than one-time series and shows this time series system's dynamic characteristics. This type is known as ARIMAX models. The ARIMAX model building mechanism is based on the studied

time series slope known as the Output series on an effective time series in the output series called the Input series.

There are some studies in the field of conversion function models. This study forecasts the sales of the Central Bank of Iraq from foreign currency using the ARIMAX (2,0,0) (0,0,2) that depends on the demand gap of the currency. Some previous studies addressed the ARIMAX model in a study regarding the forecasting of temperatures for the city of Baghdad in the presence of wind speed as a series of inputs [1], [2]. Jaraskunlanat and Kijboonchoo [3] compared the ARIMAX (1,0,0) (0,0,2) and the method of neural networks. It aims to predict the global wheat production series by relying on the global wheat price series to forecast the series global wheat production. All the previous studies relied on a study of the effect of the input series while they are in their natural condition or conditions [4].

Therefore, this study employed the ARIMAX model in studying the relationship between Dow Jones's daily values in the global market and its impact on the sharp decline in the daily prices of Brent crude in light of COVID 19 pandemic.

The model was used to estimate an appropriate conversion function to describe the relationship between index values and crude prices during the period of a pandemic to predict future index values.

II. MATERIALS AND METHOD

A. Autoregressive Integrated Moving

Autoregressive Integrated Moving Average (ARIMA) model is the most common time series model used. All models can be derived from ARIMA, whether self-regression, moving averages, or mixed models. These models consist of three parts. The first part represents the Autoregressive model and is coded by AR (p) [5]. The other part represents the Moving Average model is denoted by MA (q). The third part represents the degree of integration (Integrated) and is denoted by the symbol (d) and is defined as the differences that must be made on the time series to be stable. The mathematical formula of the model ARIMA (p, d, q) are:

$$y_t = \mu + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$$

Where $y_t = \nabla^d Y_t$, and ∇^d Represent the displacement factor (The differences d for time series Y_t), That is $y_t = \nabla Y = Y_t - Y_{t-1}$.

The ARIMA model can be rewritten as a function of the posterior displacement coefficient (B) as follows:

$$\varphi_p(B) y_t = \theta_q(B) e_t \quad (1)$$

Where $\varphi_p(B), \theta_q(B)$, the polynomials are of degree p and q, respectively, and are as follows:

$$\varphi_p(B) = 1 - \varphi B - \dots - \varphi_p B^p, \theta_q(B) = 1 - \theta B - \dots - \theta_q B^q$$

B. Autoregressive Integrated Moving Average

The ARIMAX model describes the dynamic relationship that links the input series (X_t) with the output series (Y_t). Hence, the effect of the input series on the output series is shown by the transfer function, and this effect is distributed over subsequent periods. The ARIMAX model is sometimes called a conversion function model and is expressed mathematically as shown in the following formulas [6]:

$$Y_t = w_0 X_t + w_1 X_{t-1} + \dots + w_k X_{t-k} + \epsilon_t \quad (2)$$

$$Y_t = (w_0 + w_1 B + \dots + w_k B^k) X_t + \epsilon_t$$

$$Y_t = w(B) X_t + \epsilon_t$$

Where: Y_t : Output series (dependent variable).

X_t : Input series (explanatory variable).

$w(B)$: Input series conversion function X_t , where $w(B) = \sum_{i=0}^{\infty} w_i B^i$.

w_0, w_1, \dots, w_k : Conversion function weights k.

ϵ_t : White noise, which is a time series that includes other effects on the series (Y_t), It is a series independent of the series (X_t).

There are some difficulties with the ARIMAX model defined in formula (2) is that the conversion function $w(B)$ may contain an unlimited number or a large number of parameters. The number of parameters of the input and output

series parameters is specified while another formula for the ARIMAX model is being developed. It requires fewer parameters that makes the input and output series be stable on average and invariance. If the transfers are made and the necessary differences are taken on them to convert them to a state of stability, then equation (2) becomes as follows [7]:

$$y_t = \frac{\gamma_s(B)}{\delta_r(B)} B^b x_t + \frac{\theta_q(B)}{\varphi_p(B)} e_t = \frac{\gamma_s(B)}{\delta_r(B)} x_{t-b} + \frac{\theta_q(B)}{\varphi_p(B)} e_t \quad (3)$$

Where

$$\gamma(B) = \gamma_0 - \gamma_1 B - \dots - \gamma_s B^s, \delta(B) = 1 - \delta_1 B - \dots - \delta_r B^r$$

y_t : Stationary Output Series.

x_t : Stationary Input Series.

$\frac{\gamma_s(B)}{\delta_r(B)}$: Input series transfer function.

b : Time Lag Coefficient, defined as the period (number of units of time) that precede the start of the series effect (x_t) on the series (y_t), meaning that $B^b x_t = x_{t-b}$.

s : The number of time units the series (x_t) will continue to affect the series (y_t).

r : The number of time units in which the series (y_t) is affected by its previous values.

e_t : The white noise.

C. Impulse response or Transfer function weights

The weights of the Transfer function are related to the time lag coefficient (b). It represents the change in the output series (y_t) if the input series (x_t) changes by one unit, meaning that these weights measure how the input series affects the output series. In other words, (w_0) is a measure of the influence of the current pulse response of the output series at the current value. (w_1) is a measure of the effect of the current pulse response of the output series with the value of the input series for a single period.

It should be noted that the ARIMAX model is called the causal model if it is ($w_i = 0$) then the system does not respond to the input variable except after its actual application, but if the pulse response weights are subject to absolute summation, the transform function model is called the fixed model of the series of pulse response weights are subject to absolute summation, i.e., when $\sum_{i=0}^{\infty} |w_i| < \infty$. The first method is based on an estimate of $\gamma_s(B), \delta_r(B), b$, as follows:

$$w(B) \delta_r(B) = \gamma_s(B) B^b (w_0 + w_1 B + \dots + w_k B^k) = (\gamma_0 - \gamma_1 B - \dots - \gamma_s B^s) \quad (4)$$

to obtain:

$$w_i = 0$$

$$w_i = \delta_1 w_{i-1} + \delta_2 w_{i-2} + \dots + \delta_r w_{i-r} + \gamma_0$$

$$w_i = \delta_1 w_{i-1} + \delta_2 w_{i-2} + \dots + \delta_r w_{i-r} + \gamma_{i-b}$$

$$w_i = \delta_1 w_{i-1} + \delta_2 w_{i-2} + \dots + \delta_r w_{i-r}$$

Weights $w_{b+s-r+1}, w_{b+s-1}, w_{b+s}$ and their number r are used as prime values to solve the equation:

$$\delta_r(B) w_i = 0, i > b + s$$

Therefore, the weights of the transfer function consist of:

- (b) of zero weights w_0, w_1, \dots, w_{b-1} .
- $(s - r + 1)$ of weights that have no pattern $w_b, w_{b+1}, \dots, w_{b+s-r}$.
- (r) of the initial values $w_{b+s-r+1}, w_{b+s-r+2}, \dots, w_{b+s}$.
- w_i Values $i > b + s$ Which follows the pattern in the equation $\delta_r(B)w_i = 0$.

The second method is called the direct method and depends on the cross-correlation function (CCF). It is the primary key in determining the ARIMAX model's order as it is directly related to determining and estimating the weights of the transfer function [8].

If $(t, k = 0, \pm 1, \pm 2, \pm 3, \dots)$ then the cross-correlation formula between the output series (y_t) and the input series (x_t) at displacement (k) will be [9]:

$$\rho_{xy}(k) = \frac{E(x_{t-\mu_x})(y_{t+k-\mu_y})}{\sqrt{E(x_{t-\mu_x})^2} \sqrt{E(y_{t+k-\mu_y})^2}} \quad (5)$$

$\sigma_{xy}(k)$: Covariance cross at displacement (k) .

$\sigma_x \sigma_y$: Standard deviations of the series (y_t) and (x_t) , respectively.

μ_x, μ_y : Means for the (y_t) and (x_t) series, respectively.

The relationship between the transformation function and the CCF can be found at the period $(t + k)$ by rewriting the transformation model as follows:

$$y_{t+k} = w_0 x_{t+k} + w_1 x_{t+k-1} + \dots + w_k x_{t+k} + \varepsilon_{t+k} \quad (6)$$

Then we multiply both sides of the formula (6) by (x_t) and take the expectation, and since μ_x and μ_y equal zero, then:

$$\sigma_{xy} = w_0 \sigma_{xy}(k) + w_1 \sigma_{xx}(k-1) + \dots$$

Whereas $(\sigma_{x\varepsilon} = 0, \text{ for all } k)$:

$$\rho_{xy} = \frac{\sigma_x}{\sigma_y} \{w_0 \rho_x(k) + w_1 \rho_x(k-1) + \dots\}$$

If the input series is white noise, then the weights function will be according to the following formula:

$$w_k = \frac{\rho_{xy}(k) \sigma_y}{\sigma_x} \quad (7)$$

w_k : Weights of the transfer function at displacement (k) .

$\rho_{xy}(k)$: The cross-correlation between the series (y_t) and (x_t) at displacement (k) .

D. Filtering (Whitening) ARIMAX model

Filtering or whitening is a process of purification of the time series by removing any known pattern resulting from the process of an autoregressive or moving average. Thus, the time series remains nothing but white noise and the process of filtering the two series of inputs (x_t) and outputs (y_t) of the ARIMAX model can be explained as:

$$\begin{aligned} \varphi_p(B) x_t = \theta_q(B) \tau &\Rightarrow \tau = \frac{\varphi_p(B)}{\theta_q(B)} x_t \quad \varphi_p(B) y_t = \\ \theta_q(B) \vartheta &\Rightarrow \vartheta = \frac{\varphi_p(B)}{\theta_q(B)} y_t \end{aligned} \quad (8)$$

Since τ and ϑ are the white noise limits for the input and output series, respectively. It should be noted that the filtering process itself must be applied to both the input and output series to maintain the integrity of the transformation function.

E. ARIMAX Model Building Stages

If the input string (X_t) and the output string (Y_t) are in raw form, the steps to construct the ARIMAX model can be summarized as follows:

F. Model Identification Stationary

The stationery of both the input series (X_t) and the output series (Y_t) must be confirmed by drawing for both the Autocorrelation function and the partial Autocorrelation function. The stationery is removed on average by taking the appropriate number of differences. If there is no stationary invariance, and appropriate mathematical transformation are made.

1) *Filter the input and output series* as shown in Form (8).

2) *Cross-correlation estimation*: The cross-correlation values $\rho_{\tau\vartheta}(k)$ for the two filtered series $\sigma_\tau, \sigma_\vartheta$ for different periods $(t, t + k)$ are estimated as follows:

$$\rho_{\tau\vartheta}(k) = \frac{\sigma_{\tau\vartheta}(k)}{\sigma_\tau \sigma_\vartheta} \quad (9)$$

$\sigma_{\tau\vartheta}(k)$: Cross-covariance at displacement (k) . $\sigma_\vartheta \sigma_\tau$: standard deviations of the filtered series (τ) and (ϑ) , respectively. The significance of $\rho_{\tau\vartheta}(k)$ is tested by comparing it with its standard error.

3) *Transfer function weights*: The transfer function weights w_0, w_1, \dots, w_k are estimated directly by relying on the cross-correlation coefficients of the filtered input and output string, then:

4) *Model rank determination*: The values of (b, s, r) are theoretically determined by the rules established by (Wei 2006) as follows:

- The value of b is the number of significant cross-correlations in the positive direction after displacement (m) , or the number of weights until a weight equal to zero after w_0 is reached.
- If a specific pattern appears between $(m + a, m + a + c)$ then $r = c$.
- If there is no specific pattern for the cross-correlation between the two displacements $(m, m + a)$ then $s = a$, then if the cross-correlation is significant even the displacement $m, s = m$. If the above rules cannot be applied, the model ranks can be determined by conducting a set of experiments and then choosing the values according to appropriate comparison criteria.

5) *After determining the theoretical values* for (b, s, r) , then the initial estimates $\hat{\gamma}_s$ and $\hat{\delta}_r$ can be found by equation (3), then we can get a primary estimator for the transfer function:

$$\hat{w}_k(B) = \frac{\hat{\gamma}_s}{\hat{\delta}_r} B^b \quad (10)$$

6) *Identify the ARIMA model for white noise ε_t* : After estimating the primary transfer function \hat{w}_i , it can be estimated ε_t , since:

$$\hat{\varepsilon}_t = y_t - \hat{w}(B)x_{t-b} = y_t - \frac{\hat{\gamma}_s}{\hat{\delta}_r} x_{t-b} \quad (11)$$

Then, the appropriate (ARIMA) model for white noise that can be diagnosed through the behavior functions of the

autocorrelation, partial autocorrelation, and cross-correlation is determined to obtain:

$$\varphi_p(B)\varepsilon_t = \theta_q(B)e_t \Rightarrow e_t = \frac{\varphi_p}{\theta_q} \varepsilon_t \quad (12)$$

G. Estimate Model Parameters

After specifying a prototype for the transfer function, in this step the ARIMAX parameters are estimated, that is done by rewriting equation (3) as follows:

$$\delta_r(B)\varphi_p(B)y_t = \varphi_p(B)\gamma_s(B)x_{t-b} \quad (13)$$

Or by the following brief form:

$$\omega(B)y_t = \phi(B)x_{t-b} + \lambda(B)e_t \quad (14)$$

As that:

$$\lambda(B) = \delta_r(B)\theta_q(B) \text{ and } \phi(B) = \gamma_s(B)\varphi_p(B) \text{ and } \omega(B) = \delta_r(B)\varphi_p(B)$$

$$\omega(B) = (1 - \delta_1 B - \dots - \delta_s B^s)(1 - \varphi_1 B - \dots - \varphi_p B^p)$$

$$\omega(B) = (1 - \omega_1 B - \omega_2 B^2 - \dots - \omega_{p+r} B^{p+r}) \quad (15)$$

$$\phi(B) = (1 - \varphi_1 B - \dots - \varphi_p B^p)(\gamma_0 - \gamma_1 B - \dots - \gamma_s B^s)$$

$$\phi(B) = (\phi_0 - \phi_1 B - \phi_2 B^2 - \dots - \phi_{p+s} B^{p+s}) \quad (16)$$

$$\lambda(B) = (1 - \delta_1 B - \dots - \delta_r B^r)(1 - \theta_1 B - \dots - \theta_q B^q)$$

$$\lambda(B) = (1 - \lambda_1 B - \lambda_2 B^2 - \dots - \lambda_{r+q} B^{r+q}) \quad (17)$$

Equation (18) can be written as a factorial form as follows:

$$e_t = y_t - \omega_1 y_{t-1} - \dots - \omega_{p+r} y_{t-p-r} \quad (18)$$

As λ_j, ϕ_k are functions in $\varphi_l, \theta_k, \gamma_j$ and δ_i , and the random error limit is distributed in a normal distribution with an average of zero and a variance of σ_e^2 . Through the conditional maximum likelihood function of the following series of errors, the parameters of the ARIMAX model can be estimated [10], [11]:

$$L(b, x, y, x_0, y_0, e_0) = (2\pi\sigma_e^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma_e^2}} \quad (19)$$

As x_0, y_0, e_0 are the initial values for the e_t estimate.

H. Model Diagnostics and Testing

The estimated ARIMAX model is suitable for forecasting according to the following conditions:

- If the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) coefficients for the remaining e_t model is not significant, meaning that they fall within the confidence limits [12].

- The e_t residue should be independent of the filtered series residue τ , even if the cross-correlation not significant. This is done using the (Ljung-Box) test, which is defined according to the following formula:

$$L = M(M+2) \sum_{j=0}^K n(M-j)^{-1} \rho_{\tau,e}^2(k) \quad (20)$$

As $L \sim \chi_{(K+1)-M}^2$, M is the estimated number in the ARIMAX model, so that:

$$M = n - t_0 + 1$$

I. Forecasting

The established fit and accuracy of the model in the previous step can be used to predict the future values of the output series (\hat{y}_t). The past data for each of the input and output series can measure the accuracy and efficiency of the predictive model through some criteria, including (AIC), (BIC), and (RMSE) [13], [14], [15].

III. RESULTS AND DISCUSSION

The Dow Jones index is the oldest and most famous global stock exchange index ever. It was created in 1896. Today, it is still one of the essential indicators monitored by investors in the whole world. This index includes the shares of thirty of the most prominent American companies traded on the New York Stock Exchange and NASDAQ and includes Companies in various categories such as finance and consumer goods, including Goldman Sachs, Visa, and McDonald's. The initial idea behind this indicator is that these large industrial companies should measure the American economy because it is closely related to the growth of large industries. Therefore, the case can be generalized to the market as a whole. After all, investors often prefer to use indicators at tracking the market to give them a narrower view of what is happening in the market as a whole. Dow Jones shares are classified as large-capitalized stocks, which indicates the performance of most large-capitalized shares in the market. However, shares of large-capital companies represent a minority in the stock market. However, we note the tendency of the market's performance to the background of the performance of shares of large companies.

A. Description of the Data

The research data includes the time series of the Dow Jones Index values $Y(t)$ (output series), the crude price series for the Brent crude mix $X(t)$ (input series). The data was recorded daily for the period from 1/1/2020 to 1/5/2020 at (87) views for each series. The data do not include the holidays in which trading stops. The language (R) was used to obtain the application frame results, as the two series were drawn in Figure 1. The two series have been decreasing over time, especially the Brent crude price series. Figure 2 represents the Autocorrelation coefficients for the output series data $Y(t)$ and the input series $X(t)$. These coefficients fall outside the confidence limits, meaning that the two series are not stationary.

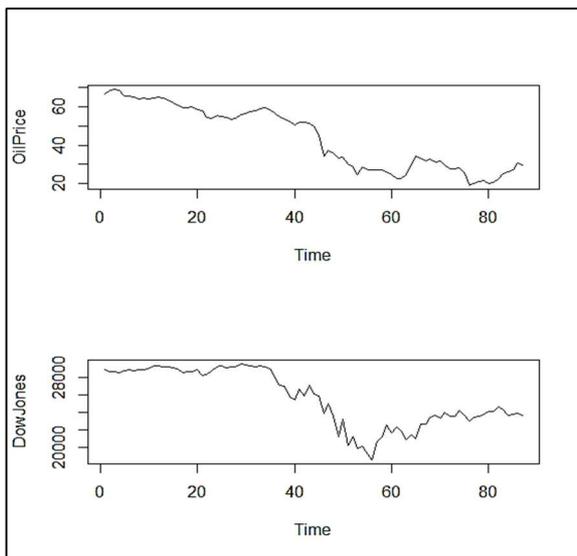


Fig. 1 The Dow Jones And Brent Daily Values Series

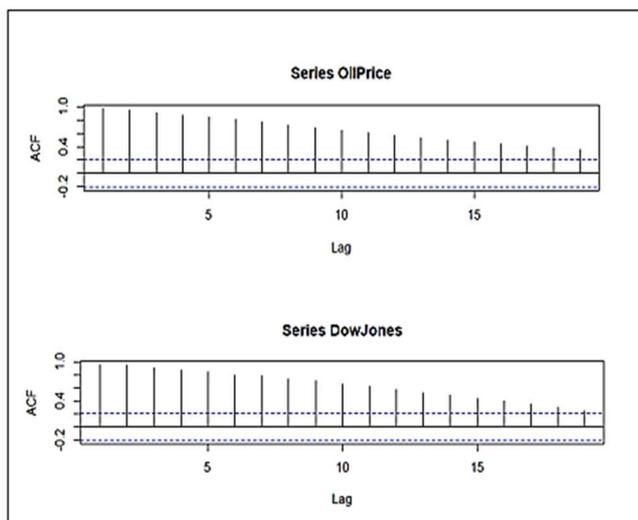


Fig. 2 Autocorrelation Coefficients for The Dow Jones And Brent Crude Series

The Augmented Dickey-Fuller test was also used at the level of significance (0.01) to analyze the stationary of both time series Dow Jones and the international oil price before and after using the estimated correlation function. The results are shown in Table (1).

TABLE I
AUGMENTED DICKEY-FULLER TEST

Time series	Augmented Dickey-Fuller (ADF)			
	The original time series		The first difference for the time series	
	Test Statistics	Sig	Test Statistics	Sig
Oil Price	-1.3738	0.8333	-3.3891**	0.04241
Dow Jones	-2.0757	0.5449	-3.574**	0.04048

***sig at 0.01, ** sig at 0.05, * sig at 0.1

It is clear from Table 1 that the results of the ADF test indicate the not stationary of the two output and input series at a significant level (0.05). The calculated statistical significance levels were more significant than (0.05).

However, the test results indicated that the time series reach the stationary stage after Taking the first differences of both series, as the value of the significance levels is less than (0.05). The output and input series are drawn after taking the first differences to confirm stationary, as shown in Figure (2). Furthermore, the ACF coefficients for the two series are plotted in Figure (3).

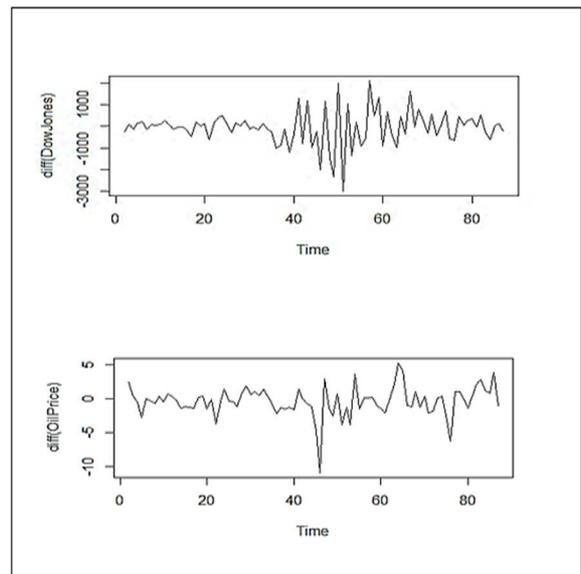


Fig. 3 Dow Jones Indexes and The Brent Crude Series After First Difference

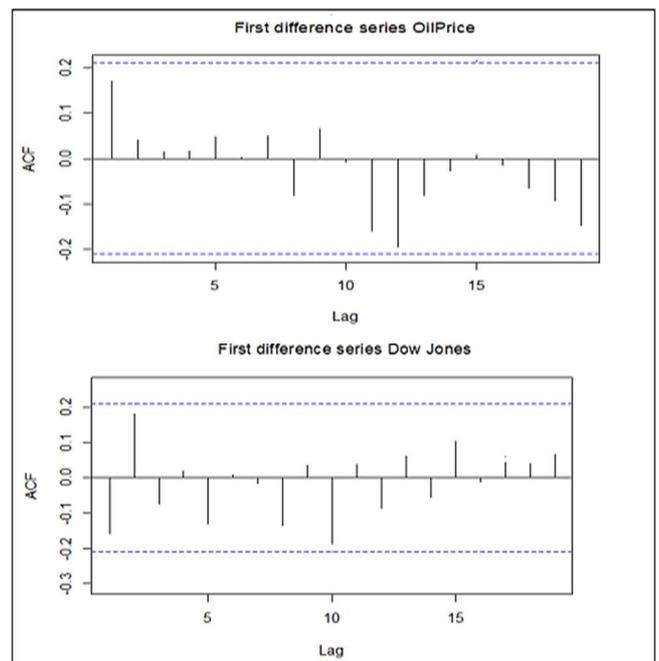


Fig. 4 The First Difference ACF For Dow Jones And Brent Crude Series

B. ARIMA Models (p, d, q)

After performing the ADF test, it was found that stationary in the input and output series is achieved at the first difference and to determine the appropriate ARIMA models. These models were differentiated through the statistical measures (AIC), (BIC), (RMSE). The results were put in Table 2 for the best models obtained for the two series prices Dow Jones and Brent crude.

TABLE II
CRITERIA FOR SUPERLATIVE BETWEEN BRENT CRUDE ARIMA MODELS

Models	AIC	BIC	RMSE	Models	AIC	BIC	RMSE
ARIMA (1,1,0)	378.9	380.2	2.165	ARIMA (0,1,2)	1380.5	1387.9	722.1
ARIMA (0,1,1)	381.3	386.2	2.182	ARIMA (1,1,2)	1382.5	1392.3	726.4
ARIMA (1,1,1)	382.7	387.1	2.187	ARIMA (2,1,2)	1384.5	1393.8	730.8
ARIMA (1,1,2)	384.9	390.8	2.195	ARIMA (2,1,1)	1385.6	1395.4	739.9
ARIMA (2,1,2)	387.8	396.4	2.203	ARIMA (1,1,1)	1387.6	1396.1	755.6
ARIMA (2,1,1)	389.6	398.3	2.225	ARIMA (0,1,1)	1391.4	1399.3	773.2

The best model being the Dow Jones Index series, is the model ARIMA (0,1,2). The best model that represents the Brent crude price series is the model ARIMA (1,1,0), as it has the lowest values of the (AIC) (BIC), (RMSE). Tables (3) and (4) represent the results of estimating the two models ARIMA (0,1,2) and ARIMA (1,1,0). The results indicate the significance of all the two models' estimated parameters, as all the values of the significance level (P-value) less than (0.05) significance level.

TABLE III
ESTIMATING OF THE DOW JONES INDEX SERIES ARIMA (0,1,2)

Parameter	Estimate	Std. Error	t	P-value
MA(1)	0.319	0.100132	3.188	0.002
MA(2)	-0.414	0.100999	-4.105	0.000

TABLE IV
ESTIMATING FOR BRENT CRUDE SERIES ARIMA (1,1,0)

Parameter	Estimate	Std. Error	t	P-value
AR(1)	0.991	0.0045	215.964	0.000

ARIMA models residuals were tested by drawing both the standardized residual, autocorrelation function, and the (Ljung-Box) statistic. It aims to ensure the two models' accuracy, as shown in Figures 5 and 6.

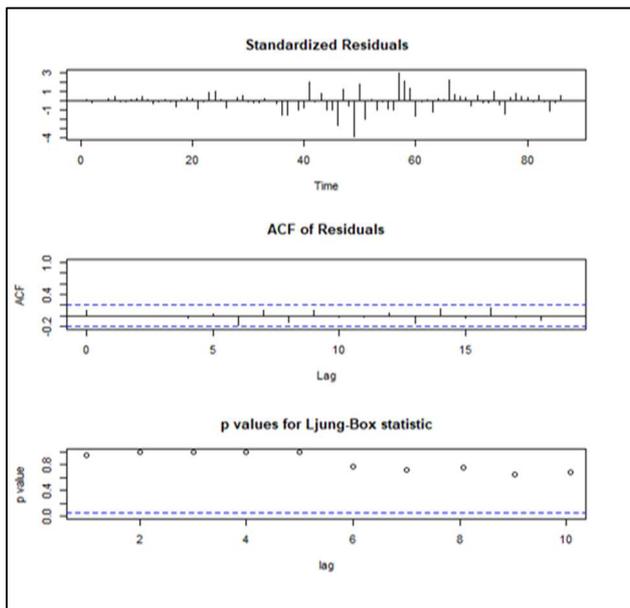


Fig. 5 Standard Errors, ACF and (Ljung-Box) For ARIMA (0,1,2) Residual

All the standard residual models' values are within the confidence limits; it was the case with the values of the

residual autocorrelation function. All the values of the significance level for the (Ljung-Box) statistics for the two models were more significant than the significance level (0.05). Thus, the residual of the two models is random changes that follow the normal distribution. Hence, the choice of the two models ARIMA (0,1,2), ARIMA (1,1,0) is appropriate for estimating the series of outputs and inputs, respectively.

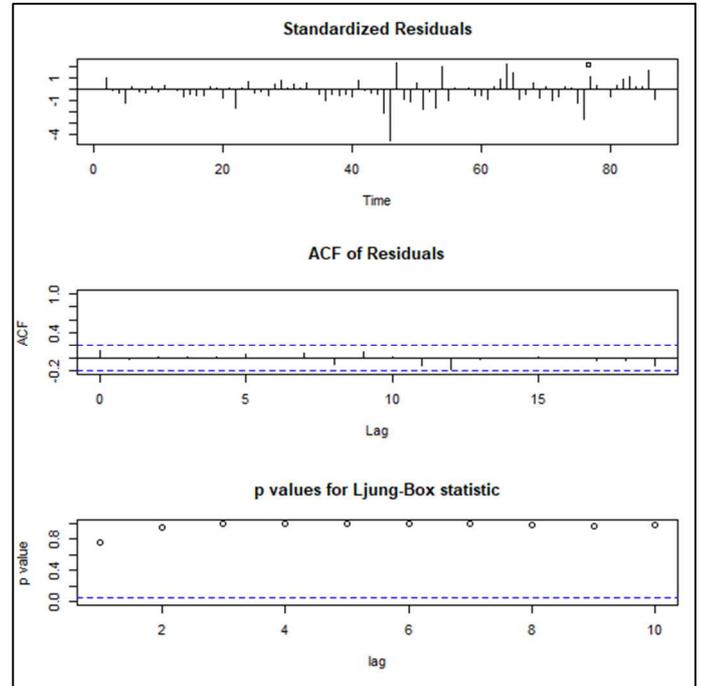


Fig. 6 Standard Residuals, ACF and (Ljung-Box) For ARIMA (1,1,0) Residual

C. ARIMAX (p, d, q) (b, r, s) Model

Determining the most appropriate model for the input series means whitening this series, here we get τ_t values as defined in formula (8), then we must whiten the output series by the same order that used in input series whitening process, thus we got ϑ_t values. Then by formula (9) we can estimate the cross-correlation coefficients between the filtered series τ_t and ϑ_t displacement (16) lags as shown in Figure (7) and Table (5).

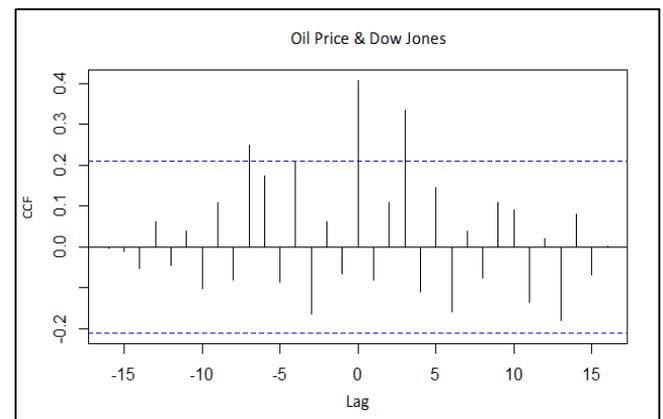


Fig. 7 Cross-Correlation Coefficients Between The τ_t And ϑ_t Filter Series

TABLE V
CROSS-CORRELATION FUNCTION (CCF) FOR FILTERED SERIES

Lag	CCF	Lag	CCF	Lag	CCF
0	0.4083	6	-0.1597	12	0.0203
1	-0.0822	7	0.0385	13	-0.1798
2	0.1088	8	-0.0747	14	0.0802
3	0.3346	9	0.1089	15	-0.0681
4	-0.1101	10	0.0904		
5	0.1439	11	-0.1351		

Table 5 above shows the time lag coefficient is either (b = 0) or (b = 3) due to the clarity of the cross-correlation's significance with the positive direction at these two displacements. However, concerning both r and s, it can be said that they cannot be decided with certainty. Therefore, it will be resorting to the experimental method of ARIMAX models of different ranks and then distinguish between these models employing RMSE and AIC. Different ARIMAX models have been tested and compared according to RMSE and AIC standards, and the results of the comparison in table (6) show the preference of ARIMAX order (b = 3, r = 1, s = 1).

TABLE VI
ARIMAX STANDARD DIFFERENTIATION CRITERIA

b, s, r	RMSE	AIC	b, s, r	RMSE	AIC
(0,0,0)	2.032	312.6	(3,0,0)	0.452	185.02
(0,0,1)	1.056	258.36	(3,0,1)	0.458	186.09
(0,0,2)	0.662	220.39	(3,0,2)	0.454	185.7
(0,0,3)	0.499	193.43	(3,0,3)	0.349	171.48
(0,1,0)	0.495	192.39	(3,1,0)	0.44	182.46
(0,1,1)	0.459	187.6	(3,1,1)	0.327	167.67
(0,1,2)	0.363	179.83	(3,1,2)	0.344	168.54
(0,1,3)	0.435	180.81	(3,1,3)	0.344	170.49
(0,2,0)	1.398	271.72	(3,2,0)	0.542	201.75
(0,2,1)	0.536	201.39	(3,2,1)	0.494	187.64
(0,2,2)	0.5	197.62	(3,2,2)	0.35	172.92
(0,2,3)	0.445	184.93	(3,2,3)	0.354	173.94
(0,3,0)	4.76	371.45	(3,3,0)	0.868	236.95
(0,3,1)	1.415	275.98	(3,3,1)	0.556	209.13
(0,3,2)	0.5536	206.38	(3,3,2)	0.549	205.13
(0,3,3)	0.5331	199.44	(3,3,3)	0.444	184.44

To directly estimate the Transfer function's weights, formula (11) was used, and we obtained the results shown in Table (7).

TABLE VII
ESTIMATED WEIGHTS OF THE TRANSFER FUNCTION

Lag	\hat{w}_i	Lag	\hat{w}_i	Lag	\hat{w}_i
0	0.135	6	-0.053	12	0.006
1	-0.027	7	0.012	13	-0.059
2	0.036	8	0.024	14	0.026
3	0.111	9	0.036	15	-0.022
4	-0.036	10	0.030		
5	0.047	11	-0.044		

Therefore, the appropriate initial transformational function model is (b=3, s=1, r=1), and it is written according to the following formula:

$$y_t - \delta_1 y_{t-1} = \gamma_0 x_t - \gamma_1 x_{t-3} \quad (21)$$

$$(1 - \delta_1 B)y_t = (\gamma_0 - \gamma_1 B)x_{t-3} \quad (22)$$

According to the estimated parameters:

$$(1 + 0.246 B)y_t = (0.135 + 0.027 B)x_{t-3}$$

By examining and analyzing the residual series of the prototype (ARIMAX) (3,1,1) in order to be diagnosed according to one of the ARIMA (p, 0, q) models. It became clear that the appropriate model for these residuals is ARIMA (0,0,1), which is shown in Figure 6. If the standardized residual values for the model within limits confidence, the values of the Autocorrelation function, and all the values of the (Ljung-Box) statistic were greater than the significance level (0.05), thus, the appropriate model for the white noise series ε_t is ARIMA (0,0,1). The formula is as follows:

$$\varepsilon_t = \frac{1}{1 - \theta_1 B} e_t = \frac{1}{1 - 0.045B} e_t \quad (23)$$

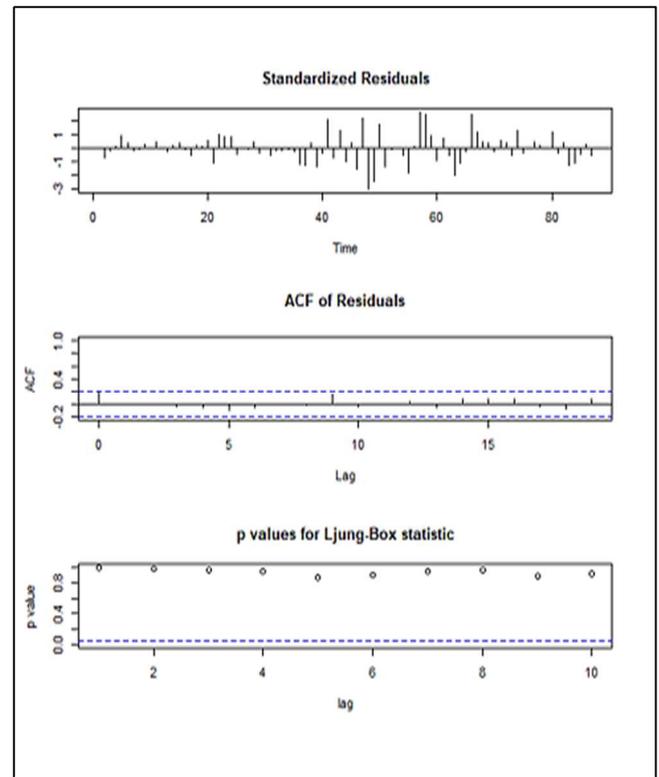


Fig.8 Standard Errors, ACF And (Ljung-Box) For ARIMA (0,0,1) Residual

From Figure 8 above, the initial model (ARIMAX) (3,1,1) was estimated according to the following formula:

$$y_t = \frac{(0.135+0.027 B)}{(1+0.246)} + \frac{1}{(1-0.045B)} e_t \quad (24)$$

According to the formula (21) and using the library (TSA) within the language (R) and based on the initial values obtained through the prototype, the final estimates of the parameters of the ARIMAX (3,1,1) (0,0,1) model have been found. The formula was as follows:

$$y_t = \frac{(0.155+0.019B)}{(1+0.217)} x_{t-3} + \frac{1}{(1+0.2105)} e_t \quad (25)$$

Figure 9 below represents the residual test for the final model. This test found the autocorrelation function's values within the confidence limits, and the residual follows the normal distribution.

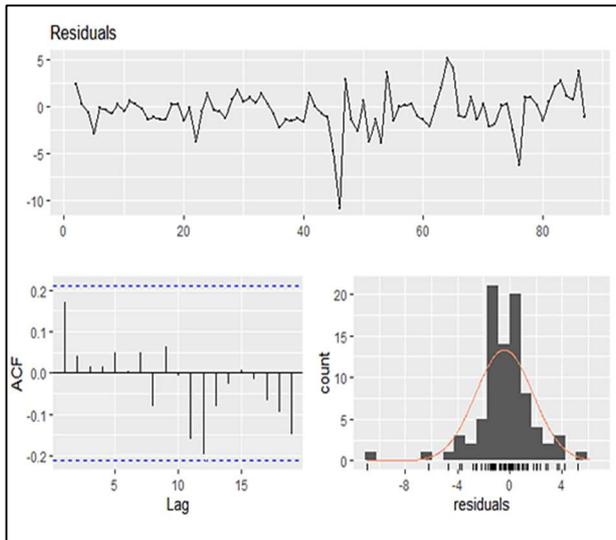


Fig. 9 ACF, Normal Distribution Fitting For ARIMAX (3,1,1)(0,0,1) Residuals

In addition to examining the independency between the residuals series and the input series τ_t , the cross-correlation function can be obtained. It was found that there were no significant correlations, and this indicates their independence, as shown in Figure 10 below.

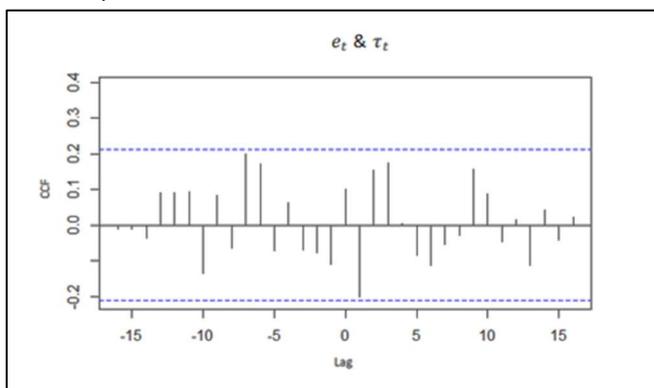


Fig. 10 CCF Between The Series e_t And The String τ_t .

Finally, a comparison between ARIMAX (3,1,1) (0,0,1) model and ARIMA (0,1,2) model were done using (AIC) and (RMSE). The results are set out in Table 8 below. The ARIMAX (3,1,1) (0,0,1) has the advantage of being the record of the two comparison criteria' lowest values.

TABLE VIII
AIC RMSE FOR ESTIMATED ARIMA AND ARIMAX MODELS

Model	AIC	RMSE
ARIMA(0,1,2)	192.44	721.99
ARIMAX(3,1,1)(0,0,1)	187.92	485

Thus, the values of the Dow Jones index were predicted according to the ARIMAX (3,1,1) and (0,0,1) models for 60 days. By the effect of Brent crude prices and through the library (Forecast) within the language (R), the results are shown in table (9) figure (11) below.

TABLE IX
PREDICTIVE VALUES FOR THE DOW JONES INDEX PRICE

Lag	D.J	Lag	D.J	lag	D.J	Lag	D.J
1	23813.40	16	24477.02	31	24961.41	46	25284.04
2	23785.70	17	24515.85	32	24987.24	47	25301.24
3	23884.71	18	24503.54	33	25012.37	48	25317.98
4	23856.21	19	24390.28	34	25036.83	49	25334.27
5	23786.70	20	24626.01	35	25060.64	50	25350.13
6	24021.59	21	24660.80	36	25083.81	51	25365.57
7	24079.26	22	24694.65	37	25106.36	52	25380.59
8	24124.40	23	24727.61	38	25128.31	52	25395.21
9	24175.22	24	24759.68	39	25149.67	54	25409.44
10	24220.42	25	24790.89	40	25170.46	55	25791.65
11	23806.27	26	24821.27	41	25190.69	56	25787.14
12	24310.81	27	24850.84	42	25210.39	57	25759.60
13	24354.41	28	24879.62	43	25229.56	58	26273.06
14	24396.21	29	24907.62	44	25248.21	59	26358.48
15	24437.28	30	24934.88	45	25266.37	60	26257.55

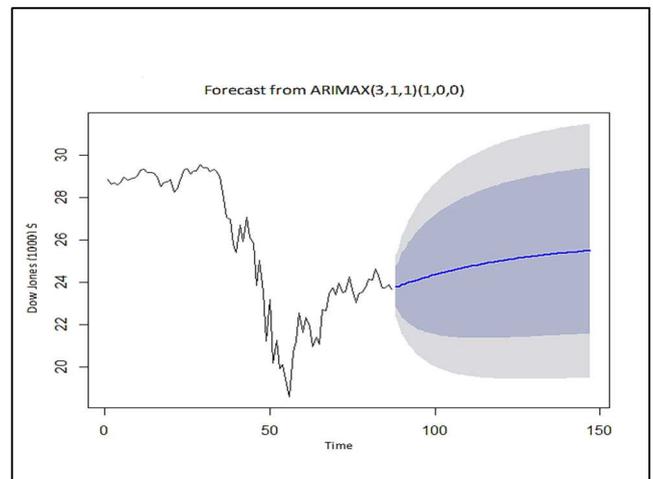


Fig. 11 Predictive Values For The Dow Jones Index Price

IV. CONCLUSION

Through the results reached in the applied frame, we can conclude the following results: The Dow Jones Indexes (Y_t) and Brent (X_t) series are not stationary on average. The first difference for the two series was taken to achieve stationary, while the two series were noted for stationary and devoid of seasonal effects. Thus, the most appropriate models for both input (X_t) and output (Y_t) series were ARIMA (1,1,0) and ARIMA (0,1,2), respectively. It turns out that the ranks for the transfer function appropriate to describe the relationship between the input series (X_t) and the output series (Y_t) are ($b = 3, s = 1, r = 1$), whereas the appropriate model for errors (white noise) the output series (Y_t) ARIMA

(0,0,1), therefore the transfer function model is ARIMAX (3,1,1) (0,0,1) according to the following formula:

$$y_t = \frac{(0.155+0.019B)}{(1+0.217)} x_{t-3} + \frac{1}{(1+0.2105)} e_t$$

The ARIMAX (3,1,1) (0,0,1) model outperformed the ARIMA (0,1,2) model according to the criteria (AIC), (RMSE). The growth of the index in the stock market has been slow based on the daily values of the Dow Jones Index. By the predictive values calculated according to the ARIMAX (3,1,1) and (0,0,1) model, it was found that the effect of the sharp decline in Brent crude prices due to the COVID 19 epidemic was slight and began fairly late. After the third day, the Dow Jones index values returned to normal nature.

REFERENCES

- [1] Elshendy, M., Colladon, A. F., Battistoni, E., & Gloor, P. A. (2018). Using four different online media sources to forecast the crude oil price. *Journal of Information Science*, 44(3), 408-421.
- [2] Alamro, R., McCarren, A., & Al-Rasheed, A. (2019, December). Predicting saudi stock market index by incorporating gdelt using multivariate time series modelling. In *International Conference on Computing* (pp. 317-328). Springer, Cham.
- [3] Jaraskunlanat, N., & Kijboonchoo, T. (2016). A Study of Factors Affecting the Gold Price in Thailand during 2005–2015. *International Research E-Journal on Business and Economics*, 2(1).
- [4] Colladon, A. F., & Scettri, G. (2019). Look inside. Predicting stock prices by analysing an enterprise intranet social network and using word co-occurrence networks. *International Journal of Entrepreneurship and Small Business*, 36(4), 378-391.
- [5] Sharaf, H. K., Ishak, M. R., Sapuan, S. M., Yidris, N., & Fattahi, A. (2020). Experimental and numerical investigation of the mechanical behavior of full-scale wooden cross arm in the transmission towers in terms of load-deflection test. *Journal of Materials Research and Technology*, 9(4), 7937-7946.
- [6] Clements, A. E., & Todorova, N. (2016). Information flow, trading activity and commodity futures volatility. *Journal of Futures Markets*, 36(1), 88-104.
- [7] Sharaf, H. K., Ishak, M. R., Sapuan, S. M., & Yidris, N. (2020). Conceptual design of the cross-arm for the application in the transmission towers by using TRIZ–morphological chart–ANP methods. *Journal of Materials Research and Technology*, 9(4), 9182-9188.
- [8] Padhi, S. S., & Pati, R. K. (2017). Quantifying potential tourist behavior in choice of destination using Google Trends. *Tourism Management Perspectives*, 24, 34-47.
- [9] Ortega, L., & Khashanah, K. (2014). A neuro-wavelet model for the short-term forecasting of high-frequency time series of stock returns. *Journal of Forecasting*, 33(2), 134-146.
- [10] Sharaf, H. K., Salman, S., Dindarloo, M. H., Kondrashchenko, V. I., Davidyants, A. A., & Kuznetsov, S. V. (2021). The effects of the viscosity and density on the natural frequency of the cylindrical nanoshells conveying viscous fluid. *The European Physical Journal Plus*, 136(1), 1-19.
- [11] Aras, H. (2015). Modelling the Gold Price in Turkish Free Market: Static Approach. *Zeszyty Naukowe Uczelni Vistula*, (41 (3)/2015 Stosunki Międzynarodowe), 103-121.
- [12] Sharaf, H. K., Salman, S., Abdulateef, M. H., Magizov, R. R., Troitskii, V. I., Mahmoud, Z. H., ... & Mohanty, H. (2021). Role of initial stored energy on hydrogen microalloying of ZrCoAl (Nb) bulk metallic glasses. *Applied Physics A*, 127(1), 1-7.
- [13] Todorova, N., & Clements, A. E. (2018). The volatility-volume relationship in the LME futures market for industrial metals. *Resources Policy*, 58, 111-124.
- [14] Jammalamadaka, S. R., Qiu, J., & Ning, N. (2019). Predicting a stock portfolio with the multivariate bayesian structural time series model: Do news or emotions matter?. *International Journal of Artificial Intelligence*, 17(2), 81-104.
- [15] Hajrahimi, Z., & Khashei, M. (2021). Parallel hybridization of series (PHOS) models for time series forecasting. *Soft Computing*, 25(1), 659-672.