Forecasting The Volatility Of OPEC Oil prices using Fractionally Integrated GARCH models

Prof. Dr. Mohammed H. Al-Sharoot

Department of Statistics

College of Administration and Economics
University Of AL-Qadisiyah-IRAQ

mohammed.AIsharoot@qu.edu.iq

Hanan A. Al-rashide

Department of Statistics

College of Administration and Economics
University Of AL- Qadisiyah-IRAQ

hananaallii7@gmail.com

Abstract

Some time series are characterized by their great volatility over time, especially time series related to the movement of the economy, and those related to the change in stock prices or the movement of financial transactions and stock markets, which are characterized by being non- stationary over time due to the change in the behavior of observations, making them suffer from the problem of Heteroscedasticity . The paper aims the use of predictive models that a time series can adapt to with large fluctuations and with long memory over time, a number of important models used to deal with FIGARCH time series when the error distribution follows the t-distribution were studied and reviewed, which were used For the first time by Researcher Engel in 1982 and developed by other researchers, the characteristics of these models were reviewed and applied for the purpose of forecasting daily oil prices according to the prices approved by OPEC for the period from 2003 to 2022, where the practical analysis of oil price data showed that the best prediction model is the ARMA model. 2,2)-FIGARCH(1,d,2) in which the error follows the t-distribution, and the best predictor performance is out of sample .

1. Introduction

long memory is one of the important topics that have received great attention by researchers in recent years, and this is what prompted statisticians to pay attention to it by studying the dynamic behavior of financial variables and employing statistical measures for the purpose of studying and analyzing the behavior of these phenomena. Perhaps the most important of these measures that Their occurrence coincides with long memory. They are the time series that are defined as a group of observations linked to each other of a phenomenon that is observed successively during a specific time period and whose emergence is from its relationship to time t, and that the aim of analyzing it is to describe the features of the phenomenon that generate this series and build a standard model to explain its behavior and adopting this model to predict values and future periods based on values for previous periods

The analysis of time series witnessed a remarkable development after the two scientists (Box-Jenkins) presented a modern methodology at the beginning of the second half of the twentieth century, which proved a high efficiency and became a real entry for the modern analysis of time series, as it includes statistical theories, methods, and graphic and computational means. Experts and researchers have taken these The methodology is a main reference for them to identify the appropriate model, estimate its parameters, diagnose and use it in predicting future observations represented by the ARMA model, which has become widely used in the analysis and modeling of linear time series. However, it cannot explain the fluctuations and changes in some phenomena, which are characterized by a large number of time fluctuations, and therefore the linear model becomes inappropriate with the real data set of the phenomenon. The errors are independent of each other, but we note in many time series that they do not fulfill the previous conditions as the variance is not fixed as in the financial time series and therefore leads to the inefficiency of the model in the forecasting process. This imposed a new challenge on scientists in finding an alternative solution, as non-linear models were proposed that take into account these hypotheses and the failure to achieve these conditions. These models and their various developments are considered one of the important ways to describe time changes, especially uncertainty or great uncertainty, which includes a large amount of uncertainty, These models take into account the treatment of the problem of fluctuations in the time series, as well as to improve the matching of the model to the data and thus give an explanation for the fluctuations that occur in the phenomena of different time series. After a series of developments, researcher Robert F. Engle (in 1982) presented a new category The models are called conditional autoregressive models with Heteroscedasticity (ARCH). If the researcher wanted through it to address the problems that the previous ARMA models suffer from, especially in the financial time series that are characterized by the speed of volatility (Volatility) associated with time, this model was generalized by Bollerslov 1986 , generalized autoregressive model conditional not to Variance smoothing (GARCH) to address the problem of ARCH model constraints on parameters, as well as solve the problem of higher-order model requirements to describe the variance series, despite the importance of models ARCH(p) and GARCH(p,q) family models in modeling non-linear time series of financial variables) . However, it has been criticized by some economists, especially in cases characterized by fluctuations in opposite directions and wide influences. These models cannot take into account these fluctuations, especially fluctuations in long memory, which led to the emergence of the FIGARCH model, which we are going to study Baillie (Bollerslev, Mikkelsen 1996, it can be used in the case where there is a slow decrease in the long-term autocorrelations of the time series). These models are called generalized partially integrated and conditional inhomogeneity of variance models, which is symbolized by the acronym FIGARCH, as these models have proven their importance in modeling time series with long memory, which is characterized by many time series and has been widely used due to its ability to characterize data in many fields. as economics and financial sciences.

2. Research problem

The problem of the study lies in the presence of fluctuations in the OPEC oil price series, which led to the no stationary of oil prices and therefore the use of ARMA models will lead to unreasonable future predictions, so plans based on these results are useless, so FIGARCH long memory models have been proposed to predict at these prices.

3. Research objective

The goal is to build the best daily oil price series forecasting model for OPEC for the period 2003 to 2022 by applying the FIGARCH long memory model.

4. Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity Models [9][15] [2]

This model is one of the extensions of the ARCH family models, which was first defined by Engle in 1982 as a linear function of the squares of random errors in the past tense, is defined follows:

$$z_t = \mu + \varepsilon_t \qquad Mean\ equation \ldots\ldots\ldots\ldots(1)$$

$$\varepsilon_t = \sigma_t\ e_t \qquad e_t \sim iid\ N(0,1)$$

$$\sigma_t^2 = \alpha_\circ + \sum_{i=1}^{p} \alpha_i\ \varepsilon_{t-i}^2 \qquad Volatility\ equation \ldots\ldots\ldots(2)$$

It is taken from this model that it requires higher ranks to describe the variance series, and that the expansion in the values of P, may produce negative values for α and this contradicts the assumptions of the model that states that the parameters are positive ($\alpha_\circ > 0$). To confront this problem, Bollerslov (1986) suggested Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, to address the problem of ARCH model's higher order, by modeling the variance of time series observations through p from the boxes of past errors and q from the values of the conditional variance in the previous period. The GARCH model can be formulated of degree (p, q).) Since (P≥1), (q ≥1) and written in the form:

$$\sigma_t^2 = \alpha_\circ + \sum_{i=1}^{p} \alpha_i\ \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j\ \sigma_{t-j}^2 \qquad \ldots\ldots\ldots\ldots (3)$$ [7]

Paraphrasing (3) of the GARCH model:

$$\sigma_t^2 = \alpha_\circ + \alpha(L)\ \varepsilon_t^2 + \beta(L)\sigma_t^2 \ldots\ldots\ldots\ldots\ldots(4)$$

Since is $v_t = \varepsilon_t^2 - \sigma_t^2$ a random variable that represents the difference between the squares of errors ε_t^2 and the unconditional variance σ_t^2 and by substituting for it, we get the following formula:

$$\varepsilon_t^2 = \alpha_o + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j\ \varepsilon_{t-j}^2 - \sum_{j=1}^{q} \beta_j\ v_{t-j} - v_t \ldots\ldots(5)$$

By reformulating equation (5) in terms of polynomials α(L), β(L), it becomes as follows:

$$[1 - \alpha(L) - \beta(L)]\varepsilon_t^2 = \alpha_\circ + [1 - \beta(L)]v_t \ldots\ldots(6)$$

Thus, the partially integrated GARCH or FIGARCH model can be obtained as d is a fractional value, 0<d<1 . Thus, the FIGARCH model can be expressed as follows:

$$[1 - \alpha(L) - \beta(L)](1 - L)^d \varepsilon_t^2 = \alpha_\circ + [1 - \beta(L)]v_t \ldots\ldots(7)$$ [17]

$$\sigma_t^2 = \alpha_\circ[1 - \beta(1)]^{-1} + \lambda(L)\varepsilon_t^2 \ldots\ldots(8)$$

$$\lambda(L) = [1 - [1 - \beta(L)]^{-1}\emptyset(L)(1 - L)^d]$$

$= \lambda_1 L + \lambda_2 L^2 + \cdots and \lambda_k \geq 0 \; for \; k = 1,2, \dots \dots \dots \dots . (9)$

5. **ARMA – FIGARCH**[15] [6]

It is clear that the mathematical representation of these models is given by two equations, one of them is for the conditional average , which represents the prediction vehicle, and the equation of the conditional variance , which is the non-predictive vehicle FIGARCH Thus, the implementation of the integrated model for univariate series analysis and for predicting time series fluctuations becomes true as follows :

$$\dots\dots . (10) z_t = \sum_{i=1}^{n} \emptyset_i \; y_{t-i} + \varepsilon_t - \sum_{j=1}^{m} \theta_j \; \varepsilon_{t-j} \quad , \varepsilon_t \sim N(0, {\sigma_t}^2)$$

$$\varepsilon_t = \sigma_t \; e_t \qquad e_t \sim iid(0,1)$$

$$\sigma_t^2 = \alpha_\circ [1-\beta(1)]^{-1} + [1-[1-\beta(L)]^{-1} \emptyset(L)(1-L)^d] \varepsilon_t^2 \dots\dots . (11)$$ [17]

6. **The Augmented Dickey-Fuller test (ADF)**[1][5][14] is used to detect the presence of a unit root in the univariate test, i.e. to test whether the time series is strong stationary or not. The ADF test is defined follows:

$$\Delta z_t = \alpha + \beta_t + \Upsilon z_{t-1} + \sum_{j=1}^{k} \delta_j \Delta \, z_{t-j} + \varepsilon_t \qquad \dots \dots (\mathbf{12})$$

the hypothesis is

$H_\circ : \Upsilon = 0$ The time series is non- stationary on mean.

$H_1 : \Upsilon \neq 0$ The time series is stationary on mean.

7. **Ljung - Box Test**[4][16][8][12]

The test was proposed by (Ljung & Box) in 1978 is used to test whether the errors of the model fitted a time series are random :

$H_\circ : \rho_1 = \rho_2 = \cdots = \rho_k \dots = \rho_m = 0 \quad ; \qquad k = 1,2, \dots . , m$

$H_1 : \rho_k \neq 0$ for some values of k .

Using the following statistics:

$$Q_M = \left(n(n+2) \sum_{k=1}^{m} \frac{\hat{\rho}_k^2}{n-k} \right) \sim \chi^2_{m-P} \dots \dots . . (13)$$

8. **ARCH Test - Lagrange Multiplier** [1][5]

proposed by Engle in 1982 to test whether the errors follow ARCH process is based on estimating the equation under study

$$\varepsilon_t^2 = \alpha_\circ + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + . \dots \dots \dots \dots . + \alpha_p \varepsilon_{t-p}^2$$

the test statistic as follows:

$$LM = ARCHtest = n \, \hat{R}^2 \sim \chi^2_{(r)} \dots \dots . (14)$$

9. **Detection of Long Memory**[2][13]

Experimental analysis and detection of the long memory property of a time series is difficult because the strong autocorrelation of long memory operations makes the statistical fluctuations very severe, so long memory tests require a large number of observations. By drawing the autocorrelation functions and the partial autocorrelation function, so there are many graphs and many statistical tests through which it can be checked whether the time series is a series with a long memory or not. We will discuss these forms as follows:

- Check the graph of the autocorrelation function ACF Plot

This study focuses on the case study that deals with the nature of time series that are characterized by the feature of long-term memory, in which the formula of the ACF function takes the following form:

$$\sum_{\boldsymbol{k}=-\infty}^{\infty} \boldsymbol{\rho}\,(\boldsymbol{k}) = \infty$$

- Using statistical tests to verify the long memory feature:

Several estimations have been proposed for Horst's decomposition for analyzing long memory in time series, the most important of which are:

1- Average R/S analysis .
2- GPH method.

In this research, we will rely on the first method, which is the statistic of the modified R/S analysis.

Rescaled - Range Analysis (R/S):

This method was presented for the first time by the researcher (1951 Hurst) to reveal the existence of the phenomenon of long memory in the time series data through the difference between the maximum and minimum subtotals of the deviations of the series values from their arithmetic mean divided by their standard deviation and symbolized by the statistic $Q_n$and calculated as follows:

$$Q_n = R/S_y = \frac{1}{S_y}\left\{\max_{1\le k\le n}\sum_{j=1}^{k}(Y_J - \bar{Y}_n) - \min_{1\le k\le n}\sum_{j=1}^{k}(Y_J - \bar{Y}_n)\right\} \ldots .(15)$$

10. Estimation[15][6]

The estimation of the parameters of the FIGARCH model is generally made using the greatest possibility method with the assumption of the standard normal distribution of e_t but sometimes the assumption of the standard normal distribution is not achieved for many applications, and therefore it is preferable to use the method of estimation with the Quasi-Maximum Likelihood Estimator, which is An iterative method characterized by not assuming the standard normality of the distribution.

The process of estimating the parameters of FIGARCH (p, d, q) using the QMLE method is considered the most common estimation method, and its main idea is to maximize the probability function depending on the sample $\{e_1, e_2, \ldots., e_n\}$ and it can be written as follows:

It is written in the case of assuming a t-distribution, where the maximum probability function is as follows:

$$log\,L(\theta) = \sum_{t=1}^{n}\left\{-\frac{1}{2}\ln\left(\frac{\pi(r-2)\Gamma(r/2)}{\Gamma(r+1/2)}\right) - \frac{1}{2}\ln({\sigma_t}^2) - \left(\frac{r+1}{2}\right)\ln\left(1+\frac{{e_t}^2}{{\sigma_t}^2(r-2)}\right)\right\} \ldots\ldots\ldots (16\,)$$ [17]

11. Model selection criteria[18]

There are several criterions to choose the best model among the proposed fitted models for the studied data, these criteria was developed to select the most common model as follows :

I - Akaikes Information Criterion (AIC) [12][14]

This criterion was introduced by Akaike in 1974 an information standard known as (AIC) used to evaluate the suitability of time series models, we choose the model that gives the least AIC . The AIC formula can be written as:

$$AIC = n\ln(\hat{\sigma}_e^2) + 2L \quad \dots\dots\dots (17)$$

n: its size sample.

k: represents the number of parameters of the model.

$\ln \hat{\sigma}_e^2$: represents the logarithm of maximum likelihood funtion.

II - Schwarz Information Criterion (SIC) [3][14]

In 1978, Schwarz and Akaike proposed another criterion for determining the degree of the model known as the Schwartz Information Criterion, (SIC) and defined as follows:

$$SIC = n\, ln(\hat{\sigma}_e^2) + L \ln(n) \quad \dots\dots\dots (18)$$

III - Hannan- Quinn Criterion (HQC)[18][2]

In 1979, this criterion was proposed by (Quinn) and (Hannan) HQC to determine the rank of the model and its formula:

$$HQC = \ln \hat{\sigma}_e^2 + 2L\, C\, ln\left(\frac{ln(n)}{n}\right) \quad , \quad C > 2 \quad \dots\dots\dots (19)$$

12. **Forecasting[15]**

Forecasting is one of the most important goals of model building in time series, as it

$$\sigma_{t+1}^2 = \alpha_\circ[1-\beta(1)]^{-1} + \lambda(L)\varepsilon_{t+1}^2$$

$$\sigma_t^2(s) = \alpha_\circ[1-\beta(1)]^{-1} + \lambda_1\sigma_t^2(s-1) + \dots + \lambda_{s-1}\sigma_t^2(1) + \lambda_s\varepsilon_t^2 + \lambda_{s+1}\varepsilon_{t-1}^2 \,..\,(\ 20\ )$$

$$\sigma_t^2(s) \approx \alpha_\circ[1-\beta(1)]^{-1} + \sum_{i=1}^{s-1}\lambda_i\sigma_t^2(s-i) + \sum_{j=0}^{M}\lambda_{s+j}\varepsilon_{t-j}^2 \ \dots\dots.(21)$$ [17]

13. **Forecasting accuracy measures**[10]

The following measures are used to measure forecast accuracy , they are very important to know the chosen model.

i- Root Mean Square Error (RMSE) [11][10]

the RMSE formula is given as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\sigma_t^2 - \widehat{\sigma^2}_t)^2} \quad \dots\dots\dots (22)$$

ii- Mean Absolute Error (MAE) [11][10]

This measure is defined as the absolute difference between the actual variance and forecast variability measures, and it has the following formula:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|\sigma_t^2 - \widehat{\sigma^2}_t\right| \quad \dots\dots\dots\dots\dots (23)$$

14. **Applied side**

This aspect will include an application and analysis of the methods presented in the theoretical side on the data of the oil price series in Iraq in US dollars for the purpose of predicting the daily global oil price series using the FIGARCH model assuming the error distribution follows the t-distribution, to choose the appropriate model to predict future fluctuations ,Where the time series of the final prices of a daily barrel of oil was obtained from the available data announced on the OPEC website published on the Internet for the period from (2/01/2003) to (31/03/2022) for the purpose of modeling it through time series models. Provided by OPEC great credit for facilitating the task of data analysis. The data were analyzed using three software packages, R program , Ox Matrices and Eviews12. Where this series of daily oil prices was drawn as shown in Figure (1):

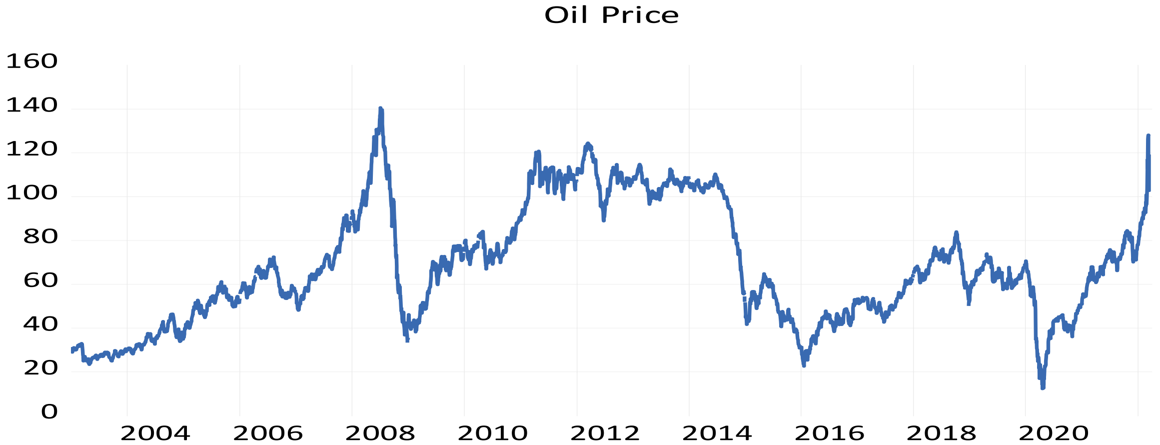


Figure (1) shows the time series of daily oil prices

It is clear from Figure (1) that the time series of oil prices is non -stationary in mean and variance has high volatility, which indicates the presence of fluctuations in the variance .

I. Returns Series

The return series y_t was calculated with the following formula:

$$y_t = \ln\frac{Z_t}{Z_{t-1}} = (\ln Z_t - \ln Z_{t-1})$$

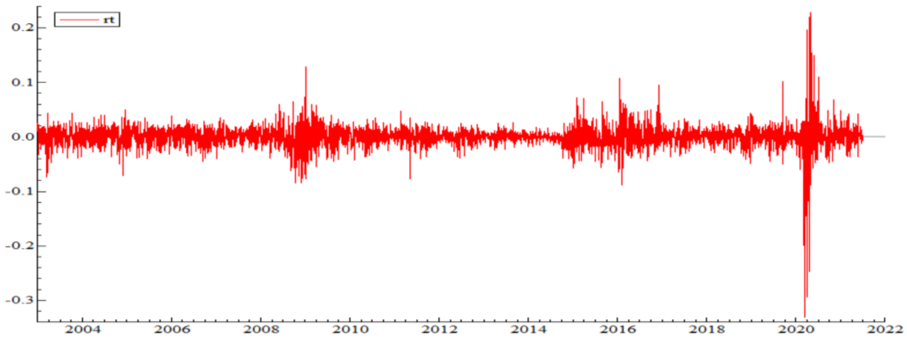


Figure (2) shows the series of returns to daily oil prices

From figure (2) we find that the series stationary in mean and contains periods of volatility, followed by periods of relative stagnation in fluctuations, and so on over time.

II. Stationary test (Unit Root test)

The unit root was tested for the series of oil prices and the series of returns for oil prices. The test results can be summarized as shown in Table (1) which displays the results of unit root tests using ADF tests with t values and probabilistic values at the 0.05 level to test the following hypothesis:

$H_\circ : \Upsilon = 0$

$H_1 : \Upsilon < 0$

Table (1) shows the Dickey-Fuller Extended Return Series Test

Stationary test (unit root) of the daily oil price series		
Probability	**t**	
0.3574	-1.8482	(Intercept)
0.6720	-1.8659	(Trend and intercept)
0.6945	0.0367	(none Trend and intercept)
Stationary test (unit root) of the return series		
Probability	**t**	
0.0000	-15.64527	(Intercept)
0.0000	-15.64439	(Trend and intercept)
0.0000	-15.63565	(none Trend and intercept)

We note from the table (1) that the probabilistic value of the developed Dickey-Fuller test for the three models is (0.3574, 0.6720, 0.6945), respectively, which is greater than (0.05), and therefore we accept the null hypothesis, which states that the time series is non-stationary on average. It is clear from the table to test the stationary (unit root) of the series of returns that the probability value of the Augmented Dickey Fuller test for the three models is less than (0.05), so we reject the null hypothesis and accept the alternative hypothesis, which indicates the stationary of the series of returns on average daily oil prices.

III. Test for the existence of autocorrelation of the original series and the series of returns

To find out that the series has autocorrelation, we will calculate its (Q) statistic and the autocorrelation and partial autocorrelation functions, as shown in the following figure (3):

Date: 07/04/22 Time: 02:00
Sample: 1/02/2003 3/31/2022
Included observations: 4964

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.999	0.999	4953.7	0.000
		2	0.997	-0.156	9890.8	0.000
		3	0.995	0.007	14811.	0.000
		4	0.993	-0.022	19714.	0.000
		5	0.991	-0.029	24599.	0.000
		6	0.989	-0.016	29464.	0.000
		7	0.987	0.053	34312.	0.000
		8	0.985	-0.030	39142.	0.000
		9	0.984	0.018	43955.	0.000
		10	0.982	0.005	48749.	0.000
		11	0.980	-0.014	53527.	0.000
		12	0.978	-0.006	58286.	0.000
		13	0.976	-0.030	63028.	0.000
		14	0.974	-0.049	67749.	0.000
		15	0.971	-0.022	72449.	0.000
		16	0.969	-0.037	77127.	0.000
		17	0.967	-0.043	81782.	0.000
		18	0.964	0.007	86413.	0.000
		19	0.961	0.028	91022.	0.000
		20	0.959	0.036	95608.	0.000
		21	0.957	-0.040	100172	0.000
		22	0.954	0.012	104714	0.000
		23	0.952	0.018	109234	0.000
		24	0.950	0.000	113733	0.000
		25	0.947	-0.010	118211	0.000
		26	0.945	-0.045	122666	0.000
		27	0.942	0.012	127098	0.000
		28	0.940	-0.004	131508	0.000
		29	0.937	0.015	135895	0.000
		30	0.935	-0.010	140259	0.000
		31	0.932	-0.002	144601	0.000
		32	0.929	-0.017	148919	0.000
		33	0.927	-0.002	153214	0.000
		34	0.924	-0.010	157485	0.000
		35	0.922	-0.001	161733	0.000
		36	0.919	-0.033	165957	0.000

Date: 09/02/22 Time: 17:58
Sample: 1/02/2003 3/31/2022
Included observations: 4964

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.133	0.133	87.304	0.000
		2	0.023	0.005	89.877	0.000
		3	0.027	0.024	93.634	0.000
		4	0.052	0.046	106.92	0.000
		5	0.063	0.051	126.71	0.000
		6	-0.096	-0.115	172.38	0.000
		7	0.070	0.098	196.71	0.000
		8	-0.004	-0.030	196.78	0.000
		9	0.033	0.037	202.25	0.000
		10	0.035	0.029	208.42	0.000
		11	-0.029	-0.033	212.75	0.000
		12	0.060	0.049	230.88	0.000
		13	0.009	0.010	231.26	0.000
		14	0.092	0.076	273.21	0.000
		15	-0.001	-0.020	273.21	0.000
		16	0.017	0.022	274.71	0.000
		17	0.061	0.036	293.52	0.000
		18	0.002	-0.002	293.55	0.000
		19	-0.016	-0.038	294.82	0.000
		20	-0.034	-0.009	300.67	0.000
		21	0.067	0.054	323.24	0.000
		22	0.005	-0.016	323.34	0.000
		23	-0.001	0.010	323.35	0.000
		24	0.037	0.026	330.28	0.000
		25	0.023	0.013	332.84	0.000
		26	0.068	0.044	355.69	0.000
		27	-0.040	-0.045	363.63	0.000
		28	0.022	0.016	365.96	0.000
		29	0.020	0.011	367.92	0.000
		30	-0.013	-0.024	368.82	0.000
		31	0.036	0.029	375.43	0.000
		32	-0.052	-0.044	388.90	0.000
		33	-0.005	-0.016	389.03	0.000
		34	-0.018	-0.005	390.71	0.000
		35	0.026	0.021	394.20	0.000
		36	-0.011	-0.023	394.85	0.000

figure (3) showing the autocorrelation, partial autocorrelation and Box-Ljung test functions of the original series of daily oil price

It is clear from the figure (3) of the autocorrelation function and the partial autocorrelation of the daily oil price series that the coefficients of the autocorrelation function in all time displacements differ significantly from zero (outside the confidence limits ((-1.96)/√4964), (+1.96)/√4964) And we note that it tends to decrease slowly towards zero as the displacement period (Lag) increases, as the behavior of the autocorrelation function suggests to the existence of long memory in the oil price series, and it is also noted that all the values of Probability of the Q statistic are less than (0.05), which leads to the rejection of the hypothesis The null which states (there is no sequential autocorrelation between random errors) and the acceptance of the alternative hypothesis which states that there is sequential autocorrelation between them and that the residuals are not distributed randomly and have a distinct shape and behavior. It is also clear from the test results for the autocorrelation function and partial autocorrelation of the series of returns that the values of the coefficients of the autocorrelation function fluctuate between positive rise and negative decline and that all coefficient values fall within confidence limits ((-1.96)/√4964), (+1.96)/ √4964) in all time shifts, and this indicates the stability of the series of returns for daily oil prices, as well as it is clear that all possibilities for the values of the statistic Q are less than (0.05), and this means rejecting the null hypothesis that states (there is no sequential autocorrelation between random errors) and accepting the hypothesis alternative that states that there is a sequential autocorrelation between them.

IV. The variance homogeneity test for the series returns

For the purpose of detecting the stability of the variance of the returns series, the Lagrange multiplier test (ARCH Test) is calculated to test the following hypotheses:

H_o : The variance is homogeneous for the return series of oil prices (no there effect of ARCH).

H_1 : Heteroscedasticity of return series for oil prices (an effect of ARCH) .

Table (2) shows the results of the ARCH LM-test .

Heteroskedasticity Test: ARCH

F-statistic	186.4279	Prob. F(1,4961)	0.0000
Obs*R-squared	179.7483	Prob. Chi-Square(1)	0.0000

From Table (2), it can be seen that the p-values in the table are less than 0.05 so we reject the null hypothesis of the returns series for oil prices which means that the data contains an ARCH effect.

V. Long Memory Tests

For the purpose of detecting the presence of long memory in the data, the R/S Statistic test was relied upon, as a program was written to calculate the two tests using the R statistical program because there is no ready program for it to test the following hypotheses:
H_o: having a long memory (decreasing non-exponential autocorrelations)
H_1: short memory (autocorrelation decreases exponentially)

Table(3) showing long memory tests

R/S Test	R/S Statistic	Bandwidth q	P-value
Returns	1.5829	25	0.2908239

From the table (3) we note that the results of the R/S test for the return series, which represents the modified R/S statistic, that the p-value is greater than 0.05, which means that the null hypothesis is accepted and the alternative hypothesis is rejected, which states that there is no short memory and therefore there is dependence between the values. The data has a long memory, therefore, based on the two tests, it is found that the chain of returns for oil prices shows the characteristics of a long memory in its volatility, and therefore the most suitable method for modeling the volatility of oil prices is by using GARCH class models that allow the long memory feature in the volatility process, which is symbolized by FIGARCH models.

VI. Estimating FIGARCH models for daily oil prices

After checking and confirming the collection of volatility through the returns chain and checking for stability using the ADF test, detecting the presence of the ARCH effect using the ARCH-LM , Ljung-Box tests, testing the presence of long memory in the oil price series and determining the FIGARCH model as an appropriate model for modeling the return chain for oil prices At this point, the QMLE method will be used to estimate the parameters of the FIGARCH model under
1- Assuming that the error distribution is a T-distribution.
The results of estimating the FIGARCH models can be presented for the values of P=1,2 and q=1,2 as shown in Tables (4).

Table (4) results of estimating the FIGARCH models (p,d,q) by the Q-MLE method for the return chain for oil prices in the case of the error distribution Student t.

نموذج FIGARCH(1,d,2) لسلسلة العودة لأسعار النفط عندما يكون توزيع الخطأ (Student-t)				
Parameter	**Estimate**	**Std. error**	**t-statistic**	**Pr(>|t|)**
Mu	0.0003418	2.67E-06	4.123	0.0000

Omega α_0	0.229677	0.093972	2.444	0.0146
d-FIGARCH	0.845060	0.11102	7.612	0.0000
ARCH(Phi1)	0.528487	0.10731	4.925	0.0000
GARCH(Beta1)	1.219124	0.13224	9.219	0.0000
GARCH(Beta2)	-0.292917	0.10629	-2.756	0.0059
Student(DF)	8.222704	0.94814	8.672	0.0000

From Tables (4), the results of estimating the FIGARCH model using the quasi-maximum possibility method (QMLE) assuming that the normal error distribution, and the t-distribution, indicate that the parameters of ARCH and GARCH are statistically significant in most cases, and that the parameters of the partial difference d, are positive and with Statistical significance at the level of 0.05 in all cases, which means that the shock of volatility will continue for a longer period.

Table (5) shows the comparison of the proposed models according to the different distribution of the error

Student t Distribution			
FIGARCH(1,d,1)	-7.213023	-7.205154	-7.210264
FIGARCH(1,d,2)	-7.213769	-7.204589	-7.210551
FIGARCH(2,d,1)	-7.213329	-7.204149	-7.210111
FIGARCH(2,d,2)	-7.213379	-7.202888	-7.209701

From the results in Table (5), we can conclude that the best model is FIGARCH(1,d,2) in the case of the Student- t Distribution of the return series for oil prices.

VII. Building ARMA(p,q)

Appropriate linear ARMA (p,q) models can be built using the daily return series of oil prices because they are stable at the 0.05 level according to the Box-Jenkins method. p and q are the most suitable, to select the best-fit linear ARMA models (q,p), using different ranks of the oil price return chain, and to select the optimal model among the candidate models, taking into account the autocorrelation and ARCH effect.

Table (6) shows a comparison of ARMA's proposed models for the return series

Model	AIC*	BIC	HQ
ARMA(2,2)	-6.617271	-6.609402	-6.614512

ARMA (1,2)	-6.615821	-6.609264	-6.613522
ARMA (2,1)	-6.614955	-6.608398	-6.612656
ARMA (1,0)	-6.612741	-6.608807	-6.611362
ARMA (1,1)	-6.612386	-6.607140	-6.610546
ARMA (2,0)	-6.612366	-6.607120	-6.610527
ARMA (0,1)	-6.612347	-6.608412	-6.610967
ARMA (0,2)	-6.612249	-6.607003	-6.610410

We note from Table (6) that the best ARMA model for the chain of return to oil prices is the ARMA(2,2) model because it has the lowest differentiation criteria (AIC, SIC, HQ) .

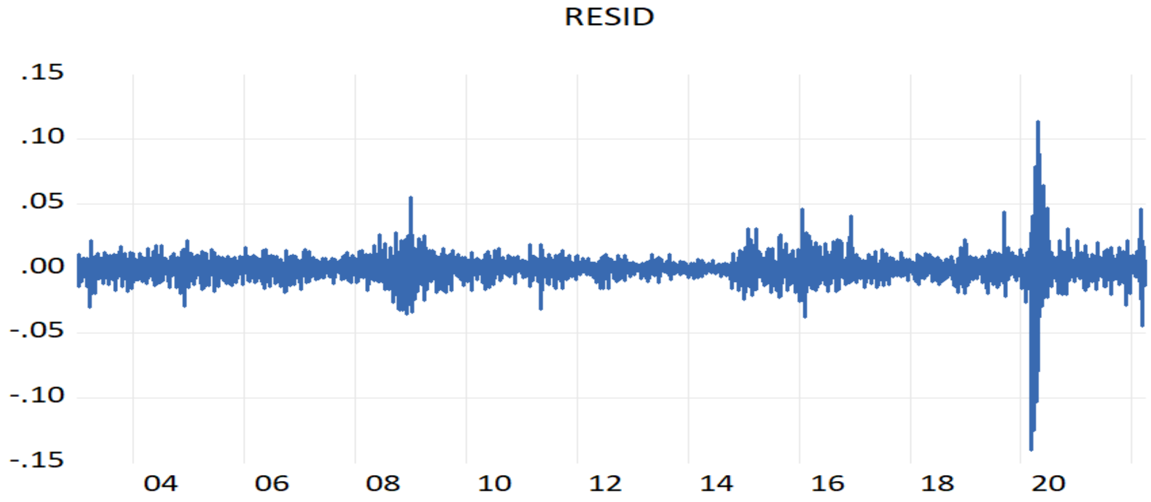


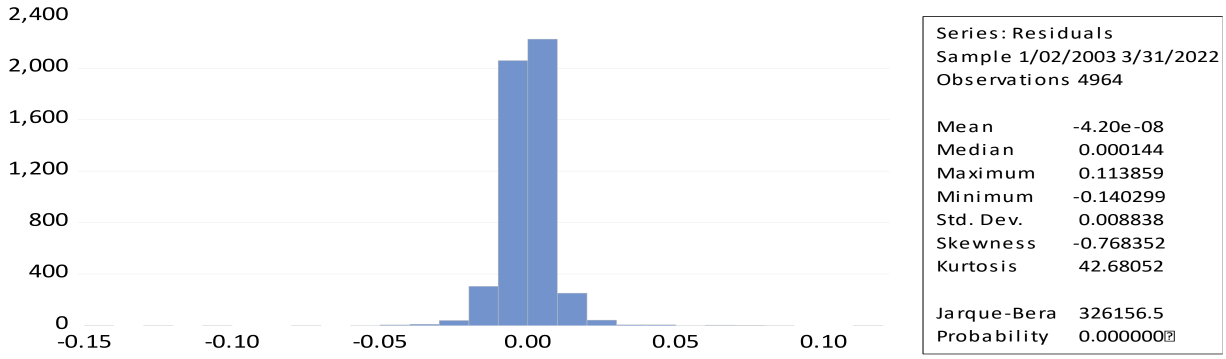
Figure (4) shows the residuals of the ARMA(2,2) model.

According to Figure (4), we see that there are periods of high volatility (large volatility) followed by periods of high volatility and periods of low volatility (small volatility) followed by periods of low volatility and … etc. for the return chain of oil prices and the rest appears to be stable and volatile. The values of the residuals indicate the heterogeneous conditional error and that the behaviors of these residuals can be represented by GARCH models, because GARCH models are used to estimate fluctuations.

VIII. Diagnosis Residual of ARMA(p,q) model.

Before starting the diagnosis stage, it can be clarified in Figure (5) the descriptive statistics of the ARMA(2,2) model, as follows:

Figure (5) shows the descriptive statistics for the series of residuals of the ARMA(p,q) model.



From Figure (5), it is noted that the cutoff value is equal to (kurtosis =42.68052), which indicates that the chain has thick ends and is characterized by flatness, and this indicates the high dispersion and its distance from the normal distribution, and this was confirmed by the Jarque-Bera test, as the p-value reached value corresponding to the test (0.0000), which indicates that the residual series called for a little deviation from the normal state at the level of significance (0.05), and therefore the residual series does not follow the normal distribution, and it is also possible to investigate the residual series for the estimated model and that it has the property of the best model through The diagnostic stage and it includes several tests, including testing the Heteroscedasticity problem (ARCH effect) and the serial correlation exists or not.

IX. ARCH-LM Test

The Lagrange multiplier test Arch-LM was used to verify the presence of an Arch effect in the residual series and that the test results were in the following table:

Table (6) Arch-LM test for residuals of the ARMA model.

Arch-LM test		
Lag	Obs*R-squared	p-value
5	658.3061	0.0000
10	774.7318	0.0000
15	915.7704	0.0000
20	1267.697	0.0000
25	1325.425	0.0000

Through the results of the table (6), it can be seen that all the p-values in the table are less than 0.05 at the slowdown period (lags = 5,10,15,20,25), so we reject the null hypothesis which states that there is no ARCH effect and thus In sum, the residual series is characterized by the presence of an effect of Heteroscedasticity of variance.

X. Ljung-Box test

With regard to detecting the randomness of the series of residuals, the Ljung-Box test was conducted by calculating the coefficients of the autocorrelation function for residuals and squares of residuals, as in the two tables below:

Table (7) shows the parameters of the Ljung-Box test.

Ljung-Box test			Ljung-Box test		
Lag	Q-Statistic	p-value	Lag	Q-Statistic	p-value
5	23.122	0.0000	5	1216.8	0.0000
10	105.07	0.0000	10	1967.7	0.0000
15	177.22	0.0000	15	2877.5	0.0000
20	202.62	0.0000	20	4166.8	0.0000
25	225.57	0.0000	25	4485.9	0.0000

From the table (7), it can be seen that the p-values are all less than 0.05, which means that the residuals of the model have a serial correlation at the slowdown period (lags = 5,10,15,20,25) and then the presence of Heteroscedasticity, which It explains that high changes in the series are followed by high changes, and at the same time low changes are followed by low changes, as well as the difficulty of determining them.

XI. ARMA-FIGARCH MODELS

Student t Distribution			
ARMA(2,2)-FIGARCH(1,d,1)	-7.265191	-7.252076	-7.260593
ARMA(2,2)-FIGARCH(1,d,2)	-7.266118	-7.251692	-7.26106
ARMA(2,2)-FIGARCH(2,d,1)	-7.26558	-7.251155	-7.260523

The study also focuses on determining the best fit non-linear ARMA-FIGARCH models for the return chain of oil prices, using the method of estimating the quasi-maximum possibility (QMLE) to estimate the equations of the conditional mean and variance of these models. Therefore, FIGARCH models are used to model fluctuations in the samples of daily returns data sets for the oil price chain, Under different error distributions (normal distribution, t distribution). We propose mixed models between linear ARMA models and nonlinear FIGARCH models in order to diagnose the degree of influence in the model. These models are taken into account and the best ones are selected from those that have the lowest value for the comparison criteria, and we will present them in the following tables:

Table (8) shows the comparison of the proposed models with the error distribution

ARMA(2,2)-FIGARCH(2,d,2)	-7.265841	-7.250104	-7.260323

From Table (8), we notice that the best model is the hybrid ARMA(2,2)-FIGARCH(1,d,2) because it has the lowest value for the comparison criteria (AIC, SIC, HQ).

XII. Choosing the right model

The best model will be selected from among the best ARMA linear models, FIGARCH nonlinear models and ARMA-FIGARCH hybrid models, as in the following table:

Table (9) shows a comparison of linear, nonlinear, and mixed models

Type Models	Models	AIC	SIC	HQ
Linear	ARMA(2,2)	-6.617271	-6.609402	-6.614512
Non-Linear	FIGARCH(1,d,2)	-7.213769	-7.204589	-7.210551
Mixed	**ARMA(2,2)-FIGARCH(1,d,2)**	-7.266118	-7.251692	-7.26106

From the results of Table (9), we note that the best model representing the return series for oil prices is the mixed model between linear and nonlinear ARMA(2,2)-FIGARCH(1,d,2) because it has the lowest values for the comparison criteria, so this model will be estimated And conducting diagnostic tests for the purpose of predicting fluctuations in daily oil prices.

XIII. **Estimating the appropriate model**

After determining the best model to represent this series, it is ARMA(2,2)-FIGARCH(1,d,2) in which the error distribution follows the t-distribution. This model will be estimated in two stages. The first stage is the use of ordinary least squares to estimate the linear part. ARMA(2,2) and then using the (QMLE) method to estimate the nonlinear part FIGARCH(1,d,2) using Ox Matrics programming as shown in the following table:

Table (10) shows the model estimate ARMA(2,2)-FIGARCH(1,d,2) in which the error follows the t-distribution

Parameter	Estimate	Std. error	t-statistic	Prob
Cst(M)	0.0003129	.6777e-005	3.228	0.0013
AR(1)	0.101845	0.39633	0.2570	0.7972
AR(2)	0.039385	0.098704	0.3990	0.6899
MA(1)	0.143232	0.39657	0.3612	0.7180
MA(2)	-0.059656	0.18099	-0.3296	0.7417
Cst(V)	0.177380	0.072026	2.463	0.0138
d-Figarch	0.843332	0.096328	8.755	0.0000
ARCH(Phi1)	0.653933	0.080690	8.104	0.0000

GARCH(Beta1)	1.392703	0.11830	11.77	0.0000
GARCH(Beta2)	-0.448875	0.097362	-4.610	0.0000
Student(DF)	7.983814	0.93097	8.576	0.0000

The result from Table (10) showed the support of QMLE estimates for the parameters in the sample for the ARMA(2,2)-FIGARCH(1,d,2) model and according to the error distribution that follows the t-distribution of the return series for oil prices. The model is statistically significant in other words, the conditional mean coefficients and the coefficients of variance are highly significant at the 0.05 level because (p values < 0.05) except for some parameters which means that the volatility is continuous, especially for the model, which is common in financial time series.

XIV. Model Diagnostic Tests

After determining the appropriate model, determining the ranks of the models, and estimating the daily oil price returns, the composition and efficiency of the model must be confirmed.

Table (11) of the Ljung-Box test and the Arch-LM test for the residuals of the ARMA(2,2)-FIGARCH(1,d,2) model.

Ljung-Box test			**Arch-LM test**		
Lag	Q-Statistic	p-value	Lag	Obs*R-squared	p-value
5	8.64965	0.0132359*	5	1.7121	0.1282
10	11.7466	0.1092051	10	1.1658	0.3089
15	13.54	0.3310399	15	0.89768	0.5665
20	29.123	0. 334086*	20	1.4696	0.0810
25	35.5359	0.0340281*	25	1.4228	0.0791

According to Table (11) and through the results of the ARCH-LM test to verify the effect of ARCH in the residuals, we conclude that the p-values > 0.05, which means accepting the null hypothesis that states “there is no effect of ARCH”, which means that there is no effect of ARCH in the residuals. Residuals of the model Based on the results of the Ljung-Box test at the significance level of 0.05 for the squared residuals, the probabilities are more than 0.05 (not significant), except for delays (5), (20) and (25), which are less than 0.05, which means that we cannot Then reject the null hypothesis, which means that there is no serial correlation in the remainders of the model

XV. Forecasting future fluctuations

After determining the appropriate model through the stage of diagnosis, estimation and verification of the accuracy of the model, where the model ARMA(2,2)-FIGARCH(1,d,2) was used to predict fluctuations, and the results were as shown in Figure (5)

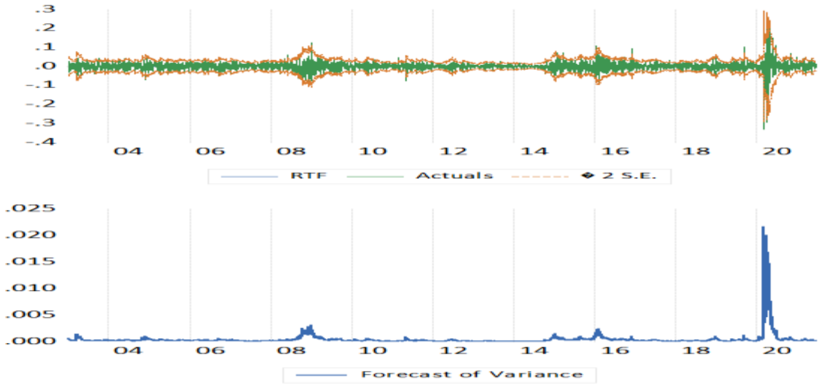


Figure (5) The graph of the series shows the returns, the predicted values, and the prediction of volatility (variance)

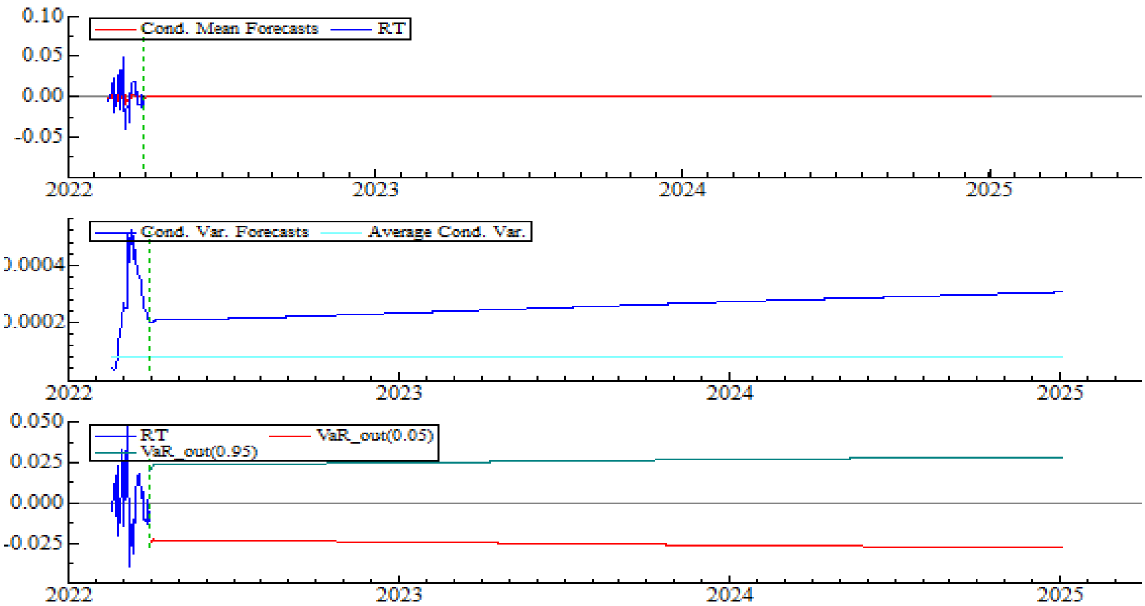


Figure (6) shows the out-of-sample forecasts for the return chain of oil prices

XVI. Prediction accuracy test criteria

The table below (12) shows the criteria by which the extent of the prediction error is measured, which were discussed in the second chapter, and are studied for the purpose of ensuring the predictive performance of the model. It has an Out of sample and the prediction error is evaluated and calculated as follows:

Table (12) shows the comparison between the forecast within and outside the sample for the best model of the oil revenue series

	In – sample predict	Out – of – sample
RMSE	0.008944	0.04587
MAE	0.005554	0.06460
MAPE	0.003498	0.02245

According to Table (12), we evaluated the predictive ability of the best model in the sample and outside the sample of the series of fluctuations in oil price returns for OPEC, as the results indicate that the relative differences between the forecast performance measures for both samples are small. Out-of-sample is more appropriate than predicting in-sample performance.

Conclusions and Recommendations

Conclusions

1. The series of oil prices is non-stationary on mean and variance.
2. The series of return for oil prices does not follow the normal distribution.
3. The series of returns to oil prices contains periods of fluctuation, followed by periods of relative stagnation over time.
4. Oil prices series of returns is stationary on mean .
5. He explained by drawing the autocorrelation function that its behavior is decreasing slowly, which suggests that the oil price series has a long memory, and this was proven by the statistical tests of the long memory R/S with the presence of dependence between the values and that the oil price series shows the characteristics of long memory in its volatility. Which called for the use of FIGARCH models .
6. The advantage of the ARMA(2,2)-FIGARCH(1,d,2) models for forecasting the future volatility of OPEC oil prices through standards AIC, SIC, H-Q and precision scales RMSE, MAE .
7. The conditional Autoregressive Heteroscedasticity models are more efficient in predicting the volatility.

Recommendations

1. Use other comparison models such as GJR-GARCH, IGARCH and NGARCH.
2. Use other methods to estimate model parameters such as QMLE.
3. Use of GARCH family models to predict other financial time series to estimate and study the behavior of these series because they have the ability to explain the behavior of these series that is characterized by Heteroscedasticity of variance.

[1] Abdullah, Suhail Najim, (2008), "Analysis of lower-order nonlinear time series models (ARCH & GARCH) using simulation." PhD thesis, University of Baghdad.

[2] Ahmed Atef Adwan, (2017), “Using ARFIMA Models in Predicting World Wheat Prices” Master’s Thesis, Al-Azhar University, Faculty of Administration and Economics.

[3] Al-Janabi Habib Kazem, Al-Sharout Muhammad Habib, (2021), "Analysis of Oil Prices in Iraq Using Parametric and Non-parametric Time Series Models", Master's Thesis, University of Al-Qadisiyah .

[4] Al-Sharout, Muhammad Habib, Omar Al-Ramadani, (2019), “Forecasting using the ARCH Generalized Model in Time Series with Application”, MA thesis, Al-Qadisiyah University.

[5] Anbar, Jinan Abdullah, (2016), "Comparing some of the Bayesian estimations with other estimations of the GARCH model (1,1) with practical application", PhD thesis, University of Baghdad.

[6] Bollerslev, T., & Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews*, *11*(2), 143-172.

[7] Bollerslev, T., (1986) " Generalized Autoregressive Conditional Heteroskedasticity " Journal of Econometrics, Vol.31, pp . 307-327.

[8] Box, G. E. P. and Pierce, D., (1970), " Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models " Journal of the American Statistical Association, Vol. (65), PP (1509-1526).

[9] Engel, R. F.(1982) " Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom," Econometric, Vol 50, No (4), July.

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

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

[12] Kazem, Buraida Burhan, (2016), "Predicting the Use of Conditional General Autoregressive Models of Seasonal Heteroscedasticity (GARCH) with Practical Application", Master's Thesis, University of Baghdad.

[13] Khelifi, Mariah, Aichouch, Riad (2017), "For forecasting oil prices in the Algerian economy for the period 1990-2016 ARFIMA using long-term memory models.

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

[15] Omar Abdel Mohsen Ali, Ahmed Yadkar, (2021), "Using QMLE and SVR methods to predict the series of fuel prices in Iraq with a FIGARCH model", PhD thesis, University of Baghdad.

[16] Sabiha Al-Saadoun, Jassem Nasser Hussein, (2020), "Diagnosing efficient time-series models with practical application", Master's thesis, Karbala University.

[17] Tayefi, M., & Ramanathan, T. V. (2012). An overview of FIGARCH and related time series models. *Austrian journal of statistics*, *41*(3), 175-196.

[18] Yadkar, Ahmed Shamar (2014), "Using Conditional Autoregressive Models of Heteroscedasticity in Predicting the Daily Closing Price of the Iraq Stock Exchange Index," Master's Thesis, Department of Statistics, College of Administration and Economics, University of Baghdad.