BENIN JOURNAL OF STATISTICS
ISSN 2682-5767
Vol. 7, pp. 55– 68 (2024)

Hybrid Fuzzy Autoregressive Integrated Moving Average Model with Adjusted Fuzzy Number

M. A. Bakawu[1, *] , **A. Abdulkadir,** [2] **U. F. Abbas,** [3] **and A. M. Baba** [4]

[1,2,3]*Department of Mathematical Sciences, Abubakar Tafawa Balewa University Bauchi, Nigeria*
[4]*Department of Mathematical Sciences, University of Health Sciences, Azare, Nigeria*

(*Received: 10 October 2023; Accepted: 26 March 2024*)

Abstract. In recent years, a number of studies have demonstrated the efficiency, adaptability and accuracy of fuzzy time series forecasting models in the quest to improve the predictive power of the extant models. However, achievement of prediction accuracy is still one of the major challenges of these techniques. In this work, we proposed a model that combines the basic concepts of Fuzzy Autoregressive Integrated Moving Average (FARIMA) and Fuzzy Regression (FR) models which help to enhance prediction accuracy by narrowing down the projection bias problems specifically associated with the FARIMA model's interval of possibility. The method is tested on real data obtained from the literature. In most of the instances, the experimental results indicated that the proposed method achieves a narrower interval of possibility compared to the existing methods in the literature.

Keywords: Adjusted Fuzzy Number, Fuzzy Forecasting Methods, Fuzzy Regression, Fuzzy Time Series

Published by: Department of Statistics, University of Benin, Nigeria

1. Introduction

The quest of the behaviour of a variable in terms of its historical trends is quite an important facet in the viability of any reasonable decision making and such processes are regarded as forecasting. The forecasting concept is one of the most important topics of research that has attracted the concern of many researchers and scientists (Alyousifi *et al.,* 2021). Forecasting activities play an indisputably pivotal role in our daily life since curiosity is human's instinct (Dong and Ma, 2021). A significant part of scientific applications requires the forecasting of natural, social and economic processes. Hence, there is an ex-

*Corresponding author. Email: maiganabk@gmail.com

tensive literature on forecasting methods and models (Silva, 2019). In modern times, the importance of forecasting in various spheres of human activity does not require additional substantiation and proof (Gorbatiuk *et al.,* 2021). In addition, Panigrahi and Behera (2020) noted that accurate forecast of any event or phenomenon significantly can influence the decision of individuals as well as organizations. Therefore, a good forecasting method is an essential tool for decision-making.

In the literature, many authors applied different forecasting techniques in the course of achieving accurate forecasting. The notable technique among others include time series forecasting (TSF). The TSF is a commonly studied forecasting method that has many applications throughout various fields. The applications of time series forecasting in the current available literature are wide and cut across almost all aspects of human endeavours. Schools enrolments, stock index, market assets, economic indicators, exchange rates, electric loads, temperature and tourism forecasting to mention a few amongst many notable areas that were subjected to time series applications over the years.

The time series forecasting methods has been undergoing evolutionary transitions over the years. Tak (2022), reported that there is a great deal of effort to improve the predictive power of extant models. In the early seventies, Box and Jenkins (1970) proposed an important improvement that turn around the time series forecasting method. Their method is known as Autoregressive Integrated Moving Average (ARIMA) model. It is one of the most result oriented, well-known and widely used time series models. The method has proved to be more applicable to time series forecasting problems (Torbat *et al.,* 2018). The ARIMA model assumes that the future values of a time series model have a clear and definite functional relationship with current, past values and white noise (Wang *et al.,* 2009). It has the advantages of fast modelling and prediction, and is widely used in the prediction of time series data (Du *et al.,* 2020).

The application of the ARIMA model in some real life situations is constrained due to underlying model assumptions. One such assumption is the linear relationship between the future, current and past values of a time series. Added to this, is the requirement of large number of historical data, at least fifty, and preferably above one-hundred observations (Torbat *et al.,* 2018; Tseng and Tzeng, 2002). Furthermore, the ARIMA models cannot deal with forecasting problems in which the historical data are linguistic values and their forecasting accuracy measures are not good enough (Chen and Phuong, 2017). These limitations among others makes unfit the ARIMA models, because the ever uncertain environment created by the technological advancement requires less of the historical trends to forecast future patterns. Consequently, the search for an efficient forecasting framework that might handle these limitations remains an issue of intense research.

Fuzzy time series (FTS) was proposed in order to overcome the downsides of the traditional time series forecasting methods such as the ARIMA. FTS models do not require restrictive assumptions and too much background knowledge of the observations Yusuf *et al.* (2017), contrary to traditional TSF methods. An examination of the exiting literature shows that numerous experimental results based on fuzzy time series models were reported to have outperformed classical time series models. Moreover, despite great improvements published in the FTS

literature in the recent years, there are still some unaddressed downsides of the models.
To address the limitations and improve the forecasting performance of both traditional and fuzzy time series models, hybridisation was considered recently. The motivation for using hybrid models according to Khashei *et al.* (2013), comes from the assumption that either one cannot identify the true data generating process or that a single model may not be sufficient to identify all the characteristics of the time series. In this case, some researchers combine Fuzzy Time Series and ARIMA model (Xie *et al.,* 2021). This hybridised model is known as Fuzzy Autoregressive Integrated Moving Average (FARIMA) model. In FARIMA models, instead of using crisp parameters, fuzzy parameters, in the form of triangular fuzzy numbers are used (Torbat *et al.,* 2017). Consequently, the use of the fuzzy parameter reduces the need for large historical data unlike ARIMA models. Tseng *et al.* (2001) first proposed the concept of FARIMA in an effort to improve foreign exchange forecasting. Thereafter, many studies were conducted and good results obtained using the FARIMA model. Xie *et al.* (2021) developed a fuzzy ARIMA correction model for transport volume forecast capable of long term prediction. Medina-Reyes *et al.* (2021) proposed a new hybrid fuzzy time series model based on Fuzzy Time Series and Fuzzy ARIMA that achieved better in-sample and out-sample accuracy tests. Torbat *et al.* (2018) using FARIMA model forecasted the Iran's steel consumption with improved accuracy. Mehdi *et al.* (2019), reported that when the sample period is shorter, the prediction using Fuzzy ARIMA model is better than other models. By using quadratic approach, Wang *et al.* (2009) demonstrated the application of FARIMA models on simulated time series data set and a better result was achieved. However, the recent hybridisation of the FARIMA method with other methods like Probabilistic Neural Network (PNN), and Sliding Window concept and integration of effect factors into Objective Function in Torbat *et al.* (2017), Mehdi *et al.* (2019) and Xie *et al.* (2021), respectively, is an indication that achievement of predictions accuracy is still one of the major challenges of the FARIMA method. Hence, a hybrid model on the basis of fuzzy regression and fuzzy ARIMA models is proposed.
Fuzzy linear regression is used in evaluating the functional relationship between the dependent and independent variables in a fuzzy environment (Alsoltany and Alnaqash, 2015). A number of studies have been conducted in order to demonstrated the applications of fuzzy regression method. Most recently, applications of fuzzy regression were proposed in Malyaretz *et al.* (2018); Lee *et al.* (2020); Taheri *et al.* (2020) and Attanayake (2021).
This paper aims to enhance the prediction accuracy through narrowing down the projection bias problems specifically associated with FARIMA model's interval of possibility. We use real data for verifying the performance of the proposed method. The data is related to gross domestic product (GDP), unemployment rate, inflation rate, and foreign direct investment (FDI). The rest of the paper is structured as follows: In section 2, Materials and Method is presented, followed by Results and Discussion in section 3. Finally, conclusion is provided in section 4.

2. Materials and Methods

2.1 *Fuzzy Regression and Fuzzy Autoregressive Integrated Moving Average*

Fuzzy linear regression is a fuzzy type of classical regression analysis in which some elements of the model are represent by fuzzy numbers (Alsoltany and Al-naqash, 2015). The functional relationship between the response and explanatory variables as reported in Tanaka *et al.* (1982) is presented as follows:

$$\tilde{Y} = \tilde{A}_0 + \tilde{A}_1 x_1 + \tilde{A}_2 x_2 + \cdots + \tilde{A}_k x_k \tag{1}$$

Where:
$\tilde{Y}$ is the fuzzy output,
$X = (x_{0i}, x_{1i}, x_{2i}, \cdots, x_{ki})^T$, k-dimensional crisp input vector,
$\tilde{A} = \tilde{A}_0, \tilde{A}_1, \tilde{A}_2, \cdots, \tilde{A}_k)^T$, fuzzy vector of coefficients presented in the form of a symmetric triangular fuzzy number denoted by $\tilde{A}_k = [\gamma_k, \beta_k]$, respectively, $\gamma_k and \beta_k$ are its center and width, while $x_{0i} = 1$, $i = 1, \dots n$.

The FARIMA model utilises same formulation as the FR model, except the explanatory variables are lagged values of the response variable and the associated residuals. Hence, the following is the generalised FARIMA (p,d,q) model:

$$\tilde{\omega}_p(L) Y_t^* = \tilde{\tau}_q(L) \epsilon_t \tag{2}$$

$$Y_t^* = \Delta^d (Y_t - \mu) \tag{3}$$

The extended form of Equation (2) is given in Equation (4):

$$\tilde{Y}_t^* = \tilde{\omega}_0 + \tilde{\omega}_1 Y_{t-1}^* + \tilde{\omega}_2 Y_{t-2}^* + \cdots + \tilde{\omega}_p Y_{t-p}^* + \epsilon_t - \tilde{\tau}_1 \epsilon_{t-1} - \tilde{\tau}_2 \epsilon_{t-2} - \cdots - \tilde{\tau}_q \epsilon_{t-q} \tag{4}$$

Where, Equation (3) is the ARIMA process of the time series Y_t, t is the time, $\Delta = 1 - L$, the difference operator, L is a lag operator; generally, $L^n Y_t = Y_{t-n}, Y_t$ are observations, while, $\tilde{\omega}_0, \tilde{\omega}_1, \tilde{\omega}_2, \cdots, \tilde{\omega}_p$ and $\tilde{\tau}_1, \tilde{\tau}_2, \cdots, \tilde{\tau}_q$ are fuzzy numbers.
The structure of the FARIMA (p,d,q) model is built on the ARIMA process of the time series Y_t. Thus, p is the order of the Autoregressive term, q is the order of the Moving Average term, while d is the differencing order needed to achieve stationarity of the time series Y_t. In addition, autocorrelation function (ACF) and partial autocorrelation function (PACF) are primary tools used to develop the structure of the ARIMA model. The sample ACF plot and the sample PACF plot are compared to the theoretical behaviour of these plots when the differencing order is known. Equation (4) is modified to obtain Equation (5):

$$\tilde{Y}_t^* = \tilde{A}_0 + \tilde{A}_1 Y_{t-1}^* + \tilde{A}_2 Y_{t-2}^* + \cdots + \tilde{A}_p Y_{t-p}^* + \epsilon_t - \tilde{A}_{p+1} \epsilon_{t-1} - \tilde{A}_{p+2} \epsilon_{t-2} - \cdots - \tilde{A}_{p+q} \epsilon_{t-q} \tag{5}$$

Where, $\tilde{A}_0, \tilde{A}_1, \tilde{A}_2, \cdots, \tilde{A}_p; \tilde{A}_{p+1}, \tilde{A}_{p+2}, \cdots \tilde{A}_{p+q}$ are fuzzy parameters.

2.2 *Determination of the Fuzzy Parameters*

A symmetrical fuzzy number A_j denoted as $\tilde{A}_j = [\gamma_j, \beta_j]$ is defined as $\mu_{A_j}(a_j) = L((a_j - \gamma_j)/\beta_j), \beta_j > 0$, where, $L(a_j)$ is a shape function, γ_j is a centre and β_j is a width of fuzzy number. According to Tanaka and Watada (1989), $L(a_j)$ is defined by:

i. $L(a_j) = L(-a_j)$,
ii. $L(0) = 1$,
iii. $L(a_j)$ is strictly decreasing function for $a_j \geq 0$,
iv. $\{a_j | L(a_j) \geq 0\}$ is a closed interval.

For each type of A_j, the membership functions are assumed triangular. By definition, it can be expressed as:

$$\mu_{\tilde{A}_j}(a_j) = \begin{cases} 1 - \frac{|\gamma_j - a_j|}{\beta_j}, & \text{if } \gamma_j - \beta_j \leq a_j \leq \gamma_j + \beta_j \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

$\beta_j > 0$.
According to the extension principle (Zadeh, 1975), the membership function of the fuzzy numbers $\tilde{Y}$ and $\tilde{Y}_t^*$ are respectively given in Equations (7) and (8):

$$\mu_Y(y) = \begin{cases} \text{Max}\left(0, 1 - \frac{|y - \sum_{j=0}^{p} \gamma_j x_{ij}|}{\sum_{j=0}^{p} \beta_j x_{ij}}\right), & \text{if } x_{ij} \neq 0 \\ 1 & \text{if} x_{ij} = 0, y \neq 0 \\ 0 & \text{if} x_{ij} = 0, y = 0 \end{cases} \tag{7}$$

The spread of $\tilde{y}$ is $\sum_{j=0}^{p} \beta_j x_{ij}$ and the middle value of $\tilde{y}$ is $\sum_{j=0}^{p} \gamma_j x_{ij}$.

$$\mu_{\tilde{Y}^*}(Y_t^*) = \begin{cases} 1 - \frac{|Y_t^* - \sum_{j=0}^{p} c_j Y_{t-j}^* - \epsilon_t + \sum_{j=p+1}^{p+q} c_j \epsilon_{t+p-j}|}{\sum_{j=0}^{p} w_j Y_{t-j}^* + \sum_{j=p+1}^{p+q} w_j \epsilon_{t+p-j}}, & \text{for } Y_t^* \neq 0, \epsilon_t \neq 0 \\ 0 & \text{Otherwise} \end{cases} \tag{8}$$

Where c_j is a centre, w_j is a width of a fuzzy number.

2.3 *Formulation of the Proposed Model*

The proposed model combined the basic concept of FR and FARIMA models in order to enhance prediction accuracy by narrowing down specifically the projection bias problems associated with FARIMA model. In other words, unlike the existing FARIMA methods, the proposed model combines the history effect and influence factors both in the objective function and constraints in the quest to improve prediction accuracy. We note that combining the history and influence factors into a single objective function may not necessarily provide new central values capable of narrowing the FARIMA interval of possibility. Therefore, a hybrid search method that is capable of generating new central values and narrowing the interval of possibility is proposed. The details of the proposed method are provided as follows:

Fitting FR and FARIMA models on an observed data are the first stage of the proposed model. The linear programming formulation (LP) of these models are presented in what follows:

Objective Function: We seek to find the coefficients $\tilde{A}_k = [\gamma_k, \beta_k]$ that minimize the spread of the fuzzy output for all data sets. Mathematically, for the FR, this becomes:

$$\text{Min S} = \sum_{i=1}^{n}\sum_{k=0}^{K} \beta_k |x_{ik}| \tag{9}$$

Similarly, for a FARIMA problem of order (p,d,q) with coefficients $\tilde{A}_j = [c_j, w_j]$, the objective function is given as in Equation (10):

$$\text{Min S} = \sum_{t=1}^{n}\sum_{j=0}^{p} w_j |\phi_{jj}||Y^*_{t-j}| + \sum_{t=1}^{n}\sum_{j=p+1}^{p+q} w_j |\rho_{j-p}||\epsilon_{t+p-j}| \tag{10}$$

Where: w_j, the width or spread around the center of the fuzzy number; ϕ_{jj}, the partial autocorrelation coefficient of time lag j; ρ_{j-p}, the autocorrelation coefficient of time lag j-p; Y^*_{t-j}, differenced time series of Y_t at time lag $t-j$; $Y^*_t = \Delta^d(Y_t - \mu)$; and ϵ_t, ARIMA residuals at time t. The residuals, ϵ_t, are assumed to be independently and identically distributed with a mean of zero and a constant variance σ^2.

Constraints: The constraints require that each observation y_i (or y_t in the case of FARIMA model) has a threshold value h in the interval $(0, 1)$ which is specified by the user of belonging to $\tilde{y}(y)$ (Taghizadeh *et al.,* 2009). This implies,

$$\tilde{y}(y_i) \geq h, i = 1, 2, \cdots, n \tag{11}$$

After separately substituting Equations (7) and (8) into Equation (11), the simplified resulting LP models along the respective objective functions (see Bakawu *et al.,* 2023) are obtained as Equations (12) and (13) for the FR and FARIMA models respectively.

$$\begin{aligned} \text{Min S} \quad &= \quad \sum_{i=1}^{n}\sum_{k=0}^{K} \beta_k |x_{ik}| \\ \text{s.t} \quad & \sum_{k=0}^{K} \gamma_k x_{ik} - (1-h)\sum_{k=0}^{K} \beta_k |x_{ik}| \leq y_i, \forall\, i = 1, \cdots, n \\ & \sum_{k=0}^{K} \gamma_k x_{ik} + (1-h)\sum_{k=0}^{K} \beta_k |x_{ik}| \geq y_i, \forall\, i = 1, \cdots, n \\ & \beta_k \geq 0, x_{i0} = 1; i = 1, \cdots, n. \end{aligned} \tag{12}$$

Where, γ_k and β_k, for $k = 0, 1, ..., K$ are the FR unknown fuzzy variables.

$$
\begin{aligned}
\text{Min S} \quad &= \quad \sum_{t=1}^{n}\sum_{j=0}^{p} w_j|\phi_{jj}||Y_{t-j}^*| + \sum_{t=1}^{n}\sum_{j=p+1}^{p+q} w_j|\rho_{j-p}||\epsilon_{t+p-j}| \\
\text{s.t} \quad & \qquad \sum_{j=0}^{p} c_j\phi_{jj}Y_{t-j}^* + \epsilon_t + \sum_{j=p+1}^{p+q} c_j\rho_{j-p}\epsilon_{t+p-j} - \\
& (1-h)\left(\sum_{j=0}^{p} w_j|\phi_{jj}||Y_{t-j}^*| + \sum_{j=p+1}^{p+q} w_j|\rho_{j-p}||\epsilon_{t+p-j}|\right) \leq Y_t^*, \forall\, t = 1, \cdots, n \\
& \qquad \sum_{j=0}^{p} c_j\phi_{jj}Y_{t-j}^* + \epsilon_t + \sum_{j=p+1}^{p+q} c_j\rho_{j-p}\epsilon_{t+p-j} + \\
& (1-h)\left(\sum_{j=0}^{p} w_j|\phi_{jj}||Y_{t-j}^*| + \sum_{j=p+1}^{p+q} w_j|\rho_{j-p}||\epsilon_{t+p-j}|\right) \leq Y_t^*, \forall\, t = 1, \cdots, n \\
& w_j \geq 0; \forall j = 0, \cdots, p+q
\end{aligned}
\tag{13}
$$

Similarly, w_j, c_j, for $j = 0, 1, ..., p + q$. are the FARIMA unknown variables, ρ_{j-p} is the autocorrelation coefficient of time lag j-p and ϕ_{jj} is the partial autocorrelation coefficient of time lag j. Based on the results in Equations (12) and (13), the relation in Equations (1) and (5) can be rewritten in Lower Bond (LB) and Upper Bond (UB) form as follows:

$$
\begin{aligned}
y_{r(LB)} &= (\gamma_0 - \beta_0) + (\gamma_1 - \beta_1)x_{1i} + (\gamma_2 - \beta_2)x_{2i} + \cdots + (\gamma_k - \beta_k)x_{ki} \\
y_{r(UB)} &= (\gamma_0 + \beta_0) + (\gamma_1 + \beta_1)x_{1i} + (\gamma_2 + \beta_2)x_{2i} + \cdots + (\gamma_k + \beta_k)x_{ki}
\end{aligned}
\tag{14a}
$$

$$
\begin{aligned}
Y_{S(LB)}^* &= (c_0 - w_0) + (c_1 - w_1)Y_{t-1}^* + (c_2 - w_2)Y_{t-2}^* + \cdots + (c_p - w_p)Y_{t-p}^* + \epsilon_t \\
&\quad -(c_{p+1} - w_{p+1})\epsilon_{t-1} - (c_{p+2} - w_{p+2})\epsilon_{t-2} - \cdots - (c_{p+q} - w_{p+q})\epsilon_{t-p} \\
Y_{S(UB)}^* &= (c_0 + w_0) + (c_1 + w_1)Y_{t-1}^* + (c_2 + w_2)Y_{t-2}^* + \cdots + (c_p + w_p)Y_{t-p}^* + \epsilon_t \\
&\quad -(c_{p+1} + w_{p+1})\epsilon_{t-1} - (c_{p+2} + w_{p+2})\epsilon_{t-2} - \cdots - (c_{p+q} + w_{p+q})\epsilon_{t-p}
\end{aligned}
\tag{14b}
$$

The interval prediction models, that is Equations (14a) and (14b) makes it possible to forecast the best and worst possible values of $\tilde{Y}$ based on predetermined values of $X = (x_{1i}, x_{2i}, \cdots, x_{ki})$ when FR model is considered or the lagged values of Y_t^* in the case of FARIMA model.

Combining the FR and FARIMA results using a probabilistic weight factors as demonstrated in Wang *et al.* (2009), generates the series y_{LB} and y_{UB}. The mathematical formulation is given in Equation (15):

$$
\begin{aligned}
y_{LB} &= \alpha_r y_{r(LB)} + \alpha_S Y_{S(LB)}^* \\
y_{UB} &= \alpha_r y_{r(UB)} + \alpha_S Y_{S(UB)}^*
\end{aligned}
\tag{15}
$$

Where, y_{LB} is the new lower control limit, y_{UB} is the new upper control limit;

while, α_r is the FR's weight and α_S is the FARIMA's weight. Therefore, the LP formulation of the proposed hybrid model that combines the history effects and other influence factors in both objective function and the constraints as in what follows:

$$\begin{aligned}
\text{Min S} \quad &= \sum_{t=1}^{n}\sum_{j=0}^{p} \alpha_j |\phi_{jj}||Y^*_{t-j}| + \sum_{i=1}^{n}\sum_{k=1}^{K} \alpha^*_k |x_{ik}| \\
\text{s.t} \quad & \sum_{j=0}^{p} \delta_j Y^*_{t-j} + \epsilon_t - (1-h)\sum_{k=1}^{K} \alpha^*_k |x_{ik}| \leq y_{LB}, \forall\, i = 1, \cdots, n; \\
& \sum_{j=0}^{p} \delta_j Y^*_{t-j} + \epsilon_t + (1-h)\sum_{k=1}^{K} \alpha^*_k |x_{ik}| \geq y_{UB}, \forall\, i = 1, \cdots, n; \\
& \alpha_j, \alpha^*_k \geq 0; j = 0, \cdots, p; k = 1, \cdots, K; t = 1, \cdots, n
\end{aligned} \tag{16}$$

Where, δ_j $(j = 0, \ldots, p)$ are the new central values used for the adjustment of the FARIMA model. In contrast to the existing models in the literature, this proposed model is capable of minimising the total spread resulting from the history effects and influence factors. Additionally, the result of the model also determines the new central values for adjusting the fuzzy numbers. The training steps and the prediction equation for the Hybrid Fuzzy Autoregressive Integrated Moving Average Model with Adjusted Fuzzy Number (FARIMA-AFN) are presented in Figure 1 and Equation (17):

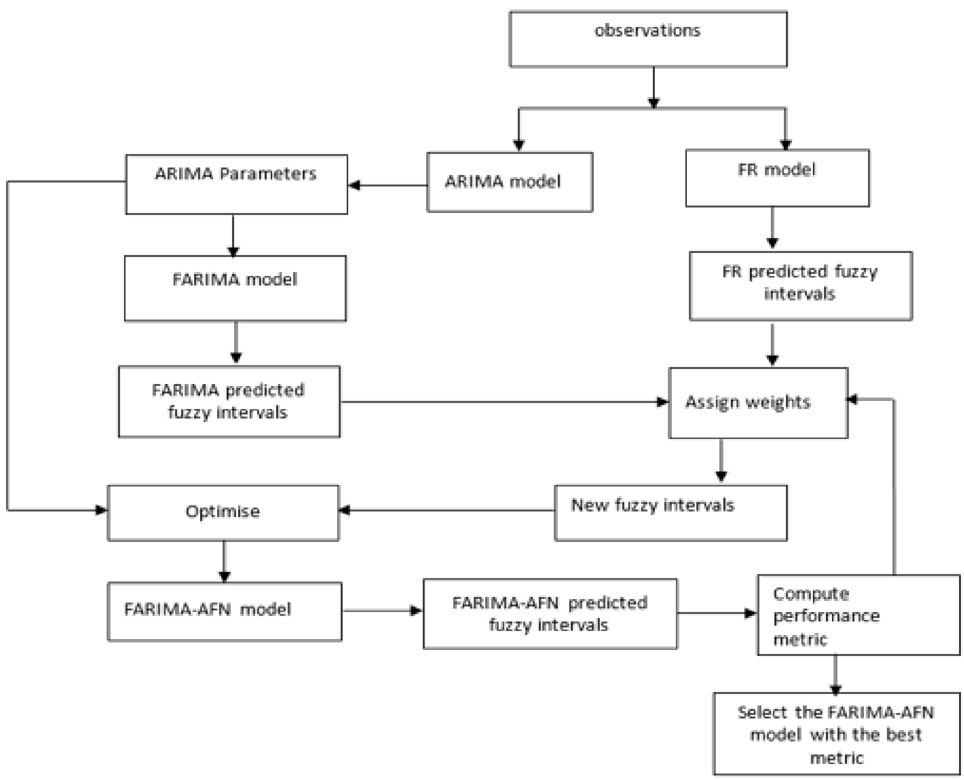

Figure 1: FARIMA-AFN training steps.

$$\tilde{Y}_t^* = (\delta_0, w_0) + (\delta_1, w_1)Y_{t-1}^* + (\delta_2, w_2)Y_{t-2}^* + \cdots + (\delta_p, w_p)Y_{t-p}^* \quad (17)$$

3. Results and Discussion

In this section, we solved the LP problems (Equations (12), (13) and (16)) with a threshold level of $h = 0$ in order to determine the minimal fuzziness of the models using real data related to GDP (Y), unemployment rate (X_1), inflation rate (X_2), and FDI (X_3) obtained in Ogosi *et al.* (2022). The empirical results are presented in four phases as in what follows:

Phase I. FR and FARIMA Parameters Estimation: the fuzzy parameters are obtained by solving Equations (12) and (13) using Tora Optimization Software (Taha, 2011). The central values and widths of each fuzzy parameter in equations (1) and (5) for $h = 0$ were obtained and presented in Tables 1 and 2 along with the corresponding lower bound (LB) and upper bound (UB) respectively.

Table 1: Central and widths values for fuzzy parameters for FR

Fuzzy parameters	Centre	Width	UB	LB
A_0	4.021	0.116	4.137	3.906
A_1	1.005	0.000	1.005	1.005
A_2	-0.145	0.022	-0.124	-0.167
A_3	0.376	0.000	0.376	0.376

Table 2: Central and widths values for fuzzy parameters for FARIMA

Fuzzy parameters	Centre	Width	UB	LB
A_0	0.074	0.000	0.074	0.074
A_1	0.990	0.007	0.983	0.997

The estimated fuzzy linear regression model for the real GDP $(\tilde{y})$ of Nigeria against the three macroeconomic factors and the FARIMA model are provided in Equations (18a) and (18b) respectively:

$$GDP(\tilde{y}) = (4.021, 0.116) + (1.005, 0.000)x_1 + (-0.145, 0.022)x_2 + (0.376, 0.000)x_3 \quad (18a)$$

$$GDP(\tilde{y}_t) = (0.074, 0.000) + (0.990, 0.007)Y_{t-1}^* \quad (18b)$$

Phase II. FARIMA-AFN Parameters Estimation: four sets of alternating weights (i.e. $0.1\&0.9$; $0.2\&0.8$; $0.3\&0.7$; $0.4\&0.6$) are examined, that is, by changing the weights α_r (FR's weight) and α_S (FARIMA's weight). The best results are obtained when $\alpha_r = 0.6$ and $\alpha_S = 0.4$. Thus, the new lower and upper control series in equation (15) is generated as follows:

$$\begin{aligned} y_{LB} &= 0.6y_{r(LB)} + 0.4Y_{S(LB)}^* \\ y_{UB} &= 0.6y_{r(UB)} + 0.4Y_{S(UB)}^* \end{aligned} \quad (19)$$

Subsequently, Equation (16) is solved to obtain the new centre values as shown in Table 3. The new centre values replace the FARIMA centre values as an adjusted fuzzy numbers. Consequently, the prediction equation for the Hybrid FARIMA-AFN is given as in Equation (20).

Table 3: Central and widths values for Hybrid FARIMA-AFN

Fuzzy parameters	Centre	Width	UB	LB
A_0	0.000	0.000	0.000	0.000
A_1	1.004	0.007	0.997	1.011

$$GDP(\tilde{y}_t) = (0.000, 0.000) + (1.004, 0.007)Y^{*}_{t-1} \tag{20}$$

In Table 3, columns 4 and 5 indicates that the real GDP can be predicted when the preceding year GPD coefficient is between 0.997 and 1.011.
Phase III. Prediction of bounds: using equation (20), the best and worst possible real GDP for the considered time range were predicted; the results are shown in Table 4 together with FR and FARIMA bounds. Figures 2, 3, and 4 represent the graphical plot of the predicted UB and LB of the FR, FARIMA and the Hybrid FARIMA-AFN models along with the actual real GPD, respectively.
Phase IV. Bound assessment: from Table 4, as well as Figures 2, 3, and 4, it can be observed that the actual GDP values are located within the predicted bounds. However, the interval of possibility based on the proposed method is narrower as compared to the other methods in most of the instances. Additionally, the proposed method achieves minimal Root Mean Square Error (RMSE) in both lower and upper bounds with the exception of FARIMA LB as presented in Table 5. This implies that the proposed method outperforms the FR and FARIMA methods. Hence, the proposed model, could be most suitable to forecast future possibility interval of the real GDP of Nigeria.

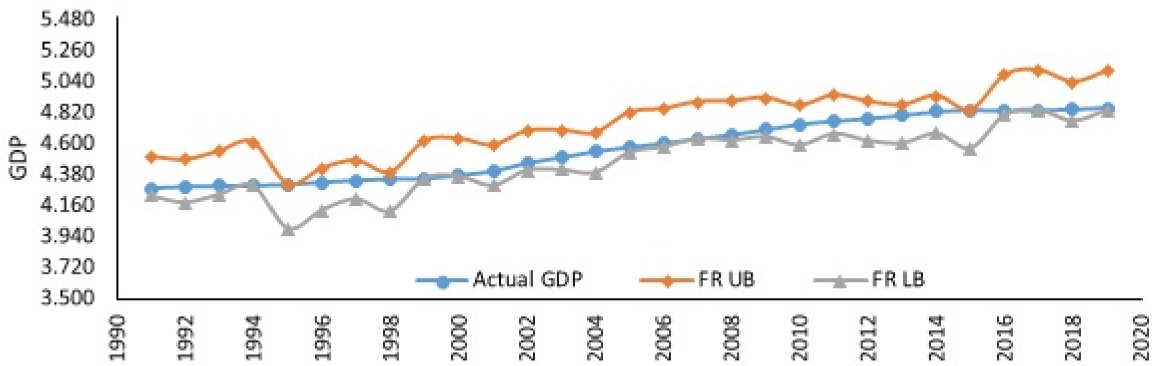

Figure 2: Actual real GDP along with UB and LB resulting from FR model

Table 4: Forecasted Intervals of Possibility for FR, FARIMA, FARIMA-AFN, and the Actual Real GDP.

Year	FR LB	FR UB	FARIMA LB	FARIMA UB	Proposed LB	Proposed UB	GDP
1991	4.507	4.228	*	*	*	*	4.283
1992	4.182	4.484	4.285	4.342	4.269	4.329	4.293
1993	4.238	4.545	4.295	4.352	4.279	4.339	4.299
1994	4.301	4.608	4.301	4.358	4.285	4.345	4.301
1995	3.998	4.309	4.303	4.360	4.287	4.347	4.309
1996	4.128	4.422	4.311	4.368	4.295	4.355	4.326
1997	4.205	4.477	4.328	4.385	4.312	4.373	4.338
1998	4.121	4.395	4.339	4.397	4.324	4.385	4.349
1999	4.351	4.618	4.350	4.408	4.335	4.396	4.351
2000	4.367	4.635	4.352	4.410	4.337	4.398	4.375
2001	4.302	4.588	4.376	4.434	4.361	4.422	4.403
2002	4.410	4.689	4.403	4.462	4.389	4.450	4.462
2003	4.414	4.694	4.461	4.521	4.448	4.510	4.501
2004	4.394	4.676	4.500	4.560	4.487	4.550	4.544
2005	4.535	4.820	4.542	4.603	4.529	4.593	4.574
2006	4.576	4.847	4.571	4.633	4.559	4.623	4.602
2007	4.633	4.896	4.599	4.660	4.587	4.652	4.633
2008	4.623	4.900	4.630	4.691	4.618	4.683	4.663
2009	4.647	4.925	4.659	4.721	4.648	4.713	4.698
2010	4.588	4.869	4.693	4.756	4.683	4.749	4.737
2011	4.668	4.944	4.732	4.795	4.722	4.788	4.760
2012	4.620	4.898	4.754	4.818	4.745	4.811	4.778
2013	4.602	4.873	4.772	4.836	4.763	4.830	4.801
2014	4.669	4.940	4.795	4.859	4.786	4.853	4.827
2015	4.567	4.839	4.820	4.885	4.811	4.879	4.839
2016	4.804	5.086	4.832	4.897	4.823	4.891	4.832
2017	4.836	5.120	4.825	4.890	4.816	4.884	4.836
2018	4.759	5.037	4.829	4.894	4.820	4.888	4.844
2019	4.837	5.114	4.837	4.902	4.828	4.896	4.854

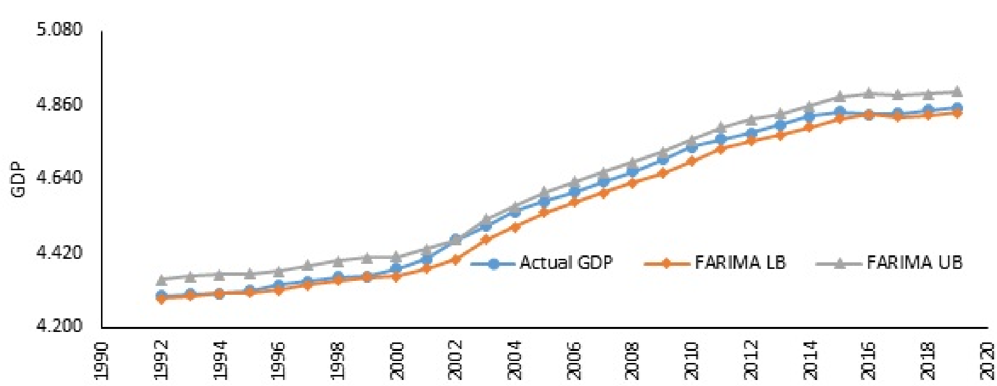

Figure 3: Actual real GDP along with UB and LB resulting from FARIMA model

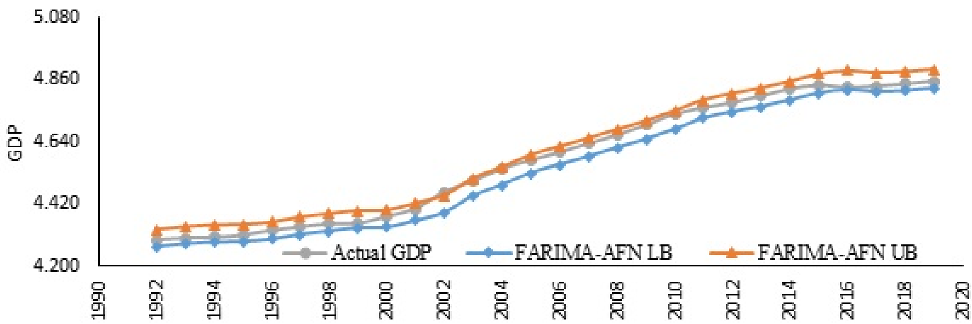

Figure 4: Actual real GDP along with UB and LB resulting from FARIMA model

Table 5: Root Mean Squares Error Values.

Bounds	FR LB	FR UB	FARIMA LB	FARIMA UB	Proposed LB	Proposed UB
RMSE	0.1300	0.2011	0.0272	0.0409	0.0380	0.0323

4. Conclusion

This study presented a hybrid method that combined the basic concepts of Fuzzy Linear Regression and Fuzzy Autoregressive Integrated Moving Average methods with a primary objective of enhancing the prediction accuracy through narrowing down the projection bias problems specifically associated with FARIMA model. Experiments were conducted using four different sets of weights to combine the FARIMA and FR results by alternating the weights. To show the forecasting performance of the proposed approach, we applied the proposed method to forecast real GDP data in Nigeria. The proposed method outperforms the considered methods in the literature in terms of predicting the real GDP. Future research is focused on hybridising FARIMA with other available tools. The results of the hybrid method are evaluated on the basis of some performance metrics.

References

Alsoltany, S. N. and Alnaqash, I. A. (2015). Estimating fuzzy linear regression model for air pollution predictions in baghdad city. Al-Nahrain Journal of Science, 18(2), 157–166.

Alyousifi, Y., Othman, M., and Almohammedi, A. A. (2021). A novel stochastic fuzzy time series forecasting model based on a new partition method. IEEE Access, 9:80236–80252.

Attanayake, A. (2021). Fuzzy linear regression: an application to heart disease. Advances and Applications in Statistics, 70(2), 219–227.

Bakawu, M., Buba, A., and Malik, A. (2023). Forecasting the real gdp in Nigeria using possibilistic linear regression model. Benin Journal of Statistics, 6, 59–69.

Box, G. E. P. and Jenkins, G. M. (1970). Time series analysis: forecasting and control.

Chen, S.-M. and Phuong, B. D. H. (2017). Fuzzy time series forecasting based on optimal partitions of intervals and optimal weighting vectors. Knowledge-Based Systems,

118, 204–216.
Dong, Q. and Ma, X. (2021). Enhanced fuzzy time series forecasting model based on hesitant differential fuzzy sets and error learning. Expert systems with applications, 166, 114056.
Du, H., Zhao, Z., and Xue, H. (2020). Arima-m: A new model for daily water consumption prediction based on the autoregressive integrated moving average model and the Markov chain error correction. Water, 12(3), 760.
Gorbatiuk, K., Hryhoruk, P., Proskurovych, O., Rizun, N., Gargasas, A., Raupelien˙e, A., and Munjishvili, T. (2021). Application of fuzzy time series forecasting approach for predicting an enterprise net income level. In E3S Web of Conferences, 280, 02007. EDP Sciences.
Khashei, M., Rafiei, F. M., and Bijari, M. (2013). Hybrid fuzzy auto-regressive integrated moving average (FARIMA) model for forecasting the foreign exchange markets. International Journal of Computational Intelligence Systems, 6(5), 954–968.
Lee, W.-J., Jhang, H.-J., and Choi, S. H. (2020). Fuzzy regression model of goal difference of the korean national football team based on elo rating and dividend. Journal of Intelligent Fuzzy Systems, 38(4), 4537–4543.
Malyaretz, L., Dorokhov, O., and Dorokhova, L. (2018). Method of constructing the fuzzy regression model of bank competitiveness. Journal of Central Banking Theory and Practice, 7(2), 139–164.
Medina Reyes, J. E., Cruz Ak´e, S., and Cabrera Llanos, A. I. (2021). New hybrid fuzzy time series model: Forecasting the foreign exchange market. Contadur´ıa y administraci´on, 66(3).
Mehdi, H., Pooranian, Z., and Vinueza Naranjo, P. G. (2019). Cloud traffic prediction based on fuzzy arima model with low dependence on historical data. Transactions on Emerging Telecommunications Technologies, 00(3), 1–19.
Ogosi, F., Andem, F., Nkanor, W., and Zibigh, P. (2022). An examination of macroeconomic determinants of economic growth in nigeria: a regression analysis model. Global Scientific Journal, 10(2), 41–58.
Panigrahi, S. and Behera, H. S. (2020). Fuzzy time series forecasting: a survey. In Computational Intelligence in Data Mining: Proceedings of the International Conference on ICCIDM 2018, 641–651. Springer.
Silva, P. (2019). Scalable models for probabilistic forecasting with fuzzy time series.
Taghizadeh, M., Shakouri, H., Menhaj, M., Mehregan, M., and Kazemi, A. (2009). Design of a multi-level fuzzy linear regression model for forecasting transport energy demand: A case study of iran. In 2009 International Conference on Computers & Industrial Engineering, 1757–1762. IEEE.
Taha, H. (2011). Operation research : an introduction. 9th Edition.
Taheri, S., Asadi, M., and Shiralipour, A. (2020). Fuzzy regression in predicting math achievement, based on philosophic-mindedness, creativity, mathematics self-efficacy, and mathematics self-concept. Fuzzy Information and Engineering, 12(3),322–334.
Tak, N. (2022). A novel arma type possibilistic fuzzy forecasting functions based on greywolf optimizer (arma-pffs). Computational Economics, 59(4), 1539–1556.
Tanaka, H., Uejima, S., and Asai, K. (1982). Linear regression analysis with fuzzy model. IEEE Trans. Systems Man Cybern, 12, 903–907.
Tanaka, H., H. I. and Watada, J. (1989). Possibilistic linear regression analysis for fuzzy data. European Journal of Operational Research, 40(3), 389–396.
Torbat, S., Khashei, M., and Bijari, M. (2018). A hybrid probabilistic fuzzy arima model for consumption forecasting in commodity markets. Economic Analysis and Policy, 58, 22–31.
Torbata, S., Khasheia, M., Bijari, M., Taherib, G., and Fakharib, M. (2017). Application of the fuzzy auto-regressive integrated moving average (farima) model for Iran's steel consumption forecasting.
Tseng, F.-M. and Tzeng, G.-H. (2002). A fuzzy seasonal arima model for forecasting. Fuzzy Sets and Systems, 126(3), 367–376.
Tseng, F.-M., Tzeng, G.-H., Yu, H.-C., and Yuan, B. J. (2001). Fuzzy arima model for forecasting the foreign exchange market. Fuzzy sets and systems, 118(1), 9–19.

Wang, M.-J., Tzeng, G.-H., and Jane, T.-D. (2009). A fuzzy arima model by using quadratic programming approach for time series data. International Journal of Information Systems for Logistics and Management, 5(1), 41–51.

Xie, Y., Zhang, P., and Chen, Y. (2021). A fuzzy arima correction model for transport volume forecast. Mathematical Problems in Engineering, 2021, 1–10.

Yusuf, S., Mohammad, A., and Hamisu, A. (2017). A novel two–factor high order fuzzy time series with applications to temperature and futures exchange forecasting. Nigerian Journal of Technology, 36(4), 1124–1134.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning—i. Information sciences, 8(3),199–249.

Appendix

Table 6: *
Data on Macroeconomic Variables: GPD(Y), Unemployment (X_1), Inflation (X_2), and FDI(X_3)

Year	**Y**	X_1	X_2	X_3	**Year**	**Y**	X_1	X_2	X_3
2019	4.854	0.908	1.057	0.519	2004	4.544	0.579	1.176	0.272
2018	4.844	0.916	1.082	0.301	2003	4.501	0.582	1.147	0.303
2017	4.836	0.924	1.218	0.544	2002	4.462	0.582	1.110	0.276
2016	4.832	0.849	1.195	0.648	2001	4.403	0.577	1.276	0.076
2015	4.839	0.634	0.955	0.486	2000	4.375	0.577	0.841	0.057
2014	4.827	0.659	0.906	0.671	1999	4.351	0.579	0.821	0.000
2013	4.801	0.568	0.928	0.745	1998	4.349	0.575	1.000	-0.523
2012	4.778	0.573	1.087	0.849	1997	4.338	0.575	0.931	-0.328
2011	4.760	0.576	1.035	0.946	1996	4.326	0.576	1.466	-0.301
2010	4.737	0.576	1.137	0.780	1995	4.309	0.575	1.862	-0.469
2009	4.698	0.571	1.099	0.932	1994	4.301	0.575	1.756	0.292
2008	4.663	0.549	1.064	0.913	1993	4.299	0.573	1.757	0.130
2007	4.633	0.553	0.732	0.781	1992	4.293	0.565	1.649	-0.046
2006	4.602	0.562	0.915	0.686	1991	4.283	0.561	1.114	-0.149
2005	4.574	0.573	1.252	0.697					

Source: Ogosi *et al.* (2022)