Prediction of OPEC oil prices using different versions of(GARCH) models in time series

Prof. Dr. Muhammad H. Al-Sharoot

Department of Statistics

College of Administration and Economics

University Of AL-Qadisiyah-IRAQ

malsharood@gmail.com

Omar M. Alramadhan

Department of Statistics

College of Administration and Economics

University Of AL- Qadisiyah-IRAQ

omar.muhy13@gmail.com

Abstract

Many time series are characterized by their large volatility over time, especially time series related to the movement of the economy، and those that relate to the change in stock prices or the movement of financial transactions and stock markets characterized by such phenomena of non-stationary over time, which makes it the problem of heterogeneity of disparities clearly. Since the time series analysis is required to construct the appropriate model for forecasting the future of the studied phenomenon is to achieve stationary on average and consistency of homogeneity of variance, The aim of this study was to study the use of predictive models that can adapt to time series with large fluctuations over time, A number of important models used to deal with heterogeneous time series in variance, GARCH, GARCH-M, and EGARCH have been studied and reviewed when the distribution of errors follows normal distribution, , Which was first used by the researcher Engle in 1982 and developed by other researchers, The characteristics of these models have been reviewed and applied for the purpose of forecasting the daily prices of oil according to the prices approved by OPEC for the period from 2003 to 2018, The study and practical analysis of oil price data showed that the best model for forecasting fluctuations in oil prices for the period mentioned is the EGARCH (2,1), By adopting some important criteria for selecting the best model such as AIC, SIC, H-QIC.

1. Introduction:

Some researchers focus on time series topics because they are important in studying the behavior of different phenomena over specific time periods through their analysis and interpretation, The topics of the time series include many areas (medical, environmental, economic, etc.) ,The definition of the time series (it is a series of observations that are arranged according to time of occurrence) There are two types, the first is Discrete Time Series and the second is the continuous time series, The aim of the time series analysis is to obtain an accurate description of the features of the phenomenon from which the time series is produced, and to construct a model to explain the behavior of that phenomenon , and predict future observations of the phenomenon studied based on what happens in the past.

The most important models applied to time series data are the ARMA models used in many different fields. To be able to use the ARMA model, there are three conditions for the random error of the model:

$$i)\ E(\varepsilon_t) = 0$$

$$ii)\ V(\varepsilon_t) = E\left(\varepsilon_t^2\right) = \sigma^2$$

$$iii)\ E(\varepsilon_t \varepsilon_s) = 0 \quad for \quad t \neq s$$

the infraction of those conditions for the existence of a particular factor may be externally or an emergency on time series need look for other models can adapt to those factors that led to the existence of differences in terms of this time series, and in particular in the time series on financial transactions.

2. Problem paper:

The problem of paper is that there are fluctuations in the OPEC crude oil price series, which has led to non- stationary oil prices and therefore the use of normal ARMA models will lead to unreasonable future predictions. The plans based on these results are therefore useless.

3. The aim:

The aim is to build the best model for forecasting the OPEC oil price series for the period from 2003 to 2018 by applying a number of different models that are used to predict in time series of fluctuations including GARCH model, GARCH-M model and EGARCH model.

4. Autoregressive Conditional Heteroscedasticity models ($ARCH_{(p)}$):[6][4]

It was the first model proposed by Robert Engle in 1982. The ARCH model is a return series with a conditional average and a conditional variation. The conditional mean of the return series μ_t is constant, the conditional variance of the return series is in the form of a model that contains an error limit and a non-stationary equation, the equations of the ARCH model are as follows.

$y_t = \mu + x_t$ …… (1)[6]

$x_t = \sigma_t * \varepsilon_t$ ... … (2) [6] , $\varepsilon_t \sim iidN(0,1)$ and

$\sigma_t^2 = \Omega + \alpha_1 x_{t-1}^2 + \alpha_2 x_{t-2}^2 + \cdots + \alpha_j x_{t-p}^2$ ……………... (3)[6]

Where σ_t^2 is the equation of volatility, which can be written in the formula below.

$\sigma_t^2 = \Omega + \sum_{j=1}^{p} \alpha_j x_{t-j}^2$ ………………… (4)[6]

Whereas $\Omega > 0$, $\alpha_j \geq 0$, $j = 1,2, \ldots\ldots,p$

and, Ω, α_j, represent the parameters of the model.

The process is in the case of stationary if and only if the total parameters of the Autoregressive are positive and less than one.

5.Generalized Autoregressive Conditional Heteroscedasticity Model ($GARCH_{(p,q)}$):[6][4]

GARCH models (p≥1) and (q≥1) can be defined as follows:

$y_t = \mu + x_t$

$x_t = \sigma_t * \varepsilon_t$, ε_t ~iidN(0,1) and

$\sigma_t^2 = \Omega + \alpha_1 x_{t-1}^2 + \alpha_2 x_{t-2}^2 + \cdots + \alpha_p x_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 + \cdots + \beta q\, \sigma_{t-q}^2$ ….. (5)[6]

As the y_t series represents a stationary return series and uncorrelated

and μ represents the average of the stationary return series

And that they are independent ε_t series and similar distribution (independent identically distribution) and keep track of the standard normal distribution with mean 0 and variance 1.

And σ_t^2 is the function of volatility, which can be written in the formula below.

$\sigma_t^2 = \Omega + \sum_{j=1}^{p} \alpha_j x_{t-p}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-q}^2$ ... ... (6)[6]

whereas.

$\Omega > 0$, $\alpha_j \geq 0$, $j = 1,2, \ldots, p$, $\beta_i \geq 0$, $i = 1,2, \ldots, q$

And, Ω, αj, βi represent the parameters of the model.

6. Generalized Autoregressive Conditional Heteroscedastic in Mean Models (GARCH-M) [2][1]

The GARCH in Mean model is a model of heterogeneity, often used in financial applications. This model distinguishes the evolution of the mean and the variance at a time and the expected return of the financial assets is linked to the expected risk of the financial asset, where conditional variation is allowed to be a determining factor for the mean equation and conditional identification is the GARCH process

GARCH-M models (p, q), (p≥1) & (q≥1) can be defined as follows:

$\mathbf{y_t = \mu + \Psi \sigma_t^2 + x_t}$ ... ... (7) [1]

or

$\mathbf{y_t = \mu + \Psi \log \sigma_t^2 + x_t}$ ... ... (8)[1]

$x_t = \boldsymbol{\sigma_t * \varepsilon_t}$ $\varepsilon_t \sim$ iid N(0,1)

$\boldsymbol{\sigma_t^2 = \Omega + \sum_{i=1}^{p} \alpha_i x_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2}$

Whereas ($\boldsymbol{\Omega > 0}$, $0 \geq \boldsymbol{\beta_j}$ $0 \geq \boldsymbol{\alpha_i}$ for i=1,2,…,p , j =1,2,…,q) Parameters represent the parameter, while Ψ represents the risk premium parameter (Risk Premium Parameter), and instead of ($\boldsymbol{\sigma_t^2}$) the conditional standard deviation can be used σ_t. . The high average yield leads to the increase in conditional variation as evidence of increased risk. These GARCH in Mean models are often used in financial applications. Time series of stocks have observed that downward movements (negative shocks) are followed by higher fluctuations than upward movements (positive shocks) with the same capacity, This is known as the Leverage effect. This means that the fluctuations before and after the shock are asymmetrical as they depend on the type of shock positive or negative.

7. Exponential Generalized Autoregressive Conditional Heteroscedastic Models (EGARCH) [1][4]

This model suggested by Nelson in (1991) and on the contrary, the classic model GARCH which assumes symmetry oscillations around the shock. As well as the positive constraint imposed on parameters, Because the EGARCH model describes the relationship between the previous values of the random error and the conditional variation logarithm, with no restrictions on transactions that ensure that there are no negative effects of conditional variation, which allows avoiding positive transaction constraints ($\boldsymbol{\beta_j}$ & $\boldsymbol{\alpha_i}$), As follows:

Let us have the EGARCH model of the class (p, q) (p ≥1) & (q ≥1). Therefore, this model can be written as follows:

$$\mathbf{y_t = \mu + x_t}$$

$$\mathbf{x_t = \sigma_t \, \varepsilon_t} \qquad \boldsymbol{\varepsilon_t \sim \text{iid } N(0,1)}$$

$$\ln(\sigma_t^2) = \Omega + \sum_{j=1}^{q} \beta_j \ln(\sigma_{t-j}^2) + \sum_{i=1}^{p} \alpha_i \left\{ \left| \frac{x_{t-i}}{\sigma_{t-i}} \right| - \sqrt{\frac{2}{\pi}} \right\} + \lambda_i \frac{x_{t-i}}{\sigma_{t-i}} \dots\dots \dots (9)$$ [1]

or

$$log(\sigma_t^2) = \Omega + \sum_{i=1}^{p} \alpha_i \, g(Z_t) + \sum_{j=1}^{q} \beta_j \log(\sigma_{t-j}^2) \dots\dots\dots(10)$$ [4]

whereas

$$\mathbf{g(Z_t) = \theta Z_t + \Upsilon(|Z_t| - E(|Z_t|)} \qquad \& \qquad \mathbf{Z_t = x_t/\sigma_t}$$

$$\mathbf{E(x_t/\sigma_t) = E\left\{\frac{|x_{t-i}|}{\sigma_{t-i}}\right\} = \sqrt{\frac{2}{\pi}}}$$

(Ω), β_j & α_i , j=1,2,…,q، i=1,2,…,p Represent the model parameters is not required to be positive, that the function g (Z_t) allows the signal size Z_t to be discrete effects from fluctuations, , And that the z_t limits are positive if the $g(z_t)$ is linear with parameters $(\boldsymbol{\theta} + \lambda)$ If z_t is negative, the $g(z_t)$is linearized by parameters $(\boldsymbol{\theta} - \lambda)$ This situation allows for asymmetry on the rise and fall in the share price, which in turn is very useful, especially in the context of bond pricing.

8. Augmented Dickey-Fuller Test [5][8]:

Augmented Dickey Fuller test (ADF) uses to detect the presence of the root of the unit in univariate test any time series whether stationary series or not, The ADF test has a regression in the first difference in the series against the series with the time Difference (p).

Using the following equation:

$$\Delta \mathbf{y_t} = \mu + \boldsymbol{\lambda_t} + \emptyset \mathbf{y_{t-1}} + \sum_{j=1}^{k} \boldsymbol{\delta_j} \Delta\, \mathbf{y_{t-j}} + \mathbf{x_t} \qquad \dots \dots (\mathbf{11})^{[8]}$$

As the y_t represents the time series to be tested, k number of Difference time, Δ represents the first differences in a series of return, x_t represents an error $x_t \sim$ iid $(0, \sigma^2)$, and $(\mu , \boldsymbol{\lambda} , \boldsymbol{\delta_j}, \emptyset)$ Symbolizes the parameters of its appreciation. The hypothesis can be tested

$\mathbf{H_\circ} : \emptyset = \mathbf{0}$ Series yield has unit root (Series yield is non-stationary)
$\mathbf{H_1}: \emptyset < \mathbf{0}$ A series yield does not has a unit root (Series yield is stationary)
Using statistics:

$$\boldsymbol{\tau} = \frac{\widehat{\emptyset}}{\mathbf{se}(\widehat{\emptyset})} \qquad \dots \dots (\mathbf{12})^{[8]}$$

The null hypothesis is rejected if the t-statistic value is greater than the statistical value of t-statistic and vice versa.

9. Ljung - Box Test [9][6]

The researchers (Ljung & Box) proposed this test in 1978, To test random errors of the time series by calculating the autocorrelation coefficients of a series of residual, the following hypothesis is tested:

$\mathbf{H_\circ} : \boldsymbol{\rho_1} = \boldsymbol{\rho_2} = \cdots = \boldsymbol{\rho_k} \dots = \boldsymbol{\rho_m} = \mathbf{0}$; $\mathbf{k = 1, 2, \dots, m}$
$\mathbf{H_1}: \boldsymbol{\rho_k} \neq \mathbf{0}$ **for some values of k**
Using statistics:

$$\mathbf{Q_{(m)}} = \mathbf{n(n+2)} \sum_{\mathbf{k=1}}^{\mathbf{m}} \frac{\widehat{\boldsymbol{\rho}}_{\mathbf{k}}^{\mathbf{2}}}{\mathbf{n-k}} \sim \boldsymbol{\chi}_{\mathbf{m-p}}^{\mathbf{2}} \qquad \dots \dots (\mathbf{13})^{[9]}$$

And that each of:
n: Represents the sample size (number of views of the time series).
m: It represents the number of back shift.
p: Number of parameters estimated in the model
$\widehat{\boldsymbol{\rho}}_{\mathbf{k}}^{\mathbf{2}}$: Represent the capabilities of the self-correlation boxes of the model's residual series.

then for string $x_t = y_t - \mu$ and x_t^2.
If p-value ≥0.05 means not rejecting the hypothesis H_0, the errors $x_t = y_t - \mu$ are random (Identically Independent Distribution) and there is (no effect ARCH) or (heteroscedasticity), and vice versa.

10. Lagrange Multiplier (*ARCH – Test*):[7][6]

It was proposed by Engle in 1982 and is used to determine whether the errors follow the ARCH process or not, which is based on the estimation of the equation under study in the form of the smallest squares and then the estimation of errors and squares for previous periods. This means that we estimate the following equation:

$$\mathbf{x_t^2 = \Omega + \alpha_1 x_{t-1}^2 + \alpha_2 x_{t-2}^2 + \ldots\ldots\ldots\ldots + \alpha_p x_{t-p}^2} \qquad \ldots\ldots(\mathbf{14})^{[7]}$$

$\mathbf{x_t} = \mathbf{y_t - \mu}$, To test (ARCH (P)) we calculate the product of the coefficient of determination resulting from this estimate used the sample size of any amount $\mathbf{TR^2}$, Which is followed by $\boldsymbol{\chi}_{\mathrm{p}}^2$, Of the degree of freedom (p) under the premise of the nuisance that the errors are homogeneous (Conditional Homoscedasticity) The small values of R^2 mean that the errors of the previous periods do not affect the current error and therefore there is no trace of the ARCH effect.

$$\mathbf{archtest = T\,\widehat{R}^2 \sim \chi_{(P)}^2} \qquad \ldots\ldots(\mathbf{15})^{[7]}$$

11. Estimation [6]:-

Can be used Maximum likelihood Method To estimate GARCH parameters (p, q) as follows:

$$\mathbf{f(x_t/F_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{1}{2}\frac{x_t^2}{\sigma_t^2}\right)} \qquad (\mathbf{16})^{[6]}$$

The natural logarithm (L) function of vector parameters $\boldsymbol{\vartheta} = (\boldsymbol{\Omega}_\circ, \boldsymbol{\alpha_1}, \ldots, \boldsymbol{\alpha_p}, \boldsymbol{\beta_1}, \ldots, \boldsymbol{\beta_q})'$

We can write as follows:

$$\mathbf{L(\vartheta) = \sum_{t=1}^{n} I_t(\vartheta)} \qquad (\mathbf{17})^{[6]}$$

the conditional logarithm of the parameter vector $\boldsymbol{\vartheta}$ is

$$\mathbf{I_t(\vartheta) = Ln\, f(x_t/F_{t-1})}$$

$$\mathbf{I_t(\vartheta) = \frac{1}{2}Ln(2\pi) - \frac{1}{2}Ln(\sigma_t^2) - \frac{1}{2}\left(\frac{x_t^2}{\sigma_t^2}\right)} \qquad (\mathbf{18})^{[6]}$$

The following derivatives are calculated:

$$\frac{\partial \mathbf{I_t}}{\partial \boldsymbol{\vartheta}} = \frac{\partial \mathbf{I_t}}{\partial \boldsymbol{\sigma_t^2}} \frac{\partial \boldsymbol{\sigma_t^2}}{\partial \boldsymbol{\vartheta}}$$

The logarithm of the conditional probability density function is derived for the variable y_t for Ω , α_i , β_j

12. Model selection criteria: [9]

To choose the best model among those proposed models for assessment and prediction of the studied data, developed choice of model data, which ideally criteria and selection of the most common model standards are:

i - Akaikes Information criterion (AIC) [9][5]

Akaike (1974) presented a standard of information known as (AIC) When the time series models in (L) are reconciled with the parameters of the time series data under consideration and to assess the suitability of those models, the AIC is calculated for each model and the model that gives the lowest value is selected. The AIC formula can be written as follows:

$$AIC = n\ln(\hat{\sigma}_e^2) + 2L \qquad \ldots\ldots (19)$$ [9]

n: represents the sample size.

$\hat{\sigma}_e^2$: The variance of the model is calculated as follows:

$$\hat{\sigma}_e^2 = \frac{1}{n-L}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2 \qquad \ldots\ldots (20)$$ [9]

L: is the rank of the model.

ii- Schwarz Information criterion (SIC): [6][5]

In 1978, Schwarz introduced a new standard known as the Schwarz standard.

$$SIC = n\,ln(\hat{\sigma}_e^2) + L\ln(n) \qquad \ldots\ldots (21)$$ [6]

n: represents the sample size.

$\hat{\sigma}_e^2$: The variance of the model is calculated as follows:

$$\hat{\sigma}_e^2 = \frac{1}{n-L}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2$$

L: is the rank of the model.

This criterion addressed the problem of over-estimation in the AIC standard, And make the Penalty of the additional parameters stronger than the penalty in the AIC standard). The penalty for this criterion is L (In) n. One of the advantages of the SIC is that it estimates the rank of a model consistently that p , q is less or equal (p_{max}, q_{max}, respectively). It is stated that with reference to the chosen AIC or SIC, $\hat{P}(\text{SIC}) \leq \hat{P}(\text{AIC})$ Remain constant even in cases of small samples. Therefore, the use of SIC results leads us to models with minimal parameters.

iii- H-Q Hannan- Quinn Criterion: [5][6]

The researchers Quinn and Hannan (1979) proposed a new criterion for determining the rank of the studied model called Hannan-Quinn Criterion (H-Q (h)) and its mathematical formula:

$$H - Q = \ln \hat{\sigma}_e^2 + 2L\, C\, ln\left(\frac{ln(n)}{n}\right) \quad , \quad C > 2 \qquad \ldots\ldots (22)$$ [6]

As the second limit above decreases as quickly as possible at the stationary of the rank due to the repeated logarithm.

13. Forecasting[7][10]:

Prediction is one of the most important objectives of model construction in time series. It represents the last stage of time series analysis that can't be accessed without passing all tests and diagnostic tests to validate the model used in prediction.

The following is a forecast prediction of the GARCH model.

In the same way for all models (EGARCH, GARCH-M)

The prediction of the GARCH model (p, q)) (where p = 1, q = 1, GARCH (1.1)) is as follows:

$$\sigma^2{}_{\mathrm{t}} = E\ (x^2{}_t\ |I_t) = \hat{\Omega} + \hat{\alpha}_1\ x^2{}_{t-1} + \hat{\beta}_1\ \sigma^2{}_{\mathrm{t-1}}$$

Predicting one future value

$$\sigma^2{}_{\mathrm{t+1}} = E(x^2{}_{t+1}|I_t) = \hat{\Omega} + \hat{\alpha}_1\ E(x^2{}_t|I_t) + \hat{\beta}_1\ \sigma^2{}_{\mathrm{t}}$$

$$\sigma^2{}_{\mathrm{t+1}} = \hat{\Omega} + \hat{\alpha}_1\ \sigma^2{}_t + \hat{\beta}_1\ \sigma^2{}_{\mathrm{t}}$$

$$\sigma^2{}_{\mathrm{t+1}} = \hat{\Omega} + (\hat{\alpha}_1 + \hat{\beta}_1)\sigma^2{}_{\mathrm{t}}$$

Prediction of value L

$$\sigma^2{}_{\mathrm{t}+l} = E(x^2{}_{t+l}|I_t) = \hat{\Omega} + \hat{\alpha}_1 E(x^2{}_{t+l-1}|I_t) + \hat{\beta}_1\ \mathrm{E}\ (\sigma^2{}_{\mathrm{t}+l-1}\ |\mathrm{I_t}\ )$$

$$\hat{\Omega} + \hat{\alpha}_1\ \sigma^2{}_{t+l-1} + \hat{\beta}_1\ \sigma^2{}_{\mathrm{t}+l-1} = \sigma^2{}_{\mathrm{t}+l}$$

$$\sigma^2{}_{\mathrm{t}+l} = \hat{\Omega} + (\hat{\alpha}_1 + \hat{\beta}_1)\sigma^2{}_{\mathrm{t}+l-1}$$

Thus, the general formula for predicting GARCH (p, q) models is as follows:

$$\sigma^2{}_{\mathrm{t}+l} = \hat{\Omega} + \sum_{i=1}^{p} \hat{\alpha}_i \sigma^2{}_{t+l-i} + \sum_{\mathrm{j}=1}^{\mathrm{q}} \hat{\beta}_j \sigma^2{}_{\mathrm{t}+l-j}$$

14. Forecasting accuracy measures[10]:

to measure prediction accuracy developed standards are called the chosen model prediction accuracy standards is the most important.

i-: Root Mean Square Error (RMSE) :- [3][10]

This criterion is defined as the square root of the squared difference between both the real variance and the estimated variance σ_t^2, Due to the absence of significant real variation, the time series observations were used x_t^2.

Thus, the RMSE formula is given as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_t^2 - \widehat{\sigma^2}_t)^2} \qquad t = 1,2,\ldots\ldots,n \qquad \ldots.(23)^{[3]}$$

whereas

$\widehat{\sigma^2}_t$ represents the estimated variance.

x_t^2 represents the actual constant.

ii-: Mean Absolute Error (MAE) [3][10]

The mean absolute error is defined as the absolute difference between real and estimated variance, and the formula of the standard is given as follows:

$$MAE = \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\left|x_t^2 - \widehat{\sigma^2}_t\right| \qquad \ldots\ldots\ldots\ldots\ldots(24)^{[10]}$$

15. Application side

The oil price series in OPEC was analyzed for the period of 2-1-2003 until 4-1-2018 on a daily basis except for non-trading days, and the observations were 3874 daily observations. y_t returns were calculated using the natural logarithm of the data according to the following equation:

$y_t = \ln(P_t / P_{t-1})$...................... (25)[6]

whereas:

P_t: is the price of a barrel of oil according to OPEC at the period t.

$P_{(t-1)}$: The price of a barrel of oil, according to OPEC at the period t-1.

Figure (1) the time series of oil prices according to OPEC data

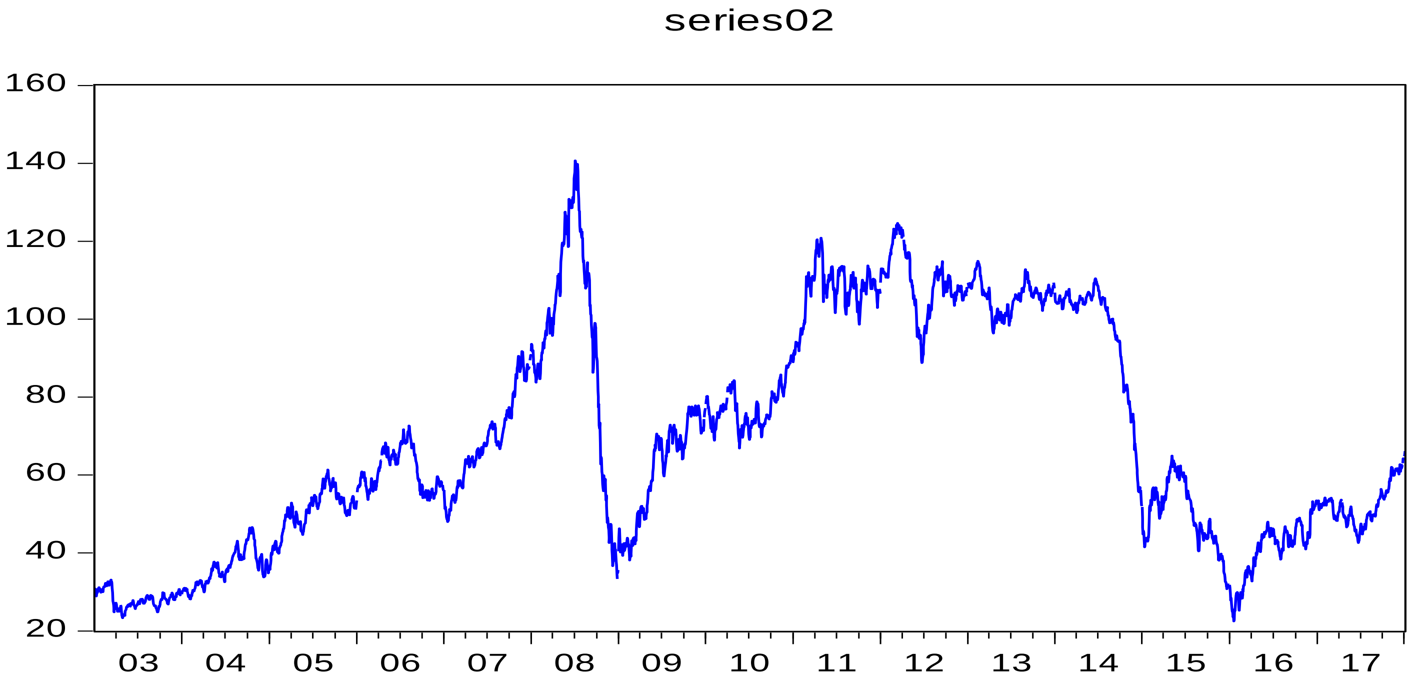


Figure (1) shows that the time series of oil prices is non- stationary and has high fluctuations, indicating fluctuations in variance

15. Augmented Dickey Fuller test for the oil prices series:

For the purpose of detecting the stationary of the time series for the price of a barrel of oil per day, the test of the Augmented Dicky Fuller was calculated and the test results as shown in Table (1)

Table (1) the Augmented Dicky Fuller test to test the stationary **of the time series**

Null Hypothesis: SERIES02 has a unit root
Exogenous: Constant
Lag Length: 2 (Automatic - based on SIC, maxlag=29)

Prob.*	t-Statistic		
0.4212	-1.719634	Augmented Dickey-Fuller test statistic	
	-3.431854	1% level	Test critical values:
	-2.862090	5% level	
	-2.567106	10% level	

We find from Table 1 that the value of p-value of 0.4212 is greater than 5% and thus we cannot reject the null hypothesis that there is a unit root in the time series that means the time series is non- stationary.

16. Test the existence of autocorrelation between the errors and the returns series

By using Box-Ljung test according to equation (13) we get the results as shown in Table (2)

Table (2) the values of autocorrelation, partial autocorrelation, and Q-Stat test values in the OPEC oil price series

Date: 04/10/18 Time: 22:25
Sample: 1/02/2003 1/04/2018
Included observations: 3874

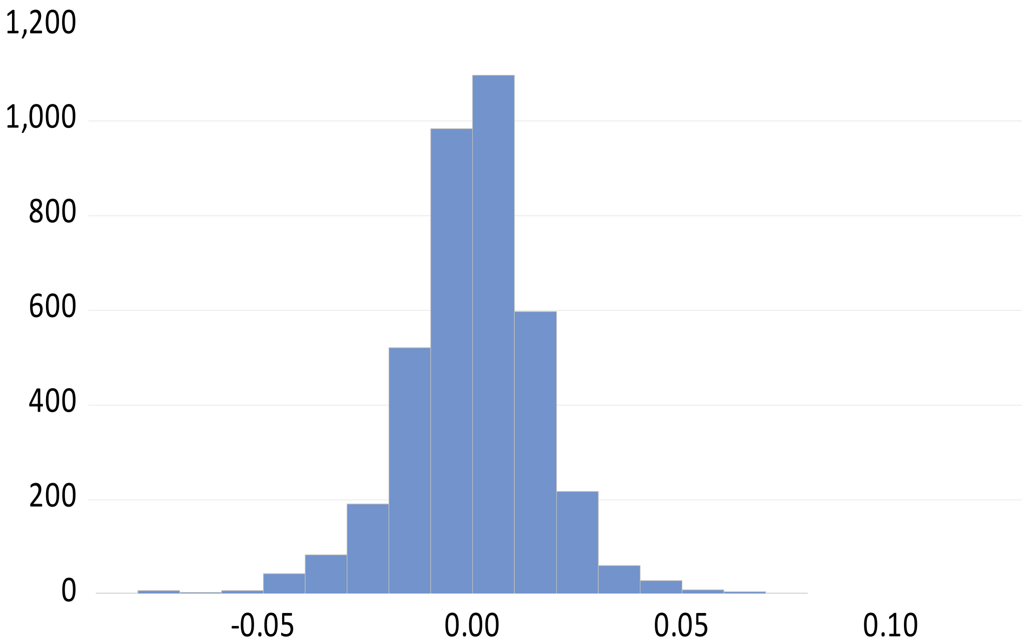
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.999	0.999	3869.9	0.000
		2	0.998	-0.185	7731.1	0.000
		3	0.997	0.027	11584.	0.000
		4	0.995	-0.018	15427.	0.000
		5	0.994	-0.025	19261.	0.000
		6	0.993	-0.009	23086.	0.000
		7	0.991	0.003	26901.	0.000
		8	0.990	-0.028	30706.	0.000
		9	0.988	0.016	34501.	0.000
		10	0.987	-0.016	38286.	0.000
		11	0.985	-0.015	42061.	0.000
		12	0.984	0.004	45825.	0.000
		13	0.982	-0.031	49579.	0.000
		14	0.981	-0.007	53322.	0.000
		15	0.979	-0.018	57054.	0.000
		16	0.978	-0.023	60774.	0.000
		17	0.976	-0.009	64483.	0.000
		18	0.974	0.009	68180.	0.000
		19	0.973	0.009	71865.	0.000
		20	0.971	-0.021	75538.	0.000
		21	0.969	-0.020	79199.	0.000
		22	0.967	0.007	82848.	0.000
		23	0.966	0.001	86485.	0.000
		24	0.964	-0.007	90109.	0.000
		25	0.962	-0.017	93721.	0.000
		26	0.960	-0.038	97320.	0.000
		27	0.958	-0.014	100906	0.000
		28	0.957	0.001	104478	0.000
		29	0.955	-0.017	108036	0.000
		30	0.953	-0.025	111580	0.000
		31	0.950	-0.001	115110	0.000
		32	0.948	-0.000	118625	0.000
		33	0.946	-0.019	122126	0.000
		34	0.944	0.004	125612	0.000
		35	0.942	-0.010	129084	0.000
		36	0.940	-0.030	132540	0.000

Table 2 shows the significance of all autocorrelations according to Q and p-values, which is less than 5%, indicating a significant correlation between the observations and thus the presence of heterogeneity of the discrepancies of the observed series

17. Series Returns:

returns were awarded a series by taking the first difference of natural logarithm data daily oil prices in the series below a summary of some measures for the series returns.

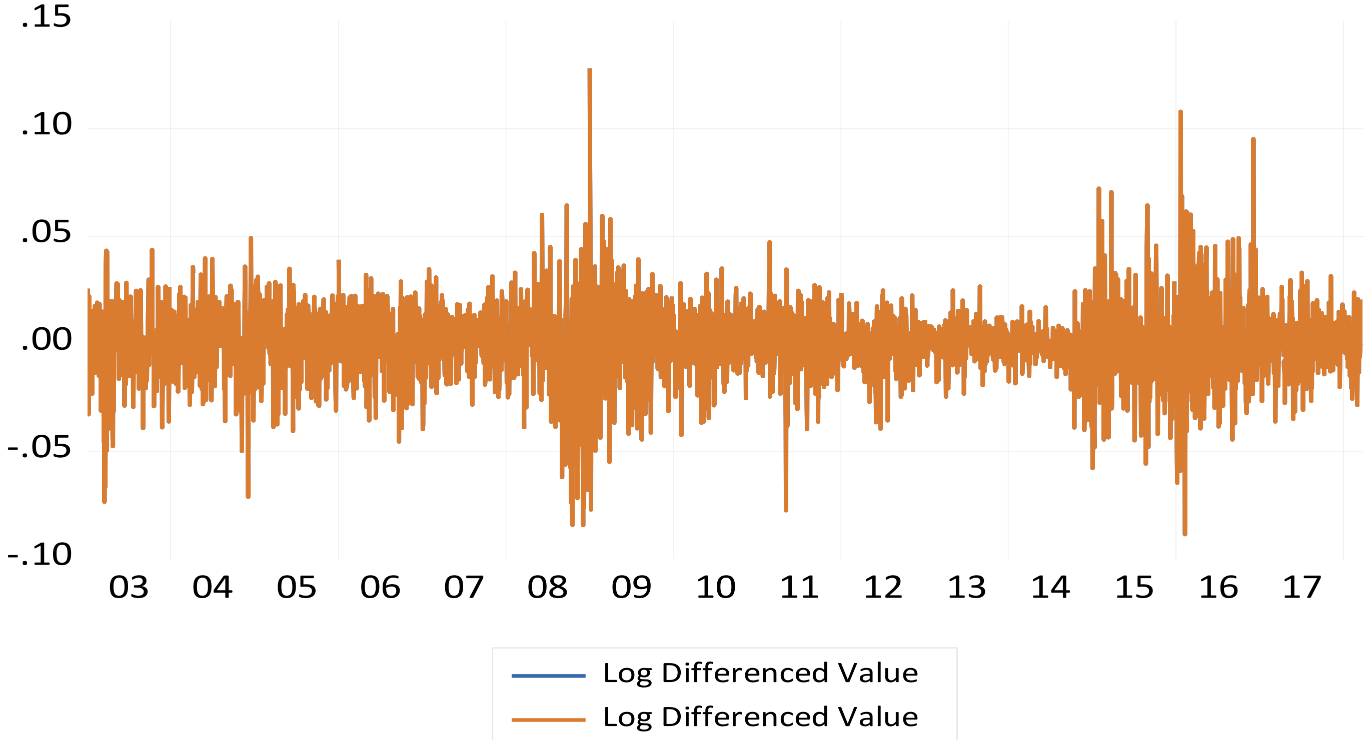
Table (3) shows some statistics about the series returns



Series: EE	
Sample 1/02/2003 1/04/2018	
Observations 3873	
Mean	0.000204
Median	0.000663
Maximum	0.128036
Minimum	-0.088521
Std. Dev.	0.016744
Skewness	-0.052740
Kurtosis	6.919979
Jarque-Bera	2481.521
Probability	0.000000

From the above indicators, the smallest value in the series of returns was -0.088521 and the largest value was (0.128036) and the average of the time series is equal to 0.000204). With a standard deviation of 0.016744, It also shows that the value of the torsion coefficient of negative (skewness = -0.05274) which shows that the distribution of revenue chain has the longest tail region of the left (skewness negative), we also note a difference splaying coefficient (kurtosis) for the value of "3" distinctive normal distribution is equal in this sample (6.919979), Which indicates that the bearers have thick limbs and characterized by flattening and this indicates dispersion and therefore different from normal distribution, This is confirmed by Jarque-Bera (which indicates that these residues do not follow the normal distribution law at a significant level of 5%). The graph of the revenue chain can be illustrated in Figure (2)

Figure (2) illustrates the series returns to the data of oil prices in OPEC bound for the period 2003-2018



From the above figure, we find that the series contains periods of volatility followed by periods of relative stagnation in fluctuations and so on as we progress in time

Table (4) Augmented Dickey Fuller test to test the stationary of returns series

Null Hypothesis: DLOG(SERIES02) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=29)

Prob.*	t-Statistic		
0.0001	-48.11514	Augmented Dickey-Fuller test statistic	
	-3.431853	1% level	Test critical values:
	-2.862090	5% level	
	-2.567106	10% level	

From the table(4) we find that the value of p-value less than 5%, which indicates the rejection of the null hypothesis, which states that the series returns are non- stationary and this avoids that the predictions are inaccurate to appear.

table (5) below shows the moral all links by calculable test Q as well as P-Value of less than 5% values, indicating the existence of a moral autocorrelation between the residuum, which confirms the existence of the heterogeneity of the errors of the series returns and shown by box-Ljung test.

Table (5) the values of autocorrelation, partial autocorrelation, and impact test values in the OPEC oil price chain

Date: 04/1
Sample: 1/
Included ob

Autocorr

Heteroskedasticity Test: ARCH

18. Homogeneity of variance test series returns

For the purpose of detecting the stationary returns a string variation was calculated multiplier test Lagrange (ARCH Test) and referred to in the theoretical side of the equation (14) and the test results as shown in Table 6.

Table (6) the ARCH test to determine the homogeneity of variances series returns oil prices in OPEC

Obs*R-squared	222.869	Prob. Chi-Square(1)	0

we see from table (6) that the value of p-value less than 5% then we reject the null hypothesis which states that the homogeneity of variances series returns.

19.Estimation:

The parameters of the studied models (GARCH, GARCH-M, EGARCH) are estimated at this stage for the purpose of determining the best model for predicting the future oil price series data using the greatest possible method. The results of model estimation are as follows:

i-Estimation of the GARCH model

By studying the functions of self-correlation and partial depending on the tests used in the diagnosis of the degree of the specimen described in the preceding paragraphs could be diagnosed with four models as shown in Table 7 was used model GARCH was estimated models described parameters and calculating the criteria for selection of the specimen is best as shown in the table (7)

Table (7) shows the GARCH models estimation, using normal distribution of errors and the criterion of choice of the best model

MODEL	μ	Ω	α1	α2	β1	β2	AIC	SIC	H-Q
GARCH (1,1)	**0.000464 (0.0231)**	**0.00000119 (0.0001)**	**0.072682 (0)**	**-**	**0.92532 (0)**	**-**	**-5.633**	**-5.626**	**-5.63**
GARCH (1,2)	**0.000503 (0.0138)**	**0.0000016 (0.0002)**	**0.111536 (0)**	**-**	**0.304601 (0.0011)**	**0.581636 (0)**	**-5.636**	**-5.628**	**-5.634**

GARCH (2,1)	**0.000544 (0.0082)**	**0.00000064 (0.0025)**	**0.16318 (0)**	**-0.109 (0)**	**0.945154 (0)**	**-**	**-5.64**	**-5.632**	**-5.637**
GARCH (2,2)	**0.000554 (0.0072)**	**0.000000453 (0.0185)**	**0.158805 (0)**	**-0.1195 (0)**	**1.172814 (0)**	**-0.21258 (0)**	**-5.64**	**-5.63**	**-5.636**

From Table (7) we find that the best model according to the selection criteria of AIC, SIC, H-QIC is GARCH (1,2). The estimated equation is:

$$y_t = \mathbf{0.000503} + \sqrt{1.6\text{E}-06 + 0.116\, x_{t-1}^2 + 0.304601\, \sigma_{t-1}^2 + \mathbf{0.581636}\, \sigma_{t-2}^2} * \varepsilon_t$$

ii-Estimation of the GARCH-M model

The GARCH-M model was applied to the four sample models described above and model parameters were estimated and the selection criteria for the best model were calculated. As shown in Table 8,

Table (8) shows the GARCH-M models estimation using normal distribution of errors and the criterion of choice of the best model

MODEL	μ	Ω	α1	α2	β1	β2	Ψ	AIC	SIC	H-Q
GARCH-M (2,1)	0.000313 (0.3049)	0.00000063 (0.0026)	0.161489 (0)	-0.107852 (0)	0.94569 (0)	-	1.380913 (0.3356)	-5.64	-5.63	-5.637
GARCH-M (2,2)	0.000305 (0.3174)	0.00000044 (0.0193)	0.157009 (0)	-0.1189 (0)	1.185295 (0)	-0.22388 (0.1129)	1.489909 (0.2992)	-5.64	-5.628	-5.636
GARCH-M (1,2)	0.000275 (0.3698)	0.00000156 (0.0002)	0.109857 (0)	-	0.306246 (0.0012)	0.581701 (0)	1.355501 (0.3433)	-5.636	-5.627	-5.633
GARCH -M(1,1)	0.000191 (0.5359)	0.00000114 (0.0001)	0.071251 (0)		0.926804 (0)		1.619678 (0.2504)	-5.632	-5.624	-5.63

From Table 8, the best model for GARCH-M models to be used is GARCH-M (1,2) according to the selection criteria of AIC, SIC, and H-QIC to exclude the first two models because their parameters are negative. The estimate is:

$$y_t = \mathbf{0.0003 + 1.3555}\, \boldsymbol{\sigma_t^2} + \sqrt{\mathbf{2E-06 + 0.1099}\, \boldsymbol{x_{t-1}^2} + \mathbf{0.306}\, \boldsymbol{\sigma_{t-1}^2} + \mathbf{0.582}\, \boldsymbol{\sigma_{t-2}^2}} * \boldsymbol{\varepsilon_t}$$

iii-Estimation of the EGARCH model

Table (9) shows that the best estimated model suitable in the EGARCH models applied according to the appropriate model selection criteria is EGARCH (2.1), which can be written as follows:

$$y_t = 0.00027 + \sqrt{e^{-0.13425+0.99414\, ln\left(\sigma_{t-j}^2\right)+0.2792\left\{\left|\frac{x_{t-i}}{\sigma_{t-i}}\right|-\sqrt{\frac{2}{\pi}}\right\}-0.17118\left\{\left|\frac{x_{t-i}}{\sigma_{t-i}}\right|-\sqrt{\frac{2}{\pi}}\right\}-0.03661\frac{x_{t-i}}{\sigma_{t-i}}}} * \varepsilon_t$$

Table (9) EGARCH models estimation using normal distribution of errors and the criterion of choice of the best model

MODEL	μ	Ω	α1	α2	β1	β2	λ	AIC	SIC	H-Q
EGARCH (2,1)	0.000266 (0.1968)	-0.134249 (0)	-0.17118 (0)	-0.036613 (0)	0.994137 (0)		0.279245 (0)	-5.647	-5.637	-5.644
EGARCH (2,2)	0.000270 (0.1908)	-0.107282 (0.0001)	-0.19937 (0)	-0.029722 (0)	1.166229 (0)	-0.17083 (0.2318)	0.286596 (0)	-5.647	-5.635	-5.643
EGARCH (1,2)	0.000275 (0.1753)	-0.257077 (0)	-0.06617 (0)		0.338944 (0)	0.649121 (0)	0.199754 (0)	-5.645	-5.635	-5.642
EGARCH (1,1)	0.000202 (0.3234)	-0.164725 (0)	-0.03916 (0)		0.991947 (0)		0.12389 (0)	-5.64	-5.632	-5.637

To select the most appropriate model for the series returns to oil prices in OPEC for the period from 2003 to 2018 by comparing the estimated models as shown in the table (10)

Table (10) shows a comparison between the best applicable GARCH family models using normal distribution of errors and the criterion of choice of the best model

MODEL	μ	Ω	α1	α2	β1	β2	Parameter	AIC	SIC	H-Q IC
EGARCH(2,1)	0.000266 (0.1968)	-0.134249 (0)	-0.17118 (0)	-0.036613 (0)	0.994137 (0)		0.279245 (0)	-5.647	-5.537	-5.644
GARCH(1,2)	0.000503 (0.0138)	0.0000016 (0.0002)	0.111536 (0)		0.304601 (0.0011)	0.581636 (0)		-5.636	-5.627	-5.634
GARCH-M(1,2)	0.000275 (0.3698)	0.00000156 (0.0002)	0.109857 (0)		0.306246 (0.0012)	0.581701 (0)	1.355501 (0.3433)	-5.636	-5.627	-5.633

We summarize from Table (10) that the EGARCH model is superior to the other models. The best model of the proposed models is EGARCH (2.1) according to the selection criteria of AIC, SIC, H-QIC

20.Check the appropriate model

After the diagnosis of the specimen and determine the degree and estimate its parameters series returns in oil prices in OPEC must ensure the efficiency of the specimen and accuracy in interpreting the time series behavior is applied through a test ARCH Test and Box-Ljung of the residuals standard boxes residuum standard.

Table (11) shows the arch-test residuals test

Heteroskedasticity Test: ARCH			
Obs*R-squared	1.897737	Prob. Chi-Square(1)	0.1683

From the above table we find that the value of p-value is greater than 5%. We can not reject the null hypothesis that the errors are homogeneous

Table (12) shows the Box-Ljung test to detect random errors

LAG	Q-Stat	Prob*
5	228.09	0
10	231.05	0
15	234.32	0
20	237.45	0
25	254.35	0
30	263.17	0
35	268.89	0

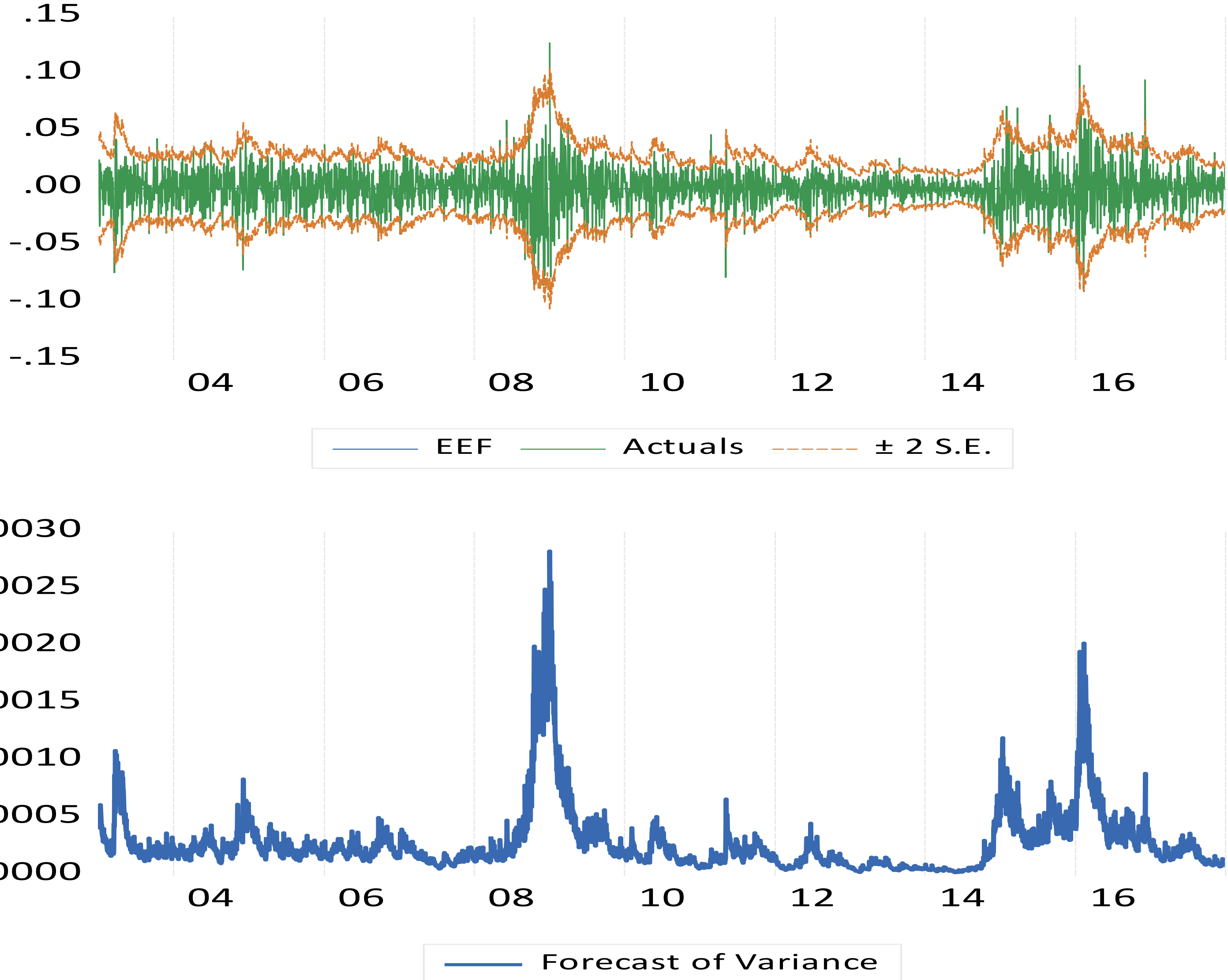
Note moral standard errors in the Offsets (5,10,15,20,25,30,35) any standard errors that are not normally distributed

Table (13) shows the Box-Ljung test to detect the standard error square error

Lag	Q-Stat	Prob*
5	8.3373	0.139
10	9.4744	0.488
15	13.657	0.552
20	17.726	0.605
25	21.351	0.673
30	25.097	0.72
35	28.897	0.757

Note that standard error boxes are not significant in displacements (5,10,15,20,25,30,35), that is mean, the standard errors are distributed naturally

Figure (3) shows the prediction of the values of the yield chain and the confidence limits of those estimates and the estimated variance



Figure(3) shows confidence limits for predictive values, true values of the yield chain, and estimated variance.

Table (14) shows a comparison between GARCH-M and EGARCH according to the criterion of choice of the best model

MODEL	RMSE	MAE
EGARCH (2,1)	0.016742	0.012057
GARCH (1,2)	0.016744	0.012054

From the table above, the EGARCH model is also superior to the GARCH model according to the MAE (RMS) criteria, which in turn indicates that the EGARCH model (2.1) is highly accurate and therefore the best model for oil price forecasting by OPEC. The GARCH (1,2) model was chosen for comparison because it is very close to the EGARCH model (2.1) according to the model selection criteria (AIC, SIC, H-QIC).

21. Conclusions and recommendations

Conclusions

1. The oil price series has been non- stationary in the middle and contrast.
2. The series of returns turned the series stationary in the center according to Dicky Fuller test.
3. The series returns is stationary according to the arch test and contains a sequential link (correlation significance).
4. The best model is the EGARCH (2.1) model, which is superior to the rest of the studied models according to AIC, SIC, H-Q
5. The selected model is superior to RMSE, MAE, compared with GARCH (1,2).
6. The models of autoregression conditional on heterogeneity of variance are more efficient in predicting fluctuations.

Recommendations

1. Use other comparison models such as ARMA-GARCH, IGARCH, NGARCH

2. Use other methods to estimate model parameters such as Quasi-maximum likelihood estimate (QMLE).

3. Using the GARCH family models to predict other financial time series to estimate and study the behavior of these strings because they have the ability to explain the behavior of these strings characterized by heterogeneity of variance.

The References:

1. Banumathy, K. and R. Azhagaiah, Modelling stock market volatility: Evidence from India. Managing Global Transitions, 2015. 13(1): p. 27.

2. Engle, R.F., D.M. Lilien, and R.P. Robins, Estimating time varying risk premia in the term structure: The ARCH-M model. Econometrica: Journal of the Econometric Society, 1987: p. 391-407.

3. Falk, M., et al., A first course on time series analysis: examples with SAS. 2006.

4. Francq, C. and J.-M. Zakoian, GARCH models: structure, statistical inference and financial applications. 2011: John Wiley & Sons.

5. Lütkepohl, H. and M. Krätzig, Applied time series econometrics. 2004: Cambridge university press.

6. عبد الله، سهيل نجم، (2008) ، "تحليل نماذج السلاسل الزمنية اللاخطية من نوع (ARCH,GARCH)للرتب الدنيا باستعمال المحاكاة". أطروحة دكتوراه، جامعة بغداد.

7. عنبر، جنان عبد الله، (2016) ،" مقارنة بعض المقدرات البيزية الحصينة مع مقدرات اخرى لأنموذج GARCH(1,1) مع تطبيق عملي"، أطروحة دكتوراه، جامعة بغداد

8. عوض الله، خلود جمال، (2016)" استخدام نماذج Holt-Winters و SARIMA في التنبؤ بالسلاسل الزمنية الموسمية"، رسالة ماجستير، جامعة الازهر-غزة

9. كاظم، بريدة برهان، (2016)،"التنبؤ باستعمال نماذج الانحدار الذاتي العامة المشروطة بعدم تجانس التباين (GARCH)) الموسمية مع تطبيق عملي"، رسالة ماجستير، جامعة بغداد

10. يادكار، احمد شامار (2014) ، "استخدام نماذج الانحدار الذاتي المشروطة بوجود عدم التجانس التباين في التنبؤ بسعر الإغلاق اليومي لمؤشر سوق العراق للأوراق المالية" رسالة ماجستير، قسم الاحصاء كلية الادارة والاقتصاد جامعة بغداد.