MJFAS
Malaysian Journal of Fundamental and Applied Sciences

RESEARCH ARTICLE

GARCH Models and Distributions Comparison for Nonlinear Time Series with Volatilities

Nur Haizum Abd Rahman[a*], Goh Hui Jia[b], Hani Syahida Zulkafli[b]

[a]Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuhraya Persiaran Tun Khalil Yaakob, 26300 Gambang, Pahang, Malaysia

[b]Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, 43400, Serdang, Selangor, Malaysia

Abstract The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model is extensively used for handling volatilities. However, with numerous extensions to the standard GARCH model, selecting the most suitable model for forecasting price volatilities becomes challenging. This study aims to examine the performance of different GARCH models in forecasting crude oil price volatilities using West Texas Intermediate (WTI) data. The models considered are the standard GARCH, Integrated GARCH (IGARCH), Exponential GARCH (EGARCH), and Golsten, Jagannathan, and Runkle GARCH (GJR-GARCH), each with normal distribution, Student's t-distribution, and Generalized Error Distribution (GED). To evaluate the performance of each model, the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) are used as the model selection criteria, along with forecast accuracy measures such as absolute error, root mean squared error (RMSE), and mean absolute error (MAE). Post-estimation tests, including the Autoregressive Conditional Heteroskedasticity Lagrange Multiplier (ARCH-LM) test and the Ljung-Box test, are conducted to ensure the adequacy of all models. The results reveal that all GARCH models are suitable for modeling the data, as indicated by statistically significant estimated parameters and satisfactory post-estimation outcomes. However, the EGARCH (1, 1) model, particularly with Student's t-distribution, outperforms other models in both data fitting and accurate forecasting of nonlinear time series.

Keywords: GARCH, volatilities, financial data, time series.

Introduction

A time series is a sequence of findings that was documented conforming to time. A scalar process which can be stated as a linear sequence of past or future values or differences is known as a linear time series. Meanwhile, any processes that cannot be expressed linearly are known as nonlinear time series [1]. In simpler words, the common perception of linear time series is that there is a relatively steady one with no unexpected outbreak occurring as a result of its system's linearity. However, for some data sets, especially financial data, a rapid outbreak of volatility exists, which is impossible to model using linear time series. Hence, nonlinear time series models are introduced to cope with data which shows nonlinear behaviours like time-varying variance (volatility), unsymmetrical cycles, and higher-moment structure.

Volatility refers to the variation or dispersion of observed returns measured over a specific unit of time [2]. According to Hung *et al.* [3], before introducing the Autoregressive Conditional Heteroskedasticity (ARCH) model, the model used to assume the volatility of predicted variables is deficient as it assumes a constant expected variance of random error. To overcome this drawback, the ARCH model was proposed in 1982 by Robert F. Engle. The ARCH model assumes a nonconstant conditional variance of random error at which it depends on the previous random errors while the unconditional variance remains constant [4]. Soon in 1986, Tim Bollerslev modified and expanded Engle's ARCH model,

***For correspondence:**
haizum@umpsa.edu.my

Received: 25 July 2023
Accepted: 10 Nov. 2023

known as the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model [5]. The GARCH model is one of the nonlinear time series models which is commonly used in financial and economic data to forecast volatilities. The term heteroskedasticity means the situation of serially autocorrelated variance over time. Bollerslev [5] explained that a GARCH model is similarly extended from the ARCH model as an Autoregressive Moving Average (ARMA) process is extended from an Autoregressive (AR) process. Thus, the GARCH model is a model which takes both past squared errors and past conditional variances into account.

With the development of time, more researchers studied the standard GARCH model and modified or extended it to fit different types of data better. Engle and Bollerslev [6] introduced the Integrated GARCH (IGARCH) model as they believe that the mean is not the only factor that should be considered in modelling. Still, the spread of data should also be considered. They proposed the IGARCH model with a persistent conditional variance which is the standard GARCH model but with a unit root. Furthermore, Nelson [7] found some disadvantages of the standard GARCH model when dealing with financial data and proposed a new model, namely the Exponential GARCH (EGARCH) model. The most distinctive contrast between EGARCH and the standard GARCH model is that the EGARCH model can respond asymmetrically to positive and negative volatility, where negative volatility brings a greater impact than positive volatility. In 1993, Lawrence Golsten, Ravi Jagannathan, and David Runkle introduced the Golsten, Jagannathan, and Runkle GARCH (GJR-GARCH) model, which detects an asymmetric response to positive and negative volatility in a similar way to the EGARCH model [8]. According to Nugroho *et al.* [9], an additional error term will be included in the current conditional variance of the GJR-GARCH model if there is negative volatility.

The development of GARCH models increases the effectiveness in predicting shifting variability or variance volatility on time series, especially in evaluating financial assets [10]. For example, using the EGARCH model, Meher *et al.* [11] researched the consequences and leverage effect of COVID-19 on the crude oil and natural gas price volatility traded on India's Multi Commodity Exchange (MCX). The EGARCH model with normal distribution, Student’s t-distribution and GED were formed for crude oil and natural gas data, respectively. The results revealed that the EGARCH model with GED was the best for crude oil data, while the EGARCH model with Student’s t-distribution was the best for natural gas data in India. It can be concluded that the outbreak of the COVID-19 pandemic contributed to the asymmetric volatility in crude oil prices but not in natural gas based on the sign of asymmetric terms.

In addition to utilising a single GARCH model for predicting oil price volatility, as previously discussed, additional studies conducted by Herrera *et al.* [12] and Hung *et al.* [13] have examined the comparative performances of various GARCH models in forecasting financial volatility. Herrera *et al.* [12] applied standard GARCH, EGARCH, GJR-GARCH, Fractionally Integrated GARCH (FIGARCH) and Markov Switching GARCH (MS-GARCH) models under normal distribution, Student’s t-distribution and GED to predict WTI crude oil price volatility on various horizons. The researchers concluded that the models with Student’s t-distribution outperformed the other distributions, and the performance of GARCH family models differed at different horizons. In addition, Hung *et al.* [13] carried out research to forecast the global oil price volatility by using three GARCH models, which were GARCH (1, 1), EGARCH (1, 1) and GJR-GARCH (1, 1) models. The estimations were made under four different types of distributions: normal distribution, Student’s t-distribution, GED and skewed Student’s t-distribution. The findings revealed that the EGARCH (1,1) model with Student’s t-distribution gave the most precise forecast after comparing the measures of forecast accuracy, including mean squared error (MSE), mean absolute error (MAE), and root mean squared error (RMSE) values for all the models.

GARCH models have proven to be versatile beyond the modelling and forecasting of financial assets. In a study conducted by Venkatareddy and Kotreshwar [14], the GARCH model was employed to analyse rainfall variations across meteorological subdivisions in India. With the exception of West UP, the data revealed significant volatility in all ten subdivisions. The authors concluded that their research has provided valuable insights for policymakers in the field of rainfall risk management. Moreover, GARCH models have also been compared in air pollution studies due to their inherent volatility. In a study by Zamrus *et al.* [15], several GARCH family models, such as Poisson integer value GARCH (INGARCH), negative binomial integer value GARCH (NBINGARCH), and integer value autoregressive conditional heteroskedasticity (INARCH), were compared in the context of the Air Pollutant Index (API) of Malaysia. The findings demonstrated that NBINGARCH (1,1) outperformed conventional ARCH family models, namely INARCH (1,0) and INGARCH (1,1), in providing more accurate predictions for the observed API at all five study areas.

Based on the mentioned literature, one of the difficulties in forecasting volatilities is deciding the most suitable GARCH models to be used. The standard GARCH model has been extended to several types, such as IGARCH, EGARCH, and GJR-GARCH models. Every GARCH model has its strengths and weaknesses, which might lead to a different performance in forecasting price volatilities, especially when different data types are used. In addition, the accuracy of each family model might also differ according to its model specifications, even when they are applied to the same data set. There is no rule of thumb on which GARCH model should be used for a certain data set. Hence, a good understanding of the features of different GARCH models and the data sets is required before identifying the most adequate model to forecast the price volatilities. The objectives of this paper are to model and forecast the nonlinear series using different GARCH models and to compare the GARCH model's accuracy in forecasting the price volatility by implementing measures of forecast accuracy.

Methodology

ARCH Effect Test

Commonly, it is considered a best practice to examine the residuals of the in-sample data to detect the existence of ARCH effects before modelling financial time series using GARCH models. The ARCH Lagrange multiplier (LM) test was introduced by Engle [4] in 1982 for this purpose. The regression model for the ARCH LM test is expressed as in Equation 1:

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-q}^2 + u_t \tag{1}$$

where α_0 is the intercept, α_i is the coefficients of the regression, q is the number of restrictions, and u_t is the error term. The hypothesis and the test statistic are given in as follows:

H_0: No ARCH (q) effect in the residuals
H_1: ARCH (q) effects in the residuals

$$LM = nR^2 \sim \chi^2(q) \tag{2}$$

where n is the number of observations, R^2 is computed from the regression. Reject the null hypothesis when p–value is lesser than the level of significance and can conclude that ARCH (q) effects exist in the residuals. This suggests that GARCH models are suitable for conducting further analysis.

Garikai [16] emphasised that another alternative to test ARCH effect is through the Ljung-Box Q statistic. The test was introduced by Ljung and Box [17] in 1978 to check the serial residuals correlation. The hypothesis is given by:

H_0: The model does not exhibit serial correlation.
H_1: The model exhibits serial correlation.

Given in Equation 3 is the test statistic.

$$Q = n(n+2) \sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-k} \tag{3}$$

where n is the number of residuals, h is the number of time lags to be tested, k is the time lag and $\hat{\rho}_k^2$ is the residual autocorrelation at lag k. Similarly, the null hypothesis is rejected if the p–value is less than the level of significance. It can be concluded that the model exhibits serial correlation, and further analysis can be done using GARCH models.

GARCH Models

Four GARCH models are considered in this study, including the standard GARCH model, the IGARCH model, the EGARCH model and the GJR-GARCH model. The smallest lag order is acceptable to capture the shifting volatility and produce adequate results [18]. Hence, the GARCH models applied in this study are defined with the smallest lag order to satisfy the parsimony concept in model building. The model specifications are briefly discussed in this section. Dritsaki [10] explained that two equations, which are the mean and the variance equation, make up the GARCH models. More attention is given to the variance equation in contrast to the mean equation since the same equation of mean is used to compare the various variances. Generally, the mean equation is given by Equation 4.

$$r_t = \mu + \varepsilon_t, \varepsilon_t = \sigma_t \eta_t \tag{4}$$

where r_t is the return at time t, μ is the mean of return, and ε_t is the return residual at time t. The return residual can be split into conditional standard deviation, σ_t following the GARCH model conditional variance equation and a stochastic piece, η_t which is identically and independently distributed.

The standard GARCH (p, q) model was introduced by Bollerslev [5] can be specified as in Equation 5. The degree of ARCH term is denoted by q meanwhile the degree of GARCH term is denoted by p.

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{5}$$

where the conditional variance denoted as σ_t^2 represents the variance at time t, which is determined by historical information up to time $t - j$. The squared residuals at time $t - i$, ε_{t-i}^2 often referred to as the ARCH term. On the other hand, σ_{t-j}^2 is indicate as the conditional variance at time $t - j$, commonly known as the GARCH term. The constant term denoted as α_0. The α_i and β_j being respective weights of ε_{t-i}^2 and σ_{t-j}^2 are parameters to be estimated through maximum likelihood estimation. The condition of $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ must be fulfilled to ensure the positivity of conditional variance, while to ensure the stationarity of conditional variance, the condition $\sum_{i=1}^{\max(q,p)} \alpha_i + \beta_i < 1$, must be fulfilled.

The GARCH (1,1) can be written as in Equation 6 that has the following conditional variance equation.

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{6}$$

Standard GARCH model with a unit root is known as the IGARCH model which was introduced by Engle and Bollerslev [6]. The general IGARCH (p, q) model can be expressed as in Equation 7.

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{7}$$

where

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j = 1 \tag{8}$$

Rearrange Equation 8 yields Equation 9.

$$\sum_{j=1}^{p} \beta_j = 1 - \sum_{i=1}^{q} \alpha_i \tag{9}$$

Substitute Equation 9 into Equation 7 yields Equation 10.

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \left(1 - \sum_{i=1}^{q} \alpha_i\right) \sigma_{t-j}^2 \tag{10}$$

The IGARCH (1,1) has the following conditional variance equation as in Equation 11.

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + (1 - \alpha_1) \sigma_{t-1}^2 \tag{11}$$

The condition $\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0$ must be fulfilled to ensure the positivity of conditional variance.

Since the standard GARCH model and IGARCH model are unable to acquire asymmetric effects, Nelson [7] introduced the EGARCH model to overcome the drawbacks. The EGARCH (p, q) model can be written as in Equation 12.

$$\ln \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q}\left[\alpha_i\left(\left|\frac{\varepsilon_{t-i}}{\sigma_{t-i}}\right| - E\left|\frac{\varepsilon_{t-i}}{\sigma_{t-i}}\right|\right) + \gamma_i \frac{\varepsilon_{t-i}}{\sigma_{t-i}}\right] + \sum_{j=1}^{p}\beta_j \ln \sigma_{t-j}^2 \tag{12}$$

The EGARCH $(1,1)$ has the following conditional variance equation as in Equation 13.

$$\ln \sigma_t^2 = \alpha_0 + \alpha_1\left(\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| - E\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right|\right) + \gamma_1 \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \beta_1 \ln \sigma_{t-1}^2 \tag{13}$$

where α_i determine the size of asymmetry effect while γ_i determine the sign of asymmetry effect. There are no sign restrictions for the parameters since the equation for the conditional variance is in log-linear form, σ_t^2 can never be negative. The EGARCH model employs a standardised value of $\varepsilon_{t-i}/\sigma_{t-i}$ instead of ε_{t-i} as Nelson [7] believed that the size and persistence of the shocks can be interpreted in a more natural way by using standardised value. If there is a positive shock, the effect on the conditional variance is $\alpha_i + \gamma_i$ whereas if there is a negative shock, the effect on the conditional variance is $\alpha_i - \gamma_i$. If $\gamma_i < 0$, this indicates that a negative shock will have a greater influence than a positive shock.

Finally, the GJR-GARCH (p, q) model which was developed by Golsten *et al.* [19] in 1993. The GJR-GARCH model can capture asymmetry effects same as EGARCH model. The model is shown below.

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q}(\alpha_i + \gamma_i N_{t-i})\,\varepsilon_{t-i}^2 + \sum_{j=1}^{p}\beta_j \sigma_{t-j}^2 \tag{14}$$

where

$$N_{t-i} = \begin{cases} 1 \text{ if } \varepsilon_{t-i} < 0, \text{negative shock} \\ 0 \text{ if } \varepsilon_{t-i} \geq 0, \text{negative shock} \end{cases} \tag{15}$$

The following conditional variance equation for GJR-GARCH (1,1) can be written as:

$$\sigma_t^2 = \alpha_0 + (\alpha_1 + \gamma_1 N_{t-1})\varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{16}$$

where γ_i represents the asymmetric response parameter. The effect on the conditional variance is $\alpha_i + \gamma_i$ if there is a negative shock, while the effect on the conditional variance is α_i if there is a positive shock. A negative shock has a greater influence on conditional variance than positive shock when $\gamma_i > 0$ which contrasts with the EGARCH model. Both conditions $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and $\sum_{i=1}^{q}(\alpha_i + \gamma_i) \geq 0$ must be fulfilled to ensure the positivity of conditional variance.

Test of Significance

The significance of every estimated parameter in each GARCH model is tested by using hypothesis testing. It is important to determine whether the estimated parameter's size is appreciably great enough to draw the conclusion that the real parameter is not zero, statistical tests are utilized. To achieve this, the estimated parameter to its standard error is contrasted. The hypothesis and the test statistic are below.

H_0 : The parameter is not statistically significant
H_1 : The parameter is statistically significant

$$t = \frac{\hat{\theta}}{se(\hat{\theta})} \tag{17}$$

where $\hat{\theta}$ is the estimated parameter. Reject the null hypothesis if the absolute value of t-statistics is greater than 2 or if the p–value is lesser than α which is the level of significance. Thus, can conclude that the parameter is statistically significant.

Model Selection

In recent years, big data analytics, machine learning and statistical learning have all placed a lot of emphasis on model selection [20]. According to Fabozzi *et al.* [21], a variety of information-based model selection approaches are available to identify the best model from a group of models, including Akaike

Information Criteria (AIC) and Bayesian Information Criteria (BIC). The AIC was introduced by Akaike [22] in 1973 which applies the log-likelihood and a penalising element according to the number of estimated parameters. It is commonly known that the fit of models may be enhanced by including additional parameters. Hence, AIC seeks to strike a compromise between the goodness of fit and the number of parameters in the model which is called the penalty term. The formula of AIC is shown in Equation 18.

Another information-based model selection technique that is applied in a Bayesian framework is BIC. BIC was developed by Schwarz [23] in 1978. BIC uses a penalising element like the AIC but has a larger impact as it is linked to the number of parameters and the sample size. Lower value of AIC and BIC is preferable that indicate better model fitting. The formula of BIC is shown in Equation 19.

$$AIC = -2\ln(L) + 2n_{par} \tag{18}$$

$$BIC = -2\ln(L) + n_{par}\ln(n) \tag{19}$$

where ln is the natural log, L is the maximum value of the likelihood function for the model, n_{par} is the total number of parameters including constant terms in the model and n is the sample size.

Dziak *et al.* [24] explained that the AIC model is not consistent and might overestimate the results while the BIC model is consistent but might underestimate the results. Generally, AIC emphasised on good prediction while BIC emphasised on parsimony. Hence, both of the AIC and BIC values are computed in this study.

Measures of Forecast Accuracy

After modelling and forecasting the nonlinear series using different GARCH models, the models are compared based on their accuracy by implementing measures of forecast accuracy on top of the model selection criteria. According to Mehdiyev *et al.* [25], by addressing future uncertainty, forecasts play a significant role in helping people make logical judgments and schedule their activities more accurately. Over the past three decades, several studies have utilised various accuracy metrics as an assessment criterion in order to quantify the effectiveness of forecasting methodologies. For instance, squared errors and absolute errors are the foundation for the computations of root mean squared error (RMSE) and mean absolute error (MAE), respectively.

Let y_t be the actual values at time t and $\hat{y}_t$ be the forecasted values at time t, the error term at time , e_t is the difference between the actual values and the forecasted values at time t. A smaller magnitude of absolute error indicates a more accurate model. The equation of the absolute error is shown in Equation 20.

$$Absolute\ Error = \sum_{t=1}^{n}|e_t| = \sum_{t=1}^{n}|y_t - \hat{y}_t| \tag{20}$$

RMSE is the square root of MSE which is the root squared value of the average of squared errors. A smaller value of RMSE is preferable as it indicates a more accurate model. The equation of the RMSE is shown in Equation 21.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{21}$$

MAE is another measure of forecast accuracy which takes the average of the absolute errors. Similar to RMSE, a smaller value of MAE is preferable as it indicates a better prediction performance. The equation of the MAE is shown in Equation 22.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \tag{22}$$

According to Koutsandreas *et al.* [26], due to the errors being squared prior to averaging, RMSE has a significantly higher weight for large errors which made it more sensitive to outliers as compared to MAE.

Post-estimation Test

According to Garikai [16], post-estimation tests are essential to confirm the effectiveness of the study's models. Specifically, the ARCH effect tests are crucial for the volatility model. Hence, both ARCH LM test and Ljung Box test will be run again to investigate the existence of ARCH effects. The test procedures for both tests are similar as discussed in section ARCH effect test. However, since the purpose of running the test again is to make sure that the corresponding models are effective and do not show any additional ARCH effects, we would expect the null hypothesis is failed to be rejected. This indicates that there is no need to utilise models that are more complex because the models in use are sufficient.

Results and Discussion

Data and Preliminary Analysis

The West Texas Intermediate (WTI) crude oil daily spot price, as reported by the U. S. Energy Information Administration, covering the period from 1 July 2002 to 30 June 2022, was used in this study. The sample is partitioned into two sections, namely, the in-sample data and the out-sample data. The in-sample data covers the period of 1 July 2002 to 31 December 2021, a total of 4897 data to estimate the GARCH models. On the other hand, the out-sample data covers the period of 3 January 2022 to 30 June 2022, a total of 124 data to compare the accuracy of the GARCH models in forecasting the price volatilities.

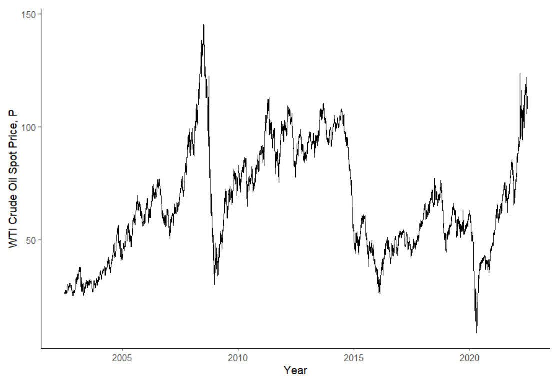

Figure 1: Plot of WTI crude oil spot price.

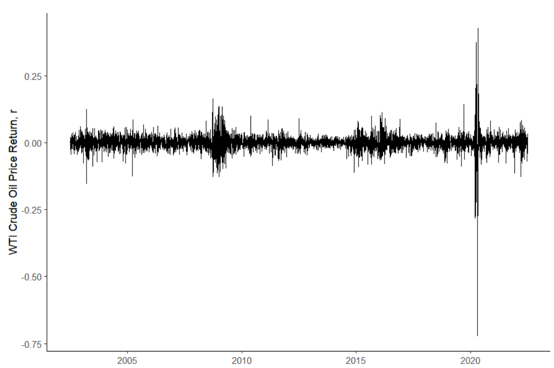

Figure 2. Plot of WTI crude oil price return.

Figure 1 presents the plot of WTI crude oil spot price from 1 July 2002 to 31 December 2021. The figure shows that the WTI crude oil spot prices are highly volatile and exhibit volatility clustering behaviour. Generally, an increase in price will continue for a period of time, and a decrease in price will also last for a period of time. Despite the desirable statistical property of being stationary, the WTI crude oil price return will be used in the study instead of the WTI crude oil spot price, to minimise the fluctuation amplitude. Let P_t be the WTI crude oil spot price at day t, the price return is given by the formula $r_t = \ln(P_t/P_{t-1})$. Figure 2 shows the plot of WTI crude oil price return which also demonstrates volatility clustering characteristics. The GARCH models that will be used in this study are useful to address volatility clustering especially in financial data.

Table 1 shows the descriptive statistics of WTI crude oil price return. The crude oil price return centred at 0.0003 and has a sample standard deviation of 0.0297. The negative value of skewness indicates that the data is not perfectly symmetrical and skewed to the left. On the other hand, since the kurtosis value is greater than three, the data is leptokurtic, which indicates a fatter and longer tail than a normal distribution.

Table 1. WTI crude oil price return descriptive statistics

Statistics	Values
Mean	0.0003
Standard deviation	0.0297
Skewness	–2.2115
Kurtosis	91.6963
ADF test	–16.3920(0.01)
Jarque-Bera test	1764261(<2.2e-16)
Langrange Multiplier test	20410(<2.2e-16)
Ljung-Box test	523.6700 (<2.2e-16)

Note: The value inside the bracket represent for *p*-value

In addition to descriptive statistics, the results of the preliminary analysis are also included in Table 1. The ADF test for stationarity is -16.3920 with p-value of 0.01. Since the p-value is less than $\alpha = 0.05$, reject the null hypothesis and conclude that the series is stationary at 5% level of significance. Due to the stationarity of the data, analysis on this time series data is meaningful and further study can be carried out. Next, the high value of Jarque-Bera (JB) statistics, 1764261 allows for the rejection of the null hypothesis that the data is normally distributed at 5% level of significance. Hence, it can be concluded that the data is not normally distributed which is in line with the high kurtosis and negative skewness of the data.

Besides, the test statistics of the LM test for the ARCH effect is 20410 with p–value $< 2.2 \times 10^{-16}$, also lesser than $\alpha = 0.05$. The null hypothesis that no ARCH effect exists in the residuals was rejected at 5% level of significance. The existence of ARCH effects supports the use of the GARCH models to deal with the non-constant variance in this study. Furthermore, the test statistics of the Ljung-Box test is 523.6700 with p–value $< 2.2 \times 10^{-16}$ which is lesser than $\alpha = 0.05$. Thus, reject the null hypothesis and can be concluded that the model exhibits serial correlation.

GARCH Models

For the purpose of this study, GARCH models known as GARCH, IGARCH, EGARCH, and GJR-GARCH were chosen as the preferred model. The GARCH model is chosen as the model is fundamental choice for modeling volatility. Meanwhile, the IGARCH model is selected because of persistent volatility in the data and the EGARCH model because the volatility responds differently to positive and negative shocks, as shown in Figure 1. The GJR-GARCH model considers both volatility asymmetries and shocks' impact on volatility. Table 2 shows the estimated results for all GARCH models based on the maximum likelihood estimation. The estimated value is presented before the p-value, which is displayed in the bracket. All the parameters are statistically significant at 1% level of significance. Small values of α_0 and highly significant α_1 and β_1 parameters of all GARCH (1, 1) models suggest that the fundamental basis of the present WTI crude oil returns volatility is the past squared errors and the past conditional variance. The overall measurement of the persistence of volatility is represented by the summation of α_1 and β_1. The summation of α_1 and β_1, which is very close to one, suggests that the volatility has a high degree of persistence. The positivity of the conditional variance of the GARCH (1, 1) models is achieved as α_0, α_1, and β_1 parameters are non-negative for all types of distributions. Furthermore, since $\alpha_1 + \beta_1 < 1$ for all types of distribution of the GARCH (1, 1) models, the conditional variance is stationary.

For the IGARCH (1, 1) models, all the parameters are statistically significant at 1% level of significance with the *p*-values ranging from 0.0000 to 0.0011, despite their error distributions. WTI crude oil current volatility depends on the past squared errors, which is evident by the non-zero significant α_1 parameter. However, it is only slightly affected by the constant since the coefficients of the significant α_0 parameter is very close to zero for all types of the IGARCH (1, 1) models. Since the IGARCH model is actually a standard GARCH model but with a unit root, the weightage of the past conditional variances, β_1 can be simply calculated by using the formula $\beta_1 = 1 - \alpha_1$. Moreover, the positive values of α_0 and α_1 for all the IGARCH (1, 1) models with different distributions warrant the positivity of the conditional variance.

Moving on to the next family of GARCH, EGARCH, at 1% level of significance, all parameters of the three EGARCH (1, 1) models are significant. A remarkable phenomenon that can be observed from the EGARCH (1, 1) models is that the estimated values of α_0 and α_1 parameters are negative with the value of -0.1212, -0.1260, and -0.1250 for normal distribution, Student's t-distribution and GED, respectively. Since the EGARCH model is log-linear, the negative values of parameters do not affect the positivity of conditional variance. As the asymmetric parameter, γ_1 is non-zero and significant, WTI crude oil return exhibits asymmetry behaviour. In other words, positive and negative shocks impact the current volatility differently. When there is a positive shock, the effect on the conditional variance is 0.0697, 0.0656 and 0.0652 for the EGARCH (1, 1) model with normal distribution, Student's t-distribution and GED, respectively. In contrast, when there is a negative shock, the effect on the conditional variance is -0.2305, -0.2118 and -0.2184 for the EGARCH (1, 1) model with normal distribution, Student's t-distribution and GED respectively. Hence, it is clearly seen that a positive shock has a greater impact on the conditional variance due to the positive γ_1 parameter in this case.

Table 2. Estimated results for GARCH models

Model	Parameters	Normal	Student's t	GED
GARCH (1,1)	α_0	0.0000*** (0.0000)	0.0000*** (0.0011)	0.0000*** (0.0000)
	α_1	0.0988*** (0.0000)	0.0914*** (0.0000)	0.0934*** (0.0000)
	β_1	0.8840*** (0.0000)	0.8908*** (0.0000)	0.8888*** (0.0000)
	v	-	6.4581*** (0.0000)	1.3503*** (0.0000)
	$\alpha_1 + \beta_1$	0.9828	0.9828	0.9828
IGARCH (1,1)	α_0	0.0000*** (0.0000)	0.0000*** (0.0000)	0.0000*** (0.0000)
	α_1	0.1097*** (0.0000)	0.1040*** (0.0000)	0.1052*** (0.0000)
	v	-	5.8581*** (0.0000)	1.3324*** (0.0000)
EGARCH (1,1)	α_0	-0.1212 ***(0.0000)	-0.1260 ***(0.0000)	-0.1250 ***(0.0000)
	α_1	-0.0804***(0.0000)	-0.0731***(0.0000)	-0.0766***(0.0000)
	β_1	0.9832***(0.0000)	0.9836***(0.0000)	0.9837***(0.0000)
	γ_1	0.1501***(0.0000)	0.1387***(0.0000)	0.1418***(0.0000)
	v	-	6.7194***(0.0000)	1.3695***(0.0000)
	$\alpha_1 + \gamma_1$	0.0697	0.0656	0.0652
	$\alpha_1 - \gamma_1$	-0.2305	-0.2118	-0.2184
GJR-GARCH (1,1)	α_0	0.0000*** (0.0000)	0.0000*** (0.0000)	0.0000*** (0.0000)
	α_1	0.0433*** (0.0000)	0.0428*** (0.0000)	0.0420*** (0.0000)
	β_1	0.8879*** (0.0000)	0.8935*** (0.0000)	0.8924*** (0.0000)
	γ_1	0.0978*** (0.0000)	0.0875*** (0.0000)	0.0915*** (0.0000)
	v	-	6.7235*** (0.0000)	1.3699*** (0.0000)
	$\alpha_1 + \gamma_1$	0.1411	0.1303	0.1335

Note: The value inside the bracket represent for *p*-value
***represent at 1% level of significance

Apart from the EGARCH model, the GJR-GARCH model can also captures asymmetry effects in the series. All the parameters are significant at 1% level of significance for the GJR-GARCH (1, 1) models regardless of their error distributions. The non-zero and positive asymmetric response parameter, γ_1 indicates that WTI crude oil return is asymmetric and a negative shock will result in a larger impact on conditional variance than a positive shock. This can be shown with the effect of positive shock on the conditional variance is 0.0433, 0.0428 and 0.0420 while the effect of negative shock on the conditional variance is 0.1411, 0.1303 and 0.1335 for GJR-GARCH (1, 1) model with normal distribution, Student's t-distribution and GED respectively. Generally, the significant γ_1 parameters in both EGARCH (1, 1) and GJR-GARCH (1, 1) models have also proven that these models perform better than the standard GARCH (1, 1) model which do not treat positive and negative shocks differently. On top of that, the positivity of the conditional variance of the GJR- GARCH (1, 1) models are achieved since $\alpha_0, \alpha_1, \beta_1$ parameters and the summation of α_1 and γ_1 are positive for all types of distributions.

Model Performance

Table 3 summarises the results of model selection criteria which are the AIC and BIC for all combinations of GARCH models. Ranking within the GARCH family models is also included in the table. Generally, a smaller value of both AIC and BIC is preferable since it indicates a better-fitted model. Based on the results, it can be seen that the choice of the best model based on the two different criteria is compromised. On top of that, the ranking reveals that the Student's t-distribution has the lowest AIC and BIC values, followed by the GED and, lastly, the normal distribution for all GARCH family models. These results are in accordance with the non-normality and heavy tail characteristics exhibited in the data. Thus, this study emphasises the significance of understanding the data distribution before modeling. Assuming the default normal distribution may not always align with the real data.

From the group of models shown in Table 3, the EGARCH model with Student's t-distribution was able to fit the data better with AIC, -4.8892 and BIC, -4.8826. In addition to model selection criteria, the forecast accuracies from all models were also identified. A total of 124 data were used to compare the accuracy of the GARCH models in forecasting the price volatilities. Table 4 summarises the results of measures of forecast accuracy for all combinations of GARCH models based on the out-sample data. Ranking within the GARCH family models is also included in the table. Models with smaller absolute error, RMSE and MAE indicate better forecast accuracy. Generally, all three criteria are similar in terms of choosing the most accurate model. For the GARCH (1, 1), EGARCH (1, 1) and GJR-GARCH (1, 1) models, Student's t-distribution has the smallest absolute error, RMSE and MAE followed by the GED and the normal distribution. On the other hand, the IGARCH (1, 1) model with GED has the smallest absolute error, RMSE and MAE, followed by the Student's t-distribution and the normal distribution. However, a similar result was found in the accuracy measure. EGARCH (1, 1) with Student's t distribution performed the lowest error indicating the best-fitted model in forecasting with the absolute error of 2.6253, RMSE of 0.0213 and MAE of 0.0212.

Table 3. Model selection criteria for GARCH models

Model	Distribution	AIC	BIC	Rank
GARCH (1,1)	Normal	-4.8093	-4.8053	3
	Student's t	-4.8746	-4.8693	1
	GED	-4.8619	-4.8566	2
IGARCH (1,1)	Normal	-4.8071	-4.8044	3
	Student's t	-4.8730	-4.8690	1
	GED	-4.8602	-4.8562	2
EGARCH (1,1)	Normal	-4.8263	-4.8209	3
	Student's t	-4.8892	-4.8826	1
	GED	-4.8757	-4.8690	2
GJR-GARCH (1,1)	Normal	-4.8212	-4.8159	3
	Student's t	-4.8814	-4.8748	1
	GED	-4.8692	-4.8626	2

Table 4. Measures of forecast accuracy for GARCH models

Model	Distribution	Absolute Error	RMSE	MAE	Rank
GARCH (1,1)	Normal	2.9444	0.0238	0.0237	3
	Student's t	2.8358	0.0230	0.0229	1
	GED	2.8509	0.0231	0.0230	2
IGARCH (1,1)	Normal	3.7509	0.0306	0.0302	3
	Student's t	3.7366	0.0305	0.0301	2
	GED	3.6769	0.0300	0.0297	1
EGARCH (1,1)	Normal	3.0073	0.0243	0.0243	3
	Student's t	2.6253	0.0213	0.0212	1
	GED	2.6318	0.0213	0.0212	2
GJR-GARCH (1,1)	Normal	2.8904	0.0234	0.0233	3
	Student's t	2.8057	0.0227	0.0226	1
	GED	2.8236	0.0229	0.0228	2

Post-estimation Test

The post-estimation test results for all combinations of GARCH models shown in Table 5 for LM test and Ljung-Box test. Different time lags have been tested to more thoroughly assess and diagnose the fitted models' adequacy. Values presented in the table are p-values at different lag orders, Lag 3, Lag 5, and Lag 7. The non-significance of the LM test with all p-values greater than $\alpha = 0.05$ at every lag of each GARCH model fails to reject the null hypothesis that no ARCH effect exists in the residuals. This implies that all of the models are sufficient. Furthermore, the Ljung-Box test also comes with a similar result: to accept the null hypothesis that the model does not exhibit serial correlation at 5% significance level. These results were shown in the table with all the p-values of each GARCH model at every lag greater than $\alpha = 0.05$. Hence, it can be concluded that all the GARCH models are adequate in modelling WTI crude oil price return and higher-order of GARCH models are not required.

Table 5. Post-estimation test results for GARCH model

	GARCH (1,1)			
	Lag	Normal	Student's t	GED
LM test	Lag[3]	0.1923	0.2212	0.2110
	Lag[5]	0.3596	0.4148	0.3961
	Lag[7]	0.5782	0.6443	0.6233
Ljung-Box test	Lag[1]	0.2805	0.2788	0.2801
	Lag[2*(p+q)+(p+q)-1][2]	0.3732	0.3687	0.3710
	Lag[4*(p+q)+(p+q)-1][5]	0.4946	0.4922	0.4936
	IGARCH (1,1)			
	Lag	Normal	Student's t	GED
LM test	Lag[3]	0.1202	0.1395	0.1281
	Lag[5]	0.2228	0.2693	0.2449
	Lag[7]	0.3667	0.4464	0.4052
Ljung-Box test	Lag[1]	0.3806	0.3768	0.3889
	Lag[2*(p+q)+(p+q)-1][2]	0.4701	0.4627	0.4743
	Lag[4*(p+q)+(p+q)-1][5]	0.5646	0.5613	0.5685
	EGARCH (1,1)			
	Lag	Normal	Student's t	GED
LM test	Lag[3]	0.2269	0.2760	0.2583
	Lag[5]	0.5434	0.6058	0.5855
	Lag[7]	0.7629	0.8190	0.8041
Ljung-Box test	Lag[1]	0.2332	0.2113	0.2194
	Lag[2*(p+q)+(p+q)-1][2]	0.3745	0.3415	0.3540
	Lag[4*(p+q)+(p+q)-1][5]	0.4879	0.4621	0.4718
	GJR-GARCH (1,1)			
	Lag	Normal	Student's t	GED
LM test	Lag[3]	0.1445	0.1586	0.1543
	Lag[5]	0.2835	0.3163	0.3071
	Lag[7]	0.3818	0.4376	0.4200
Ljung-Box test	Lag[1]	0.2154	0.2225	0.2206
	Lag[2*(p+q)+(p+q)-1][2]	0.3306	0.3387	0.3371
	Lag[4*(p+q)+(p+q)-1][5]	0.4614	0.4702	0.4682

Final Model Selection

Given that all the GARCH models are statistically significant based on the hypothesis testing and effective based on the post-estimation test results, all models are suitable for the crude oil price return. However, the models show differences in performance based on the model selection criteria and measures of forecast accuracy. Table 6 below presents the overall ranking of GARCH models based on the two indicators.

Table 6. Overall ranking based on model selection criteria and measures of forecast accuracy.

Rank	Model			
	Model Selection Criteria		Measures of Forecast Accuracy	
1	EGARCH (1, 1)	Student's t	EGARCH (1, 1)	Student's t
2	GJR-GARCH (1, 1)	Student's t	GJR-GARCH (1, 1)	Student's t
3	GARCH (1, 1)	Student's t	GARCH (1, 1)	Student's t
4	IGARCH (1, 1)	Student's t	IGARCH (1, 1)	GED

Across the GARCH family models, the best fitted model based on the model selection criteria is the EGARCH (1, 1) model with Student's t-distribution followed by the GJR-GARCH (1, 1) model with Student's t-distribution, the standard GARCH (1, 1) model with Student's t-distribution and lastly the IGARCH (1, 1) model with Student's t-distribution. On the other hand, based on the analysis of the measures of forecast accuracy across the GARCH family models, it can be concluded the most accurate model is the EGARCH (1, 1) model with Student's t-distribution followed by the GJR-GARCH (1, 1) with Student's t-distribution, the standard GARCH (1, 1) model with Student's t-distribution and lastly the IGARCH (1, 1) model with GED. Hence, it can be deduced that the EGARCH (1, 1) model outperformed the other three GARCH family models, followed by the GJR-GARCH (1, 1) model, the standard GARCH (1, 1) model and lastly the IGARCH (1, 1) model.

Conclusions

This study demonstrated experimental research on the application of nonlinear time series models. GARCH model is a nonlinear time series models that is well known with its ability to capture volatilities in data. Hence, the standard GARCH model and three other extended GARCH models which are the IGARCH model, the EGARCH model and the GJR- GARCH model were used in this study. Apart from normal distribution, Student's t-distribution and GED were also considered in this study. Data set used in this study is WTI crude oil price data. Instead of using the daily crude oil spot price, the analysis utilised the crude oil price return to reduce the amplitude of fluctuation. The data is then fitted into the 12 combinations of four types of GARCH models, each with three error distributions.

All the parameters in 12 combinations are statistically significant at 1% level of significance. Hence, the estimated parameters of each combination were used to forecast the price return for 124 days ahead. The performance of the GARCH models were compared by using model selection criteria and also the forecast accuracy measures. Results showed that the EGARCH (1, 1) model with Student's t-distribution has the lowest AIC and BIC values as well as the lowest absolute error, RMSE and MAE. Since the post-estimation results are also satisfactory, it can be concluded the EGARCH (1, 1) model with Student's t-distribution outperformed the other 11 combinations. In conclusion, the EGARCH (1, 1) model with Student's t-distribution performs the best in both modelling WTI crude oil price volatility and the simulation study. The result is reasonable since the EGARCH model takes asymmetric effect into account and the data exhibits fat tail characteristics.

The area of study that can be done with nonlinear time series models, specifically the GARCH models is infinite. For this study, only four of the univariate GARCH family models were considered. Hence, in the future, bivariate or even multivariate GARCH family models can be included to compare their efficiency. In future, comparison of the performance of GARCH models across different data sets can be conducted.

Conflicts of Interest

The authors declares that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

We would like to express our gratitude to the editor and referees for their valuable comments.

References

[1] R.S. Tsay, R. Chen. (2018). *Nonlinear Time Series Analysis*. Wiley, Hoboken.

[2] R. Bhowmik, S. Wang. (2020). Stock market volatility and return analysis: A systematic literature review. *Entropy, 22*, 522.

[3] N. Hung, N. Thach, A. Le. (2018). GARCH models in forecasting the volatility of the World's oil prices. *Stud. Comp. Intell.,* 673-683.

[4] R. F. Engle. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, *50*, 987-1007.

[5] T. Bollerslev. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econom., 31*, 307-327.

[6] R.F. Engle, T. Bollerslev, Modelling the Persistence of Conditional Variances, Econom. Rev. 5 (1986) 1-50.

[7] D. B. Nelson. (1991). Conditional heteroskedasticity in asset returns: A new approach, *Econometrica, 59*, 347-370.

[8] V. Naimy, O. Haddad, G. Fernández-Avilés, R. El Khoury. (2021). The predictive capacity of GARCH-type models in measuring the volatility of crypto and world currencies. *PLoS One, 16.*

[9] D. B. Nugroho, D. Kurniawati, L. P. Panjaitan, Z. Kholil, B. Susanto, L. R. Sasongko. (2019). Empirical performance of GARCH, GARCH-M, GJR-GARCH and log-GARCH models for returns volatility. *J. Phys. Conf. Ser.*, *1307,* 012003.

[10] C. Dritsaki. (2017). An empirical evaluation in GARCH volatility modeling: Evidence from the Stockholm stock exchange. *J. Math. Financ., 7*, 366-390.

[11] B. Meher, I. Hawaldar, L. Mohapatra, A. Sarea. (2020). The impact of COVID-19 on price volatility of crude oil and natural gas listed on multi commodity exchange of India. *Int. J. Energy Econ. Policy, 10*, 422-431.

[12] A. M. Herrera, L. Hu, D. Pastor. (2018). Forecasting crude oil price volatility. Int. J. *Forecast., 34*, 622-635.

[13] N. Hung, N. Thach, A. Le. (2018). GARCH models in forecasting the volatility of the world's oil prices. *Stud. Comp. Intell.,* 673-683.

[14] B. Venkatareddy, K. G. (2020). Rainfall indexation for evaluating rainfall risk profile of Indian subcontinent. *Indian. J. Finance, 14*, 40-52.

[15] N. Zamrus, M. H. Mohd Rodzhan, N. N. Mohamad. (2022). Forecasting model of air pollution index using generalized autoregressive conditional heteroskedasticity family (GARCH). *Mal. J. Fund. Appl. Sci., 18*, 184-196.

[16] B. Wellington Garikai. (2019). Stock market volatility analysis using GARCH family models: Evidence from Zimbabwe stock exchange. MPRA Paper No. 94201.

[17] G. M. Ljung, G. E. P. Box. (1978)z. On a measure of lack of fit in time series models. *Biometrika*, *65*, 297-303.

[18] L. Schmidt. (2021). Volatility forecasting performance of GARCH models: A study on nordic indices during COVID-19. Master Thesis, Umeå University.

[19] L. R. Glosten, R. Jagannathan, D. E. Runkle. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *J. Finance., 48,* 1779-1801.

[20] H. Pham. (2019). A new criterion for model selection. *Mathematics*, *7*, 1215.

[21] F. Fabozzi, S. Focardi, T. Rachev, B. Arshanapalli. (2014). Appendix E: Model selection criterion: AIC and BIC, in the basics of financial econometrics: Tools, concepts, and asset management applications, 399-403.

[22] H. Akaike. (1973). Information theory and an extension of the maximum likelihood principle. Selected Papers of Hirotugu Akaike. 199-213.

[23] G. Schwarz. (1978). Estimating the dimension of a model. *Ann. Stat., 6*, 461-464.

[24] J. J. Dziak, D. L. Coffman, S. T. Lanza, R. Li, L. S. Jermiin. (2020). Sensitivity and specificity of information criteria. *Brief Bioinform.,* 21553-565.

[25] N. Mehdiyev, D. Enke, P. Fettke, P. Loos. (2016). Evaluating forecasting methods by considering different accuracy measures. *Procedia Comput. Sci., 95*, 264-271.

[26] D. Koutsandreas, E. Spiliotis, F. Petropoulos, V. Assimakopoulos. (2021). On the selection of forecasting accuracy measures. *J. Oper. Res. Soc., 73*, 1-18.