FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 7 No. 3, June (Special Issue), 2023, pp 270 - 274
DOI: https://doi.org/10.33003/fjs-2023-0703-1868

DEVELOPING THE HYBRID ARIMA- FIGARCH MODEL FOR TIME SERIES ANALYSIS

***[1]Bawa, M. U., [2]Dikko, H. G., [2]Garba, J. [3]Sadiku, S., [2]Tasi'u, M.**

[1]Department of Mathematics, School of Science, Abubakar Tafawa Balewa University, Bauchi, Nigeria.
[2]Department of Statistics, Ahmadu Bello University, Zaria, Nigeria
[3]Department of Medicine, Ahmadu Bello University, Zaria, Nigeria

*Corresponding authors' email: musausmanbawa@gmail.com

ABSTRACT

This study takes into account the newly developed hybrid ARIMA-FIGARCH. We use the daily price index of the S&P 500. The data employed for this study was secondary in nature for all the variables and was obtained from the publications of the Central Bank of Nigeria Bulletin, the National Bureau of Statistics, and the World Bank Statistics Database, dated January 2005 to December 2020. Also, the result of the Jarque-Bera test indicated that the p-values for all variables were less than the alpha level of significance (0.05). Hence, we would reject the null hypothesis that the data for all variables are normally distributed. Also, unit root tests were conducted using ADF and KPSS tests. The result of the ADF test shows that the variable is stationary at a level of 5% significance. That means the variables are integrated in order zero, i.e., 1 (0). And for the KPSS test, 0.881749 is greater than 0.463000, indicating that it is not significant at level 1, indicating that it is not stationary, whereas KPSS is 0.011158, which is less than 0.463000, indicating that it is stationary at level 1. The unit root test is necessary in order to determine the nature of the series and to avoid getting spurious results. We estimate the fractional difference order, d, by the Geweke and Porte-Hudak (GPH) method for testing the present and long memory of the series. The results show that the value of (d) for the S&P 500 price was found to be (0.043621), which falls within the range of 0d0.5, indicating the presence of long-memory processing of the data. Descriptive statistics yield a measure of central tendency as well as a measure of dispersion. Because of the high volatility in macro variables, the reported series are not bell-shaped. Most series are right-skewed and volatile. Finally, we found that the proposed ARIMA-FIGARCH model shows greater consistency because, as the sample size increased, the performance measure decreased and approached zero.

Keywords: FIGARCH, ARIMA hybridizations

INTRODUCTION

The ARIMA-FIGARCH model, is the combines the Autoregressive Integrated Moving Average (ARIMA) model with the FIGARCH model. The ARIMA model is suitable for modeling stationary time series data, while the FIGARCH model can capture the volatility clustering and long memory often observed in financial time series data.

These hybrid models have been applied in various fields, such as finance, economics, and engineering. For instance, these models have been used for forecasting stock prices, volatility, and risk management in the finance industry. Despite the growing interest in hybrid time series models, there is a lack of empirical studies that compare the performance of different hybrid models on financial data. Therefore, this study aims to develop the hybrid of ARIMA-FIGARCH models on financial data. We hope to contribute to the existing literature on time series modeling and provide insights for practitioners in the finance industry.

Baillie et al. (1996) extended the Garch family model and proposed the FIGARCH model. *Engle* (2013) came to the conclusion that GARCH-family models have been well-liked by academics due to their capacity to reproduce some of the typical stylized characteristics of financial time series, such as volatility clustering. *Bollerslev* (1986) demonstrates how GARCH-family models incorporate the characteristic of time-varying volatility over a long period of time and offer a strong in-sample estimate. The FIGARCH model was specifically developed to model the long memory property in addition to the above-mentioned features. The clustering of volatility in financial time series reveal that GARCH family models provide a strong in-sample estimate and account for the feature of time-varying volatility over a long period of time. In addition to the previously mentioned characteristics, the FIGARCH model is especially made to simulate the long memory property.

The volatility of daily futures returns is found to be well described by the FIGARCH model, with relatively similar estimates of the long memory parameter across commodities. The conditional means of the daily returns are close to being uncorrelated, with small departures from martingale behavior being represented by low-order moving average models. We also estimate FIGARCH models for high-frequency commodity futures returns based on intraday tick data. These high-frequency commodity returns are dominated by strong intraday periodicity, which is thought to be the result of repeated trading day cycles caused by the institutional features of futures exchanges where trades are conducted. The intraday periodicity is removed using a deterministic Flexible Fourier Form (FFF) filter. The filtered high-frequency futures returns are also well described by the FIGARCH process.

The FIGARCH model has already been effectively used to analyze a variety of economic variables, including exchange rates, stock returns, inflation rates, and others; see, for instance, Baillie et al. (1996). The concept has only frequently been applied to commodities, though. Although they did not explicitly estimate FIGARCH models, Chen (2009) found strong evidence of long memory in daily commodity futures prices by examining the daily volatilities of multiple agricultural commodity futures returns, a stock index return, currencies, metals, and heating oil. The main goal of establishing the FIGARCH model was to develop a more flexible class of conditional variance processes that are able to explain and depict the temporal dependencies in financial market volatility that have been observed.

In 1976, Box-Jenkins developed an ARIMA model, which he named the Box-Jenkins Methodology after themselves, as a forecasting tool for financial and economic variables. Aiming to identify an autoregressive integrated moving average (ARIMA) (p, d, q) model that satisfies the stochastic process and can be estimated using the Box-Jenkins approach, the Box-Jenkins methodology was developed in 1976. The Box-Jenkins technique is a three-stage iterative modeling approach that includes model identification, parameter estimates, and model checking. A modification is performed to the data series to remove the trend, which is a need for an ARIMA model when the observed time series exhibits both trend and non-seasonal behavior.

Although ARIMA is a prominent forecasting technique, it cannot handle volatile data. The FIGARCH model will be hybridized with the ARIMA in order to improve it. While ARCH-based models have been used in the case of financial time series that have been proved to reveal volatility clustering, ARIMA models have been widely used for forecasting different types of time series to capture the long-term trend. In order to understand and reflect the observed temporal dependencies in financial market volatility as well as to capture persistent volatility and long-memory behaviors in the data series, we are employing the FIGARCH model, which represents a more flexible class of conditional variance processes in this study.

MATERIALS AND METHODS

ARIMA process

The Autoregressive integrated moving average (ARIMA) (p, d, q) model of the time series y_t, t = 1, 2…n was developed by Box and Jenkins (1976), which is still widely used .Tasi'u (2022). It is defined as

$$\varphi(B)(1-L)^d(y_t-\mu)=\theta(B)\varepsilon_t \quad (1)$$

where y_t and ε_t denote time series and random error terms, respectively, at time t.

The backward shift operator is B. The $\varphi(B)$ and $\theta(B)$ are of order p and q and defined as

$$\varphi_p(B)=1-\sum_{i=1}^{p}\varphi_i L^i \quad \text{and} \quad \theta_q(B)=1-\sum_{j=1}^{q}\theta_j L^j \ .$$

Then L is the difference operator defined as

$$\Delta y_t=y_t-y_{t-1}=(1-L)y_t$$

Also, $\varphi_1, \varphi_2, \dots \varphi_p$ *and* $\theta_1, \theta_{2,\dots\dots\dots}\theta_q$ Are the parameters of autoregressive and moving average terms with order p and q respectively

ARCH (p) Process

We obtain the ARCH (p) process if ε_t^2 follows an AR (p) process is giving by

$$h_t=\alpha_0+\sum_{i=1}^{p}\alpha_i\,\varepsilon_{t-1}^2 \quad (2)$$

Where $\alpha_i \geq 0$, i = 1,..., q

GARCH Process

The GARCH (p, q) process introduced by Bollerslev (1986) and Taylor (1986) (independently of each other) is given along with the volatility equation. The generalized ARCH or GARCH model is a pared-down alternative to the ARCH (p) model. It is given by

$$h_t=\alpha_0+\sum_{i=1}^{p}\alpha_i\,\varepsilon_{t-1}^2+\beta_j h_{t-1} \quad (3)$$

Where the ARCH term is ε_{t-1}^2 and the GARCH term is h_{t-1}

In general, a GARCH (p.q) model includes p(ARCH) terms and q (GARCH) terms.

where p > 0, q > 0, $\alpha_0 > 0$, $\alpha_i \geq 0$, i = 1,...,q, $\beta j \geq 0$, j = 1,...,p and α(L) and β(L) are lag operators such that $\alpha(L)=\alpha_1 L+\alpha_2 L^2+\dots+\alpha_q L^q$ and $\beta(L)=\beta_1 L+\beta_2 L^2+\dots+\beta_p L^p$

For p = 0, the process reduces to an ARCH(q), and when p = q = 0, the process is simply white noise. An equivalent ARMA-type representation of the GARCH (p, q) process is given by the

$$[1-\alpha(L)-\beta(L)]\varepsilon_t^2=\alpha_0+[1-\beta(L)]v_i$$

.

IGARCH Process

Engle and Bollerslev (1986) analyze a type of GARCH model known as the IGARCH model, which has no unconditional variance. This occurs when $\sum_{i=1}^{q}\alpha_i+\sum_{j=1}^{p}\beta_j=1$ in a GARCH (p, q) model. The IGARCH is defined as follows:

$$\Phi(L)(1-L)\varepsilon_t^2=\alpha_0+[1-\beta(L)]v_t \quad (4)$$

Similar to ARIMA models a key feature of IGARCH models is that the impact of past squared

Shocks $V_i=\varepsilon_{t-i}^2-h_{t-i}$ for i >0 on ε_t^2 is persistent

We see that a GARCH (p, q) process may also be expressed as an ARMA (m, p) process. In ε_t^2, by writing

$$(L)(1-L)\varepsilon_t^2=\alpha_0+[1-\beta(L)]v_t$$

Where m = max {p, q} and $V_i=\varepsilon_t^2-h_t$. The { V_i } process can be interpreted as the "innovation" for the conditional variance, as it is a zero-mean martingale. Therefore, an integrated GARCH (p, q) process can be written as

$$(1-\alpha(L)-\beta(L))(1-L)\,\varepsilon_t^2=\alpha_0+[1-\beta(L)]v_t$$

The fractionally integrated GARCH or FIGARCH class of models is obtained by replacing the first difference operator. (1 − L) in (6) with the fractional differencing operator $(1-L)^d$, where d is a fraction 0 < d < 1. Thus, the FIGARCH class of models can be obtained by considering.

$$(1-\alpha(L)-\beta(L))(1-L)^d\varepsilon_t^2=\alpha_0+[1-\beta(L)]v_t \quad (5)$$

Assume the return can be used in calculating the volatility of any given asset return. Thus, we may use as a measure of volatility which could be written as

$$r_t=\mu+\sqrt{h_t\varepsilon_t^2} \quad (6)$$

r_t − is the return on day i, μ −is the average return and h_t - is the variance used as a volatility measure.

Proposed ARIMA-FIGARCH Model

Using a two-step process, the ARIMA and FIGARCH models will be combined. The linear data of a time series are modeled using the best ARIMA model in the first phase. The residuals sequence from the fitted ARIMA model is modeled using the FIGARCH technique in the second phase. The FIGARCH processes described by will be followed by the error term of the ARIMA model in this technique

$$\varphi_p(L)(1-L)^d y_t=\theta_q(L)\varepsilon_t$$

where

$$(L)=1-\varphi_1 L-\varphi_2 L^2\dots-\varphi_p L^P \quad \text{and} \quad \theta_q(L)=1+\theta_1 L+\theta_2 L^2+\dots+\theta_q L^q$$

$$(1-\varphi_1 L-\varphi_2 L^2\dots-\varphi_p L^P)(1-L)^d y_t=(1+\theta_1 L+\theta_2 L^2+\dots+\theta_q)\varepsilon_t$$

$$.e(1-L)^d y_t=\sum_{i=1}^{p}\varphi_i(1-L)^d y_{t-i}+\sum_{j=1}^{q}\theta_j\varepsilon_{t-j}+\varepsilon_t\sim ARIMA(p,d,q) \quad (7)$$

Thus, we may use as a measure of volatility which could be written as

$$r_t=\mu+\sqrt{h_t\varepsilon_t^2}$$

$r_t -$ is the return on day i, $\mu -$is the average return and h_t - is the variance used as a volatility measure.

$$\varepsilon_t = \frac{(r_t-\mu)}{\sqrt{h_t}} \quad (8)$$

where $\varepsilon_t = \frac{(r_t-\mu)}{\sqrt{h_t}}$

The hybrid of ARIMA (p, d, q)-FIGARCH (p, d, q) this is given as

$$(1-L)^d y_t = \frac{(r_t-\mu)}{\sqrt{\alpha_0^* + \delta(L)\varepsilon_t^2}} + \sum_{i=1}^{p} \varphi_i (1-L)^d y_{t-i} + \sum_{j=1}^{q} \theta_j \sqrt{\frac{(r_{t-j}-\mu)^2}{\alpha_0^* + \delta(L)\varepsilon_{t-j}^2}} \quad (9)$$

The conditional variance dynamics of the component h_t of a FIGARCH process is given by Baillie et al. (1996)

$$h_t = \alpha_0^* + \delta(L)\,\varepsilon_t^2 \quad (10)$$

where $\alpha_0^* = \alpha_0(1-\beta(L))^{-1}$ and $\delta(L) = (1-(1-(\beta(L))^{-1}\phi(L)(1-L)^d$ for the FIGARCH (p, d, q), must be non-negative, i.e., $\lambda_k \geq 0$ for k = 1, 2 $where$ 0<d<1, is the fractional differencing operators, and its value depends on the decay rate of a shock to conditional volatility. ε_t is random error.

Then $\varphi_1, \varphi_2 \dots, \varphi_p$ are the autoregressive coefficients that attempt to predict an output of a system based on the previous outputs and $\theta_1, \theta_2 + \cdots + \theta_q$ are the moving averages coefficients. ε_t^2 is a persistant shocks of long memory , and y_t is the deterministic functions of r_t where historical observations for estimated $(r_t-\mu)^2$ is not available.

Parameter estimation of Hybrid ARIMA-FIGARCH Model Using MLE methods.

The hybrid ARIMA-FIGARCH model is a nonlinear time series model that combines a linear ARIMA model with the conditional variance of a FIGARCH model. We consider the following ARIMA-FIGARCH of eqn (2.9)

$$(1-L)^d y_t = \sum_{i=1}^{p} \varphi_i (1-L)^d y_{t-i} + \sum_{j=1}^{q} \theta_j \frac{(r_{t-j}-\mu)}{\sqrt{\alpha_0^* + \delta(L)\varepsilon_{t-j}^2}} + \frac{(r_t-\mu)}{\sqrt{\alpha_0^* + \delta(L)\varepsilon_t^2}}$$

$$y_t = \sum_{i=1}^{p} \varphi_i y_{t-i} + \frac{1}{(1-L)^d} \sum_{j=1}^{q} \theta_j \sqrt{\frac{(r_{t-j}-\mu)^2}{\alpha_0^* + \delta(L)\varepsilon_{t-j}^2}} + \frac{1}{(1-L)^d} \sqrt{\frac{(r_t-\mu)^2}{\alpha_0^* + \delta(L)\varepsilon_t^2}} \quad (11)$$

The estimation procedures of the ARIMA and FIGARCH models are based on the Maximum Likelihood Method. The nonlinear Marquardt's algorithm (Marquardt, 1963) is used to estimate parameters in a logarithmic likelihood function. The logarithmic likelihood function has the following equation:

$$ln\, L\, [(y_t), \Theta = \sum_{t=1}^{T} \left\{ ln[\, D(z_t(\Theta)), v] - \frac{1}{2} ln[\, y_t(\Theta)] \right\} \quad (12)$$

where $\Theta$is the vector of the parameters that have to be estimated for the proposed model, z_t denoting their density function, $D(z_t(\Theta), v)$, is the log-likelihood function of [y$t(\Theta)$], for a sample of T's observations. The maximum likelihood estimator $\hat{\Theta}$for the true parameter vector is found by maximizing eqn (2.12) .

It is very important to know that equation (2.12) cannot be solved explicitly due to its complex nature. We therefore resolved to use a numerical method. In this case, we adopted the quasi-maximum likelihood estimator to estimate the parameters of the developed model.

Simulation

Using data of sizes 200, 500, and 1000, we imulate the ARIMA-FIGARCH models in this section. It is assumed that the error term in the simulation would behave normally, with mean= 0 and variance = 1. Two widely used accuracy measures (lost functions) were used in the evaluation: meansquare error (MSE) and mean absolute percentage error (MAPE). In a Monte Carlo simulation, we first assess the MLE's performance with finite samples (see Ghysels, 2012).

There were two stages of simulation for theARIMA and FIGARCH models. The best ARIMA model is used in the first phase to model a time series' linear data. In the second stage, the nonlinear designs of the residual sequences from the fitted ARIMA model are modeled using the FIGARCH method

. .

Table 1: Summary of Simulated Results of the ARIMA-FIGARCH model.

Sample size	ARIMA-FIGARCH	
	MSE	**MAPE**
200	5.26	6.05
500	2.00	2.05
1000	0.67	0.05

Based on the table above, the proposed ARIMA-FIGARCH model shows consistency than the existing models, because as the sample size increased, the performance measure decreased and it approach zero. And also, the real data analyst shows the proposed model performs better due to least MSE and MAPE are least value.

RESULTS AND DISCUSSION

Data

We use the daily price index of the S&P 500 to analyst the proposed model. The data employed for this study was secondary in nature for all the variables and was obtained from the World Bank Statistics Database dated January 2005 to December 2020

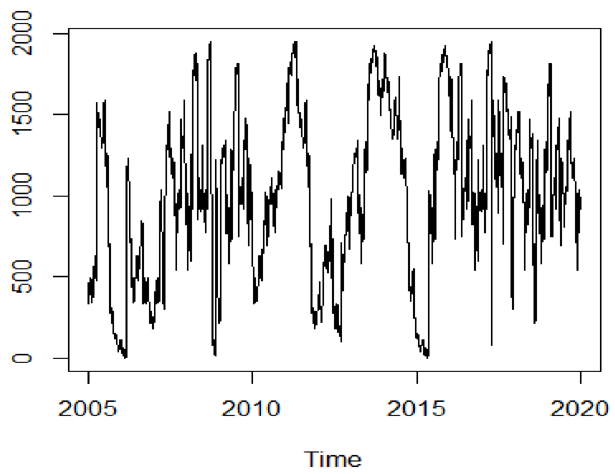
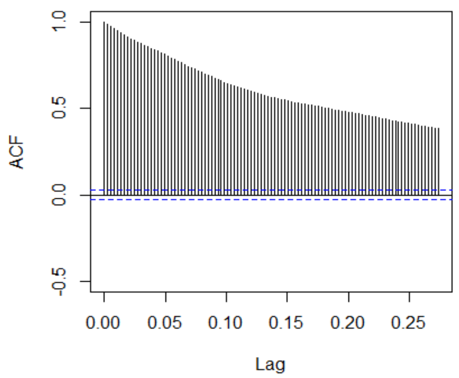


Figure 1: S&P 500 Time Plot Test the Present of Variability in data series

The plot of the S&P 500 series gives a good picture of its main characteristics. The typical behavior of the data series is that it fluctuates over time, such that during some periods of the year the amounts are low and in other periods they are high. This characteristic indicates that the daily S&P 500 varies throughout the period and that the variability is not constant..



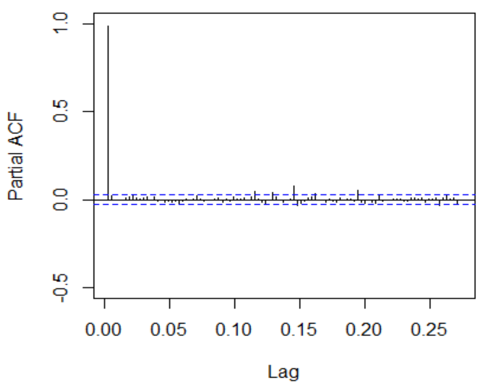


Figure 2: ACF and PACF PLOTS

The ACF plot dies down slowly. The autocorrelation function provides a measure of temporal correlation between data points with different time lags. For a purely random event, all autocorrelation coefficients *(r)* are zero, apart from *r(0),* which is equal to 1.

Table 2: Results for estimated model parameters: ARIMA (p, d, q) for S&P 500 Price.

Models	Coefficient	Estimate Error	Log-like hood	RMSE	MSE
ARIMA (1, 1, 1)	AR1	0.3153	91.71	1.43	2.7
	MA1	0.6522			
ARIMA (2, 0, 1)	AR1	0.1613	71.13	2.26	3.3
	AR2	0.0852			
	MA1	0.3607			
ARIMA (1, 2,0)	AR1	0.4645	62.12	2.23	4.7
ARIMA (2,1,1)	**AR1**	**0.2352**	**61.13**	**0.13**	**1.3**
	AR2	0.7865			
	MA1	0.5463			
ARIMA (1,0,1)	AR1	0.4982	65.4	1.24	1.43
	MA1	0.2314			

Based on the table above, ARIMA (2, 1, 1) is the best one among the five models because it has the lowest MAPE and MSE (0.13 and 1.3, respectively), so we can use it to fit the remaining Garch-family model.

Table 3: **The results of the estimated ARIMA-FIGARCH model for samples of S&P 500 price**

Models	Coefficient	Estimate	Log-like hood	MSE	**MAPE**
ARIMA (2,1,1)	**AR1**	**0.2352**	**61.13**	**1.3**	**0.13**
	AR2	0.7865			
	MA1	0.543			

FIGARCH	μ	0.0000	290.77	2.0078	2.07
	ω	0.00132			
	α_1	0.012912			
	β_1	0.02998			
	δ	0.048580			

The above findings suggest that the estimated parameter mu is not significant. Additionally the alpha and beta parameters' sum is quite close to unity ($\alpha + \beta$ = 0.51159), indicating the return's highpersistence. Additionally, the findings indicate that the delta is significant, indicating that the model has a long memory, and that the coefficient gamma is not significant, indicating that the sign of the innovation has no effect on the volatility of returns

CONCLUSION

The Jarque-Bera test result showed that all of the variables' p-values were below the alpha level of significance (0.05). As a result, we would reject the null hypothesis that the data are normally distributed across all variables. ADF and KPSS tests were also used to do unit root tests. At a level of 5% significance, the ADF test's result reveals that the variable is stationary. The variables are therefore integrated in order zero, or 1. (0). KPSS is 0.011158, which is less than 0.463000, suggesting that it is The unit root test is required to identify the type of the series and minimize misleading findings. In order to test the series' present and long-term memory, we estimate the fractional difference order, d, using the Geweke and Porte-Hudak (GPH) approach. The findings reveal that the S&P 500 price's value of d was determined to be (0.043621), which is within the range of 0d0.5, indicating that long-memory processing of the data was used. Both a measure of central tendency and a measure of dispersion are produced by descriptive statistics. The reported series are not bell-shaped because macro variables have a high level of volatility. Most series are turbulent and right-skewed. Last but not least, we found that the proposed ARIMA-FIGARCH model is more reliable.

REFERENCES

Box, G.E.P., Jenkins, G.M. (1976), Time Series Analysis. Forecasting and Control. San Francisco: Holden-Da Dark, F.(1976). Investigations of financial exchange instability changes. In1976 Proceedings of the Meetings of the Business and Economic Statistics Section; *American Statistical Association: Washington, DC, USA, 177-181*

Baillie R.T., Long memory processes and fractional integration in econometrics, *Journal of Econometrics*, 1996, Vol. 73, pp. 5-59

Ghysels, E., Santa-Clara, P. & Valkanov, R. (2006). Predicting volatility: Getting the most out of return data sampled at different frequencies. Journal of Econometrics, 131(1-2), 59-95.

Engle, R. F., Ghysels, E. and Sohn, B.: 2013, Stock market volatility and macroeconomic fundamentals, Review of Economics and Statistics 95, 776–797.

Bollerslev, T.: 1986, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31, 307–327

Chen, X. and Ghysels, E. (2009). News - good or bad - and its impact on predicting future volatility. Review of Financial Studies (forthcoming).

Tasi'u, M. (2022). *Development of symmetric non-linear state-space models*. Unpublish Ph.D. Thesis of Department of Statistics, Ahmadu Bello University, Zaria, Nigeria.

Engle, R. F. and Kelly, B.: 2012, Dynamic equicorrelation, Journal of Business and Economic Statistics 30, 212–228