International Journal of Mathematical Analysis and
Optimization: Theory and Applications
Vol. 2020, No. 1, pp. 689 - 706

Some Forecast Asymmetric GARCH Models for Distributions with Heavy Tails

J. N. Onyeka-Ubaka[1*], U. J. Anene [2]

1*, 2, Department of Mathematics, University of Lagos, Nigeria.
*Corresponding author: jonyeka-ubaka@unilag.edu.ng

Article Info
Received: 05 April 2019 Revised: 25 April 2020
Accepted: 26 April 2020 Available online: 18 May 2020

Abstract

Crude oil prices are influenced by a number of factors that are far beyond the traditional supply and demand dynamics such as West Texas Intermediate (WTI), Brent and Dubai. The high frequency crude oil data exhibit non-constant variance. This paper models and forecasts the exhibited fluctuations via asymmetric GARCH models with the three commonly used error distributions: Student's t distribution, normal distribution and generalized error distribution (GED). The Maximum Likelihood Estimation (MLE) approach is used in the estimation of the asymmetric GARCH family models. The analysis shows that volatility estimates given by the exponential generalized autoregressive conditional heteroskedasticity (EGARCH) model exhibit generally lower forecast errors in returns of WTI oil spot price while the asymmetric power autoregressive conditional heteroskedasticity (APARCH) model exhibits lower forecast errors in returns of Brent oil spot price, therefore they are more accurate than the estimates given by the other asymmetric GARCH models in each returns. The results obtained from the volatility forecasts seem to be useful to oil future traders and policy makers who need to perceive "apriori" the effects of news on return volatilities before executing their trading, investments and political strategies for the economic wellbeing of the country.

Keywords and Phrases Stylized facts, Asymmetric GARCH, Oil price volatility, Volatility estimate, Error distribution
MSC2010: 62M10, 91B79.

1 Introduction

Crude oil is the raw oil from which petroleum products such as gasoline are derived. In recent years, crude oil price has become one of the major economic challenges facing most countries in the world especially those in Africa including Nigeria. World crude oil prices are established in accordance with three market traded benchmarks, namely: West Texas Intermediate (WTI), Brent and Dubai, and are quoted at premiums or discounts to these benchmark prices [1]. WTI is a light sweet crude oil and actually lighter than Brent Crude oil. WTI contains about 0.24% sulphur and

International Journal of Mathematical Analysis and
Optimization: Theory and Applications
Vol. 2020, No. 1, pp. 689 - 706

are considered the world's most actively traded energy products. Its properties and production site make it ideal for being refined in the United States and in the Gulf Coast regions. They are preferred because of low sulphur content and its relatively high energy efficiency of high-quality products such as gasoline, diesel fuel, heating oil and jet fuel. The price of a barrel (159 litres) of oil depends on both its grade and location. Grade is determined by its specific gravity and its sulphur content. The Energy Information Administration (EIA) uses the imported refinery acquisition cost, the weighted average cost of all oil imported into the US, as its world oil price [1].

Crude oil prices just like financial markets exhibit empirical characteristics: (i) leptokurtic, i.e. they have heavy tails and a higher peak than a normal distribution; (ii) equity returns are typically negatively skewed and (iii) squared return series shows significant autocorrelation, i.e. volatilities tends to cluster. Considering the first two facts, it is important to examine which probability density function capture heavy tails and asymmetry best. The third fact requires to correctly specify conditional mean and conditional variance equations from GARCH family models. We equally know crude oil prices measured over short time intervals, i.e. daily or weekly, are characterized by high kurtosis and kurtosis is both a measure of peakedness and fat tails of the distribution. Modelling and forecasting the volatility (conditional variance) of these non-constant empirical regularities are challenging tasks. The emerging financial markets are becoming more sophisticated and also, the need for accurate volatility forecasting and estimation are becoming increasingly more important [2, 3]. Volatility forecasts are important for many financial decisions such as the issues for policy makers, option traders and investors to effectively price, speculate, and hedge before executing their trading and investments strategies. The policy makers require the crude oil price dynamics for the economic and political growth of the country. The deep understanding of the results produced by GARCH and Asymmetric GARCH models are needed for out-of-sample forecast in the oil market. Thus, in this paper, a family of generalized autoregressive conditional heteroskedasticity (GARCH) models: exponential GARCH (EGARCH), Threshold GARCH (TGARCH), Glosten, Jogannathan Ronkle GARCH (GJRGARCH), asymmetric power GARCH (APARCH) and Non-linear Asymmetric GARCH (NAGARCH) which use past variances and past variance forecasts to forecast future variances are employed to capture volatility clustering and predict the periods of fluctuations in the future.

Several scholars have worked on how oil price shocks affect economic growth in a body of research. Examples include [4, 5] and [6] who clearly demonstrated that positive oil price shocks induce a slow-down in aggregate measures of growth or employment and that negative oil price shocks lead to an increase in aggregate measures of growth or unemployment. [7, 8, 9] and [10] focused on how shocks to volatility of crude oil prices affect future oil price and found out that negative and positive news have a different impact on oil price volatility. [11, 12], in their works investigated the linear and nonlinear causal linkages between spot and futures prices of different maturities of West Texas Intermediated (WTI) crude oil and observed that spot and future prices (oil shocks) exhibit asymmetric GARCH effects and significant higher order conditional moments. Also, [13] in analysing oil prices examine 43 stock markets and find that the volatility of oil prices has a negative impact on international stock market returns. In many cases, the first-order GARCH family models have been extensively proven to be appropriate for modelling and forecasting financial time series as observed by [14, 15]. But in these studies little or no attention has been given to their suitable error distributions, hence, the paper focuses on modelling and forecasting asymmetric GARCH models using distributions with heavy tails.

(a) GARCH (p, q) Model [16]

In order to model in a parsimonious way the conditional heteroskedasticity, [16] proposed the generalized ARCH model, i.e GARCH(p, q):

$$\sigma_t^2 = \omega + \alpha(L)\epsilon_t^2 + \beta(L)\sigma_t^2 \tag{1.1}$$

International Journal of Mathematical Analysis and Optimization: Theory and Applications
Vol. 2020, No. 1, pp. 689 - 706

where $\alpha(L) = \alpha_1 L + \alpha_2 L^2 + ... + \alpha_q L^q$, $\beta(L) = \beta_1 L + \beta_2 L^2 + ... + \beta_p L^p$. Recall that $\epsilon_t^2 = z_t^2 \sigma_t^2$ and z_t^2 is independent and identically distributed with standard normal. And also, $\epsilon_t \sim N(0, \sigma_t^2)$, and the conditional function of ϵ_t is given by

$$f(\epsilon_t : v) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(\frac{(\epsilon_t)^2}{2\sigma_t^2}\right) \tag{1.2}$$

where σ_t^2 is the conditional variance, and also the function of the unknown parameters. An effective GARCH $(1,1)$ [17], the conditional variance is modelled as

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \tag{1.3}$$

When $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, then ϵ_t is stationary if and only if $\alpha + \beta < 1$. The non-negativity constraints on ω, α and β guarantee the positivity of σ_t^2. If the condition that $\alpha + \beta < 1$ is fulfilled, then the unconditional variance in GARCH $(1,1)$ is obtained as:

$$\begin{aligned}
\sigma^2 = E[\sigma_{t-1}^2] &= E[\omega + \alpha\epsilon_t^2 + \beta\sigma_t^2] \\
&= \omega + \alpha E[\epsilon_t^2] + \beta E[\sigma_t^2] \\
&= \omega + \alpha + \beta\sigma^2 \\
&= \omega(1 - \alpha - \beta)^{-1}
\end{aligned}$$

The h-step ahead forecasts from GARCH $(1,1)$ are obtained recursively as [17]:

$$\begin{aligned}
\sigma_{t+h}^2 &= \omega + (\alpha + \beta)\sigma_{t+h-1}^2 \\
&= \sigma^2 + (\alpha + \beta)(\sigma_{t+h-1}^2 - \sigma^2) \\
&= \sigma^2 + (\alpha + \beta)^{h-1}(\sigma_{t+h-1}^2 - \sigma^2)
\end{aligned}$$

As $h \to \infty$, it is clear that the volatility forecast approaches the unconditional variance σ^2 and $\alpha + \beta$ dictates the speed of the mean reversion just like our GRACH model above. GARCH models assume that only the magnitude and not the positivity or negativity of anticipated excess returns determines the conditional variance. Importantly, GARCH models cannot explain the observed covariance between ϵ_t^2 and ϵ_{t-j}^2 as it can only be achieved if the conditional variance is expressed as an asymmetric function of ϵ_{t-j}^2.

(b) Asymmetric GARCH Models

(i) The EGARCH Model [18]
The exponential GARCH (EGARCH) model was proposed by [18] with the aim to capture the asymmetries in the volatility changes noted by [19]. The EGARCH model is generally specified as:

$$\ln(\sigma_t^2) = \omega + \sum_{i-1}^{p} \beta_j \ln(\sigma_{t-i}^2) + \sum_{i-1}^{q} \alpha_i \left[\phi z_{t-i} + \gamma\left(|z_{t-i}| - E|z_{t-i}|\right)\right] \tag{1.4}$$

The left hand side is the log of the variance series. This makes the leverage effect exponential and therefore the parameters ω, β_j, α_i are not restricted to be non negative, $\alpha_1 = 1$, $E|z_t| = (\frac{2}{\pi})^{\frac{1}{2}}$ when $z_t \sim NID(0,1)$. Let define $g(z_t) = \phi z_t + \gamma\left[|z_t| - E|z_t|\right]$ by construction $\{g(z_t)\}_{t=-\infty}^{\infty}$ is a zero-mean, *i.i.d.* random sequence. The components of $g(z_t)$ are ϕz_t and $\gamma(|z_t| - E|z_t|)$, each with mean zero. If the distribution of z_t is symmetric, the components are orthogonal, but not independent. Over the range $0 < z_t < \infty$, $g(z_t)$ is linear in z_t with slope $\phi + \psi$,, and over the range $-\infty < z_t < 0$,

$g(z_t)$ is linear with slope $\phi - \gamma$. The term $\gamma(|z_t| - E|z_t|)$ represents a magnitude effect. If $\gamma > 0$ and $\phi = 0$, the innovation in $\ln \sigma^2_{t+1}$ is positive (negative) when the magnitude of z_t is larger (smaller) than its expected value. If $\gamma = 0$ and $\phi < 0$, the innovation in conditional variance is now positive (negative) when returns innovations are negative (positive). A negative shock to the returns which would increase the debt to equity ratio and therefore increase uncertainty of future returns could be accounted for when $\alpha_i > 0$ and $\phi < 0$ [20].
The EGARCH $(1, 1)$ variance equation with a normal distribution is stated by [21] as:

$$\ln \sigma^2_t = \omega + \beta \ln(\sigma^2_{t-1}) + \gamma \frac{\epsilon_{t-1}}{\sqrt{\sigma^2_{t-1}}} + \alpha \left[\frac{\epsilon_{t-1}}{\sqrt{\sigma^2_{t-1}}} - \sqrt{\frac{2}{\pi}} \right] \tag{1.5}$$

where ω is the intercept for the variance; β is the coefficient for the logged GARCH term; α is the magnitude; $\ln(\sigma^2_{t-1})$ is the logged GARCH term; γ is the scale of the asymmetric leverage parameter that quantifies the degree of the volatility leverage effect in the model if $\gamma \neq 0$. $\left[\frac{\epsilon_{t-1}}{\sqrt{\sigma^2_{t-1}}} - \sqrt{\frac{2}{\pi}} \right]$ is the parameter that takes into account the absolute value of the last period's volatility shock. It replaces the regular ARCH term; $\frac{\epsilon_{t-1}}{\sqrt{\sigma^2_{t-1}}}$ is the last period's shock which is standardized. Given that the model uses the log of the variance, this means that even if the parameters are negative, the variance will still be positive. Thus, the model is subjected to the non-negativity constraints [18, 21].

(ii) The Glosten - Jagannathan - Runkle model [22]
The GJR-GARCH model is designed in a way that allows the model to account for the potential larger impact of negative shocks on return volatility. It models the variance directly and does not use the natural logarithm as in the EGARCH model. The GJR-GARCH model is stated as:

$$\sigma^2_t = \omega + \sum_{i=1}^{p} \beta_i \sigma^2_{t-i} + \sum_{i=1}^{q} \left(\alpha_i \epsilon^2_{t-i} + \gamma_i S^-_{t-i} \epsilon^2_{t-i} \right) \tag{1.6}$$

where

$$S^-_t = \begin{cases} 1, & \text{if} \quad \epsilon_t < 0 \\ 0, & \text{if} \quad \epsilon_t \geq 0. \end{cases}$$

The conditional variance in the GJR-GARCH (1, 1) is given as:

$$\sigma^2_t = \omega + \alpha_1 \epsilon^2_{t-1} + \beta \sigma^2_{t-1} + \alpha_2 I_{t-1} \epsilon^2_{t-1} \tag{1.7}$$

where σ^2_t is the conditional forecasted variance; ω is the intercept for the variance; $\alpha_1 \epsilon^2_t$ is the variance that depends on the previous lag error terms; α_2 is the scale of the asymmetric volatility; β is the coefficient for yesterday's forecasted variance; α^2_{t-1} yesterday's forecasted variance; I_{t-1} is a dummy variable that is only activated if the previous shock is negative ($\epsilon_{t-1} < 0$), allowing the GJR-GARCH to take leverage effect into consideration [22]. The sign of leverage effect is opposite compared to that of EGARCH model. If $\alpha_2 = 0$, no asymmetric volatility; if $\alpha_2 > 0$, negative shocks will increase the volatility more than the positive shocks and if $\alpha_2 < 0$, positive shocks increase the volatility more than negative shocks.

International Journal of Mathematical Analysis and Optimization: Theory and Applications
Vol. 2020, No. 1, pp. 689 - 706

(iii) Threshold GARCH (TGARCH) Model [23]
The threshold GARCH (TGARCH) is similar to the GJR GARCH model except that the standard deviation instead of the variance is included in the specification. TGARCH is modelled as:

$$\sigma_t = \omega + \sum_{i=1}^{q} \alpha_i \left(|\epsilon_{t-i}| + \gamma_i I_{t-i}\epsilon_{t-i}\right) + \sum_{i=1}^{p} \beta_j \sigma_{t-j} \tag{1.8}$$

where $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, $-1 < \gamma_i < 1$ and

$$I_{t-i} = \begin{cases} 1, & \text{if} \quad \epsilon_{t-i} < 0 \\ 0, & \text{if} \quad \epsilon_{t-i} \geq 0. \end{cases} \tag{1.9}$$

(iv) The Asymmetric Power ARCH [24]
The asymmetric power ARCH (APARCH) model introduced by [24] aims to reproduce both the leverage effects and the sample autocorrelation of absolute squared returns. APARCH is modelled as:

$$\begin{aligned} r_t &= \mu + \epsilon_t \\ \epsilon_t &= \sigma_t z_t \quad z_t \sim N(0,1) \\ \sigma_t^\delta &= \omega + \sum_{i=1}^{q} \alpha_i \left(|r_{t-i}| + \gamma_i \epsilon_{t-i}\right)^\delta + \sum_{i=1}^{p} \beta_j \sigma_{t-j}^\delta \end{aligned}$$

where $\omega > 0$; $\delta \geq 0$; $\alpha_i \geq 0$ $i = 1, ..., q$; $-1 < \gamma_i < 1$ $i = 1, ..., q$; $\beta_j \geq 0$ $j = 1, ..., p$.
This model nests at least seven ARCH type models, according to the estimated parameters. In model (1.12) parameter δ plays the role of a Box-Cox transformation of the time-varying conditional standard deviation σ_t, while γ_i reflects leverage effect, i.e. asymmetric information influence. The parameter, β, in the mean conditional equation, determines volatility of stock, i.e. stock risk. If $\beta > 1$, the stock is highly risked. If $0 < \beta < 1$, the stock is lowly risked, while free-risk stock assumes that $\beta = 0$. The Box-Cox transformation for a positive random variable γ_t:

$$\gamma_t^{(\lambda)} = \begin{cases} \dfrac{\gamma_t^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log \gamma_t & \lambda = 0. \end{cases} \tag{1.10}$$

The asymmetric response of volatility to positive and negative "shocks" is the well-known leverage effect. This generalized version of APARCH model includes the below models as special cases: ARCH(q) model, just let $\delta = 2$ and $\gamma_i = 0$, $i = 1, ..., q$, $\beta_j = 0$, $j = 1, ..., p$; GARCH(p, q) model just let $\delta = 2$ and $\gamma_i = 0$, $i = 1, ..., q$; Taylor/Schwert's GARCH in standard deviation model just let $\delta = 1$ and $\gamma_i = 0$, $i = 1, ..., q$; and GJR model just let $\delta = 2$.

(v) Non-linear Asymmetric GARCH Model [24]
NAGARCH model was proposed by [24]. NAGARCH $(1, 1)$ is defined by

$$\sigma_t^2 = \omega + \alpha\sigma_{t-1}^2(\epsilon_{t-1} - \gamma)^2 + (\sigma_{t-1}^2) \tag{1.11}$$

σ_t^2 is the conditional volatility; ω is the intercept for the variance; α is the coefficient for the ARCH term; α_{t-1}^2 is lagged conditional volatility and ϵ_{t-1}^2 is the residuals (measures the shock) of the time series.

International Journal of Mathematical Analysis and Optimization: Theory and Applications
Vol. 2020, No. 1, pp. 689 - 706

2 Methodology

Price-influencing events may be normally distributed, but the likelihood of said events being reported in news increases with the magnitude of the impact of the event [25]. In this paper, we use the maximum likelihood estimation (MLE) approach to estimate unknown parameters that define a probability density function (pdf): location parameter (mean), scale parameter (variance or standard deviation) and shape parameter (skewness and/or kurtosis) for all the forms of the GARCH models using the Gaussian (Normal) and the non-Gaussian (Student's t and Generalized Error Distribution) distributions to allow the model fit both the tails and the central part of the conditional distribution present in high frequency energy time series data.

Since the paper is dealing with dependent variables, we use a juxtaposition of calibration and maximum likelihood via the Bayes' rule to get the joint distribution as

$$f(y_1, y_2, ..., y_t : \theta) = f(y_1|X_1 : \theta)f(y_2|y_1X_1, X_2 : \theta)...f(y_t|y_{t-1}...y_1X_{t-1}...X_1 : \theta) \tag{2.1}$$

Taking natural log of equation, we have

$$\begin{aligned}\ln f(y_1, y_2, ..., y_t : \theta) = \ln f(y_1|X_1 : \theta) + \ln f(y_2|y_1X_1, X_2 : \theta)\\ +... + \ln f(y_t|y_{t-1}...y_1X_{t-1}...X_1 : \theta)\end{aligned}$$

This likelihood function provides a systematic way to adjust the parameters ω, α_1, β_1 to give the best fit. The conditional likelihood function is given as

$$L(y_1, ..., y_n|\omega) = \prod_{t=1}^{n} g(y_t, \mu_t(\alpha), \sigma_t(\omega)) \tag{2.2}$$

where $g(y_t, \mu_t(\alpha), \sigma_t(\omega))$ denotes the conditional density function for the random variables y_t with mean μ_t and standard deviation σ_t and $\omega = (\alpha, \theta)$ is the parameter vector to be estimated; α corresponds to the set of parameter in the conditional mean assumed, in what follows to be an ARMA (k, l) model and $\theta = (\alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p)$.

The numerical method used to estimate the parameters of the models with a larger order is the score algorithm. In this case the first and second partial derivatives of the likelihood must be calculated. Assuming the conditional density and the $\mu_t(\theta)$ and $\sigma_t^2(\theta)$ functions are differentiable for all $\psi \in \Theta \times H \equiv \Psi$, the maximum likelihood estimator is the solution to

$$S_T(y_T, y_{T-1}, ..., y_1; \psi) \equiv \sum_{t=1}^{T} s_t(y_t; \psi) \tag{2.3}$$

where $s_t \equiv \dfrac{\partial l_t(y_t; \psi)}{\partial \psi}$ is the score vector for the t^{th} observation and $\epsilon_t(\theta) \equiv y_t - \mu_t(\theta)$.

(i) Normal distribution
The normal distribution is uniquely determined by its first two moments [26]. Hence, only the conditional mean and variance parameters enter the log-likelihood function

$$L(\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\left\{\log\left(\sigma_t^2 + \frac{\epsilon_t^2}{\sigma_t^2}\right)\right\} \tag{2.4}$$

where n is the sample size. To obtain an analytical or numerical solution of the MLE, we need to know the first-order derivative of $\dfrac{\partial L(\theta)}{\partial \theta}$ with respect to θ and solve $\dfrac{\partial L(\theta)}{\partial \theta} = 0$. Assuming

International Journal of Mathematical Analysis and Optimization: Theory and Applications
Vol. 2020, No. 1, pp. 689 - 706

$\psi = 0$, we have the score functions as

$$\frac{\partial L(\theta)}{\partial \theta} = \sum_{t=1}^{n} \frac{\epsilon_t}{\sigma_t^2}\frac{\partial \mu_t}{\partial \theta} + \frac{1}{2}\sum_{t=1}^{n}\frac{1}{\sigma_t^2}\left(\frac{\epsilon_t^2}{\sigma_t^2} - 1\right)\frac{\partial \sigma_t^2}{\partial \theta} \tag{2.5}$$

The Hessian matrix is given by

$$\begin{aligned}\frac{\partial^2 L(\theta)}{\partial\theta\partial\theta'} &= -\sum_{t=1}^{n}\frac{1}{\sigma_t^2}\frac{\partial \mu_t}{\partial \theta}\frac{\partial \mu_t}{\partial \theta'} - \sum_{t=1}^{n}\frac{\epsilon_t}{\sigma_t^4}\frac{\partial \mu_t}{\partial \theta}\frac{\partial \sigma_t^2}{\partial \theta'} - \sum_{t=1}^{n}\frac{\epsilon_t}{\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \theta}\frac{\partial \mu_t}{\partial \theta'}\\ &+\sum_{t=1}^{n}\frac{1}{\sigma_t^4}\left(\frac{1}{2} - \frac{\epsilon_t^2}{\sigma_t^2}\right)\frac{\partial \sigma_t^2}{\partial \theta}\frac{\partial \sigma_t^2}{\partial \theta'}\end{aligned}$$

Now, if the paper assumes that the innovations $(\epsilon_t)_{t\in Z}$ have conditional non-Gaussian distributions:

(ii) The Student-t distribution

$$f\left[z_t(\theta);\eta\right] = (2\pi)^{-\frac{1}{2}}(s^2)^{-\frac{1}{2}}\left(\tfrac{v}{2}\right)^{-\frac{1}{2}}\Gamma\left(\tfrac{v+1}{2}\right)\Gamma^{-1}\left(\tfrac{v}{2}\right)\left\{1+\tfrac{(x-\mu)^2}{vs^2}\right\}^{-\frac{(v+1)}{2}} \tag{2.6}$$

where μ is location parameter, s^2 is scale parameter and v is a shape parameter, or degrees of freedom and $\Gamma(.)$ is gamma function. Standard t-distribution assumes $\mu = 0$, $s^2 = 1$, with integer v.

The degrees of freedom can be estimated using method of moments, which means that kurtosis and degrees of freedom are closely related: $k = \dfrac{6}{v-4} + 3$ for every $v > 4$. So, when empirical distribution is leptokurtic, then Student's t-distribution with parameter $4 < v \leq 30$ should be used to allow heavy tails of high kurtosis distribution. The first two central moments are given as: $\mu_2 = E\left[(x-\mu)^2\right] = \frac{s^2 v}{v-2}$ and $\mu_4 = E\left[(x-\mu)^4\right] = \frac{3s^4v^2}{(v-2)(v-4)}$ with excess kurtosis (greater than 3):

$$k^* = \tfrac{\mu_4}{\mu_2^2} - 3 = \tfrac{6}{v-4} \quad if \;\; v < \infty \;\; and \;\; v \neq 4. \tag{2.7}$$

Hence, we may apply method of moments to get consistent estimators: $\hat{\mu}_2 = \frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n}$; $\hat{\mu}_4 = \frac{\sum_{i=1}^{n}(x_i-\overline{x})^4}{n}$; $k^* = \frac{\hat{\mu}_4}{\mu_2^2} - 3$; $\hat{v} = 4 + \frac{6}{k^*}$ and $\hat{s}^2 = \left(\frac{3+\hat{k}^*}{3+2.\hat{k}^2}\right).\hat{\sigma}^2$, where variance from sample $\hat{\sigma}^2$ is biased estimator of scale parameter s^2.
The MLE estimator $\hat{\theta}^2$ maximizes the log-likelihood function l_t given by

$$\begin{aligned} l_t &= n\left[\log\Gamma\tfrac{(v+1)}{2} - \log\Gamma\left(\tfrac{v}{2}\right) - \tfrac{1}{2}\log\pi(v-2)\right]\\ &\quad -\tfrac{1}{2}\sum_{t=1}^{n}\left\{\log(\sigma_t^2) + (v+1)\log\left[1+\tfrac{\epsilon_t^2}{\sigma_t^2(v-2)}\right]\right\}\end{aligned}$$

where $2 < v \leq \infty$ and *Gamma* is the Euler gamma function defined by $\Gamma(x) = \int\limits_0^\infty t^{x-1}e^{-1}dt$. When $v \to \infty$, we have the normal distribution, so that the smaller the value of v the fatter the tails. This means that for large v, the product $\left(\frac{v}{2}\right)^{-\frac{1}{2}}\Gamma\left(\frac{v+1}{2}\right)\Gamma^{-1}\left(\frac{v}{2}\right)$ tends to unity, while the right-hand bracket in (2.8) tends to $e^{\frac{1}{2\sigma^2}(x-\mu)^2}$.

International Journal of Mathematical Analysis and Optimization: Theory and Applications
Vol. 2020, No. 1, pp. 689 - 706

The score function is given

$$\begin{aligned}\frac{\partial l_t}{\partial \theta} &= \sum_{t=1}^{n}\left[\frac{v+1}{v-2}\frac{\epsilon_t}{\sigma_t^2}\left(1+\frac{\epsilon_t^2}{\sigma_t^2}\right)^{-1}\right]\frac{\partial \mu_t}{\partial \theta} \\ &+\frac{1}{2}\sum_{t=1}^{n}\left[\frac{v+1}{v-2}\frac{\epsilon_t^2}{\sigma_t^2}\left(1+\frac{\epsilon_t^2}{\sigma_t^2(v-2)}\right)^{-1}\right]\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta}\end{aligned}$$

and the Hessian matrix is given by

$$\begin{aligned}\frac{\partial^2 l_t}{\partial \theta \partial \theta'} &= \frac{v+1}{v-2}\sum_{t=1}^{n}\left(1+\frac{\epsilon_t^2}{(v-2)\sigma_t^2}\right)^{-1}\frac{\epsilon_t}{\sigma_t^2}\left[\frac{2}{v-2}\left(1+\frac{\epsilon_t^2}{(v-2)\sigma_t^2}\right)^{-1}\frac{\epsilon_t}{\sigma_t^2}-1\right]\frac{\partial \mu_t}{\partial \theta}\frac{\partial \mu_t}{\partial \theta'} \\ &+\frac{v+1}{v-2}\sum_{t=1}^{n}\left(1+\frac{\epsilon_t^2}{(v-2)\sigma_t^2}\right)^{-1}\frac{\epsilon_t}{\sigma_t^4}\left[\frac{1}{v-2}\left(1+\frac{\epsilon_t^2}{(v-2)\sigma_t^2}\right)^{-1}\frac{\epsilon_t}{\sigma_t^2}-1\right]\frac{\partial \mu_t}{\partial \theta}\frac{\partial \sigma_t^2}{\partial \theta'} \\ &+\frac{v+1}{v-2}\sum_{t=1}^{n}\left(1+\frac{\epsilon_t^2}{(v-2)\sigma_t^2}\right)^{-1}\frac{\epsilon_t}{\sigma_t^4}\left[\frac{1}{v-2}\left(1+\frac{\epsilon_t^2}{(v-2)\sigma_t^2}\right)^{-1}\frac{\epsilon_t}{\sigma_t^2}-1\right]\frac{\partial \sigma_t^2}{\partial \theta}\frac{\partial \mu_t}{\partial \theta'} \\ &+\frac{1}{2}\sum_{t=1}^{n}\frac{1}{\sigma_t^4}\left[1+\frac{v+1}{v-2}\frac{\epsilon_t^2}{\sigma_t^2}\left(1+\frac{\epsilon_t^2}{(v-2)\sigma_t^2}\right)^{-1}\right] \\ &\left[\frac{\epsilon_t^2}{(v-2)\sigma_t^2}\left(1-\frac{\epsilon_t^2}{(v-2)\sigma_t^2}\right)^{-1}-2\right]\frac{\partial \sigma_t^2}{\partial \theta}\frac{\partial \sigma_t^2}{\partial \theta'}\end{aligned}$$

(iv) The Generalized error distribution (GED)
The GED is a symmetrical unimodal member of the exponential family. The domain of the probability density function (pdf) is $x \in [-\infty, \infty]$ and the distribution is defined by three parameters, $\mu \in (-\infty, \infty)$, which locates the mode of the distribution, $\sigma \in (0, \infty)$, which defines the dispersion of the distribution and $k \in (0, \infty)$, which controls the skewness. The probability distribution function, $F(x)$, is given by

$$dF(x|\mu,\sigma,v) = \frac{e^{\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{\frac{1}{v}}}}{2^{v+1}\sigma\Gamma(v+1)}dx \tag{2.8}$$

If we choose $v = \frac{1}{2}$, then (2.13) is recognized as the pdf for the univariate Normal distribution, that is, $G(\mu, \sigma^2, \frac{1}{2}) = N(\mu, \sigma^2)$. If we use $v = 1$, then (2.13) is recognized as the pdf for the Double- Exponential or Laplace distribution, that is, $G(\mu, \sigma^2, 1) = N(\mu, 4\sigma^2)$. If the limit $v \to 0$, the pdf tends to the Uniform distribution, $U(\mu-\sigma, \mu+\sigma)$. This shows that the GED allows one to maintain the same mean and variance, but vary the distribution's shape (via the parameter v) as required. The GED is useful because it can be smoothly transformed from normal distribution into a leptokurtic distribution ("fat tails") or even into a platykurtic distribution ("thin tails"). This also allows us to use the maximum likelihood ratio test to test the hypothesis as to whether the GARCH process innovations are *i.i.d.* normal. The kurtosis of the GED is given by

$$E\left[\epsilon_t^4\right] = \Gamma\left(\frac{1}{v}\right)\Gamma\left(\frac{5}{4}\right)\left[\Gamma\left(\frac{3}{v}\right)\right]^2 \tag{2.9}$$

which is greater than three if $v < \frac{1}{2}$.

International Journal of Mathematical Analysis and Optimization: Theory and Applications
Vol. 2020, No. 1, pp. 689 - 706

In GED, the score function is

$$\frac{\partial l_t}{\partial\theta}=\frac{1}{2}\sum_{t=1}^{n}\frac{v}{|\lambda v|}\left(\frac{\epsilon_t}{\sigma_t}\right)^{v}\frac{1}{\epsilon_t}\frac{\partial\mu_t}{\partial\theta}+\frac{1}{2}\sum_{t+1}^{n}\frac{1}{\sigma_t^2}\left[\frac{1}{2}\frac{1}{|\lambda v|^{v}}\left(\frac{\epsilon_t}{\sigma_t}\right)^{v}-1\right]\frac{\partial\sigma_t^2}{\partial\theta} \tag{2.10}$$

and the Hessian matrix is equal to

$$\begin{aligned}\frac{\partial l_t}{\partial\theta\partial\theta'} &= \frac{v}{|\lambda v|}\left(\frac{v-3}{2}\right)\sum_{t=1}^{n}\frac{2}{\sigma_t^2}\left(\frac{\epsilon_t^2}{\sigma_t^2}\right)^{\frac{v}{2}-1}\frac{\partial\mu_t}{\partial\theta}\frac{\partial\mu_t}{\partial\theta'}\\ &\quad -\frac{1}{2}\frac{v}{|\lambda v|}\sum_{t=1}^{n}\frac{\epsilon_t}{\sigma_t^4}\left(\frac{\epsilon_t^2}{\sigma_t^2}\right)^{\frac{v}{2}-1}\left[1+\left(\frac{v}{2}-1\right)\frac{\epsilon_t^2}{\sigma_t^2}\right]\frac{\epsilon_t^2}{\sigma_t^2}\\ &= -\frac{1}{2}\frac{v}{|\lambda v|^{v}}\sum_{t=1}^{n}\frac{\epsilon_t}{\sigma_t^4}\left(\frac{\epsilon_t^2}{\sigma_t^2}\right)^{\frac{v}{2}-1}\left[1+\left(\frac{v}{2}-1\right)\frac{\epsilon_t^2}{\sigma_t^2}\right]\\ &\quad -\frac{1}{2}\sum_{t=1}^{n}\left[\frac{1}{4}\frac{v(v+2)}{|\lambda v|^{v}}\left(\frac{\epsilon_t}{\sigma_t}\right)^{v}-1\right]\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'}\end{aligned}$$

Using (2.6) and (2.16), we have

$$\begin{aligned}\frac{\partial\sigma_t^2}{\partial\theta} &= \left(1,\epsilon_{t-1}^2,\sigma_{t-1}^2,\sigma_{t-1}\epsilon_{t-1}\right)+2\sigma_1\epsilon_{t-1}\frac{\partial\epsilon_{t-1}}{\partial\theta}+\beta_1\frac{\partial\sigma_{t-1}^2}{\partial\theta}\\ &\quad +c_1\left(\sigma_{t-1}\frac{\partial\epsilon_{t-1}}{\partial\theta}+\frac{1}{2}\frac{\epsilon_{t-1}}{\sigma_{t-1}}\right)\frac{\partial\sigma_{t-1}^2}{\partial\theta}\end{aligned}$$

The main feature of the selected distributions is that the maximum likelihood estimators achieve optimal accuracy, in the sense that they are asymptotically consistent and achieve Cramer-Rao lower bound [27]. When the errors have heavy tails that can be described by a student t-distribution (t-AFTER) and error distribution (g-AFTER), the forecast combination adapted, following the spirit of the AFTER strategy by [28] should be used.

Let $\|X\|=\sum_{i=1}^{n}|x_i|$ be the l_1-norm of vector $X=(x_1,...,x_n)$. Let $W_i=(W_{i,1},...,W_{i,J})$ be a vector of combination weights of $\hat{Y}_i$. It is assumed that and $\sum_{j=1}^{J}W_{i,j}=1$ and $W_{i,j}\geq 0$ for any $i\geq i_0$, $1\leq j\leq J$. Let $W_{i_0}=(w_1,...,w_J)$ be the initial weight vector. The combined forecast for y_i from the combination method is $\hat{y}_i=(\hat{Y}_i,W_i)$. The general form of W_i for the AFTER approach is $W_i=\frac{I_{i-1}}{\|I_{i-1}\|_1}$, where $I_{i-1}=(I_{i-1,1},...,I_{i-1,J})$ and for any $1\leq j\leq J$, $l_{i-1,j}$, $l_{i-1,j}=w_j\prod_{i'\geq i_0}^{i-1}\frac{1}{\hat{S}_{i',j}}h\left(\frac{y_i'-\hat{y}_{i',j}}{\hat{s}_{i',j}}\right)$, where $\hat{s}_{i',j}$ is an estimate of s_i' from the j^{th} forecaster at time point $i'-1$. Suppose at each time period $i\geq 1$, there are J forecasters available for predicting y_i and the forecast combination starts at $i_0\geq 1$, where i_0 is assumed to be large enough, say 10.
Let $\Omega=(v_1,...,v_k)$ be a set of degrees of freedom for Student's t distributions. Let $w_{j,k}(w_{j,k}\geq 0$ and $\sum_{k=1}^{K}\sum_{j=1}^{J}w_{j,k}=1)$ be the initial combination weight of the forecaster j under the degrees of freedom v_k. Let the combining weight of $\hat{Y}_i$ from a t-AFTER method be $W_i^{A_t}$ and the combined forecast be $\hat{y}_i^{A_t}$. The $W_i^{A_t}$ and $\hat{y}_i^{A_t}$ are obtained through the following steps:

(i) Estimate (e.g. by MLE) s_i for each $v_k\in\Omega$ and for each candidate forecaster. The estimate for s_i from the j^{th} forecaster given v_k is denoted as $\hat{s}_{i,j,k}$.

IJMAO

INTERNATIONAL JOURNAL OF MATHEMATICAL ANALYSIS AND OPTIMIZATION: THEORY AND APPLICATIONS
VOL. 2020, NO. 1, PP. 689 - 706

(ii) Calculate $W_i^{A_t}$ and $\hat{y}_i^{A_t}$:

$$W_i^{A_t} = \frac{I_{i-1}^{A_t}}{\|I_{i-1}^{A_t}\|_1}, \quad \hat{y}_i^{A_t} = \left(\hat{Y}_i, W_i^{A_t}\right) \tag{2.11}$$

where $I_{i-1}^{A_t} = (I_{i-1,1}^{A_t}, ..., I_{i-1,J}^{A_t})$ and for any $1 \leq j \leq J$ and any $i \geq i_0 + 1$, $l_{i-1,j}^{A_t} = \sum_{k=1}^{K} l_{i-1,j,k}^{A_t}$ with

$$l_{i-1,j,k}^{A_t} = w_{j,k} \prod_{i' \geq i_0}^{i-1} \frac{1}{\hat{S}_{i',j,k}} f_t\left(\frac{y_i' - \hat{y}_{i',j}}{\hat{s}_{i',j,k}} | v_k\right), \tag{2.12}$$

where $f_t(.|v)$ is the pdf of the student's t distribution with degrees of freedom v.

The performance of the forecast combination method, g-AFTER, for situations when there is a lack of strong or consistent evidence on the tail behaviours of the forecast errors due to shortage of data and/or evolving data-generating process is provided from the theorem that allows the random errors to be from one of the three popular distribution families (normal, double-exponential and scaled student's t). Let the combining weight $\hat{Y}_i$ from the g-AFTER be $W_i^{A_g}$ for any $i > i_0$, $W_i^{A_g}$ and the associated combined forecast $\hat{y}_i^{A_g}$ are

$$W_i^{A_g} = \frac{I_{i-1}^{A_g}}{\|I_{i-1}^{A_g}\|_1}, \quad \hat{y}_i^{A_g} = \left(\hat{Y}_i, W_i^{A_g}\right), \tag{2.13}$$

where $I_{i-1}^{A_g} = (I_{i-1,1}^{A_g}, ..., I_{i-1,J}^{A_g})$ and for $1 \leq j \leq J$.

$$l_{i-1,j}^{A_g} = (l_{i-1,j}^{A_2} + c_1 l_{i-1,j}^{A_1}) + c_2 l_{i-1,j}^{A_t} \tag{2.14}$$

where $l_{i-1,j}^{A_2}, l_{i-1,j}^{A_1}$ and $l_{i-1,j}^{A_i}$ are from the L_{2-}, L_{1-} and t-AFTER, respectively and c_1 and c_2 are non-negative constants that control the relative importance of the L_{2-}, L_{1-} and t-AFTERs in the g-AFTER. , L_2-AFTER is when the random errors in the data generating process follow a normal distribution or a distribution close to a normal distribution, The L_1-AFTER method is the double-exponential distribution with scale parameter 1 and location parameter 0 which was designed for robust combination when the random errors have occasional outliers. The g-AFTER satisfies the following conditions:

(i) There exists a constant $\tau > 0$ such that for any $i \geq i_0$,

$$P(\sup_{1 \leq j \leq J} |\hat{y}_{i,j} - m_i|/s_i \leq \sqrt{\tau}) = 1.$$

(ii) Suppose the random errors have zero mean and are from of the three families (normal, double-exponential and scaled student's t), and there exists a constant $0 < \gamma_2 \leq 1$ such that for any with probability 1, we have $\gamma_2 \leq \frac{\hat{s}_i}{s_i} \leq \frac{1}{\gamma_2}$, where s_i is the actual conditional scale parameter at time point i and $\hat{s}_i$ refers to any estimate of s_i used in the g-AFTER.

(iii) When the random errors in the true model follow a scaled student's t distribution with degrees of freedom v, assume there exist positive constants $\underline{v}$, λ and $\overline{v}$ such that $v \leq \min_{v_k \in \Omega}(v_k, v) - 2 \leq \overline{v}$, $\max |v_k - v| \leq \lambda$.

International Journal of Mathematical Analysis and Optimization: Theory and Applications
Vol. 2020, No. 1, pp. 689 - 706

Let $w_j^{A_2}$ and $w_j^{A_1}$ be the initial combination weights of the forecaster j in the L_{2-}, L_{1-}AFTERS respectively and $w_{j,k}^{A_t}$ be the initial combination weight of the j^{th} forecaster under the degrees of freedom v_k in the t-AFTER. Let $\hat{w}_{i,j}^{A_2} = \frac{l_{i-1,j}^{A_2}}{\|l_{i-1}^{A_g}\|}$, $\hat{w}_{i,j}^{A_1} = \frac{c_1 l_{i-1,j}^{A_1}}{\|l_{i-1}^{A_g}\|}$ and $\hat{w}_{i,j,k}^{A_t} = \frac{c_2 l_{i-1,j}^{A_t}}{\|l_{i-1}^{A_g}\|}$, where $l_{i-1,j,k}^{A_t}$ is defined in (2.19) and $l_{i-1}^{A_g}$ is defined in (2.21), so, $\hat{w}_{i,j}^{A_2}, \hat{w}_{i,j}^{A_1}$ and $\hat{w}_{i,j,k}^{A_t}$ are weights of the density estimates under normal, double-exponential and scaled student's t with degrees of freedom v_k in the g-AFTER procedure at time point $i-1$ from the j^{th} forecast, respectively.
Let $G = \sum \left(w_j^{A_2} + c_1 w_j^{A_1} + c_2 \sum_k w_{j,k}^{A_t}\right)$, where c_1 and c_2 are defined in (2.21) and Let q_i be the pdf of ϵ_i at time i and its estimator from g-AFTER procedure be

$$\hat{q}_i^{A_g} = \sum_{j=1}^{J} \left(\hat{W}_{i,j}^{A_2} \frac{1}{\hat{\sigma}_{i,j}} f_N \left(\frac{\hat{y}_{i,j} - y_i}{\hat{\sigma}_{i,j}} \right) + \hat{W}_{i,j}^{A_1} \frac{1}{\hat{d}_{i,j}} f_{DE} \left(\frac{\hat{y}_{i,j} - y_i}{\hat{d}_{i,j}} \right) \right)$$
$$+ \sum_{k=1}^{K} \hat{W}_{i,j,k}^{A_t} f_t \left(\frac{\hat{y}_{i,j} - y_i}{\hat{s}_{i,j,k}} | v_k \right),$$

where $\hat{q}_i^{A_g}$ is the mixture estimator of q_i from the g-AFTER procedure, $\hat{\sigma}_{i,j}$ is the sample standard deviation in the L_2-AFTER assuming the random error are independent and identically distributed, $\hat{s}_{i,j}$ used in the L_1-AFTER, denoted as $\hat{d}_{i,j}$ is the mean of $\{|y_{i'} - \hat{y}_{i',j}|\}_{i'=1}^{i-1}$. The L_1-AFTER method was designed for robust combination when the random errors have occasional outliers.

Theorem 1 [29]: If conditions (ii) and (iii) hold, then for $\hat{y}_i^{A_g}$ from g-AFTER procedure, we have

$$\frac{1}{n} \sum_{i=i_0+1}^{i_0+n} E\left(D\left(q_i \| \hat{q}_i^{A_g}\right)\right) \leq \inf_{1 \leq j \leq J} \left(\frac{B_1}{n} \sum_{i=i_0}^{i_0+n} E\left(\frac{m_i - \hat{y}_{i,j}}{\hat{\sigma}_i^2} \right) + R \right),$$

where

$$R = \begin{cases} \log \frac{G}{w_j^{A_2}} + \frac{B_2}{n} \sum\limits_{i=i_0+1}^{i_0+n} E\left(\frac{(\hat{\sigma}_{i,j} - \sigma_i)^2}{\hat{\sigma}_i^2} \right) & \text{under } normal \;\; errors, \\ \frac{\log \frac{G}{c_1 w_j^{A_1}}}{n} + \frac{B_2}{n} \sum\limits_{i=i_0+1}^{i_0+n} E\left(\frac{(\hat{d}_{i,j} - d_i)^2}{d_i^2} \right) & \text{under } double-exponential \;\; errors, \\ \inf\limits_{1 \leq k \leq K} \left(\log\left(\frac{G}{c_2 w_{j,k}^{A_t}} \right) + \frac{B_2}{n} \sum\limits_{i=i_0+1}^{i_0+n} E\left(\frac{(\hat{B}_{i,j,k} - B_i)^2}{B_i^2} \right) + B_3 \left| \frac{v - v_k}{v} \right| \right), & \text{under} \\ scaled \; t \; errors \end{cases}$$

where $D(f\|g) = \int f \log \frac{f}{g}$ is the Kullback-Leibler divergence between two density functions f and g. So, $E\left(D\left(q_i \| \hat{q}_i^{A_g}\right)\right)$ is a measure of the performances of $\hat{q}_i^{A_g}$ as estimate of q_i under Kullback-Leibler divergence at time point i. If condition (i) also holds, then
$\frac{1}{n} \sum\limits_{i=i_0}^{i_0+n} E\left(\frac{m_i - \hat{y}_i^{A_g}}{\hat{\sigma}_i^2} \right) \leq C \inf\limits_{1 \leq j \leq J} \left(\frac{B_1}{n} \sum\limits_{i=i_0}^{i_0+n} E\left(\frac{(m_i - \hat{y}_{i,j})^2}{\sigma_i^2} \right) + R \right)$ where C, B_1, B_2 and B_3 are constants depending on τ, γ_2 and parameters in Condition (iii).

When strong evidence is shown that the errors are highly heavy-tailed, Ω can be very small with only small degrees of freedom and $c_2 w_{j,k}^{A_t}$ in G can be relatively large (relative to $w_j^{A_2}$ and $c_1 w_j^{A_1}$). The more information in the tails of the error distribution that is available, the more efficient the allocation of the initial weights can be.

IJMAO

INTERNATIONAL JOURNAL OF MATHEMATICAL ANALYSIS AND OPTIMIZATION: THEORY AND APPLICATIONS
VOL. 2020, NO. 1, PP. 689 - 706

3 Results and Discussion

3.1 Data

Each empirical time series comprises of weekly observations covering Jan 12, 1986 to Jan 12, 2018 (1672) of the West Texas Intermediate Cushing (WTI) oil spot price and May 15, 1987 to Jan 12, 2018 (1601) of Brent oil spot price sourced from the website of United States Energy information Administration, a central online data warehouse are analyzed using asymmetric GARCH models. The period selected includes the dynamics in oil prices from above 100 (USD) per barrel in late 2014 to less than $30(USD)$ per barrel in early 2016 came amid over supply of crude oil at a time when global demand for crude oil stagnated [30]. In each case, weekly returns are computed as logarithmic price (P_t) relatives as: Given a series of stock/oil prices $(p_0, p_1, ..., p_t)$, return series are computed as follows:

$$R_t = \log(P_t) - \log(P_{t-1}) \tag{3.1}$$

where P_t = Observed weekly price at time t and R_t= compounded weekly returns. Analyses are written using the $R-i386$ 3.4.4 programming language under the Maximum Likelihood Estimation.

3.2 Preliminary Tests

A time plot graph(s) of WTI and Brent Crude oil spot price in USD/Barrel displayed in Figures 1 and 2 follows an upward trending behaviour indicating non-stationarity of the series. The plots also show a sharp fall at about the last quarter of 2008 which persisted till 2010. The series are transformed using logarithmic approach to get the return series and plotted in Figures 3 and 4. The plots show that the returns were more volatile over some time periods and became very volatile toward the end of the sample period. This pattern of alternating quiet and volatile periods of substantial duration is referred to as volatility clustering [31]. A visual inspection shows clearly, that the mean process for the different oil prices are not statistically significantly different from zero, but the variance changes over time, so the return series is not a sequence of independently and identically distributed ($i.i.d.$) random variables [32]. A simple test to investigate the leverage effect is to calculate first-order autocorrelation coefficient between lagged returns and contemporary squared returns: $\sum_{t=2}^{n} r_t^2 r_{t-1} / \sqrt{\sum_{t=2}^{n} r_t^4 \sum_{t=1}^{n} r_{t-1}^2}$.

Table 1: Summary Statistics for the compounded returns R_t of WTI and Brent spot oil prices

Series Statistics	**WTI OIL RETURNS**	**BRENT OIL RETURNS**
Mean	0.0008243	0.0005372
Std Dev.	0.04208652	0.04341357
Skewness	-0.1962921	-0.1467255
Kurtosis	5.92395	6.286493
Jarque-Bera (Probability)	0.0000	0.0000
σ	1	1
ADF test (Probability)	0.0100	0.0100
σ	1	1
$cov(r_t^2, r_{t-1})$	-0.065172	-0.073240

From Table 1, the mean weekly returns of the both series (WTI and BRENT) are positive (0.0005372 and 0.008243) and they are close to zero. These show that positive changes in the oil prices indices

IJMAO

International Journal of Mathematical Analysis and Optimization: Theory and Applications
Vol. 2020, No. 1, pp. 689 - 706

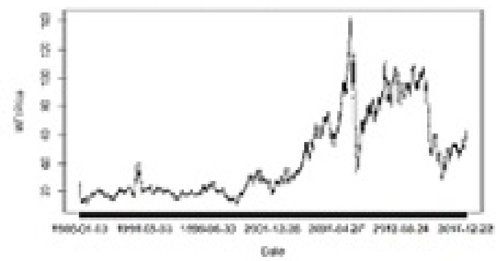

Fig. 1: Plot of WTI oil spot prices

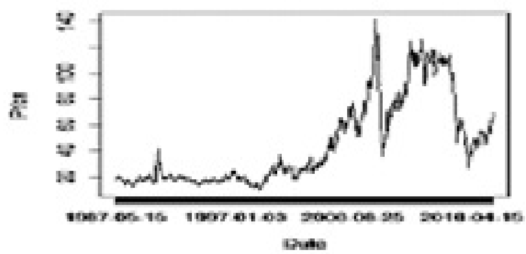

Fig. 2: Plot of Brent oil spot prices

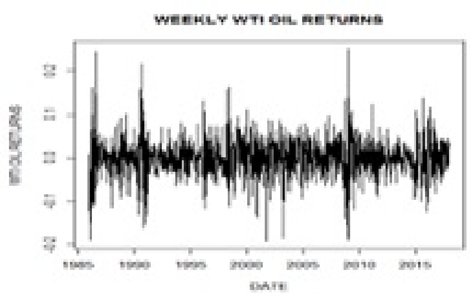

Fig. 3: Plot of the transformed weekly WTI oil returns

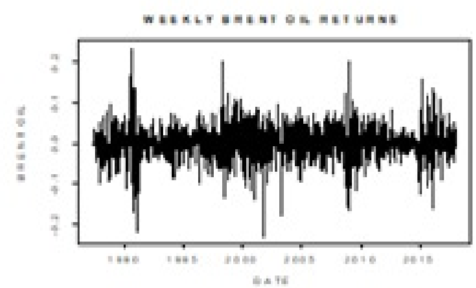

Fig. 4: Plot of the transformed weekly Brent oil returns

are more dominant than negative changes. The skewness coefficients are negative for all series suggesting that they have long left tail. The kurtosis coefficients on the other hand are very high (greater than 3), a reflection that the distributions of the two sets of real data are highly leptokurtic. The values of skewness and kurtosis show that the assets are far from being unconditionally normally distributed, thus supporting the conjecture that more flexible distributional assumptions can be conducive to enhanced model performance [33]. The p-value corresponding to the Jarque-Bera normality test is zero at 5% level suggesting that the test is significant for all series. The test gave a value of $\sigma = 1$ which indicates that the series r_t does not come from a normal distribution; in favour of $\sigma = 0$ which indicates that the series r_t comes from a normal distribution with unknown mean and variance. The test results imply that the two series exhibit non-normal behaviour. The Augmented Dickey-Fuller (ADF) test rejects the unit root null hypothesis in all data sets. This is indicated by the minimal p-values at 5% level and the values of σ. The test returns a value of $\sigma = 1$ which indicates rejection of the unit root in favour of the trend-stationary alternative. $\sigma = 0$ indicates failure to reject the unit root null hypothesis. The series statistics show strong serial correlations in both levels of the return series.

Their positive kurtosis coefficient indicates that the tails of the histogram (Figures 5 and 6) of returns of WTI and Brent oil prices are fatter than the tails of a normal distribution (leptokurtic). Therefore, the returns of weekly WTI and Brent oil series are approximated using Student's t-distribution with 5 degrees of freedom. It means that higher probabilities are assigned to extreme values of the distribution, and empirical evidence for heavy tails can be found in high kurtosis.

International Journal of Mathematical Analysis and Optimization: Theory and Applications
Vol. 2020, No. 1, pp. 689 - 706

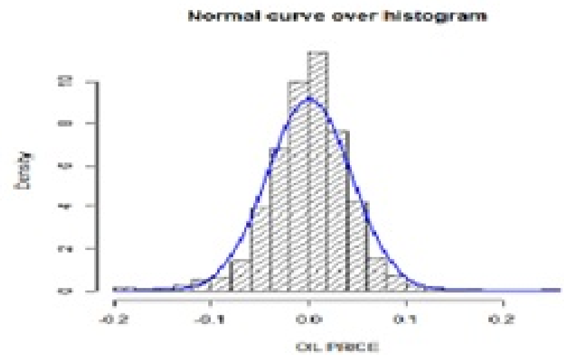

Fig. 5: Histogram and Normal Curve of WTI oil returns

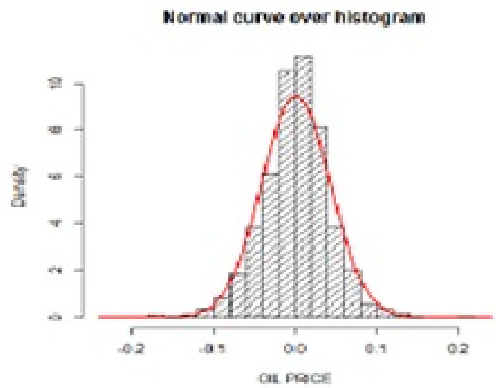

Fig. 6: Histogram and Normal Curve of Brent oil returns

3.3 Estimation of Parameters on Returns of Brent and WTI Oil Prices

We estimate the GARCH model and asymmetric models to explain conditional variance and volatility clustering for each of the series: GARCH(1,1), EGARCH(1,1), GJRGARCH(1,1), APARCH(1,1), TGARCH (1,1) and NAGARCH(1,1). For each oil returns, the parameter estimates are reported in Tables 2 and 3.

Table 2: Parameters estimates of asymmetric GARCH models; AIC and Log-Likelihood function for Brent Oil Returns

Table 3: Parameters estimates of asymmetric GARCH models, AIC and Log-Likelihood function for WTI Oil Returns

Tables 2 and 3 report the parameter estimates of all conditional volatility models employed in the analysis and information criteria and the log likelihood function for the estimate. Except GJRGARCH (the coefficients that reflects the leverage effects γ), the estimates of other models are significant at 1% level suggesting the strong validity of models. For the returns of Brent and WTI oil, EGARCH models show a positive and significant leverage effects γ parameter showing that future price volatility is greatly influenced by past positive events. GJRGARCH and TGARCH leverage effects are positive and significant in both oil series returns, attesting that bad news increases volatility. APARCH models estimates confirm that the asymmetric effects are present. The leverage coefficients are also positive and statistically significant, showing that positive innovations would imply a higher conditional variance than negative innovation of the same magnitude. Also the sum of the ARCH and the GARCH coefficients of each of the GARCH models are very close to one, indicating that volatility shocks are quite persistent. We can, therefore, attest that asymmetric effects are indeed present in WTI and Brent return series returns and performed better than a simple GARCH (1, 1) model in explaining the conditional volatility for the considered series.

Table 2 also shows that TGARCH-STD, followed closely by APARCH-STD model has the lowest values for AIC and the highest value for log-likelihood function. These criteria reveal that the TGARCH-STD is the better estimate for volatility of returns of Brent oil series compared to other models. In Table 3, EGARCH-STD, followed closely by NAGARCH-STD model has the lowest values for AIC and the highest value for log-likelihood function. These criteria reveal that the EGARCH-STD is the better estimate for the conditional volatility of the series returns of WTI compared to the other models. The results, therefore, reveal that the asymmetric GARCH models clearly outperformed GARCH (1, 1) models under the three error distributions.

3.4 Out-of-Sample Forecast (t-AFTER and g-AFTER)

Table 2: Asymmetric GARCH models of Brent oil returns

	NORMAL	NORMAL	GED	GED	STD	STD
MODEL	MSE	MAE	MSE	MAE	MSE	MAE
GARCH	0.00138492	0.02677674	0.001389167	0.02684516	0.00138786	0.02682452
EGARCH	0.001383527	0.02675223	0.001387628	0.02682074	0.001386556	0.02680363
GJRGARH	0.001384876	0.026776	0.001389193	0.02684556	0.001387796	0.02682347
APARCH	0.001383715	0.02675553	0.00138798	0.02682646	0.001386895	0.02680914
TGARCH	0.001383709	0.02675543	0.001387981	0.02682648	0.001386899	0.02680922
NAGARCH	0.00138369	0.02675521	0.001388261	0.02683099	0.001387262	0.02681501

Table 3: Asymmetric GARCH models of WTI oil return

	NORMAL	NORMAL	GED	GED	STD	STD
MODEL	MSE	MAE	MSE	MAE	MSE	MAE
GARCH	0.001283268	0.02639347	0.001287621	0.02643919	0.001285604	0.02641641
EGARCH	0.001280298	0.02636687	0.001284731	0.02640764	0.001283195	0.02639273
GJRGARH	0.001281199	0.02637424	0.001285974	0.02642031	0.001284336	0.02640389
APARCH	0.001279798	0.02636264	0.001284751	0.02640785	0.001283388	0.02639468
TGARCH	0.0012798	0.02636265	0.00128475	0.02640783	0.001283365	0.02639444
NAGARCH	0.001280498	0.02636856	0.001285362	0.02641401	0.001283872	0.02639944

The volatility forecasts obtained from Table 4 for the GARCH and symmetric GARCH models for the 1-week horizon show that EGARCH-Normal performed better than the other models because it has the lowest values in both mean absolute error (MAE) and mean square error (MSE). Therefore we can now say that at 1- week horizon EGARCH-Normal yields more accurate forecast for the volatility of the Brent oil returns series. In Table 5, APARCH-Normal outperformed the other models in terms of its MAE and MSE values. The paper, therefore, upholds APARCH-Normal as the best accurate forecast for the volatility of the returns of WTI oil prices at 1- week horizon. When the true random errors have tails significantly heavier than normal and double-exponential, they could be assumed to be from a scaled student's t distribution with unknown v and a (general) t-AFTER is more reasonable. In this case, $I^{A_g}_{i-1,j} = I^{A_t}_{i-1,j}$. When the random errors have heavy tails, the t- and g-AFTER provide more accurate forecasts than the L_2- and L_1- AFTER consistently. When the tails of the random errors distributions are not only mildly heavy, the g-AFTER is better than the t-AFTER in terms of forecast accuracy.

The best model for the forecast of the volatility is also plotted in Figures 7 and 8 for WTI and Brent respectively. Importantly, the MAE and MSE statistics clearly identifies the EGARCH and APARCH models as superior forecasting of Brent and WTI oil spot prices. Nevertheless, caution should be used in the interpretation of these results, as a change in sample size, rolling window, forecast horizon and frequency of observations could greatly impact the above findings.

International Journal of Mathematical Analysis and
Optimization: Theory and Applications
Vol. 2020, No. 1, pp. 689 - 706

Fig. 7: APARCH(1, 1) Model 1-Week Forecasts WTI oil price

Fig. 8: EGARCH(1, 1) Model 1-Week Forecasts Brent oil price

4 Conclusion

The analyses reveal that GARCH and asymmetric GARCH models are appropriate forecast tools when modelled under the selected error distributions: normal, student t and generalized error distributions for capturing volatility clustering, heavy tails, serial correlation, leverage effects and news asymmetry in the weekly return series of Brent and WTI oil spot prices. Also the sum of the ARCH and the GARCH coefficients of each of the GARCH models are very close to one, indicating that volatility shocks are quite persistent. The AIC and likelihood function criteria reveal that the TGARCH-STD and EGARCH-STD are the better estimate for volatility of returns of Brent and WTI oil series respectively when compared with the other models. Evaluating relative out-of-sample performances, EGARCH-Normal and APARCH-Normal performed better in returns of Brent and WTI oil spot prices respectively. The results obtained from the volatility forecasts seem to be useful to oil future traders especially when there is evidence of high standard deviation in the descriptive statistics of the return series. The policy makers on the other hand need to perceive "apriori" the effects of natural catastrophes and political or financial crises news far deeper than any good news on return volatilities before executing their investments and political strategies for the economic well being of the citizenry of any country.

References

[1] International Energy Agency. *Oil Market Report* (2006).

[2] Knight, J. & Satchell, S. Forecasting Volatility in the Financial Markets. *University of California, Butterworth Heinemann* (1998).

[3] Poon, S. H. & Granger, C. Forecasting volatility in financial markets. *A review, Journal of Economic Literature* **41**, 478–539 (2003).

[4] Hamilton, J. D. Oil and the Macro-economy since World War II. *Journal of Political Economy* **91**, 228–248 (1983).

[5] Lee, K. & Ni, S. On the dynamic effects of oil price shocks: a study using industry level data. *Journal of Monetary Economics* **49**, 823–852 (2002).

[6] Olugbenga, F. & Kehinde, O. S. Modeling the Impact of Oil Price Volatility on Investment Decision marking in marginal field's Development in Nigeria. *British Journal of Economics, Management and Trade* **17**, 1–16 (2017).

[7] Narayan, P. K. & Narayan, S. Modelling oil price volatility. *Energy Policy* **35**, 6549–6553 (2007).

[8] Olowo, R. A. Modeling naira/dollar exchange rate volatility: application of GARCH and asymmetric models. *International Review of Business Research Papers* **5**, 377–398 (2009).

[9] Hassan, S. A. Modelling Asymmetric Volatility in Oil Prices. *The Journal of Applied Business Research* **27**, 71–78 (2011).

[10] Chhatwal, H., Puri, H. & Purohit, H. An Empirical Investigation of Volatility of Indian Spot and Future Prices of Crude Oil. *Metamorphosis* **12**, 54–66 (2013).

[11] Rotemberg, J. J. & Woodford, M. Imperfect competition and the effects of energy price increases. *Journal of Money, Credit, and Banking* **28**, 549–577 (1996).

[12] Salisu, A. A. & Fasanya, I. O. Comparative Performance of Volatility Models for Oil Price. *International Journal of Energy Economics and Policy* **2**, 167–183 (2012).

[13] Abduikareem, A. & Abdulhakeem, K. A. Comparative Performance of Volatility Models for Oil Price. *International Journal of Energy Economics and Policy* **2**, 167–183 (2012).

[14] Hsieh, D. Chaos and Non-linear Dynamics: Application to Financial Markets. *Journal of Finance* **46**, 1839–1877 (1999).

[15] Hansen, P. R. & Lunde, A. A forecast comparison of volatility models: does anything beat a GARCH (1, 1). *Journal of Applied Econometrics* **20**, 873–889 (2005).

[16] Bollerslev, T. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* **31**, 307–328 (1986).

[17] Bolleslev, T., Engle, R. F. & Nelson, D. B. *ARCH Models in Handbook of Econometrics* **4** 2959–3038 (1994).

[18] Nelson, D. B. *Conditional Heteroskedasticity in assets returns: A new approach, the modelling of financial time series* **59** 347–370 (1991).

[19] Black, F. Studies of Econometric stock market volatility changes. 1976 Proceedings of the American Statistical Association, Business and Economic Section. *American Statistical Association* 177–181 (1976).

[20] Onyeka-Ubaka, J. N. & Abass, O. On Optimal Estimate Functions for Asymmetric GARCH Models. *International Journal of Mathematical Analysis and Optimization Theory and Applications* **2018** 291–311 (2018).

[21] Brooks, C. Introductory Econometrics for Finance. *Cambridge University Press* (2008).

[22] Glosten, L., Jagannathan, R. & Runkle, D. Relationship between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance* **48** 1779–1801 (1993).

[23] Zakoian, J. M. Threshold Heteroskedasticity Models. *Journal of Economic Dynamics and Control* **19** 931–944 (1994).

[24] Ding, Z., Granger, C. & Engle, R. A long memory property of stock returns and a new model. *Journal of Empirical Finance* **1** 83–106 (1993).

[25] Dhesi, G., Shakeel, B. & Austoos, M. Modelling and Forecasting the Kurtosis and Returns Distributions of Financial Markets: Irrational fractional Brownian motion model approach. *Annals of Operations Research* 1–14 (2019).

[26] Dallah, H., Okafor, R. O. & Abass, O. A Model-Based Bootstrap Method for Heteroskedasticity Regression Models. *Journal of Scientific Research and Development* **9** 9-22 (2004).

[27] Onyeka-Ubaka, J. N., Okafor, R. O. & Abass, O. Trading Volume Volatility in Nigeria Banking Sector. *Benin Journal of Statistics* **1** 21-32 (2018).

[28] Yang, Y. Combining Forecast Procedures: Some Theoretical Results. *Econometric Theory* **20** 76-222 (2004).

[29] Wei, X. & Yang, Y. Robust Forecast Combinations. *Journal of Econometrics* **166** 224-236 (2012).

[30] Onyeka-Ubaka, J. N., Agwuegbo, S. O. N. & Abass, O. Symmetric Volatility Forecast Models for Crude Oil Price in Nigeria. *Proceedings of UNILAG Research Conference* **5** 987-1000 (2017).

[31] Mandelbrot, B. The Variation of Certain Speculative Prices. *Journal of Business* **36** 394-419 (1963).

[32] Onyeka-Ubaka, J. N. & Abass, O. Central Bank of Nigeria (CBN) and the Future of Stocks in the Banking Sector. *American Journal of Mathematics and Statistics* **3** 407-416 (2013).

[33] Braione, M. & Scholtes, N. K. Forecasting Value-at-Risk under different Distributional Assumptions *Econometrics* **4** 1-27 (2016).