Eur. J. Stat. 5 (2025) 2
doi: 10.28924/ada/stat.5.2

Assessing the Performance of the MARFIMA Model Using Simulated and Real Life Data

A. Bello[1*], M. Tasi'u[2], H.G. Dikko[2], B.B. Alhaji[3]

[1]Department of Mathematical Sciences, Gombe State University, Gombe State, Nigeria
bellobajogagsu@gmail.com

[2]Department of Statistics, Ahmadu Bello University, Zaria, Kaduna State, Nigeria
dagastatistician@gmail.com, hgdikko@yahoo.com

[3]Department of Mathematics, Nigerian Defence Academy, Kaduna, Kaduna State, Nigeria
bbukar@nda.edu.ng

*Correspondent Author: bellobajogagsu@gmail.com

ABSTRACT: A modified autoregressive fractional integrated moving average MARFIMA (p, d, q) is presented in this study to describe time series data that are nonstationary and have a fractional difference value of $1<d<1.5$. Data from ARFIMA simulations are used to assess the performance of the MARFIMA model. The autoregressive fractional integrated moving average ARFIMA model and the MARFIMA model's performance were also compared in a number of applications. Using the Akaike Information Criterion (AIC), Schwartz Bayesian Information Criterion (SBIC), root mean square error (RMSE), and normalized mean square error (NMSE), the best model was chosen, and its performance was evaluated using a variety of forecast accuracy metrics. Results indicated that across four distinct financial and economic data sets, which include the price of crude oil, the Nigerian stock market, the Nigerian all-shares index, and the Nigerian food and beverage index, the MARFIMA model performed better than the ARFIMA model. The research provides a more robust method for modeling and forecasting long memory data. The study has also contributed to existing literature on the most appropriate method for modelling long memory associated with financial and economic data.

1. Introduction

The idea of long memory features refers to the relationships or interdependencies between data items that have been gathered over an extended period of time. Long-term memory features were characterized in the studies by [6] and [7] as the progressive reduction in the graphical depiction of the autocorrelation function within a dataset. They proposed using fractional differencing in mean models when extended memory is found in time series data as a result of this phenomenon. The Autoregressive Tempered Fractional Integrated

Received: 21 Nov 2024.

Key words and phrases. Long memory; Simulations; Crude oil price; Nigerian stock exchange; Nigerian all-shares index; Nigerian food and beverage index; ARFIMA model; MARFIMA model.

Moving Average (ARTFIMA) model introduced by [9] and the Autoregressive Fractional Integrated Moving Average (ARFIMA) model proposed by [6] are noteworthy examples of long-memory mean models found in the literature. Other models in this category are the Beta-ARFIMA (βARFIMA) model by [12], the ARFURIMA model by [13], and the Semiparametric Fractional Autoregressive Moving average (SEMIFARMA) model by [1] etc.

Research has demonstrated that serial correlation is often seen in residuals produced from nonstationary mean models with long memory properties, such as ARFIMA, SARFIMA and ARTFIMA, as well as other mean models. Research by [18] as well as [3] has documented this finding.

Therefore, when dealing with time series data, previous models were unable to handle scenarios in which the fractional differencing value (d) could take any value greater than zero. However, because the MARFIMA model employs sequence fractional differencing, it is capable of handling this data when modeling extended memory in the mean. We introduce MARFIMA (p,d,q), a recently built and updated fractionally differenced model. This model's primary objective is to assess and address the problem of noise in large data sets, which has the potential to warp modeling approaches in terms of mean time series exhibiting properties of long-term memory.

The following are the fundamental attributes of the MARFIMA model that were determined and reported in this research. The fundamental characteristics of the Modified-ARFIMA (p,d,q) process and a sequential differencing filter for the Modified-ARFIMA model were deduced in Section 2. Furthermore, Section 2 provides a brief overview of the ARFIMA (p,d,q) model. ARFIMA model simulated data is used to evaluate the performance of the MARFIMA model. In Section 3, various applications were also shown to evaluate its qualities utilizing financial and econometric data (e.g., crude oil price, Nigerian Stock exchange, Nigerian All shares index, and Nigerian Food & beverage index).

2. The Material and Method

2.1 Data

In this part of the study, employed a simulation sample data for 1000, 2000, 5000, and 10000 were generated using ARFIMA (p, d, q) model. Also some applications were presented using financial and econometrics data to assess the performance of the developed MARFIMA (p, d, q) model.

2.2 ARFIMA (p, d, q) model

The Autoregressive Fractional Integrated Moving Average (ARFIMA) model is a time series model that combines autoregressive (AR), moving average (MA), and fractional differencing techniques to capture the long memory dependence often observed in financial and economic data.

ARFIMA models are particularly useful for time series data exhibiting long memory behavior, where past values have a persistent impact on future values. This makes them applicable in various fields such as finance, economics, and climatology, where traditional models may fail to adequately capture the underlying dynamics.

Meanwhile, the ARFIMA model with 0 < d < 1 as presented by [6] and [7] have differenced value of d to be fractional. Therefore the general form of ARFIMA model can be presented as follow;

$$\phi(L)(1-L)^d Y_m = \theta(L)\varepsilon_m \tag{1}$$

Where $\phi(L)=1-\phi_1 L-\phi_2 L^2-\cdots-\phi_p L^p$ and $\theta(L)=1-\theta_1 L-\theta_2 L^2-\cdots-\theta_q L^q$ are characteristic polynomials of AR and MA process, d is the fractional differencing filter, L is the backward shift operator, Y_m is the series and ε_m is the error term or white noise.

2.3 MARFIMA Model

Given a time series $Y_m, \ldots, Y_M$, the observations are assumed to be trendy, non-stationary and long memory. Also, they are assumed to have positive autocorrelation and long memory denoted as d and d is assumed to be in the range 1 < d < 1.5.

i. Consider a series {Ym}, m= 1, . . . , M. Estimate d by applying [5] semiparametric method for estimating long memory value. The GPH semiparametric model can be defined by:

ii. $$\hat{d}_{GPH} \to M\left(d, \frac{\pi^2}{6\sum_{i=1}^{m}(y_i-\bar{y})^2}\right) \tag{2}$$

Where M is the sample size, $\bar{y}$ is the mean of y_m and π is constant and d is fractional differencing order define in (ii). Also See [5] for the derivation of equation (2)

(2) Is the differencing operator,

$$\nabla Y_m = Y_m - Y_{m-1} \tag{3}$$

Therefore,

$$d(\nabla Y_{m-1}) = d(Y_{m-1} - Y_{m-2}) \tag{4}$$

minus equation (4) from equation (3) become

$$Q_m = \nabla Ym - d(\nabla Y_{m-1})$$

Where Q_m is the filter, ∇ is the differencing operator and d is a sequence fractional differencing operator [4].

Note,

$$Q_m = Y_m - dY_{m-1} - Y_{m-1} + dY_{m-2} \tag{5}$$

The method for obtaining the fractional filters to induce nonstationarity is as follows:

$$Q_m = \{(1 - L)(1 - dL)\}Y_m = Y_m - Y_{m-1} - d(Y_{m-1} - Y_{m-2}) \tag{6}$$

$$Q_{m-1} = \{(1 - L)(1 - dL)\}Y_{m-1} = Y_{m-1} - Y_{m-2} - d(Y_{m-2} - Y_{m-3})$$

$$\vdots$$

$$Q_{m-n} = \{(1 - L)(1 - dL)\}Y_{m-n} = Y_{m-n} - Y_{m-n-1} - d(Y_{m-n-1} - Y_{m-n-2}) \tag{7}$$

Therefore, the n^{th} sequence fractional series can be written as [13]

$$= Y_{m-n} - Y_{m-n-1} - d(Y_{m-n-1} - Y_{m-n-2})$$

$$=(L^n - dL^{n+1} - L^{n+1} + dL^{n+2})Y_m \tag{8}$$

Finally, the proposed fractional filter has the following general form:

$$Q_m = \sum_{n=0}^{N}(L^n - dL^{n+1} - L^{n+1} + dL^{n+2})Y_m \tag{9}$$

The ARMA model of [16] which became known to researches in time series is given as follows in the book [2].

$$\phi(L)Y_m = \theta(L)\varepsilon_m \tag{10}$$

$$\phi(L)(Q_m)Y_m = \theta(L)\varepsilon_m \tag{11}$$

where φ(L) and θ(L) are characteristic polynomials of AR and MA process, Q_m is the sequence fractional filter, Y_m is the series, L is the lag operator, and $\varepsilon_m \sim WN(0, \sigma^2)$. The lag representation of the proposed Modified-ARFIMA model shown below in equation (12)

$$\phi(L)\{(1-L)(1-dL)\}Y_m = \theta(L)\varepsilon_m \tag{12}$$

Where $\phi(L) = 1-\phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ and $\theta(L) = 1-\theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q$ are characteristic polynomials of AR and MA process, d is the fractional differencing filter, L is the backward shift operator, and $\varepsilon_m \sim WN(0, \sigma^2)$.

2.4 MARFIMA model Properties

To derived the properties of the Modified-ARFIMA model such as the mean, variance, autocovariance, autocorrelation and spectral density. The model in equation (12), is the Modified-ARFIMA (p,d,q) model. For convenience, let consider the Modified-ARFIMA (1, d, 1) model.

$$\{(1 - L)(1 - dL)\}Y_m = \varphi_1 Y_{m-1} + \varepsilon_m - \theta_1\varepsilon_{m-1} \tag{13}$$

$$(1 - dL - L + dL^2)Y_m = \varphi 1 Y_{m-1} + \varepsilon_m - \theta_1\varepsilon_{m-1} \tag{14}$$

$$Ym - dY_{m-1} - Y_{m-1} + dY_{m-2} = \varphi_1 Y_{m-1} + \varepsilon_m - \theta_1\varepsilon_{m-1} \tag{15}$$

$$Ym = (\varphi_1 + d + 1)Y_{m-1} - dY_{m-2} + \varepsilon_m - \theta_1\varepsilon_{m-1} \tag{16}$$

2.4.1 The Mean of the Modified-ARFIMA Model

To obtain the mean, let recalled equation (16) and take the expectation simplify as follows:

$$E(Y_m) = (\phi_1 + d + 1)E(Y_{m-1}) - dE(Y_{m-2}) + E(\varepsilon_m) - \theta_1 E(\varepsilon_{m-1}) \tag{17}$$

Reference to the definition of a white noise process [4]

$\varepsilon_m \sim \left(0, \sigma^2\right)$

$E(Y_m) = E(Y_{m-1}) = \cdots = E(Y_{m-k}) = \mu$ and $E(\varepsilon_m) = E(\varepsilon_{m-1}) = \cdots = E(\varepsilon_{m-k}) = 0$, the mean of the Modified-ARFIMA model is

$$\mu = 0 \tag{18}$$

2.4.2 The Variance of the Modified-ARFIMA Model

To obtain the variance, let recalled equation (16) and take the variance simplify as follows:

$$\mathrm{Var}(Y_m) = \mathrm{Var}((\varphi_1 + d + 1)Y_{m-1}) + \mathrm{Var}(-dY_{m-2}) + \mathrm{Var}(\varepsilon_m) + \mathrm{Var}(-\theta_1\varepsilon_{m-1}) \quad (19)$$

$$\sigma^2 = (\phi_1 + d + 1)^2\sigma^2 + d^2\sigma^2 + \sigma_\varepsilon^2 - \theta_1\sigma_\varepsilon^2$$

$$\sigma^2 = \frac{(1+\theta_1^2)\sigma_\varepsilon^2}{(1-(\phi_1 + d + 1)^2 - d^2)} \quad (20)$$

2.4.3 The Autocovariance of the Modified-ARFIMA Model

Equation (16) is multiplied by Ym, Ym-1, Ym-2, and Ym-k to obtain the auto-covariance γ_0, γ_1, γ_2, and γ_k, which is also referred to as variance in auto-covariance structure. It is best to interpret the expectation as follows [4]:

$$E(Y_mY_m) = (\phi_1 + d + 1)E(Y_mY_{m-1}) - dE(Y_mY_{m-2}) + E(\varepsilon_mY_m) - \theta_1E(\varepsilon_{m-1}Y_m) \quad (21)$$

$$\gamma_\circ = (\phi_1 + d + 1)\gamma_1 - d\gamma_2 + [1 - \theta_1(\phi_1 + d + 1) - \theta_1]\sigma_\varepsilon^2 \quad (22)$$

$$E(Y_{m-1}Y_m) = (\varphi_1 + d + 1)E(Y_{m-1}Y_{m-1}) - dE(Y_{m-1}Y_{m-2}) + E(\varepsilon_mY_{m-1}) - \theta_1E(\varepsilon_{m-1}Y_{m-1}) \quad (23)$$

$$\gamma_1 = (\phi_1 + d + 1)\gamma_\circ - d\gamma_1 - \theta_1\sigma_\varepsilon^2 \quad (24)$$

$$E(Y_{m-2}Y_m) = (\phi_1 + d + 1)E(Y_{m-1}Y_{m-2}) - dE(Y_{m-2}Y_{m-2}) + E(\varepsilon_mY_{m-2}) - \theta_1E(\varepsilon_{m-1}Y_{m-2}) \quad (25)$$

$$\gamma_2 = (\phi_1 + d + 1)\gamma_1 - d\gamma_\circ$$

.

.

.

$$E(Y_mY_{m-k}) = (\phi_1 + d + 1)E(Y_{m-1}Y_{m-k}) - dE(Y_{m-2}Y_{m-k}) + E(\varepsilon_mY_{m-k}) - \theta_1E(\varepsilon_{m-1}Y_{m-k}) \quad (26)$$

$$\gamma_k = (\phi_1 + d + 1)\gamma_{k-1} - d\gamma_{k-2} \quad (27)$$

2.4.4 The Autocorrelation Function of the Modified-ARFIMA Model

$$\rho_1 = \frac{(\phi_1 + d + 1)\gamma_0 - d\gamma_1 - \theta_1\sigma_\varepsilon^2}{(\phi_1 + d + 1)\gamma_1 - d\gamma_2 + [1 - \theta_1(\phi_1 + d + 1) - \theta_1]\sigma_\varepsilon^2} \quad (28)$$

$$\rho_2 = \frac{(\phi_1 + d + 1)\gamma_1 - d\gamma_0}{(\phi_1 + d + 1)\gamma_1 - d\gamma_2 + [1 - \theta_1(\phi_1 + d + 1) - \theta_1]\sigma_\varepsilon^2} \quad (29)$$

.

.

.

$$\rho_k = \frac{(\phi_1 + d + 1)\gamma_{k-1} - d\gamma_{k-2}}{(\phi_1 + d + 1)\gamma_1 - d\gamma_2 + [1 - \theta_1(\phi_1 + d + 1) - \theta_1]\sigma_\varepsilon^2} \quad \text{For } K=0,1,2,\ldots \quad (30)$$

2.4.5 The Spectral Density Function of the Modified-ARFIMA Model

To derive the spectral density function of the Modified-ARFIMA model recall from equation (12)

$$\phi(L)\{(1 - L)(1 - dL)\}Y_m = \theta(L)\varepsilon_m \tag{31}$$

the spectral density for the above equation (31) can be represented as this

$$S(\lambda) = \frac{\sigma^2}{2\pi}\left|\frac{\theta(e^{-i\lambda})}{\phi(e^{-i\lambda})(e^{-i\lambda})d(e^{-i\lambda})}\right|^2 \tag{32}$$

$$= \frac{\sigma^2\left|\theta(e^{-i\lambda})\right|^2}{2\pi\left|\phi(e^{-i\lambda})\right|^2\left|(e^{-i\lambda})\right|^2\left|d(e^{-i\lambda})\right|^2} \tag{33}$$

$$\left|\theta(e^{-i\lambda})\right|^2 = 1+\theta^2 - 2\theta\cos(i\lambda)$$

$$\left|\phi(e^{-i\lambda})\right|^2 = 1+\phi^2 - 2\phi\cos(i\lambda)$$

$$\left|(e^{-i\lambda})\right|^2 = 2 - 2\cos(i\lambda)$$

$$\left|d(e^{-i\lambda})\right|^2 = 1+d^2 - 2d\cos(i\lambda)$$

The spectral density for the Modified-ARFIMA model is presented in equation (34) below

$$S(\lambda) = \frac{\sigma^2[1+\theta^2 - 2\theta\cos(i\lambda)]}{2\pi[1+\phi^2-2\phi\cos(i\lambda)][2-2\cos(i\lambda)][1+d^2-2d\cos(i\lambda)]} \tag{34}$$

3. Parameter estimation of the Modified-ARFIMA (p, d, q) model

Consider series Y= $(y_1, y_2, \ldots, y_m)^d$, where $y_1, \ldots, y_m$ is a long memory process. In order to obtain the estimates of the Modified-ARFIMA model, the series Y_m is filtered by the non-power operator ε_m, we can estimate our parameters of the propose Modified-ARFIMA model from the spectral density of equation (34) above using Whittle log-likelihood estimator [14].

$$S_\lambda(\theta,\phi,d) = \frac{\sigma^2\left|\theta(e^{-i\lambda})\right|^2}{2\pi\left|\phi(e^{-i\lambda})\right|^2}[2-2\cos(i\lambda)]^{-1}[1+d^2-2d\cos(i\lambda)]^{-1} \tag{35}$$

Then,

$$\frac{\partial \log S_\lambda(\theta_j,\phi_j,d)}{\partial\theta_j,\partial\phi_j,\partial d} = \log[\sigma^2(\left|\theta_j(e^{-i\lambda j})\right|^2) - \log[2\pi(\left|\phi_j(e^{-i\lambda j})\right|^2] - \log[2-2\cos(i\lambda j)] - \log[1+d^2-2d\cos(i\lambda j)] \tag{36}$$

Estimation of $\theta,\phi,$ and d can be obtained by log-likelihood of equation (36) and this is referred as Whittle log- likelihood estimation [17]. We have

i. $$\frac{\partial \log S_\lambda(\theta_j,\phi_j,d)}{\partial\theta_j/\partial\phi_j,\partial d} = \log\left[\sigma^2\left(\left|\theta_j(e^{-i\lambda j})\right|^2\right)\right] \tag{37}$$

$$\log(\sigma^2) + \frac{\partial}{\partial\theta_j}\log\left(\left|\theta_j(e^{-i\lambda j})\right|\right) = 0 + \frac{\partial}{\partial\theta_j}\log\left(\left|\theta_j(e^{-i\lambda j})\right|\right)$$

$$= \frac{e^{-i\lambda_j}}{(1+\theta_1 e^{-i\lambda_1}+\theta_2 e^{-i\lambda_2}+\cdots+\theta_j e^{-i\lambda_j}+\cdots+\theta_q e^{-i\lambda_q})} + \frac{e^{i\lambda_j}}{(1+\theta_1 e^{i\lambda_1}+\theta_2 e^{i\lambda_2}+\cdots+\theta_j e^{i\lambda_j}+\cdots+\theta_q e^{i\lambda_q})} \tag{38}$$

$$= \theta_j^{-1}\left[e^{-i\lambda_j}\times e^{-i\lambda_j} + e^{i\lambda_j}\times e^{i\lambda_j}\right]$$

$$\frac{\partial y_j}{\partial\theta_j} = \theta_j^{-1}\left[2\cos(2\lambda_j)\right]$$

$$\hat{\theta}_j = e^{y_j-c} - e^{4\cos^2(\lambda_j)}\times e^{-2} \tag{39}$$

Where y_j is the indigenous variable will assume values between 0 and ∞: $0 < y_j < \infty$ for all $_j = 1, 2, \ldots,$. $-\pi < \lambda j < \pi$ and c for zero mean process c= 0.

ii. $$\frac{\partial \log S_\lambda(\theta_j,\phi_j,d)}{\partial\phi_j/\partial\theta_j,\partial d} = \log[2\pi(\left|\phi_j(e^{-i\lambda j})\right|^2] \tag{40}$$

$$\frac{\partial}{\partial\phi_j}\log(\left|\phi_j(e^{-i\lambda j})\right|^2) = \frac{\partial}{\partial\phi_j}\log(\phi_j(e^{-i\lambda j})) + \frac{\partial}{\partial\phi_j}\log(\phi_j(e^{i\lambda j}))$$

$$= \frac{e^{-i\lambda j}}{(1+\phi_1 e^{-i\lambda 1}+\phi_2 e^{-i\lambda 2}+\cdots+\phi_j e^{-i\lambda j}+\cdots+\phi_p e^{-i\lambda p})} + \frac{e^{i\lambda j}}{(1+\phi_1 e^{i\lambda 1}+\phi_2 e^{i\lambda 2}+\cdots+\phi_j e^{i\lambda j}+\cdots+\phi_p e^{i\lambda p})} \tag{41}$$

$$= \phi_j^{-1}[e^{-i\lambda j}e^{-i\lambda j} + e^{i\lambda j}e^{i\lambda j}]$$

$$\frac{\partial y_j}{\partial \phi_j} = \phi_j^{-1}[2\cos(2\lambda j)]$$

$$\hat{\phi}_k = e^{y_k - c} - e^{4\cos^2(\lambda_k)} \times e^{-2} \tag{42}$$

Where y_j is the indigenous variable will assume values between 0 and ∞: $0 < y_j < \infty$ for all $j = 1, 2, ..., .$ $-\pi < \lambda j < \pi$ and c for zero mean process c= 0.

iii.
$$\frac{\partial S_\lambda(\theta_j, \phi_j, d)}{\partial d / \partial \theta_j, \partial \phi_j} = -\log[2 - 2\cos(\lambda j)] - \log[1 + d^2 - 2d\cos(\lambda j)] \tag{43}$$

$$= 0 - \frac{\partial}{\partial d}[log(1 + d^2 - 2d\cos(\lambda j)]$$

$$\frac{\partial y_j}{\partial d} = \frac{2d - 2\cos(\lambda j)}{[1 + d^2 - 2d\cos(\lambda j)]} \tag{44}$$

$$\frac{\partial H}{\partial d} = 2d - 2\cos\lambda j, \partial H = (2d - 2\cos\lambda j)\partial d$$

$$\int \partial y_j = \int \frac{\partial H}{H} = \int \frac{1}{H}\partial H$$

$$y_j = \ln(1 + d^2 - 2d\cos\lambda j) + c$$

$$d^2 - 2d\cos\lambda j + 1 - e^{y_j - c} = 0$$

$$\hat{d} = \cos\lambda j \pm \sqrt{\cos^2\lambda j + e^{y_j - c} - 1} \tag{45}$$

4 Model selection method

The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), also known as the Schwarz-Bayesian Information Criterion (SBIC), are commonly used statistical measures for model selection in the context of regression analysis and other statistical modeling techniques. These criteria aim to balance the goodness of fit of a statistical model with the complexity of the model. Lower values of AIC and BIC indicate a better balance between goodness of fit and model complexity [10].

The Akaike Information Criteria is

$$AIC = M\ln\left[\frac{\hat{\sigma}_e^2}{M}\right] + 2P \tag{46}$$

Where M is the number of observations, $\hat{\sigma}_e^2$ is the variance of the error term, and P is the number of parameters of the model.

The Bayesian information criteria is an extension of the AIC that imposes a large penalty for additional coefficients. It is given as:

$$SBIC = M \ln\left[\frac{\hat{\sigma}_e^2}{M}\right] + P + P\ln(M) \quad (47)$$

Where $\hat{\sigma}_e^2$ is the variance of the error term, *ln(M)* where ***M*** is the number of observations in the dataset and P is the number of parameters of the model.

4.1 Measures of Forecast Accuracy

4.1.1 Root mean square error (RMSE)

The estimate for predicting error deviation is known as the root mean square error, or RMSE. To compute it, take the square root of the difference between the historical and anticipated observations, squared and averaged over the sample. An improved model estimate is shown by a reduced RMSE.

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \quad (48)$$

Where m is the number of observations, y_i is the actual observed value, $\hat{y}_i$ is the predicted value.

4.1.2 Normalize mean square error (NMSE)

Normalized mean square error (NMSE) is an alternative to the Mean Square Error (MSE), which accounts for the volume of data being examined. This normalization is useful when comparing models between Data sets or when there is a large variation in the variable scales.

$$NMSE = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{y_i - \hat{y}_i}{std(y)}\right|^2 \quad (49)$$

Where m is the number of observations, y_i is the actual observed value, $\hat{y}_i$ is the predicted value and $Std = \sqrt{\sum_{i=1}^{M}\frac{(y_i - \mu)^2}{M}}$ is the standard deviation of the actual values.

5. Results and Discussion

Results using Jarque-Bera, ADF, KPSS, and Portmanteau tests on the original series shows that the series is nonstationary and has a large amount of autocorrelation. The significance level was set at 5%. Initially, discrepancy lowers autocorrelation and increases stationarity. Stationarity and normalcy are further improved by second differencing, and all tests provide strong evidence for the conclusion that the series are stationary and more closely resemble a normal distribution.

Table 1 Model Selection criteria for Simulated data for n= 1000, 2000, 5000, and 10000

MARFIMA(p, d, q)			ARFIMA(p, d, q)		
1000					
Model	AIC	SBIC	Model	AIC	SBIC
MARFIMA (1, 1.0744, 1)	-1012.575	-997.8531	ARFIMA(1, 0.0331, 1)	6999.123	2766.723
MARFIMA (1, 1.0744, 2)	-953.7391	-934.108	ARFIMA(1, 0.0331, 2)	5890.819	2758.723
MARFIMA (2, 1.0744, 1)	-668.2608	-648.6298	ARFIMA(2, 0.0331, 1)	2866.417	2758.747
MARFIMA (2, 1.0744, 2)	-1283.586	-1259.047	ARFIMA(2, 0.0331, 2)	2812.741	2756.723
2000					
MARFIMA (1, 1.0401, 1)	-2236.276	-2219.475	ARFIMA(1, 0.0461, 1)	5902,934	5916.803
MARFIMA (1, 1.0401, 2)	-1457.672	-1435.27	ARFIMA(1, 0.0461, 2)	5759.35	5772.803
MARFIMA (2, 1.0401, 1)	-1680.969	-1658.567	ARFIMA(2, 0.0461, 1)	5628.575	5640.805
MARFIMA (2, 1.0401, 2)	-1493.651	-1465.649	ARFIMA(2, 0.0461, 2)	5609.482	5620.803
5000					
MARFIMA (1, 1.0224, 1)	-4079.572	-4060.021	ARFIMA(1, 0.0270, 1)	14836.08	14853.55
MARFIMA (1, 1.0224, 2)	-6436.672	-6410.604	ARFIMA(1, 0.0270, 2)	14424.07	14424.07
MARFIMA (2, 1.0224, 1)	-4581.74	-4555.673	ARFIMA(2, 0.0270, 1)	14137.17	14153.55
MARFIMA (2, 1.0224, 2)	-5022.148	-4989.563	ARFIMA(2, 0.0270, 2)	14091.46	14105.55
10000					
MARFIMA (1, 1.0211, 1)	-10516.12	-10494.49	ARFIMA(1, 0.0510, 1)	30192.04	30207.54
MARFIMA (1, 1.0211, 2)	-7254.419	-7225.578	ARFIMA(1, 0.0510, 2)	28652.1	28667.63
MARFIMA (2, 1.0211, 1)	-9971.506	-9942.665	ARFIMA(2, 0.0510, 1)	28536.39	28547.63
MARFIMA (2, 1.0211, 2)	-6581.702	-6545.651	ARFIMA(2, 0.0510, 2)	28538.39	28547.63

This table 1 presents a comparison of model selection criteria, specifically the Akaike Information Criterion (AIC) and the Schwarz Bayesian Information Criterion (SBIC), for different configurations of MARFIMA (Modified Autoregressive Fractionally Integrated Moving Average) and ARFIMA (Autoregressive Fractionally Integrated Moving Average) models based on simulation data samples (1000, 2000, 5000, and 10000). These criteria are used to assess the quality of statistical models, with lower values typically indicating a better model fit to the data. The MARFIMA model demonstrates superior performance over the ARFIMA model across all examined metrics and sample sizes (1000, 2000, 5000, and 10000). It shows better fit (lower AIC and SBIC).

Table 2 Forecast performance Measures for Simulated data for n= 1000, 2000, 5000, and 10000

1000					
Model	RMSE	NMSE	Model	RMSE	NMSE
MARFIMA (1, 1.0744, 1)	0.6006	0.0807	ARFIMA(1, 0.0331, 1)	9.4621	1.0592
MARFIMA (1, 1.0744, 2)	0.6179	0.0668	ARFIMA(1, 0.0331, 2)	5.4527	1.0528
MARFIMA (2, 1.0744, 1)	0.7127	0.0336	ARFIMA(2, 0.0331, 1)	1.0127	0.9993
MARFIMA (2, 1.0744, 2)	0.5235	0.2766	ARFIMA(2, 0.0331, 2)	0.9846	0.9994
2000					
MARFIMA (1, 1.0401, 1)	0.5706	0.9209	ARFIMA(1, 0.0461, 1)	1.0567	0.9996
MARFIMA (1, 1.0401, 2)	0.6929	0.9481	ARFIMA(1, 0.0461, 2)	1.0381	0.9997
MARFIMA (2, 1.0401, 1)	0.6553	0.9346	ARFIMA(2, 0.0461, 1)	0.9893	0.9995
MARFIMA (2, 1.0401, 2)	0.6863	0.9454	ARFIMA(2, 0.0461, 2)	0.9812	0.9995
5000					
MARFIMA (1, 1.0224, 1)	0.6645	0.9458	ARFIMA(1, 0.0270, 1)	1.0645	0.9998
MARFIMA (1, 1.0224, 2)	0.5248	0.8109	ARFIMA(1, 0.0270, 2)	1.0432	0.9999
MARFIMA (2, 1.0224, 1)	0.6318	0.9194	ARFIMA(2, 0.0270, 1)	1.0011	0.9998
MARFIMA (2, 1.0224, 2)	0.6045	0.8923	ARFIMA(2, 0.0270, 2)	0.9893	0.9998
10000					
MARFIMA (1, 1.0211, 1)	0.5908	0.9234	ARFIMA(1, 0.0510, 1)	1.0908	0.9999
MARFIMA (1, 1.0211, 2)	0.6954	0.9349	ARFIMA(1, 0.0510, 2)	2.1612	1.0002
MARFIMA (2, 1.0211, 1)	0.6071	0.9059	ARFIMA(2, 0.0510, 1)	1.4998	1.0002
MARFIMA (2, 1.0211, 2)	0.7191	0.9408	ARFIMA(2, 0.0510, 2)	1.4969	1.0002

Table 2 presents the forecast performance measures for two types of models, MARFIMA (Modified Autoregressive Fractionally Integrated Moving Average) and ARFIMA (Autoregressive Fractionally Integrated Moving Average), across four different sample sizes (1000, 2000, 5000, and 10000). The performance of these models is evaluated using two metrics: Root Mean Square Error (RMSE), and Normalized Mean Square Error (NMSE). MARFIMA models generally offer more accurate and higher predictive accuracy (lower RMSE and NMSE). This suggests that for the given data and across the range of parameter configurations tested, MARFIMA is the more efficient and accurate model for forecasting.

5 Application

This section presents the application of the proposed Modified Autoregressive Fractional Integrated Moving Average MARFIMA model by using financial and econometrics data such as crude oil price, Nigerian stock exchange, Nigerian all shares index, and Nigerian food and beverage index.

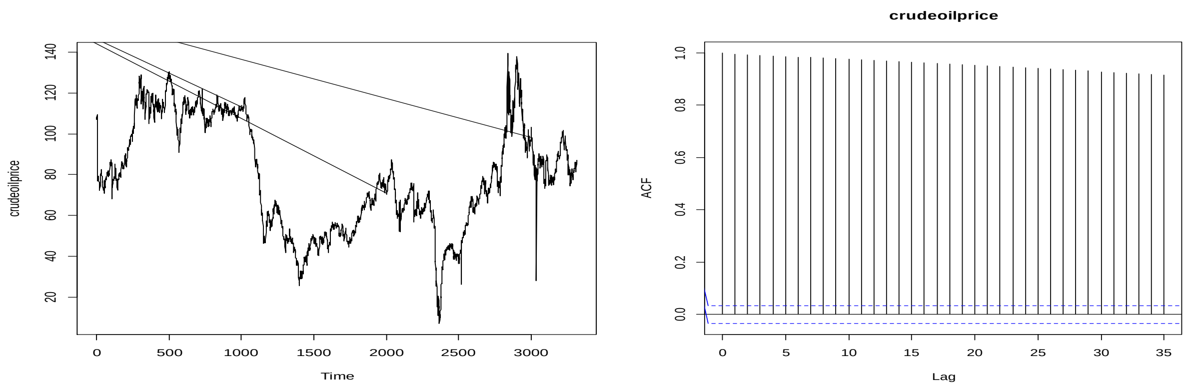

Figure 1: Plot of daily Crude oil price and its Autocorrelation function

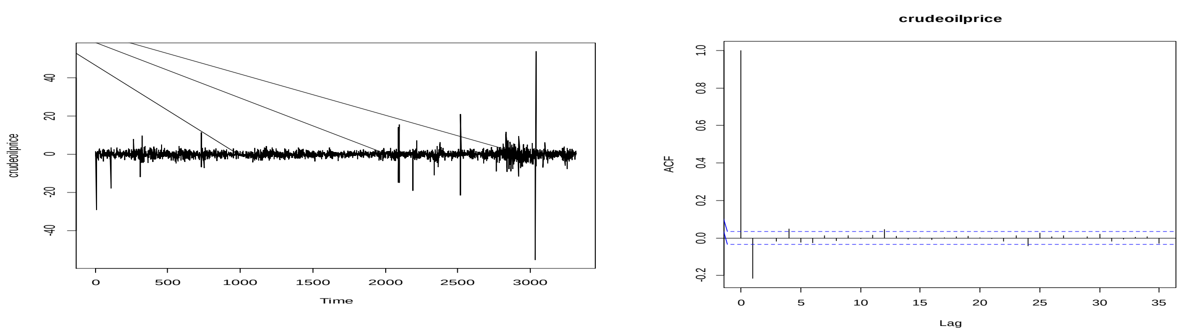

Figure 2: Plot of the sequence fractional difference for daily Crude oil price and its Autocorrelation function

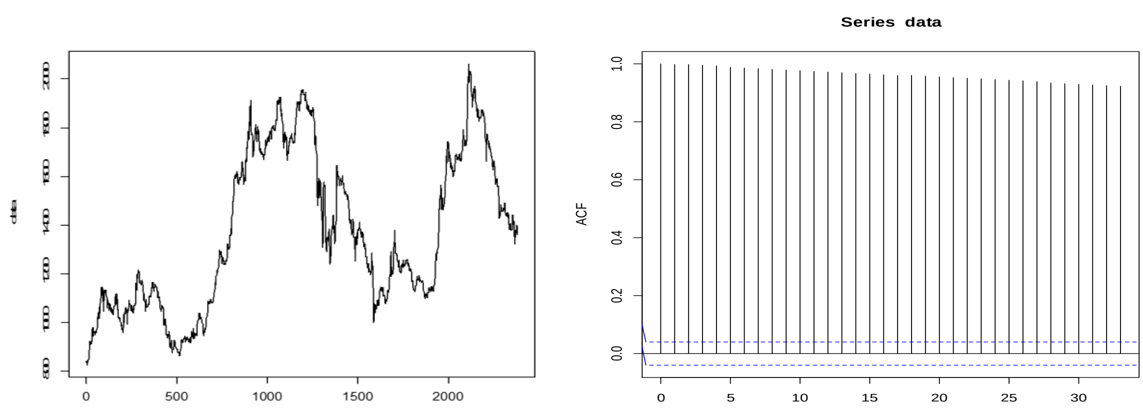

Figure 3: Plot of daily Nigerian stock exchange and its Autocorrelation function

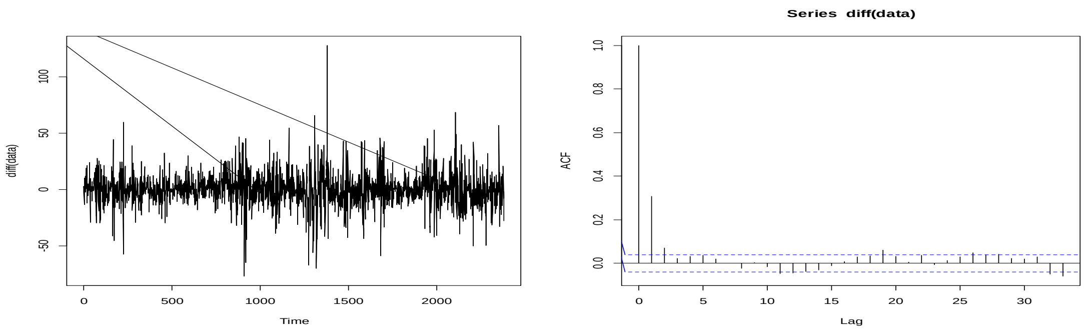

Figure 4: Plot of the sequence fractional difference for daily Nigerian stock exchange and its Autocorrelation function

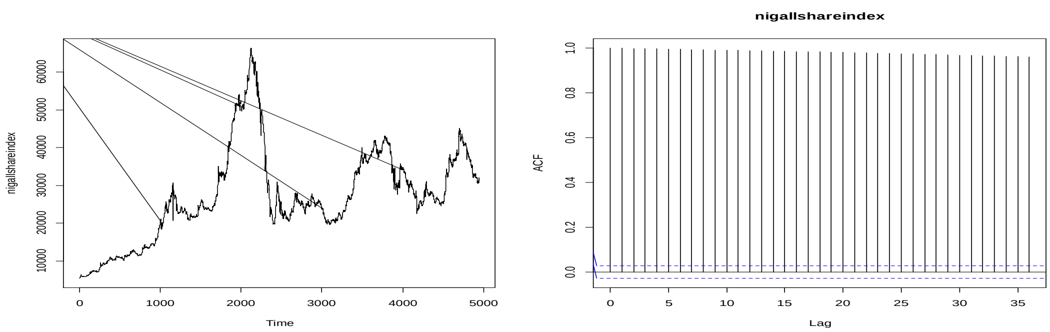

Figure 5: Plot of daily Nigerian all shares index and its Autocorrelation function

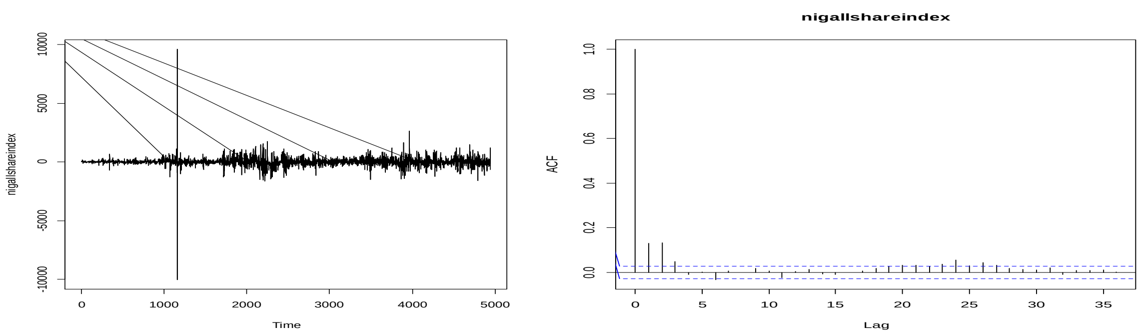

Figure 6: Plot of the sequence fractional difference for daily Nigerian all shares index and its Autocorrelation function

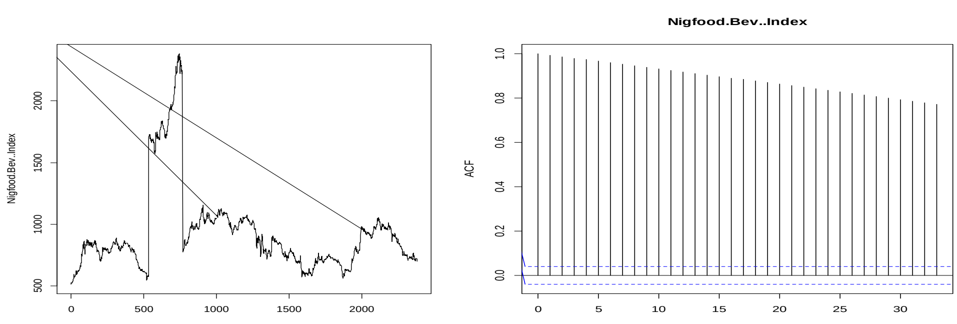

Figure 7: Plot of daily Nigerian food and beverage index and its Autocorrelation function

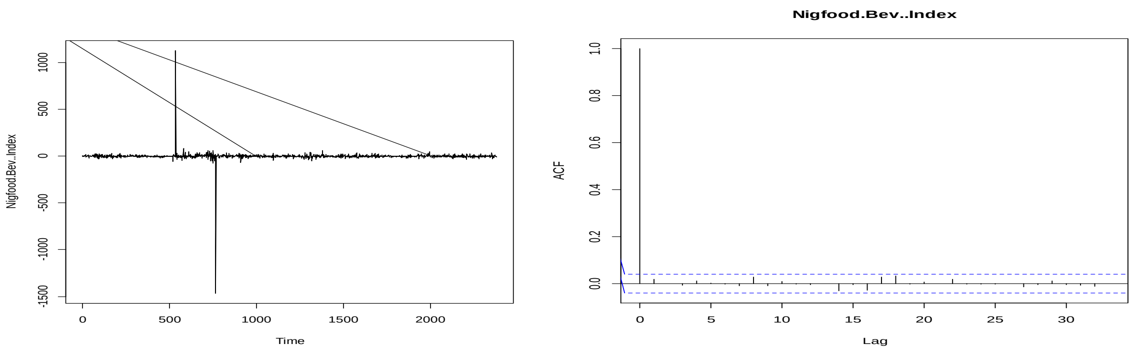

Figure 8: Plot of the sequence fractional difference for daily Nigerian food and beverage index and its Autocorrelation function.

Figure 1, 3, 5, and 7 are time series plots of the studied series exhibited nonstationary deterministic trends and all the ACF showed a very slow decay in the long range dependence or long term with positive autocorrelations, which provided evidence of the long memory process. While figure 2, 4, 6, and 8 indicate the plots series attains stationarity and all the ACF indicate stationarity which showed that, the long memory disappear.

Table 3 Stationary test for Crude oil price, Nigerian Stock exchange, Nigerian All shares index, and Nigerian Food & beverages index.

Stationary test	Crude oil Price	Nigerian Stock exchange	Nigerian All shares index	Nigerian Food & beverages index
Original Series				
Portmanteaus	3290.3(2.2e-16)	2375.5(2.2e-16)	4938.5(2.2e-16)	2351.2(2.2e-16)
ADF	-2.1249(0.5255)	-1.5631(0.7633)	-1.6065(0.745)	-3.1637(0.09415)
KPSS	7.7776(0.01)	8.0104(0.01)	14.278(0.01)	3.5009(0.01)
Jarque-Bera	135.67(2.2e-16)	176.84(2.2e-16)	293.53(2.2e-16)	4235.4(2.2e-16)
First difference				
Portmanteaus	153.22(2.2e-16)	224.04(2.2e-16)	83.773 (2.2e-16)	0.71719(0.3971)
ADF	-13.996(0.01)	-13.722(0.01)	-15.864(0.01)	-13.288(0.01)
KPSS	0.084986(0.1)	0.17481(0.1)	0.19416(0.1)	0.062858 (0.1)
Jarque-Bera	5135304(2.2e-16)	3789.8 (2.2e-16)	10994512(2.2e-16)	115521487(2.2e-16)

Table 3 presents results from stationary tests for four economic indicators: Crude Oil Price, Nigerian Stock Exchange, Nigerian All Shares Index, and Nigerian Food & Beverages

Index. The tests are performed on both the original series and the first differences of the series. The original series are nonstationary with significant autocorrelation, nonstationarity, and deviation from normality. For the first differences, all tests indicate stationarity and normality, strongly supporting the conclusion that the series are stationary and more closely follow a normal distribution.

Table 4 Model Selection criteria for Crude oil price, Nigerian Stock exchange, Nigerian All shares index, and Nigerian Food & beverage index.

MARFIMA(p, d, q)			ARFIMA(p, d, q)		
Model	AIC	SBIC	Model	AIC	SBIC
Crude oil price					
MARFIMA (1, 1.0960, 1)	6611.236	6629.556	ARFIMA(1, 0.0204, 1)	14897.15	14914.32
MARFIMA (1, 1.0960, 2)	4236.245	4260.671	ARFIMA(1, 0.0204, 2)	14888.62	14902.32
MARFIMA (2, 1.0960, 1)	4670.429	4694.855	ARFIMA(2, 0.0204, 1)	14899.03	14914.32
MARFIMA (2, 1.0960, 2)	2881.506	2912.038	ARFIMA(2, 0.0204, 2)	14889.44	14902.32
Nigerian Stock exchange					
MARFIMA (1, 1.0483, 1)	11867.13	11884.46	ARFIMA(1, 0.0461, 1)	19147.94	19163.33
MARFIMA (1, 1.0483, 2)	11209.15	11232.25	ARFIMA(1, 0.0461, 2)	19149.77	19163.33
MARFIMA (2, 1.0483, 1)	11821.16	11844.26	ARFIMA(2, 0.0461, 1)	19152.22	19165.33
MARFIMA (2, 1.0483, 2)	9968.224	9997.102	ARFIMA(2, 0.0461, 2)	19151.71	19163.33
Nigerian All shares index					
MARFIMA (1, 1.1645, 1)	58294.99	58314.51	ARFIMA(1, 0.4590, 1)	72067.25	72085.52
MARFIMA (1, 1.1645, 2)	54406.42	54432.44	ARFIMA(1, 0.4590, 2)	72048.13	72065.52
MARFIMA (2, 1.1645, 1)	55032.52	55058.55	ARFIMA(2, 0.4590, 1)	72038.01	72045.52
MARFIMA (2, 1.1645, 2)	53225.23	53257.76	ARFIMA(2, 0.4590, 2)	72038.17	72045.52
Nigerian Food and beverage index					
MARFIMA (1, 0.9493, 1)	17690.78	17708.11	ARFIMA(1, 0.0172, 1)	24280.29	24303.33
MARFIMA (1, 0.9493, 2)	15997.18	16020.28	ARFIMA(1, 0.0172, 2)	24279.67	24283.33
MARFIMA (2, 0.9493, 1)	16623.23	16646.33	ARFIMA(2, 0.0172, 1)	24279.66	24283.33
MARFIMA (2, 0.9493, 2)	15550.56	15579.44	ARFIMA(2, 0.0172, 2)	24253.65	24263.33

Table 4 shows the model selection criteria, specifically the Akaike Information Criterion (AIC) and Schwarz Bayesian Information Criterion (SBIC), for different models applied to four economic indicators: Crude Oil Price, Nigerian Stock Exchange, Nigerian All Shares Index, and Nigerian Food & Beverage Index. Both the MARFIMA (Modified Autoregressive

Fractionally Integrated Moving Average) and ARFIMA (Autoregressive Fractionally Integrated Moving Average) models are considered with varying orders (p, d, q). Across all economic indicators, the MARFIMA models with fractional integration orders and autoregressive orders of 2 consistently outperform other specifications based on lower AIC and SBIC values. These criteria suggest that the MARFIMA (2, 1.0960, 2), MARFIMA (2, 1.0483, 2), MARFIMA (2, 1.1645, 2), and MARFIMA (2, 0.9493, 2) models are preferred for forecasting the respective economic indicators.

Table 5 Parameter estimation for Crude oil price, Nigerian Stock exchange, Nigerian All shares index, and Nigerian Food & beverage index.

		Crude oil price			
MARFIMA (1, 1.0960, 2)	**Parameter estimate**	**P-value**	**MARFIMA (2, 1.0960, 2)**	**Parameter estimate**	**P-value**
φ_1	-0.5474	0.0142	φ_1	-0.7408	0.0162
θ_1	-1.9940	0.0011	φ_2	-0.3247	0.0160
θ_2	0.9945	0.0010	θ_1	-1.9989	0.0101
			θ_2	0.9992	0.1000
		Nigerian Stock exchange			
MARFIMA (1, 1.0483,2),	Parameter estimate	P-value	MARFIMA (2, 1.0483, 2)	Parameter estimate	P-value
φ_1	-0.2442	0.0199	φ_1	-0.4070 -	0.0182
θ_1	-1.9799	0.0046	φ_2	-0.2597	0.0184
θ_2	0.9805	0.0046	θ_1	-1.9978	0.0102
			θ_2	0.9999	0.2010
		Nigerian All shares index			
MARFIMA (1, 1.1645, 2)	Parameter estimate	P-value	MARFIMA (2, 1.1645, 2)	Parameter estimate	P-value
φ_1	-0.4735	0.0126	φ_1	-0.6130	0.0138
θ_1	-1.9886	0.0022	φ_2	-0.2519	0.0138
θ_2	0.9890	0.0022	θ_1	-1.9922	0.0011
			θ_2	0.9929	0.0010
		Nigerian Food & beverage index			
MARFIMA (1, 0.9493, 2)	Parameter estimate	P-value	MARFIMA (2, 0.9493, 2)	Parameter estimate	P-value
φ_1	-0.4740	0.0183	φ_1	-0.6089	0.0196
θ_1	-1.9898	0.0040	φ_2	-0.2833	0.0195
θ_2	0.9906	0.0040	θ_1	-1.9898	0.0029
			θ_2	0.9900	0.0029

The table 5 presents parameter estimates for Modified-Autoregressive Fractionally Integrated Moving Average (MARFIMA) models applied to four different financial indices: Crude oil price, Nigerian Stock Exchange, Nigerian All Shares Index, and Nigerian Food & Beverage Index. The results indicating the strength and significance of autoregressive effects, moving average effects, and fractional differencing parameters. These estimates are crucial for understanding the dynamics and predictive power of the models applied to each financial index.

Table 6 Forecast performance Measures for Crude oil price, Nigerian Stock exchange, Nigerian All shares index, and Nigerian Food & beverage index.

Crude oil price					
Model	RMSE	NMSE	Model	RMSE	NMSE
MARFIMA (1, 1.0960, 1)	2.2069	0.4937	ARFIMA(1, 0.0204, 1)	2.2838	0.9997
MARFIMA (1, 1.0960, 2)	1.8916	0.2194	ARFIMA(1, 0.0204, 2)	2.2836	0.9997
MARFIMA (2, 1.0960, 1)	2.0195	0.3554	ARFIMA(2, 0.0204, 1)	2.2843	0.9997
MARFIMA (2, 1.0960, 2)	1.5416	0.1821	ARFIMA(2, 0.0204, 2)	2.2836	0.9997
Nigerian Stock exchange					
MARFIMA (1, 1.0483, 1)	12.0556	0.6752	ARFIMA(1, 0.0461, 1)	13.4409	0.9996
MARFIMA (1, 1.0483, 2)	10.4960	0.3774	ARFIMA(1, 0.0461, 2)	13.44	0.9996
MARFIMA (2, 1.0483, 1)	11.93484	0.5601	ARFIMA(2, 0.0461, 1)	13.44	0.9996
MARFIMA (2, 1.0483, 2)	8.0857	0.3196	ARFIMA(2, 0.0461, 2)	13.44	0.9996
Nigerian All shares index					
MARFIMA (1, 1.1645 , 1)	362.719	0.5418	ARFIMA(1, 0.4590, 1)	373.21	0.9998
MARFIMA (1, 1.1645, 2)	244.752	0.2660	ARFIMA(1, 0.4590, 2)	352.39	0.9998
MARFIMA (2, 1.1645, 1)	260.747	0.4040	ARFIMA(2, 0.4590, 1)	352.05	0.9998
MARFIMA (2, 1.1645, 2)	217.155	0.2404	ARFIMA(2, 0.4590, 2)	351.99	0.9998
Nigerian Food & beverage index					
MARFIMA (1, 0.9493, 1)	40.9340	0.5644	ARFIMA(1, 0.0172, 1)	39.46	0.99958
MARFIMA (1, 0.9493, 2)	28.6754	0.2674	ARFIMA(1, 0.0172, 2)	39.45	0.99958
MARFIMA (2, 0.9493, 1)	32.7026	0.4274	ARFIMA(2, 0.0172, 1)	39.45	0.99958
MARFIMA (2, 0.9493, 2)	26.0984	0.2384	ARFIMA(2, 0.0172, 2)	39.22	0.99958

Table 6 presents forecast performance measures for different models applied to four different financial and economic indicators: Crude Oil Price, Nigerian Stock Exchange, Nigerian All Shares Index, and Nigerian Food & Beverage Index. The models used include MARFIMA (Modified Autoregressive Fractionally Integrated Moving Average) and

ARFIMA (Autoregressive Fractionally Integrated Moving Average). The forecast performance is evaluated based on two measures: Root Mean Squared Error (RMSE) and Normalized Mean Squared Error (NMSE). The MARFIMA models with fractional integration orders and autoregressive orders of 2 tend to outperform other specifications, as indicated by lower RMSE and NMSE values. These models demonstrate better accuracy in forecasting the respective financial and economic indicators, suggesting that they capture the underlying dynamics of the data more effectively.

6. Conclusion

Results obtained have shown MARFIMA model's robustness and superiority over the ARFIMA model in handling both simulated and real-life time series data using RMSE and NMSE as performance evaluation metrics. Using various tests and criteria, MARFIMA model demonstrated better fit and forecasting accuracy, making them a preferable choice for analysing financial and economic time series data. The application of these findings to real-world financial indices further validates the practical utility of the MARFIMA model in forecasting, making it an invaluable tool for financial analysts and researchers in modeling and predicting market dynamics.

Competing interests: The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] J. Beran, D. Ocker, SEMIFAR Forecasts, with Applications to Foreign Exchange Rates, J. Stat. Plan. Inference 80 (1999), 137–153. https://doi.org/10.1016/S0378-3758(98)00247-X.
[2] G.E.P. Box, G.M. Jenkins, Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, 1970.
[3] G. Duppati, A.S. Kumar, F. Scrimgeour, L. Li, Long Memory Volatility in Asian Stock Markets, Pac. Account. Rev. 29 (2017), 423–442. https://doi.org/10.1108/PAR-02-2016-0009.
[4] J.D. Cryer, K. Chan, Time Series Analysis: With Applications in R, 2nd ed, Springer, New York, 2008.
[5] J. Geweke, S. Porter-Hudak, The Estimation and Application of Long Memory Time Series Models, J. Time Ser. Anal. 4 (1983), 221–238. https://doi.org/10.1111/j.1467-9892.1983.tb00371.x.
[6] C.W.J. Granger, R. Joyeux, An Introduction to Long-Memory Time Series Models and Fractional Differencing, J. Time Ser. Anal. 1 (1980), 15–29. https://doi.org/10.1111/j.1467-9892.1980.tb00297.x.
[7] J.R.M. Hosking, Fractional Differencing, Biometrika, 68 (1981), 165-176.
[8] J.D. Cryer, K.-S. Chan, Time Series Analysis: With Applications in R, Springe, New York, 2015. https://doi.org/10.1007/978-0-387-75959-3.

[9] M.M. Meerschaert, F. Sabzikar, M.S. Phanikumar, A. Zeleke, Tempered Fractional Time Series Model for Turbulence in Geophysical Flows, J. Stat. Mech. Theor. Exp. 2014 (2014), P09023. https://doi.org/10.1088/1742-5468/2014/09/P09023.

[10] Y. Musa, M. Tasi'u, A. Bello, Forecasting of Exchange Rate Volatility between Naira and US Dollar Using GARCH Models, Int. J. Acad. Res. Bus. Soc. Sci. 4 (2014), 369-381. https://doi.org/10.6007/IJARBSS/v4-i7/1029.

[11] S. Porter-Hudak, An Application of the Seasonal Fractionally Differenced Model to the Monetary Aggregates, J. Amer. Stat. Assoc. 85 (1990), 338–344. https://doi.org/10.1080/01621459.1990.10476206.

[12] G. Pumi, M. Valk, C. Bisognin, F.M. Bayer, T.S. Prass, Beta Autoregressive Fractionally Integrated Moving Average Models, J. Stat. Plann. Inference 200 (2019), 196–212. https://doi.org/10.1016/j.jspi.2018.10.001.

[13] R.A. Rahman, S.A. Jibrin, A Modified Long Memory Model for Modeling Interminable Long Memory Process, in: L.-K. Kor, A.-R. Ahmad, Z. Idrus, K.A. Mansor (Eds.), Proceedings of the Third International Conference on Computing, Mathematics and Statistics (iCMS2017), Springer, Singapore, 2019: pp. 235–243. https://doi.org/10.1007/978-981-13-7279-7_29.

[14] F. Sabzikar, A.I. McLeod, M.M. Meerschaert, Parameter Estimation for ARTFIMA Time Series, J. Stat. Plann. Inference 200 (2019), 129–145. https://doi.org/10.1016/j.jspi.2018.09.010.

[15] M. Tasi'u, A. Usman, A.I. Ishaq, D. Hamisu, Modelling and Forecasting of Nigerian Naira to Saudi Riyal Exchange Rate Using ARIMA Framework, in: 2022 International Conference on Data Analytics for Business and Industry (ICDABI), IEEE, Sakhir, Bahrain, 2022: pp. 333–335. https://doi.org/10.1109/ICDABI56818.2022.10041592.

[16] P. Whittle, Hypothesis Testing in Time Series Analysis. Thesis, Uppsala University, 1951.

[17] W.W.S. Wei, Time Series Analysis: Univariate and Multivariate Methods, Pearson Addison Wesley, Boston Munich, 2006.

[18] J. Zhou, C. He, Modeling S & P 500 Stock Index Using ARMA-Asymmetric Power ARCH models, Thesis, Högskolan Dalarna, 2009.