Muhammad Iraqi Journal of Science, 2021, Vol. 62, No. 7, pp: 2307-2326
DOI: 10.24996/ijs.2021.62.7.19

ISSN: 0067-2904

Mixing ARMA Models with EGARCH Models for Modeling and Analyzing the Time Series of Temperature

Abduljabbar Ali Mudhir Muhammad
Kirkuk Education Directorate

Received: 23/6/2020 Accepted: 18/10/2020

Abstract

In this article, our goal is mixing ARMA models with EGARCH models and composing a mixed model ARMA(R,M)-EGARCH(Q,P) with two steps. The first step includes modeling the data series by using EGARCH model alone, interspersed with steps of detecting the heteroscedasticity effect, estimating the model's parameters, and testing the adequacy of the model. Also, we predict the conditional variance and verify its convergence to the unconditional variance value. The second step includes mixing ARMA with EGARCH and using the mixed (composite) model in the modeling of time series data, predicting future values, and then assessing the prediction ability of the proposed model by using the prediction error criterion.

Keywords: Time Series, ARMA, GARCH, EGARCH, Mixed ARMA-EGARCH.

مزج نماذج ARMA بنماذج EGARCH واستخدامها في نمذجة وتحليل المتسلسلة الزمنية لدرجات الحرارة

عبدالجبار علي مظهر محمد
مدرس مديرية تربية كركوك ، كركوك، العراق

الخلاصة

في هذا المقال سيكون هدفنا مزج نماذج ARMA بنماذج EGARCH وتكوين انموذج مختلط ARMA(R,M)−EGARCH(Q,P) بمرحلتين، الاولى تتضمن نمذجة متسلسلة البيانات باستخدام انموذج EGARCH فقط تتخللها خطوات الكشف عن تأثير عدم التجانس ثم تقدير معلمات الانموذج وفحص ملائمة الانموذج لمتسلسلة البيانات وكذلك التنبؤ بقيم التباين المشروط وتحقق اقتراب هذه القيم من قيمة التباين غير المشروط. المرحلة الثانية تتضمن مزج نماذج ARMA بِـ EGARCH واستخدام الانموذج المركب (المختلط) في نمذجة بيانات المتسلسلة الزمنية والتكهن بقيم مستقبلية لها ثم يليها تقييم القدرة التنبؤية للأنموذج المقترح باستخدام معايير ومقاييس خطأ التنبؤ.

Introduction

As a general description of this research, we will show how to mix the equation model of ARMA model with the variance equation of the conditional variance model and compose a mixed model that controls the linear and nonlinear behaviors of the series that contains volatility in its data.

The theories that were standing for most of the studies related to time series in the fifties, sixties, and seventies of the last century adopted the homoscedasticity hypothesis. However, the remarkable changes and volatilization in natural and unnatural phenomena brought this hypothesis to criticism. A

*Email: gabbarfalcon@st.tu.edu.iq

new hypothesis, called the heteroscedasticity hypothesis, was created and many mathematical models were suggested for linear and nonlinear time series, such as ARCH models and its generalized (GARCH) models for Engle 1982[1] and Bollerslev 1986[2], respectively, TAR models for Tong 1983, and STAR models for Tsay 1989, along with many other models. Nevertheless, the autoregressive conditional heteroscedasticity (ARCH) model has a great importance globally because it has been used in modeling and analyzing the behavior of big changes in modern phenomena, such as the accompanying risks of returns for the financial time series, as in those of gold and oil prices, volume trading, and inflation. However, there are a lot of efforts in modeling this behavior, most of which being successful, but the ARCH model was the most favorable due the simplicity of its formulas and interpretation of the analysis of time series that has fluctuations in its data.

The ARCH models were suggested by Engle in 1982. In 1986, Bollerslev generalized these models in his doctorate thesis and submitted the generalized (ARCH) models. Then, in 1991, Nelson proposed the exponential model for the generalized (ARCH), which is called the EGARCH model. Later, many models were introduced as an expansion of ARCH and GARCH models. All these models perform a specific work that is to make the variance steady and unchanging with time, except that these models cannot be used to predict the future of the series, only if they are mixed with ARMA models, forming what is known as Mixed ARMA(R,M)-GARCH(Q,P) models.

Our study will mix ARMA model with EGARCH model, which is one of conditional variance models that process the volatility in data of some time series. This volatility makes the variance conditional and changing with time, while it is supposed to be steady. As known for most researchers, the modeling of time series is an expressing operation of the series under study, as a function depending on its past values. This function is often expressed as a difference equation or differential equation plus a term of white noise. The process of mixing EGARCH with ARMA consists of two equations; the first is a mean equation which will be as an ARMA model that allows controlling the linear behavior of the series. The second represents a variance equation which will be as an EGARCH model that allows controlling the nonlinear behavior in the heteroscedasticity of the variance. The real goal of mixing these models is to enhance the prediction process for time series containing volatility in its data. This volatility makes the variance conditional and changing with time, which causes high prediction errors and unacceptable results. Also, since the purpose of creating the conditional variance models is to control changing of the variance and making it steady, the mixed model will confer good forecasted results and acceptable prediction errors.

This study came as an expansion of a previous study submitted conducted by Al-Bazzoni (2013)[3] in her master's thesis, in which she analyzed the time series of temperatures for Mosul city, north Iraq, using the GARCH(Q,P)-AR(P) model, a mixed model of GARCH and autoregressive (AR) models that provided close and acceptable predictions. Likewise, this paper can be considered as a comparative study with a master's thesis submitted by Al Obaidy in 2015[4], in which he proposed the STSech-AR(P) model for analyzing and predicting the time series of temperatures for Kirkuk city, north Iraq. However, there are many local and global researches that studied the ARCH and GARCH models and their extensions, many of which introduced predictions of conditional variance, but they did not process the predictions of the future values of the series.

We gave some definitions to clarify the equations of the models. We mentioned the methodology used by Box-Jenkins in modeling the time series and descripted the used data and the method of mixing the two equations. Also, our research consists of using EGARCH model with fitting, checking the adequacy, and forecasting the conditional variance. We then apply the mixed model to predict the future of the temperature series, providing important figures and tables, and finish the paper with several conclusions.

Some definitions and basic concepts

Definition: (Ramzan, Ramzan and Zahid, 2012)(Senaviratna and Cooray, 2017)(Francq and Zokoian, 2010)(Box, Jenkins and Reinsel, 1994)[5-8]

Autoregressive and moving average (ARMA) is a mixed model of autoregressive AR(R) and moving average MA(M) models, denoted by ARMA(R,M), such that R and M are positive numbers, with the existence of real coefficients, namely $(a_1, a_2, \dots, a_R)$ and $(b_1, b_2, \dots, b_M)$, and a constant C, such that:

$$y_t = c + a_1 y_{t-1} + \cdots + a_R y_{t-R} + x_t + b_1 x_{t-1} + \cdots + b_M x_{t-M} \quad \dots \quad 1.1$$

which can be written as:

$$y_t - \sum_{i=1}^{R} a_a y_{t-i} = c + x_t + \sum_{j=1}^{M} b_j x_{t-j}$$

and by using lag operator written as:

$$a(L)y_t = c + b(L)x_t$$

such that:

$$a(L) = 1 - a_1 L - \cdots - a_R L^R$$
$$b(L) = 1 + b_1 L + \cdots + b_M L^M$$

There are certain requirements for the matching between the stationary condition for AR(R) model and invertible condition for MA(M) model. That is, for the stationarity condition of ARMA model, all the roots of polynomials $a(L)$ and the polynomial $b(L)$ should lie outside the unit circle.

Definition: (Malmsten, 2004)(Moffat, Akpan and Abasiekwere, 2017)(Alexander, 2008)(Teräsvirta, 2006)(TSAY, 2010)(Nelson, 1991)(Mohammad and Mudhir, 2020)**[9-15]**

The exponential generalized autoregressive conditional heteroscedasticity model, denoted by EGARCH(Q,P), was introduced by Nelson in 1991 with the following formula:

$$y_t = \mu_t + x_t \qquad x_t \sim iid\ D(0, \sigma^2) \qquad \ldots\ 1.2$$
$$x_t = \sigma_t z_t \qquad z_t \sim iid\ D(0,1) \qquad \ldots\ 1.3$$
$$\log \sigma_t^2 = \omega + \sum_{i=1}^{Q} \alpha_i \log \sigma_{t-i}^2 + \sum_{j=1}^{P} \delta_j\ g(z_{t-j}) \qquad \ldots\ 1.4$$

such that:

$$g(z_{t-j}) = \gamma_j(z_{t-j}) + \beta_j(|z_{t-j}| - \mathrm{E}|z_{t-j}|)$$

where $(i = 1, \ldots, Q)$, $(j = 1, \ldots, P)$, ω is a constant and $\alpha_i, \beta_j, \gamma_j$ are parameters of GARCH, ARCH, and leverage, respectively. z_t is a sequence of independent variables with zero mean and variance equal to one. The stationary condition of EGARCH model is expressed so that the summation of all roots of the polynomial $(1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_Q L^Q)$ lies outside the unit circle, which is expressed as $\sum_{i=1}^{Q} \alpha_i < 1$.

Definition: (Francq and Zokoian, 2010)(Jelonek, 2017)[7,16]

The mixed ARMA(R,M)-EGARCH(Q,P) model is defined by the formula:

$$y_t = \mu + \sum_{i=1}^{R} a_i y_{t-i} + x_t + \sum_{j=1}^{M} b_j x_{t-j} \qquad \ldots\ 1.5 \qquad Mean\ Eq.$$

such that $a_i, i = 1, \ldots, R$; b_j , $j = 1, \ldots, M$ are real numbers, μ is a constant, and x_t is the white noise term with $x_t \sim iid\ N(0, \sigma_t^2)$.

$$x_t = \sigma_t z_t \qquad z_t \sim iid\ N(0,1)$$
$$\log \sigma_t^2 = \omega + \sum_{i=1}^{Q} \log \alpha_i\, \sigma_{t-i}^2 + \sum_{j=1}^{P} \gamma_j(z_{t-j})$$
$$+ \sum_{j=1}^{P} \beta_j[|z_{t-j}| - \mathrm{E}|z_{t-j}|] \qquad \ldots 1.6 \qquad Variance\ Eq.$$

with ω , $\alpha_i, \gamma_j, \beta_j$, $i = 1, \ldots, Q$, $j = 1, \ldots, P$ are real numbers.

Box-Jenkins Methodology in Time Series Modeling: (Box, Jenkins and Reinsel, 1994)(Box, Jenkins and Reinsel, 2015)(Makridakis, Harvard and Hyndman, 1998)**[8,17,18]**

The Box-Jenkins methodology is a process of five steps. The first step is ensuring the stability of the data series, such that if the series is not stable about its mean and variance, which is supposed to be steady, then we must use a transformation, such as returns, difference, or logarithm transformations. The second step is diagnosing the model, i.e. whether we use the autoregressive (AR) model, the moving average (MA) model, a mixed (ARMA) model, or may be a conditional variance model if the Ljung-Box Q-test shows the heteroscedasticity effect in variance. The third step is estimating the parameters of the used model. The forth step is testing the adequacy of the used model, which is achieved by inferring the residuals and ensuring that they are uncorrelated, homoscedastic, and normally distributed. The fifth step is forecasting the future and testing the predictive ability by using the prediction error criterion, which is achieved by dividing the data under study into two groups; the first group is called (in sample) and used in all previous steps, while the second group is called (out sample) and left to compare it with the predicted values.

Data Description

The data used in this study represent monthly mean temperature records for Kirkuk city from January 1980 until November 2016 with 444 observations. We divided the data set into two groups; the first group represents data until April 2016, which is used in creating the time series and fitting the model. The second group represents eight months data that are left for the comparison with the predicted values and testing the prediction performance of the model.

Materials

The method used in the present study is not different from the Box-Jenkins methodology because the structure of GARCH models is similar to that of the ARMA models. The modeling operation consists of two parts. The first part is to use EGARCH model in creating the time series and ensuring the convergence of conditional variance to the unconditional variance value. This is performed by testing the heteroscedasticity effect, fitting the model, estimating its parameters, testing its adequacy, and forecasting the conditional variance to observe its approach. The second part consist of applying the mixed ARMA-EGARCH model and choosing its best rank by using AIC and BIC criteria, then predicting eight future values. We used MATLAB R2016a program and attached the used data in appendix (1). Also we attached a programing text as (m.file) of modeling the first part in appendix (2) and the second part in appendix (3). Finally, appendix (4) represents a programing text used to examine the stationarity conditions of the models.

Using EGARCH Model and Forecasting the Conditional Variance

The data attached in appendix (1) were entered as a vertical vector of (Nx1) to create the time series, as shown in Figure-1.

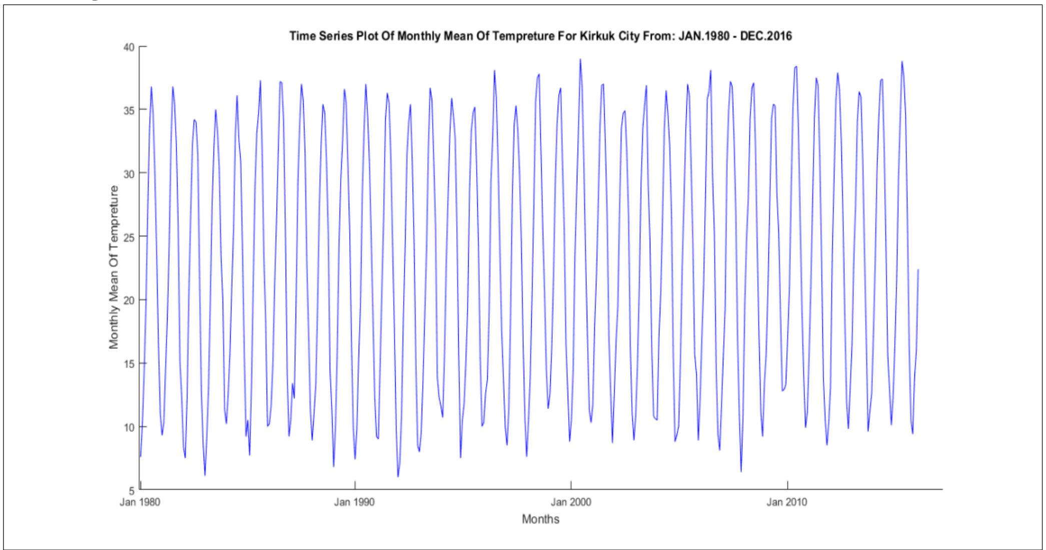

Figure 1 -Time series plot of monthly mean of temperature.

The original series was transformed to the returns series by using the formula:

$$r_t = \ln d_t - \ln d_{t-1} \;=\; \ln\left(\frac{d_t}{d_{t-1}}\right)$$

where d_t is temperature of the month t, r_t is return series.

However, the series became stable about its mean after being transformed to the returns series, but it still had volatility in its data. This caused heteroscedasticity effects, as appears in Figure-2. To control this behavior, we used the conditional variance models.

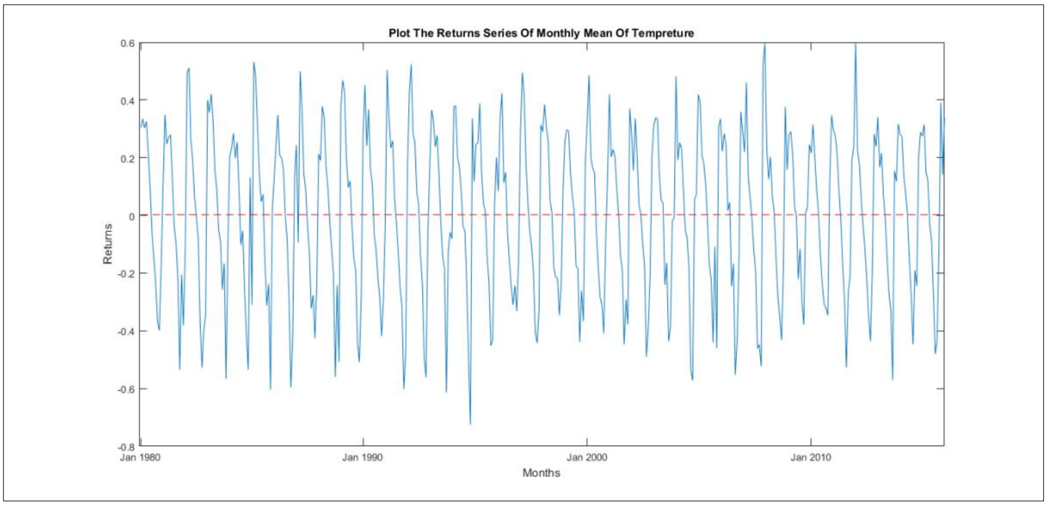

Figure 2- Plot of the return series.

We then found the squared errors series and plotted the autocorrelation function (ACF) and the partial autocorrelation function (PACF) to check the heteroscedasticity effect, as shown in Figure-4.

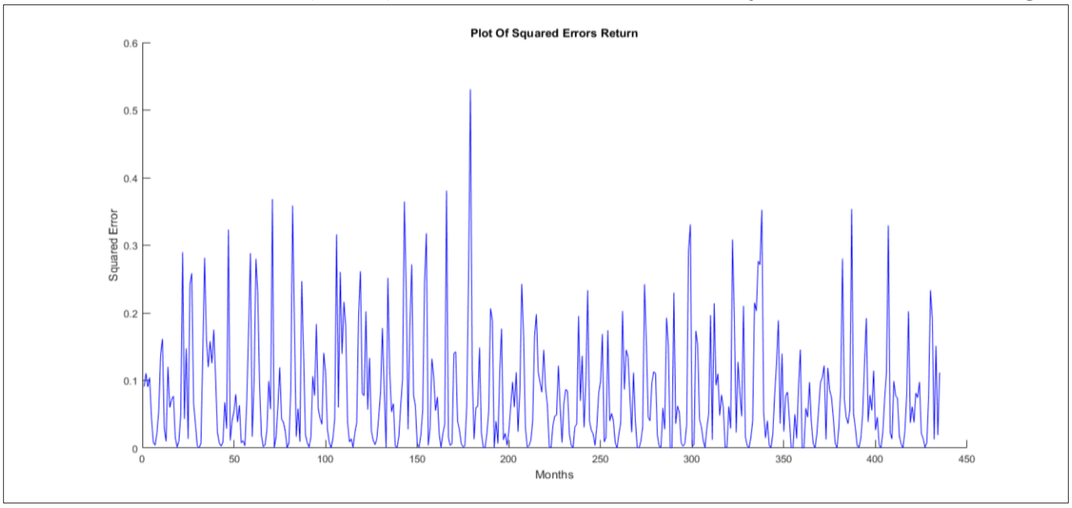

Figure 3- Plot of squared errors return.

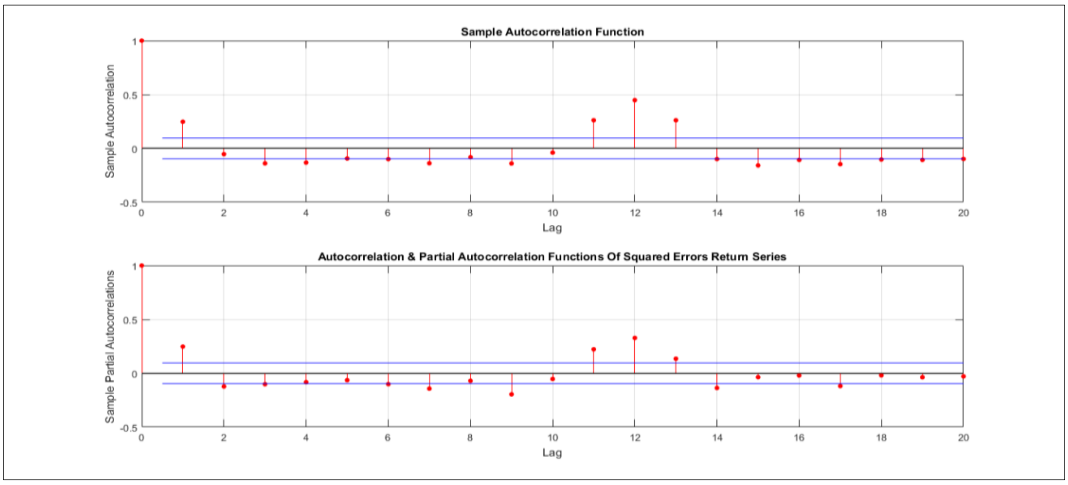

Figure 4- ACF and PACF of equared errors.

Figure-4 shows that the coefficients of ACF and PACF lay outside the boundaries of the trust interval, which represents a visual detection for the heteroscedasticity effect. To check it numerically, we used Ljung-Box Q-test in MATLAB program. If the result is (*h=1*), this means the acceptation to reject the null hypothesis H_0 and take the alternative hypothesis H_1. This refers to the existence of a heteroscedasticity effect and that the ***p-value*** will be less than ($\alpha = 0.05$), as shown in Table-1.

Table 1: Ljung-Box Q-test results

	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8	Lag9	Lag10
$h - value$	1	1	1	1	1	1	1	1	1	1
$p - value$	0	0	0	0	0	0	0	0	0	0
$Q - test$	186.3	230.8	233.5	305.2	474.3	684.5	851.3	928.9	931.4	982.0
$critical\ value$	3.841	5.991	7.814	9.487	11.07	12.59	14.06	15.50	16.91	18.30

Model Fitting and its Parameters Estimation

In this step, we fit EGARCH model for the return series and estimate the parameters of the model. Then we choose the best rank of the model based on the lowest values of AIC and BIC criteria, as shown in Table-2.

Table 2: Parameters, AIC and BIC values of the EGARCH model orders.

EGARCH(Q,P)	ω	α_i			γ_j			β_j			AIC	BIC
		α_1	α_2	α_3	γ_1	γ_2	γ_3	β_1	β_2	β_3		
EGARCH(0,1)	-2.6	-	-	-	0.67	-	-	-0.22	-	-	105	117.2
EGARCH(0,2)	-2.5	-	-	-	0.2	-0.5	-	-0.05	-0.65	-	77.2	97.6
EGARCH(0,3)	-2.6	-	-	-	0.3	-0.5	-0.4	0.09	-0.4	-0.5	20.1	48.6
EGARCH(1,1)	-6.6	0.67	-	-	0.61	-	-	-0.4	-	-	55.5	71.8
EGARCH(1,2)	-1.2	0.5	-	-	0.3	-0.7	-	-0.07	-0.5	-	21	45
EGARCH(1,3)	-1.9	0.2	-	-	0.3	-0.5	-0.4	0.02	-0.4	-0.2	18	51
EGARCH(2,1)	-0.6	1.5	-0.7	-	-0.1	-	-	-0.1	-	-	34.2	54.6
EGARCH(2,2)	-1.5	0.8	-0.4	-	0.2	-0.7	-	0.2	-0.5	-	12	41
EGARCH(2,3)	-1.4	1.2	-0.7	-	0.2	-1	0.6	0.1	-0.3	-0.09	7.5	44
EGARCH(3,1)	-4.1	0.3	-0.3	0.5	0.3	-	-	0.09	-	-	87	111
EGARCH(3,2)	-0.6	1.6	-1.5	0.6	0.1	-0.3	-	-0.08	-0.2	-	7.5	40
EGARCH(3,3)	-1.1	1.5	-1.5	0.5	0.1	-0.3	0.2	-0.2	0.3	-0.5	-14	26

Table-2 shows that the best model is EGARCH(3,3) and its equation is written as:

$$x_t = \sigma_t z_t \qquad z_t \sim iid\ N(0,1)$$

$$\begin{aligned}\log \sigma_t^2 = &-1.18205 + 1.55242\ \log \sigma_{t-1}^2 \\ &- 1.5 \log \sigma_{t-2}^2 + 0.52666 \log \sigma_{t-3}^2 + 0.199012[|z_{t-1}| - \mathrm{E}|z_{t-1}|] \\ &- 0.398616[|z_{t-2}| - \mathrm{E}|z_{t-2}|] + 0.229099[|z_{t-3}| - \mathrm{E}|z_{t-3}|] - 0.294789(z_{t-1}) \\ &+ 0.357283(z_{t-2}) - 0.517839(z_{t-3})\end{aligned}$$

with standard errors and t-statistic as listed in table 3.

EGARCH(3,3) Conditional Variance Model

Conditional Probability Distribution : Gaussian

Table 3: Parameter values of EGARCH(3,3) model.

Parameter	Value	Standard Error	t - Statistic
Constant	-1.18205	0.359252	-3.29031
GARCH(1)	1.55242	0.152433	10.1842
GARCH(2)	-1.5	0.17014	-8.81628
GARCH(3)	0.52666	0.145101	3.6296
ARCH(1)	0.199012	0.208635	0.953876
ARCH(2)	-0.398616	0.287039	-1.38871
ARCH(3)	0.229099	0.225754	-1.01482
Leverage(1)	-0.294789	0.169753	-1.73658
Leverage(2)	0.357283	0.256345	1.39376
Leverage(3)	-0.517839	0.188835	-2.74228

Also, when examining the stationary condition, we see that the model is stationary, as follow:

$$\sum_{i=1}^{3} \alpha_i = 1.55242 - 1.5 + 0.52666 = 0.579 < 1$$

and the unconditional variance value can be calculated as :

$$\sigma_x^2 = exp\left(\frac{\omega}{1-\sum_{i=1}^{3}\alpha_i}\right) = exp\left(\frac{-1.18205}{1-0.579}\right) = 0.0603102.$$

Checking the adequacy of the model

As in the Box-Jenkins methodology, we tested the adequacy of the selected model to ensure its efficiency in analyzing and explaing the data series. Then, we could use the model for forecasting, which was achieved by finding the standard residuals series, by using the following formula, and checking it visually and numerically:

$$R_t = \frac{r_t}{\sigma_t}$$

where R_t is standard residuals, r_t is white noise residual of return series, and σ_t is squared root of inferred conditional variance. Figure-5 represents a visual test plot of squared standard residuals (R_t) and its distribution line, compared with the normal distribution line and plot of ACF and PACF.

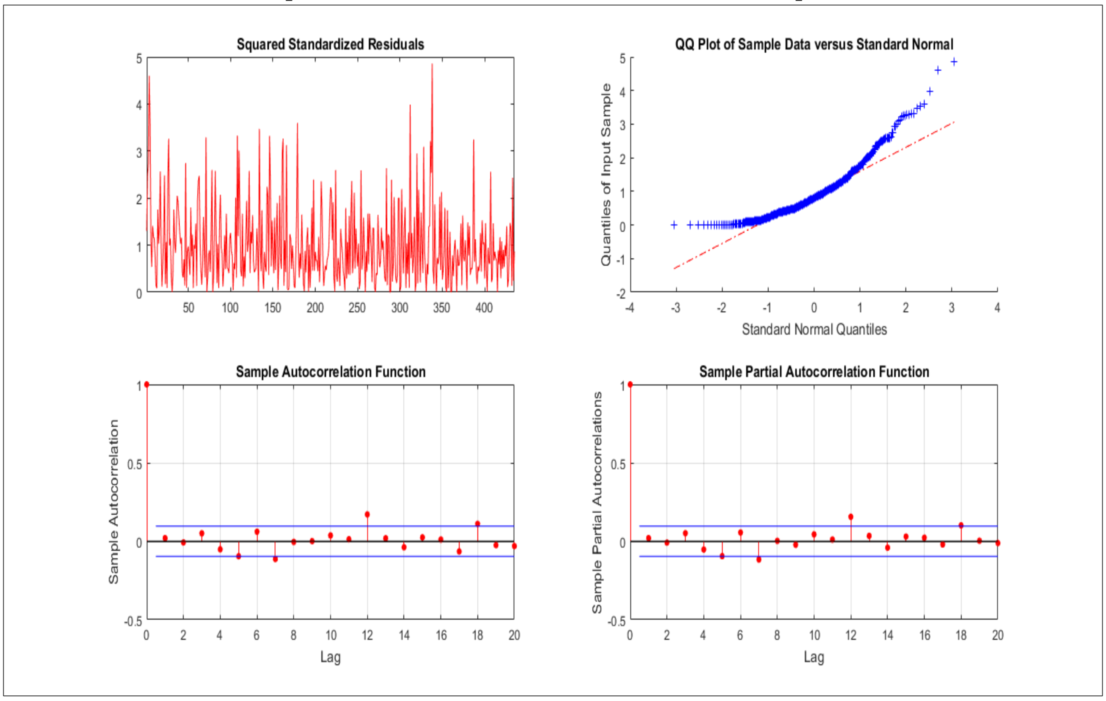

Figure 5- Plot of squared standard residuals.

Table-4 presents results of the numerical test and Ljung-Box test for squared standard residuals series with $h=0$, which implies unaccepting the rejection of the null hypothesis (H_0), i.e. accepting it. This refers to the removal of the heteroscedasticity effect, where the p-value is higher than $\alpha = 0.05$.

Table 4: Ljung-Box test for squared standard residuals**.**

	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8	Lag9	Lag10
$h-value$	0	0	0	0	0	0	0	0	0	0
$p-value$	0.675	0.903	0.723	0.648	0.2612	0.228	0.054	0.086	0.128	14.440
$Q-test$	0.174	0.202	1.325	2.477	6.494	8.133	13.82	13.83	13.83	14.44
$critical\ value$	3.841	5.991	7.814	9.487	11.07	12.59	14.06	15.50	16.91	18.30

These results show that the residuals series of EGARCH(3,3) model is uncorrelated and homoscedastic. That is, the variance equation of EGARCH(3,3) is adequate.

Convergence of the Conditional Variance

Now, we can use EGARCH(3,3) model and forecast the conditional variance. We plotted the inferred and the forecasted conditional variance in Figure-6. The results clearly demonstrate the steadiness of the conditional variance and the removal of the heteroscedasticity effect.

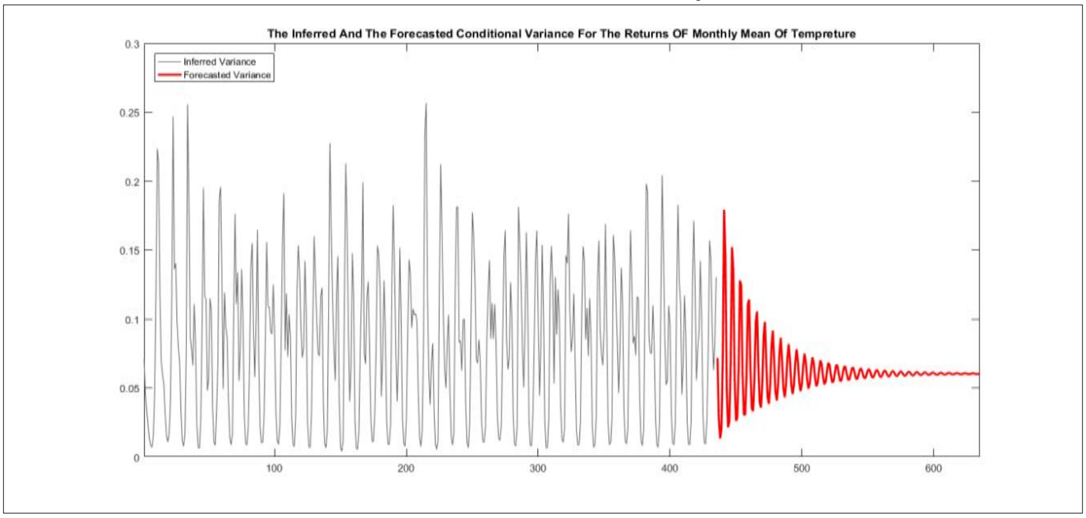

Figure 6-: Plot of the inferred and the forecasted conditional variance.

Figure-7 shows the results of the convergence of the conditional variance to the unconditional variance value, which is fixed at a equal to 0.06031.

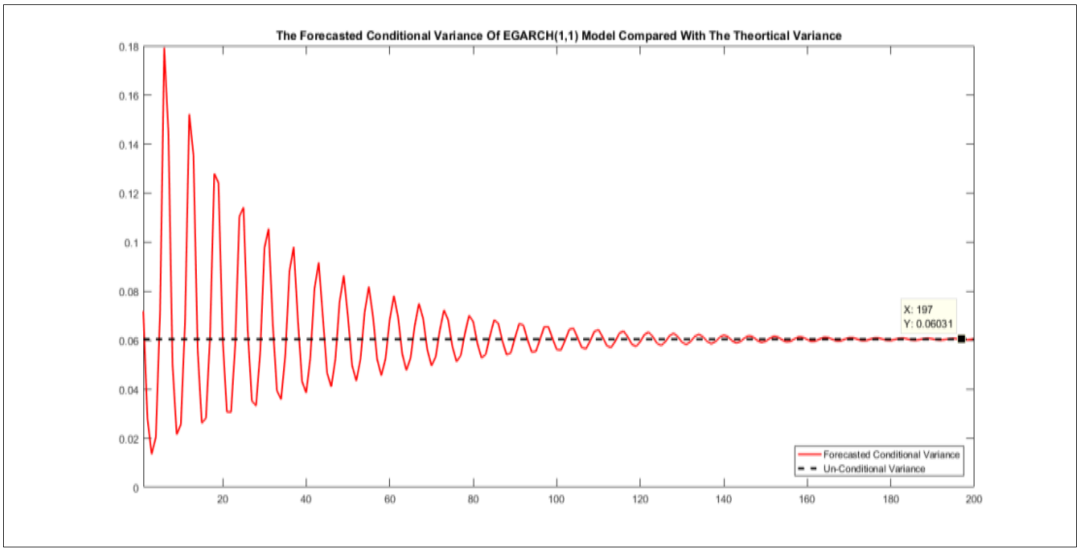

Figure 7- Plot of the forecasted conditional variance

Applying the Mixed ARMA(R,M)-EGARCH(Q,P) Model

The EGARCH model was mixed with the ARMA model to be used for the prediction of future values of time series under study. We attached data of the text program in appendix (3), which represents this process. Data were entered and transformed to the return series, as we did before. We fitted the mixed model and found its parameters until the fourth rank, considering the stationarity of ARMA and EGARCH models separately by another text program attached in appendix (4). This included finding roots of polynomial $(1-\sum_{i=1}^{R} a_i L^i)$ such that (a_i) represents autoregressive (AR) coefficients and L is the lag operator. It also included finding roots of polynomial $(1+\sum_{i=1}^{M} b_i L^i)$ such that (b_i) represents moving average (MA) coefficients. To check the stationary condition of ARMA model, all roots of the two polynomials must lie outside the unit circle, or we can find the roots of the characteristic equation of the two polynomials which is supposed to lie inside the unit circle, that is, the absolute value of the roots which is less than one. Also, the EGARCH stationary condition must be verified, which is also achieved by the program attached in appendix (4). Beside that, we selected the best model with the best rank, based on AIC and BIC criteria. Table (5) provides the stationarity statement along with AIC and BIC values of the models until the fourth rank (ARMA(4,4)-EGARCH(3,3)..).

Table 5: The stationarity, AIC and BIC values of tests.

Model	Stationarity	AIC	BIC
ARMA(0,1)-EGARCH(3,3)	yes	-205.9446	-157.0404
ARMA(0,2)-EGARCH(3,3)	yes	-274.5847	-221.6052
ARMA(0,3)-EGARCH(3,3)	yes	-315.3907	-258.3358
ARMA(0,4)-EGARCH(3,3)	yes	-359.1872	-298.0570
ARMA(1,0)-EGARCH(3,3)	yes	-383.5985	-334.6944
ARMA(2,0)-EGARCH(3,3)	yes	-384.5214	-331.5449
ARMA(3,0)-EGARCH(3,3)	yes	-414.7844	-357.7844
ARMA(4,0)-EGARCH(3,3)	yes	-441.7856	-380.6554
ARMA(1,1)-EGARCH(3,3)	yes	-371.7142	-294.2826
ARMA(1,2)-EGARCH(3,3)	yes	-389.7796	-312.3481
ARMA(1,3)-EGARCH(3,3)	yes	-407.0524	-329.6208
ARMA(1,4)-EGARCH(3,3)	yes	-415.9260	-338.4944
ARMA(2,1)-EGARCH(3,3)	yes	-373.8480	-296.4165

ARMA(2,2)-EGARCH(3,3)	yes	-654.0337	-576.6021
ARMA(2,3)-EGARCH(3,3)	yes	-639.1614	-561.7298
ARMA(2,4)-EGARCH(3,3)	yes	-714.8210	-637.3894
ARMA(3,1)-EGARCH(3,3)	yes	-547.5221	-470.0905
ARMA(3,2)-EGARCH(3,3)	yes	-653.8175	-576.3859
ARMA(3,3)-EGARCH(3,3)	yes	-655.5629	-578.3131
ARMA(3,4)-EGARCH(3,3)	yes	-639.2609	-561.8294
ARMA(4,1)-EGARCH(3,3)	yes	-576.2726	-498.8410
ARMA(4,2)-EGARCH(3,3)	yes	-654.3702	-567.9386
ARMA(4,3)-EGARCH(3,3)	no	-649.5050	-572.0734
ARMA(4,4)-EGARCH(3,3)	yes	-699.7428	-592.3112

Table-5 shows that the better model, with the lowest value of AIC and BIC criteria, is ARMA(2,4)-EGARCH(3,3) and the formula of the model will be as follows:

$$y_t = 0.00093 + 1.71632 y_{t-1} - 0.97824 y_{t-2} + x_t - 2x_{t-1} + 0.99737 x_{t-2} + 0.43746 x_{t-3} - 0.40035 x_{t-4} \qquad x_t \sim iid\ N(0, \sigma_t^2) \qquad Mean\ Eq.$$

$$x_t = \sigma_t z_t \qquad z_t \sim iid\ N(0,1)$$

$$\log \sigma_t^2 = -2.05536 + 0.904027 \log \sigma_{t-1}^2 + 0.446547 \log \sigma_{t-2}^2 - 0.8044221 \log \sigma_{t-3}^2 + 0.451993(|z_{t-1}| - \mathrm{E}|z_{t-1}|) - 0.4239(|z_{t-2}| - \mathrm{E}|z_{t-2}|) + 0.0314529(|z_{t-3}| - \mathrm{E}|z_{t-3}|) + 0.0099(z_{t-1}) + 0.0327769(z_{t-2}) + 0.0943251(z_{t-3}) \qquad Variance\ Eq.$$

The mean equation represents ARMA(2,4) model and the stationary condition was verified, since all roots of the polynomial for AR(2), which is:

$1 - 1.71632\,L + 0.97824\,L^2 = 0,$

lie outside the unit circle, as follow:

$$r_1 = 0.8772 + 0.5027i \ \Rightarrow\ |r_1| = 1.0111$$
$$r_2 = 0.8772 - 0.5027i \ \Rightarrow\ |r_2| = 1.0111$$

Also, all roots of the polynomial for MA(4), which is:

$$1 - 2\,L + 0.99737\,L^2 + 0.43746\,L^3 - 0.40035\,L^4 = 0,$$

lie outside the unit circle, as follow:

$$m_1 = -1.9180 \ \Rightarrow\ |m_1| = 1.9180$$
$$m_2 = 1.1083 \ \Rightarrow\ |m_2| = 1.1083$$
$$m_3 = 0.9512 + 0.5198i \Rightarrow\ |m_3| = 1.0840$$
$$m_4 = 0.9512 - 0.5198i \ \Rightarrow\ |m_4| = 1.0840$$

The variance equation represents the EGARCH(3,3) model, where its stationary condition was verified because:

$$\sum_{i=1}^{3} \alpha_i = 0.904027 + 0.446547 - 0.804221 = 0.5464 \ < 1$$

Now, we must infer the standard residuals series of the model ARMA(2,4)-EGARCH(3,3) and plot the series and the Quantle-Quantle function (QQ-plot). It appears from Figure-8 that the residuals series takes a straight line beside the normal distribution line. This implies that it is normally distributed. Figure-9 shows the plot of the curve of kernel distribution, compared with the curve of normal kernel distribution. Also, we plotted the histogram of the residual series.

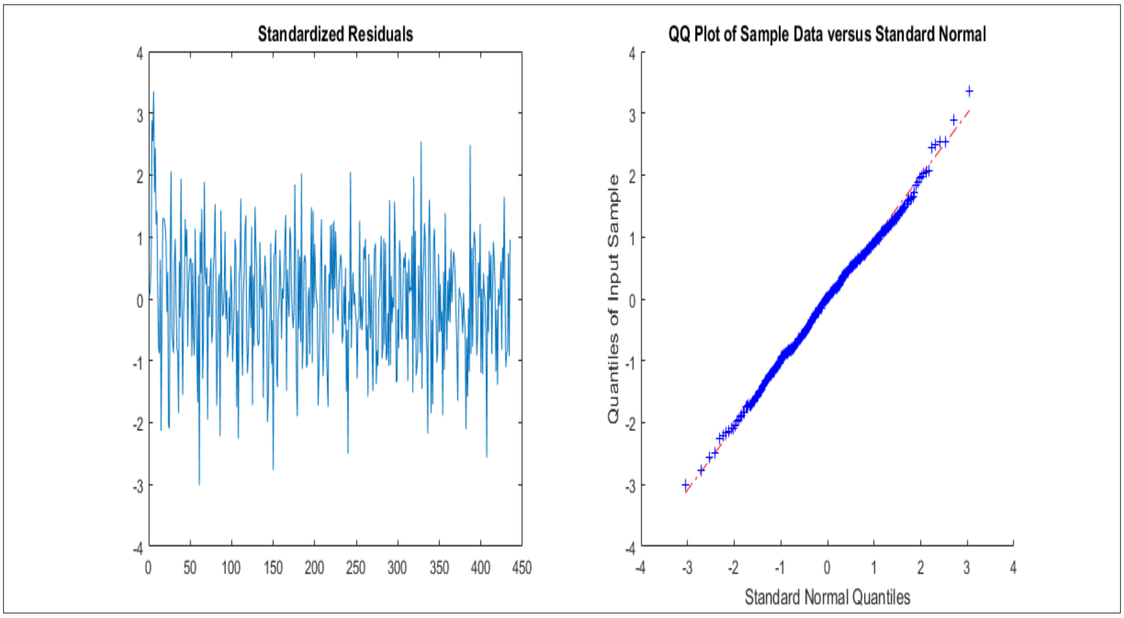

Figure 8- Plot the series and the Quantle-Quantle function of the residuals series.

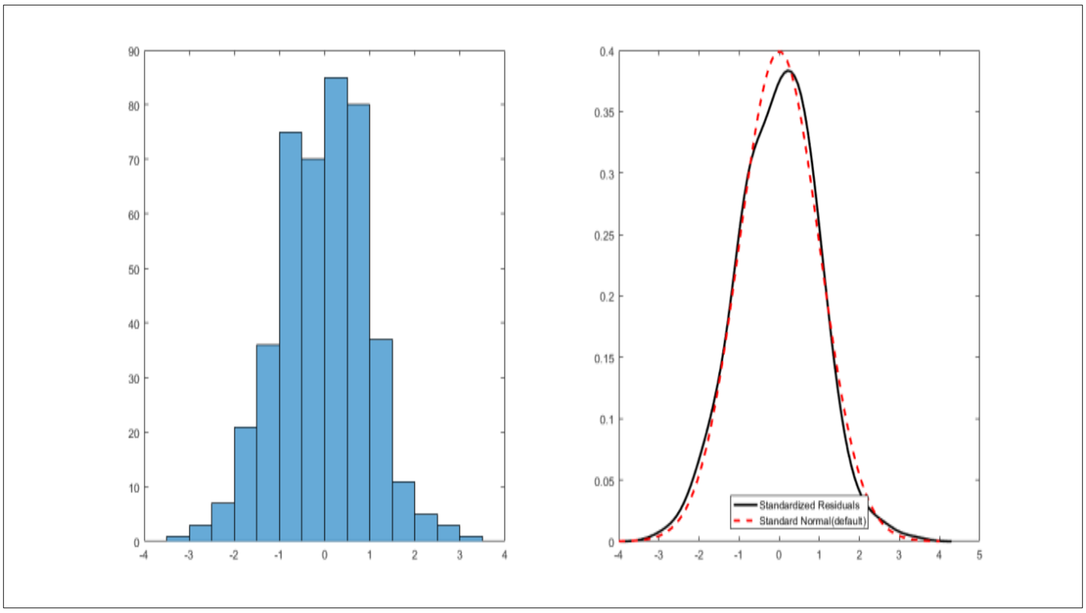

Figure 9- Plot of the curve of kernel distribution and the histogram of the residual series.

We can calculate the unconditional mean and unconditional variance of the model as follow:

$$\mu = \frac{Eq.1\ Constant}{1-\sum AR.coefficient} = \frac{0.000934623}{1-1.71632+0.978243} = 0.0036$$

$$\sigma^2 = exp\left(\frac{Eq.2\ Constant}{1-\sum GARCH\ coefficient}\right) = exp\left(\frac{-2.05536}{1-0.904027-0.446547+0.804221}\right) \cong 0.0108$$

The standard deviation value is equal to $\sigma = \sqrt{\sigma^2} = 0.1038$.

Figure (10) shows the convergence of the predicted return series for 250 future values along with its steadying at the value of unconditional mean. The predicted series lies inside the trust interval, bounded by $(y - \frac{1.96}{\sqrt{MSE}}, y + \frac{1.96}{\sqrt{MSE}})$. The predicted values of mean square error (MSE) converge to zero as the prediction steps increase. Also, we noticed the converge of the forecasted conditional variance and its steadying at the unconditional variance value.

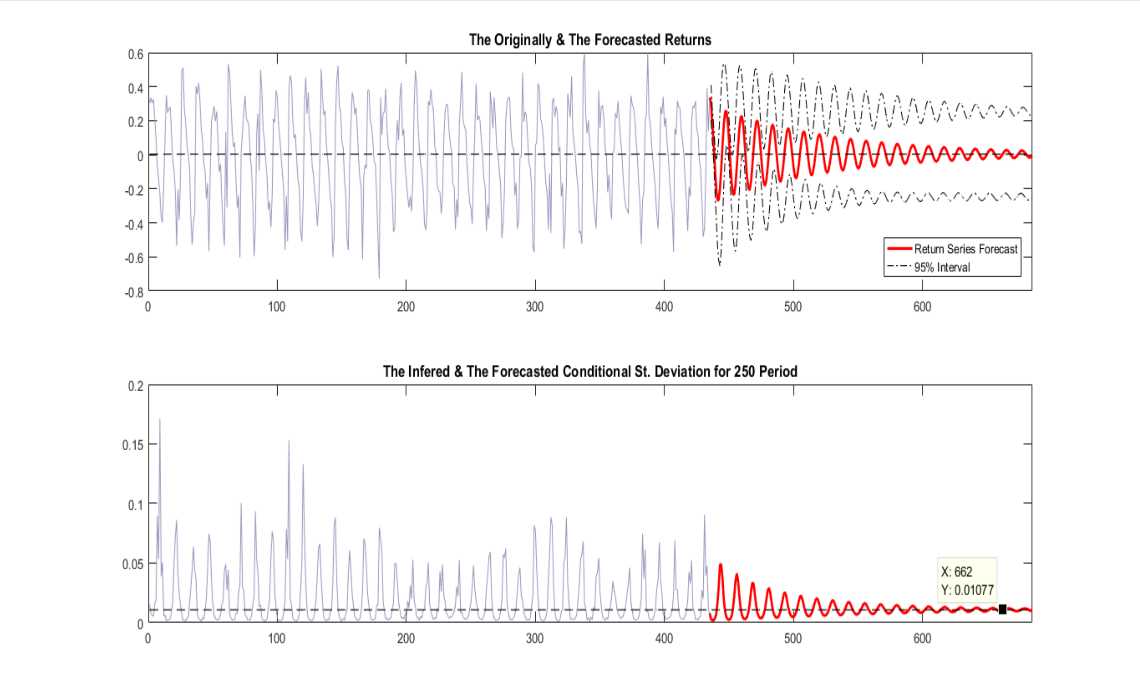

Figure 10- Plot of the convergence of the predicted return series.

Result and Discussion

After fitting, estimating, and testing the adequacy of the mixed model, we can use it in predicting the future of the series. We predicted eight future values, which is the same as the number of data that we left as a validating set. Then, we compared the predicted results with real values to asses and calculate the prediction error visually, by plotting the predicted and the real series, and then numerically, by using the prediction error criteria. We used the best three models in prediction, based on AIC and BIC criterions value as mentioned in table (5), which are ARMA(2,4)-EGARCH(3,3), ARMA(3,3)-EGARCH(3,3), and ARMA(4,4)EGARCH(3,3), to compare the prediction performance and estimate the magnitude of difference between the real values and the predicted results.

This process is necessary since it is worth noting that, although the model is good in adequacy, there is no guarantee that it will be a good predictor, as it may lack a good predictive performance. Table (6) presents the predicted values of the three chosen models, along with the values of prediction mean squared error (PMSE) and prediction mean absolute error (PMAE). We noticed that ARMA(4,4)EGARCH(3,3) model gives the lowest prediction error compared with the other models. However, this does not negate that the predictions of the other two models are acceptable.

Table 6: The real and predicted values of the three chosen models, along with the values of the prediction error criterions.

Predicted steps	months	Real data	ARMA(2,4)-EGARCH(3,3)	ARMA(3,3)-EGARCH(3,3)	ARMA(4,4)-EGARCH(3,3)
1	MAY	28.9	30.4	29.5	29.2
2	JUNE	35.2	36.3	35.7	35.8
3	JULY	38.3	39.4	38.5	39.1
4	AUG.	39.1	36.8	36.5	37.3
5	SEP.	31.8	30.3	31.2	31.8
6	OCT.	27.5	23.3	25.2	25.1
7	NOV.	19.2	17.8	20.2	19.4
8	DEC.	9.4	14.6	17.0	15.6

MSE			6.6042	8.1065	5.4287
MAE			2.0345	1.7238	1.3860

The three following figures (11, 12, and 13) illustrate the plots of real and predicted series of the three models.

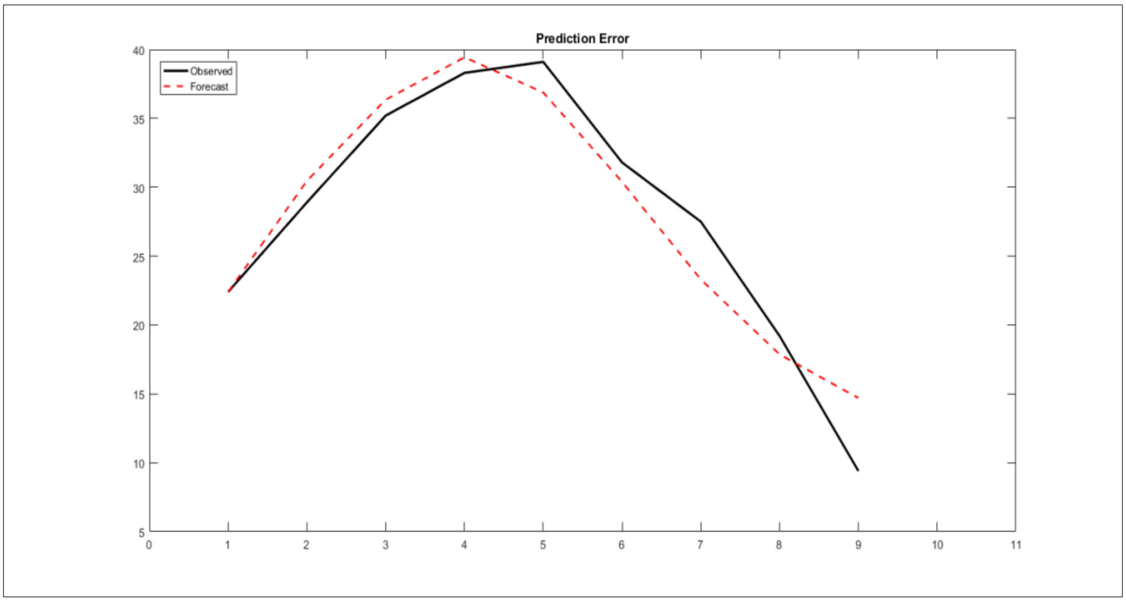

Figure 11 – ARMA(2,4)-EGARCH(3,3)

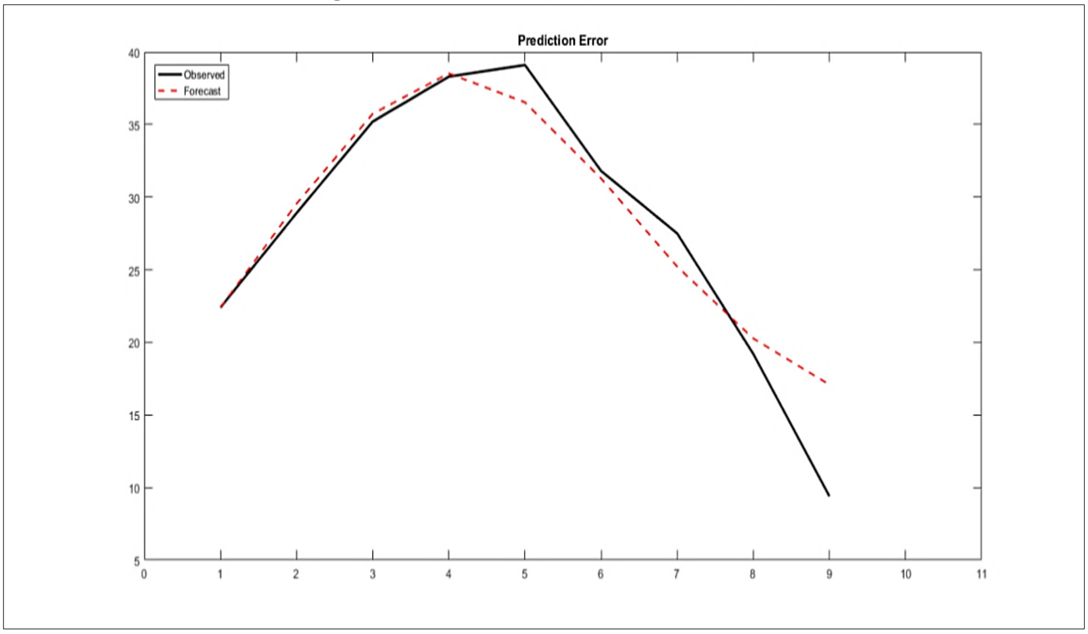

Figure 12 -ARMA(3,3)-EGARCH(3,3)

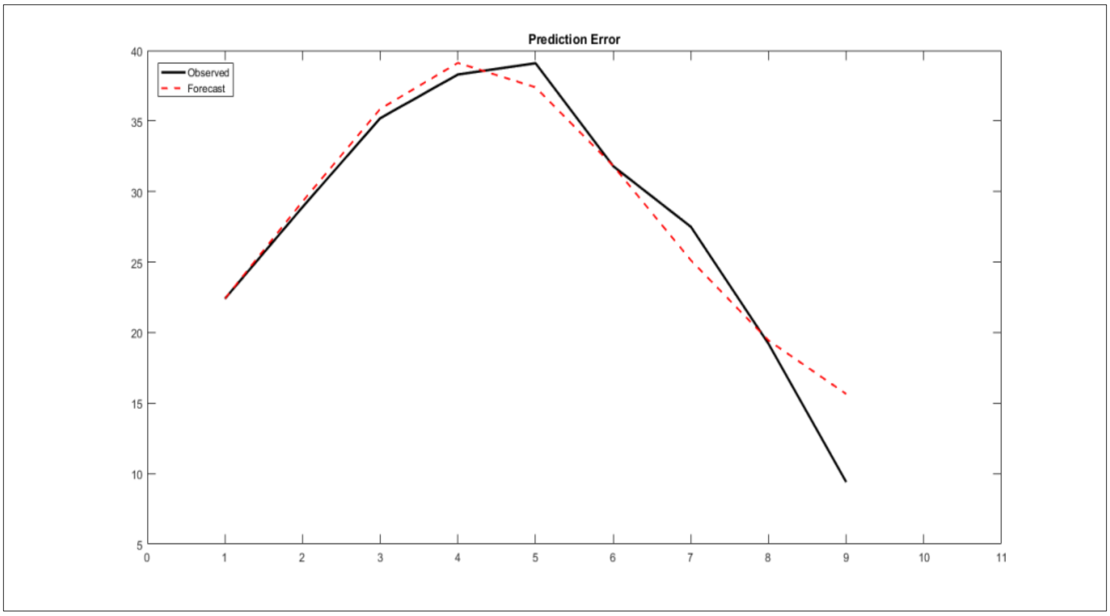

Figure 13 –ARMA (4,4)-EGARCH(3,3)

Conclusions

A mixed model of ARMA and EGARCH was suggested to be used in modeling and analyzing mean temperature time series. We also provided definitions and sufficient explanations on modeling and analyzing methodology. We showed that the EGARCH model cannot be used alone to predict future of the series, unless we mix it with ARMA, since the EGARCH model only forecasts the conditional variance. Also, forecasting by using ARMA model alone for data series containing volatility gives an unacceptable predicted result with high prediction error. Three of the best ranks of the mixed model were tested and its stationarity was verified. The results revealed that the best model, based on prediction error value, is ARMA(4,4)-EGARCH(3,3), with comparable predictive values and very acceptable prediction error.

In conclusion, the proposed model can be used with another data series but may need certain transformations that can be found in Box-Cox transformations. Also, the used methodology, which is known globally for many researchers and we mentioned before with its complete steps and procedures, enables the proposed model to control the linear and non-linear behavior of the series data. It has a prediction ability that allows its usage in predicting the future of temperature series.

Acknowledgement

I offer my full respect and thanks to Asst. Prof. Dr. Azhar Abbas Mohammad, the teacher in the mathematical department, college of computer sciences and mathematics of Tikrit University.

References

1- Engle, R. F. **1982**. ‘Autoregressive Conditional Heteroscedasticity with Estimate of the variance of united Kingdom Inflation’, *Econometrica*, **50**(4): 987–1007.

2- Bollerslev, T. **1986**. ‘Generalized Autoregressive Conditional Heteroskedasticity’, *Journal of Econometrics*, **31**: 307–327.

3- Al-Bazzoni, M. K. G. **2013**. Using of AR-GARCH Mixed Models in Analysis and Modeling The Time Series for Monthly Temprature Mean in Mosul City. Tikrit Universiyu.

4- Al-Obaidy, A. M. H. A. **2015**. A Study On Stability Conditions Of Suggested Model (STSechAR(p)) With Application. Tiktit Univercity.

5- Ramzan, Shahla, Ramzan, Shumila and Zahid, F. M. **2012**. ‘Modeling And Forecasting Exchange Rate Dynamics In Pakistan Using Arch Family Of Models’, *Electronic Journal of Applied Statistical Analysis*, **5**(1): 15–29. doi: 10.1285/i20705948v5n1p15.

6- Senaviratna, N. A. M. R. and Cooray, T. M. J. A. **2017**. ‘Forecasting Gold Prices In Sri Lanka Using Generalized Autoregressive Conditional Heteroskedasticity Approach’, *International Research Journal of Natural and Applied Sciences*, **4**(7): 99–110.

7- Francq, C. and Zokoian, J.-M. **2010**. *GARCH Models Structure, Statistical Inference and*

Financial Applications. 1st edn. John Wiley & Sons Ltd Registered.
8- Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. **1994**. *Time Series Analysis Forecasting and Control*. 3rd Ed. Englewood Cliffs, NJ : Prentic Hall.
9- Malmsten, H. **2004**. *Evaluating exponential GARCH models*. No 564. Stockholm.
10- Moffat, I. U., Akpan, E. A. and Abasiekwere, U. A. **2017**. 'A time series evaluation of the asymmetric nature of heteroscedasticity : an EGARCH approach', *International Journal of Statistics and Applied Mathematics*, **2**(6): 111–117.
11- Alexander, C. **2008**. *Market Risk Analysis Quantitative Methods in Finance*. 1st Ed. John Wiley & Sons Ltd.
12- Teräsvirta, T. **2006**. *An Introduction to Univariate GARCH Models*. No. 646. Stockholm.
13- TSAY, R. S. **2010**. *Analysis of Financial Time Series*. 3rd Ed. John Wiley & Sons, INC**.**
14- Nelson, D. B. **1991**. 'Conditional Heteroskedasticity in Asset Returns : A New Approach', *Econometrica*, **59**(2): 347–370.
15- Mohammad, A. A. and Mudhir, A. A. **2020**. 'Dynamical Approach in Studying Stability Condition Of Exponential (GARCH) Models', *Journal of King Saud University - Science*. Elsevier-King Saud University, **32**: 272–278. doi: 10.1016/j.jksus.2018.04.028.
16- Jelonek, P. Z. **2017**. *Econometrics in MATLAB : ARMAX , pseudo ex-post forecasting , GARCH and EGARCH*. Warwick.
17- Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. **2015**. *Time Series Analysis Forecasting and Control*. 5th Ed. Englewood Cliffs, NJ : Prentic Hall.
18- Makridakis, S., Harvard, S. C. W. and Hyndman, R. J. **1998**. *FORECASTING Methods and Applications*. 3rd Ed. John Wiley & Sons, INC.

Appendixes
Appendix (1)

STATION:-
KIRKUK Mean Air Temperature C°

YEAR	JAN.	FEB.	MAR.	APR.	MAY	JUN.	JUL.	AUG.	SEP.	OCT.	NOV.	DEC.
1980	7.6	10.3	14.4	19.5	27.0	33.5	36.8	34.6	29.9	23.7	16.4	11.0
1981	9.3	10.3	14.6	18.7	24.5	32.4	36.8	35.5	32.2	25.8	15.1	12.3
1982	8.4	7.5	12.3	20.5	26.6	32.2	34.2	34.0	31.4	21.7	12.8	8.6
1983	6.1	9.1	13.0	19.8	27.4	32.0	35.0	33.3	30.4	23.5	19.9	11.3
1984	10.2	12.5	15.8	21.0	25.6	33.0	36.1	32.6	30.9	23.8	15.7	9.2
1985	10.5	7.7	13.1	21.2	28.8	33.1	34.7	37.3	31.7	23.2	18.3	10.0
1986	10.2	11.6	14.9	21.1	26.0	31.7	37.2	37.1	34.1	25.6	14.1	9.2
1987	10.5	13.4	12.2	20.1	29.3	33.8	37.0	35.7	31.5	22.8	17.3	11.3
1988	8.9	11.0	13.3	19.4	27.2	32.2	35.4	34.7	30.9	25.2	14.4	11.3
1989	6.8	9.9	15.8	24.2	29.5	32.5	36.6	35.5	30.6	25.3	16.3	9.8
1990	7.4	9.8	15.4	19.6	28.3	33.2	37.0	34.6	30.8	24.9	18.7	12.3
1991	9.2	9.0	14.9	20.9	26.4	34.2	36.3	35.5	31.2	24.3	17.7	9.7
1992	6.0	7.1	10.9	18.4	24.4	31.4	34.0	35.4	30.9	24.4	14.9	8.5
1993	8.0	9.5	13.7	19.0	24.1	31.8	36.7	35.7	30.9	25.7	13.9	12.3
1994	11.6	10.7	15.6	22.8	27.9	33.0	35.9	34.5	32.7	25.6	15.5	7.5
1995	10.5	11.8	15.1	19.4	28.6	33.3	34.7	35.2	30.5	24.2	15.4	10.0
1996	10.3	12.6	13.7	19.2	29.3	32.8	38.1	35.9	30.7	24.0	17.6	13.8
1997	9.9	8.5	11.4	18.7	28.3	33.8	35.3	33.4	30.1	24.7	16.5	10.6
1998	7.6	10.4	13.9	20.4	27.6	35.6	37.5	37.8	31.6	25.5	20.5	14.5
1999	11.4	12.5	16.2	21.8	29.2	33.8	36.1	36.7	30.8	25.6	16.5	12.7
2000	8.8	10.5	14.4	23.4	28.6	33.7	39.0	36.8	30.8	23.2	17.0	11.3
2001	10.3	11.7	17.8	21.8	27.4	33.6	36.9	37.0	32.2	26.6	17.0	12.7
2002	8.7	12.6	16.7	19.5	27.3	33.5	34.7	34.9	31.7	26.6	16.3	11.0
2003	8.9	10.9	14.9	20.9	29.2	33.6	35.4	36.9	29.0	24.6	15.9	10.8
2004	10.6	10.5	17.0	20.6	26.5	33.4	36.5	34.7	32.1	26.7	15.6	8.8
2005	9.3	10.0	15.2	22.5	27.7	33.2	37.0	36.1	30.5	24.4	15.7	14.1
2006	8.9	12.1	16.9	21.1	28.0	35.8	36.4	38.1	29.8	25.2	14.5	9.4
2007	8.1	11.6	15.5	19.3	30.6	34.9	37.2	36.8	32.6	26.8	16.9	10.8
2008	6.4	10.8	19.6	24.7	28.0	34.3	36.7	37.1	32.1	24.3	17.1	11.1
2009	9.2	13.4	15.7	20.8	27.8	34.3	35.4	35.3	28.3	25.2	18.7	12.8
2010	12.9	13.3	17.0	21.1	28.9	35.2	38.3	38.4	33.4	26.3	19.3	14.0
2011	9.9	11.1	15.7	21.1	27.8	34.4	37.5	36.9	31.1	23.2	13.7	10.5
2012	8.5	10.3	13.1	23.8	29.9	35.7	37.9	36.6	32.6	26.1	18.4	11.9
2013	9.8	13.0	16.5	23.2	27.4	34.0	36.4	36.0	31.0	23.7	17.0	9.6
2014	11.2	12.6	17.3	22.9	30.1	34.5	37.3	37.4	32.1	24.4	15.6	12.9
2015	10.1	12.3	16.4	21.6	29.6	34.4	38.8	37.6	34.5	26.2	16.2	10.5
2016	9.4	13.9	16.0	22.4	28.9	35.2	38.3	39.1	31.8	27.5	19.2	9.4

Appendix (2)

```
Data = [enter data here];
figure(1)
hold on
xlabel('Months');
h = gca;
h.XTick = [1 121 242 363];
h.XTickLabel = {'Jan 1980','Jan 1990','Jan 2000',...
    'Jan 2010'};
ylabel('Monthly Mean Of Temperature');
title('Time Series Plot Of Monthly Mean Of Temperature For Kirkuk City From: JAN.1980 - DEC.2016')
plot(Data,'b');
hold off
r=price2ret(Data);
N =length(r);
meanR = mean(r);
error = r - mean(r);
squerror = error.^2;
figure(2)
plot(r)
hold on
plot(meanR*ones(N,1),'--r')
xlim([0,N])
xlabel('Months');
h = gca;
h.XTick = [1 121 242 363];
h.XTickLabel = {'Jan 1980','Jan 1990','Jan 2000',...
    'Jan 2010'};
ylabel('Returns');
title('Plot The Returns Series Of Monthly Mean Of Temperature')
hold off
figure(3)
hold on
xlabel('Months');ylabel('Squared Error');
title('Plot Of Squared Errors Return')
plot(squerror,'b');
hold off
figure(4)
subplot(2,1,1)
autocorr(squerror)
subplot(2,1,2)
parcorr(squerror)
title('Autocorrelation & Partial Autocorrelation Functions Of Squared Errors Return Series')
[h,pValue,Qstat,cValue] = lbqtest(r,'Lags',[1:10])
Q=3;
P=3;
Mdl=egarch(Q,P);
[EstMdl,EstParamCov,LogL,info] = estimate(Mdl,r);
numParams = sum(any(EstParamCov));
[AIC,BIC] = aicbic(LogL,numParams,N)
rng default;
V=infer(EstMdl,r);
StdRes=r./sqrt(V);
SquStdRes=StdRes.^2;
```

```
figure(5)
subplot(2,2,1)
plot(SquStdRes,'r')
xlim([1,N])
title('Squared Standardized Residuals')
subplot(2,2,2)
qqplot(SquStdRes)
subplot(2,2,3)
autocorr(SquStdRes)
subplot(2,2,4)
parcorr(SquStdRes)
[h,pValue,Qstat,cValue]=lbqtest(SquStdRes,'lags',[1:10])
Vf = forecast(EstMdl,200,'y0',r);
figure(6)
plot(V,'Color',[.4,.4,.4])
hold on
plot(N+1:N+200,Vf,'r','LineWidth',2)
xlim([1,N+200])
legend('Inferred Variance','Forecasted Variance','Location','Northwest')
title('The Inferred And The Forecasted Conditional Variance For The Returns OF Monthly Mean Of Temperature')
hold off
UnConVar=exp(EstMdl.Constant/(1-EstMdl.GARCH{1}-EstMdl.GARCH{2}-EstMdl.GARCH{3}))
figure(7)
plot(Vf,'r','LineWidth',1.5)
hold on
plot(ones(200,1)*UnConVar,'k--','LineWidth',2)
xlim([1 200]);
title('The Forecasted Conditional Variance Of EGARCH(3,3) Model Compared With The Theortical Variance')
legend('Forecasted Conditional Variance','Un-Conditional Variance','Location','SouthEast')
hold off
```

Appendix (3)

```
Data = [enter data here];
r=price2ret(Data);
N =length(r)
R=2;D=0;M=4;Q=3;P=3;
Mdl = arima(R,D,M);
Mdl.Variance = egarch(Q,P);
[EstMdl,EstParamCov,LogL,info]= estimate(Mdl,r);
numParams = sum(any(EstParamCov));
[AIC,BIC] = aicbic(LogL,numParams,N)
rng 'default';
[E0,V0,~] = infer(EstMdl,r);
stres=E0./sqrt(V0);
figure(8)
subplot(1,2,1)
plot(stres)
title('Standardized Residuals')
subplot(1,2,2)
qqplot(stres)
```

```
figure(9)
subplot(1,2,1)
histogram(stres)
x = -4:.05:4;
[f,xi] = ksdensity(stres);
subplot(1,2,2)
plot(xi,f,'k','LineWidth',2);
hold on
plot(x,normpdf(x),'r--','LineWidth',2)
legend('Standardized Residuals','Standard Normal(default)','Location','South')
hold off
[Y,MSE,V] = forecast(EstMdl,250,'Y0',r,'E0',E0,'V0',V0);
upper = Y + 1.96*sqrt(MSE);
lower = Y - 1.96*sqrt(MSE);
figure(10)
subplot(2,1,1)
plot(r,'Color',[.6,.6,.75])
hold on
UnConMean=0.000934623/(1-1.71632+0.978243);
p1=plot(N(end):N(end)+250,[r(end);Y],'r','LineWidth',2);
p2=plot(N+1:N+250,lower,'k-.','LineWidth',0.5);
plot(N+1:N+250,upper,'k-.','LineWidth',0.5);
plot(ones(685,1)*UnConMean,'k--','LineWidth',0.5)
xlim([0,N+250]);
title('The Originally & The Forecasted Returns')
legend([p1,p2],'Return Series Forecast','95% Interval','Location','SouthEast','Orientation','vertical')
hold off
subplot(2,1,2)
UnConVar=exp(-2.05536/(1-0.904027-0.446547+0.804221));
plot(V0,'Color',[.6,.6,.75])
hold on
plot(N(end):N(end)+250,[V0(end);V],'r','LineWidth',2);
plot(ones(685,1)*UnConVar,'k--','LineWidth',0.5)
xlim([0,N+250])
title('The Infered & The Forecasted Conditional St. Deviation for 250 Period')
hold off
disp(UnConVar)
STDVriance = sqrt(UnConVar)
disp(UnConMean)
rng default;
[Y1,MSE1,V1]=forecast(EstMdl,8,'Y0',r,'E0',E0,'V0',V0);
[YSim,ESim,VSim] = simulate(EstMdl,8,'NumPaths',20000,'Y0',r,'E0',E0,'V0',V0);
Real = [22.4; 28.9; 35.2; 38.3; 39.1; 31.8; 27.5; 19.2; 9.4];
Fdata=ret2price(Y1,22.4)
sim=mean(YSim,2);
Sdata=ret2price(sim,22.4)
figure(11, 12 and 13)
plot(Real,'k','LineWidth',2)
hold on
plot(Fdata,'r--','LineWidth',1.5)
xlim([0,11])
title('Prediction Error')
legend('Observed','Forecast','Location','NorthWest')
hold off
MSE = mean((Real-Fdata).^2)
```

MSEsim = mean((Real-Sdata).^2)
MAE = mean(abs(Real-Fdata))
MAEsim = mean(abs(Real-Sdata))

Appendix (4)

EGARCH stationary text program

- x1=1.61895; x2= -1.5; x3=0.643516;
- p = [x3 x2 x1 -1];
- r = roots(p)
- Absr=abs(r)
- A = LagOp({1,-x1,-x2,-x3},'Lags',[0,1,2,3]);
- [indicator,eigenvalues] = isStable(A)
- AbsE=abs(eigenvalues)
- Sum=x1+x2+x3

ARMA stationary text program

- A1=1.71632; A2=-0.978243; A3=0; A4=0;
- B1=-2; B2=0.997374; B3=0.437468; B4=-0.400353;
- p = [A4 A3 A2 A1 -1];
- r = roots(p)
- Absr=abs(r)
- pB = [B4 B3 B2 B1 1];
- rB = roots(pB)
- AbsrB=abs(rB)
- A = LagOp({1,-A1,-A2,-A3,-A4},'Lags',[0,1,2,3,4]);
- [indicator,eigenvalues] = isStable(A)
- AbsE=abs(eigenvalues)
- AB = LagOp({1,B1,B2,B3,B4},'Lags',[0,1,2,3,4]);
- [indicatorB,eigenvaluesB] = isStable(AB)
- AbsEB=abs(eigenvaluesB)