

On Order and Regime Determination of SETAR Model in Modelling Nonlinear Stationary Time Series Data Structure: Application to Lafia Rainfall Data, Nasarawa State, Nigeria

Nicholas Pindar Dibal^{1,*}, Akeyede Imam², Mustafa Babagana Abubakar³

¹Department of Mathematical Sciences, University of Maiduguri, Maiduguri, Nigeria

²Department of Mathematics, Federal University Lafia, Lafia, Nigeria

³Directorate of Academic Planning, Federal University Lafia, Lafia, Nigeria

Email address:

pnidibal@unimaid.edu.ng (N. P. Dibal), imamakeyede@gmail.com (A. Imam), abubakarbabagana001@gmail.com (M. B. Abubakar)

*Corresponding author

To cite this article:

Nicholas Pindar Dibal, Akeyede Imam, Mustafa Babagana Abubakar. On Order and Regime Determination of SETAR Model in Modelling Nonlinear Stationary Time Series Data Structure: Application to Lafia Rainfall Data, Nasarawa State, Nigeria. *American Journal of Theoretical and Applied Statistics*. Vol. 10, No. 2, 2021, pp. 89-98. doi: 10.11648/j.ajtas.20211002.11

Received: January 9, 2021; **Accepted:** January 16, 2021; **Published:** March 3, 2021

Abstract: The linear time series model refers to the class of models for which fixed correlation parameters can fully explain the dependency between two random variables, but many real-life circumstances, such as monthly unemployment results, supplies and demands, interest rate, exchange rate, share prices, rainfall, etc., violate the assumption of linearity. For fitting and forecasting of nonlinear time series data, the self-exciting threshold autoregressive (SETAR) model was suggested. Using R to generate random nonlinear autoregressive data, a Monte Carlo simulation was performed, the SETAR model was fitted to the simulated data and Lafia rainfall data, Nasarawa State, Nigeria to determine the best regime orders and/or scheme number to make future forecast. Using Mean Square Error (MSE) and Akaike Information Criteria (AIC), the relative performance of models was examined. At a specific autoregressive order, regime order, sample size and step ahead, the model with minimum criteria was considered as the best. The results show that the best autoregressive and regime orders to be chosen are 3rd and 2nd [SETAR (3, 2)] respectively for fitting and forecasting nonlinear autoregressive time series data with small and moderate sample sizes. As the sample size increases, the output of the four models increases. Finally, it is shown that when sample size and number of steps forward are increased, the efficiency and forecasting capacity of the four models improves.

Keywords: SETAR Model, Regime Order, Autoregressive Order, Nonlinear Time Series

1. Introduction

In many real-life circumstances, time series data arises naturally, such as economics, monthly unemployment data, supplies and demands, interest rates, exchange rates, share prices, rainfall, etc. These variables are known to be responsive and affected by many variables, responding quickly to any external intervention, resulting in sudden and drastic behavioral changes.

Linear relationships are the first approximation used to characterize any relationship, according to Akeyede *et al.* [1], but there is no unique way of describing what a linear relationship is in terms of the underlying essence of the quantities. The nonlinear model is the model class for which

the functional form of the dependency between two random variables is more general than linear equation and/or can change over time.

Nonlinear time series models have a much broader variety of potential dynamics for series, such as economic and financial data, rainfall data, etc. Compared to linear models, they are able to capture asymmetry, jump, wave, and other nonlinear behaviors. The self-exciting autoregressive threshold (SETAR) model is suggested, this class of nonlinear models are increasingly used in time series analysis to describe and forecast different empirical phenomena in an observed time series as it is helpful in capturing nonlinear dynamics, this could be seen in the works of the time series (Tong and Yeung [16], Watier and Richardson [20],

Grabowski *et al.* [9]).

In relation to the traditional linear modeling approach, the essence of these classes of models is examined using simulated data and rainfall data for Lafia obtained from the Nasarawa State Meteorological Department, Nigeria. The statistical characteristics and forecast performance of the 2-regime SETAR models with long-term storage in the first and short-term storage in the second were extensively examined in order to locate the threshold parameter using stock indices and individual asset prices (Tong [15]; Hansen [10]; Clements and Smith [2]; De Gooijer [3]; Dufrenot *et al.* [5]).

The SETAR models introduced by Tong [12] that were more thoroughly developed in the seminal paper by Tong and Lim [13] belong to the class of nonlinear models that have been increasingly used in the study and forecasting of time series as they are useful for adequately capturing nonlinear dynamics (Grabowski *et al.* [9]). In terms of expansion of autoregressive models, they can be considered, allowing for changes in the parameters of the model by regime switching behavior. The SETAR model in the space of the threshold variable is a piecewise linear autoregressive model. There are many approaches to SETAR model estimation that differ in their ability to estimate the hyper-parameters and to measure SETAR models of high order. The methods widely used are; the approach of Tsay, which emphasizes graphical analyses to define thresholds and, Hansen's methodology covers 2 and 3 order models in depth, which helps the thresholds to be calculated.

The success of SETAR models is because, compared to many other nonlinear time series models, they are relatively easy to specify, estimate, and interpret (Tong [17]). In the variable's relationship, the common empirical time series modeling assumes linearity and stationarity. Applied time series analysis, however, finds it difficult to assume this linearity in data simply because arguments have been raised that nonlinear specification can reflect data generation processes more realistically (Franses and van Dijk [7]). In the analysis of nonlinear time series results, this research therefore determines the best scheme and autoregressive orders for SETAR models.

2. Methodology

Using the Monte Carlo algorithm, we generate random data on nonlinear autoregressive processes from normal

distribution to test the output of the proposed SETAR model. The importance of the choice of the proposed model is demonstrated by an empirical application on the Lafia rainfall data from Nasarawa, Nigeria. The simulation is the realization of a simple two-stage SETAR model produced to identify nonlinear phenomena and sample analytical processes that have been performed using R statistical software for the following sample sizes: 20, 40, 60, 80, 100, 120, 140, 160, 180 and 200. For all possible parameters, the parameters of the SETAR (p, d) order model (p) and the regime number (d) have been defined.

In modeling rainfall results, the empirical application will demonstrate the relevance of the choice of process. Mean Square Error (MSE) and Akaike Knowledge Criteria (AIC) were used to determine the consistency of fit for each model, and the results of the study were presented in Tables 1-3 along with their respective graphs in Figures 1-9. Simulation was done with the trigonometric function given as follows under the assumption of stationarity from second-order nonlinear autoregressive processes;

$$Y_t = \phi_1 \sin(Y_{t-1}) + \phi_2 \cos(Y_{t-2}) + e_t \quad (1)$$

The current value of the Y_t series is a nonlinear combination of its self's most recent past values plus an e_t "innovation" concept that integrates something new in the t series that is not explained by the previous values. For each t , therefore, we assume that e_t is independent of $Y_t, Y_{t-1}, Y_{t-2}, \dots$ (Akeyede *et al.*, 2015).

2.1. The SETAR Model

Order p 's self-exciting threshold Autoregressive (SETAR) model belongs to the autoregressive threshold (TAR) family, which is important for nonlinear time series modeling. Such models are a relatively simple relaxation of standard autoregressive linear models that allow a number of states to be approximated linearly. According to Tong [15], by decomposing the one-dimensional Euclidean space into k regimes with a linear autoregressive model in each regime, the threshold principle allows the analysis of the complex stochastic system; this method makes the model nonlinear for at least two regimes but remains locally linear (Gibson and Nur [8]). In this research, the class of SETAR models considered are classified as;

$$\text{SETAR}(2, 2), Y_t = \begin{cases} \phi_0^1 + \phi_1^1 Y_{t-1} + \phi_2^1 Y_{t-2} + e_t^1, & \text{if } Y_{t-d} \leq r \\ \phi_0^2 + \phi_1^2 Y_{t-1} + \phi_2^2 Y_{t-2} + e_t^2, & \text{if } Y_{t-d} > r \end{cases} \quad (2)$$

$$\text{SETAR}(2, 3), Y_t = \begin{cases} \phi_0^1 + \phi_1^1 Y_{t-1} + \phi_2^1 Y_{t-2} + e_t^1, & \text{if } Y_{t-d} \leq r_1 \\ \phi_0^2 + \phi_1^2 Y_{t-1} + \phi_2^2 Y_{t-2} + e_t^2, & \text{if } r_1 \leq Y_{t-d} \leq r_2 \\ \phi_0^3 + \phi_1^3 Y_{t-1} + \phi_2^3 Y_{t-2} + e_t^3, & \text{if } Y_{t-d} > r_2 \end{cases} \quad (3)$$

$$\text{SETAR}(3, 2), Y_t = \begin{cases} \phi_0^1 + \phi_1^1 Y_{t-1} + \phi_2^1 Y_{t-2} + \phi_3^1 Y_{t-3} + e_t^1, & \text{if } Y_{t-d} \leq r \\ \phi_0^2 + \phi_1^2 Y_{t-1} + \phi_2^2 Y_{t-2} + \phi_3^2 Y_{t-3} + e_t^2, & \text{if } Y_{t-d} > r \end{cases} \quad (4)$$

$$\text{SETAR}(3, 3), Y_t = \begin{cases} \phi_0^1 + \phi_1^1 Y_{t-1} + \phi_2^1 Y_{t-2} + \phi_3^1 Y_{t-3} + e_t^1, & \text{if } Y_{t-d} \leq r_1 \\ \phi_0^2 + \phi_1^2 Y_{t-1} + \phi_2^2 Y_{t-2} + \phi_3^2 Y_{t-3} + e_t^2, & \text{if } r_1 \leq Y_{t-d} \leq r_2 \\ \phi_0^3 + \phi_1^3 Y_{t-1} + \phi_2^3 Y_{t-2} + \phi_3^3 Y_{t-3} + e_t^3, & \text{if } Y_{t-d} > r_2 \end{cases} \quad (5)$$

where the superscripts indicate states of the world or regimes in the models. It is assumed that a linear autoregressive process follows the dynamic behavior of the time series variable within each regime, the subscript in the models indicates the autoregressive order, r is the threshold value, y_{t-d} is the threshold variable that governs the transition between the two regimes with d being the delay parameter that is a positive integer ($d < p$) and $e_t^{(j)}$ white noise processes that are random variables with zero mean and constant variance distributed separately and identically i.e. $e_j \sim iidN(0, \sigma_j^2)$, $\sigma_j^2 < \infty$.

The threshold parameters satisfy the constraint $-\infty = r_0 < r_1 < r_2 \dots < r_{k-1} < r_k = \infty$. The mechanism that is operating at any moment depends on the measurable past history of $\{Y_t\}$ itself and, in particular, on the importance of $(t-d)$. In equation (2) to (5), Tong and Lom [13] referred to the system as self-exciting threshold autoregressive models. The benefit of using SETAR models is their ability to generate some widely observed phenomena that are not capable of being captured by simple linear models that provide irreversibility, hops and limit intervals, such as the autoregressive moving average (ARMA) model.

2.2. Parameters Estimation

The most commonly used approach for parameter estimation under the SETAR model is conditional least squares (CLS), according to Gibson and Nur [8]. In this method, in order to obtain parameter estimates, the predictive number of square errors is minimized. First, let $E(Y_t) < \infty$, $t = 1, 2, \dots$ and $\phi = (a_0, a_1, \dots, a_p; b_0, b_1, \dots, b_p; \sigma^2)$. This provides the requisite details for the estimation of the conditional least squares, and it is possible to estimate for ϕ by minimizing the residual number of squares in relation to ϕ such that:

$$Q_N(\phi) = \sum_{j=1}^N [Y_j - E_\theta(Y_j | B_{j-1})]^2 \quad (6)$$

Firat [6] defined the necessary parameter estimation steps for the SETAR model as follows;

Step 1

As a first step, it is assumed that the d and γ , the delay and threshold parameter values are known. The observed values are divided into small sub-groups based on these assumptions, and the AIC data criterion for each sub-group is determined at the level of $p_i (i = 1, \dots, k)$ and is shown as follows;

$$AIC(\hat{p}_i) = \min[AIC(k_i)], i = 1, 2 \quad (7)$$

The p_i value of each scheme is obtained in this situation, using $\min AIC(k_i)$ in exchange for the constant values d and γ .

Step 2

The d value is kept constant in the second step (it is presumed to correspond to a certain value; in other words, it

is known), and the threshold parameters that minimize the value of the AIC data criterion are checked this time. Among the other threshold parameters, this is the γ value that minimizes the AIC (d, γ) value, which is shown in Tong [14] as follows;

$$AIC(d_0, \hat{\gamma}) = \min[AIC(d_0, \gamma)] \quad (8)$$

Step 3

The p_i and γ values are calculated in the first two steps. The value of d will be calculated in the remaining third stage. From the d option, the d value that minimizes the NAIC(d) value will be found in the k number. The model will be calculated by conditioning on the above-mentioned parameters after the 3 measures are evaluated using the data criterion.

2.3. Test of Stationarity

Stationarity in the study of time series data is a critical principle. Generally speaking, if there is no systematic change in the mean (no trend) and variance, and if periodicity is removed, a time series is said to be stationary. In other words, the properties of one segment of the data are just like those of every other section (Tsay [19]). To highlight these important associated statistical characteristics, consider the simple p^{th} autoregressive [AR(p)] model to demonstrate these relevant statistical properties associated with the autoregressive unit root test,

$$y_t = \phi y_{t-1} + e_t \quad (9)$$

where $e_t \sim N(0, \sigma^2)$ and $H_0: |\phi| = 1$ vs $H_1: |\phi| < 1$ then the test statistic is,

$$t_{\phi=1} = \frac{\hat{\phi}-1}{SE(\hat{\phi})} \quad (10)$$

where $\hat{\phi}$ is the least square approximation and the normal standard error estimate is $SE(\hat{\phi})$ and the measurement is a one-sided left tail test. If y_t is stationary (i.e. $|\phi| < 1$), it can be demonstrated that $t_{\phi=1} \sim N(0, 1)$. Similarly, Dickey and Fuller [4] developed the unit root test in which the null hypothesis is $\phi = 0$ against the alternative hypothesis of $\phi < 0$. A value for the statistics for the test

$$DF = \frac{\hat{\phi}}{SE(\hat{\phi})} \quad (11)$$

compared to the relevant critical value for the test of Dickey and Fuller, this is computed. If the test statistic is lower than the critical value, the $\phi = 0$ null hypothesis is rejected and there is no unit root presented.

2.4. Test of Nonlinearity

Before applying possible nonlinear model, we first perform linearity test against nonlinearity, this pretesting for

nonlinearity is important to help protect against over-fitting the data. In this analysis, we therefore consider the modification by Tsay [18] of Keenan's One-Degree Test for Nonlinearity (Keenan [11]), where the F-Test is modified by replacing the aggregated quantity Y_t^2 with the disaggregated variable $Y_{t-i}, Y_{t-j}; i, j = 1, 2, \dots, M$ where M is specified in the test of Keenan. The protocol for the F-test is as follows:

- i) Regress Y_t on $(1, Y_{t-1}, \dots, Y_{t-m})$ and then calculate the fitted values (\hat{Y}_t) and the residuals (\hat{e}_t) for $t = M + 1, \dots, n$, hence the regression model is;

$$Y_t = W_t \phi + e_t \quad (12)$$

where $W_t = (1, Y_{t-1}, \dots, Y_{t-m})$ and $\phi = (\phi_0, \phi_1, \dots, \phi_m)'$.

- ii) Regress vector Z_t on $(1, Y_{t-1}, \dots, Y_{t-m})$ and calculate the residuals (\hat{Y}_t) for $t = M + 1, \dots, n$. In this step, the multivariate regression model is;

$$Z_t = W_t H + Y_t \quad (13)$$

where Z_t is a dimensional vector defined by $Z_t' = \text{vech}(U_t', U_t)$, with $U_t = (Y_{t-1}, \dots, Y_{t-m})$ and vech denoting the half stacking vector.

- iii) Fit $\hat{e}_t = \hat{Y}_t \beta + \varepsilon_t, t = M + 1, \dots, n$ and then define

$$\hat{F} = \left\{ \frac{\sum \hat{Y}_t^T \hat{e}_t (\sum \hat{Y}_t^T \hat{Y}_t)^{-1} \sum \hat{Y}_t^T \hat{e}_t}{m} \right\} / \left[\frac{\sum \varepsilon_t^2}{n-M-m-1} \right] \quad (14)$$

where the summation is over t from $M + 1$ to n and \hat{F} is asymptotically distributed as $F_{m, n-M-m-1}$

2.5. Forecasting

Different procedures exist to forecast the linear and nonlinear time series models and the SETAR model forecasts

could be interpreted as enhancing the ARIMA model in order to compare these forecasts. Rising values could suggest an attempt by the model to capture the variance in the series more efficiently and represent overall movements in the process. In general, it should be noted that if a model better describes the characteristics of time series with an in-sample fit, there is no guarantee that it will also make better forecasts (Franses and Dijk [7]). Computing point estimates from models of nonlinear time series require complicated computations.

Consider Y_t , a nonlinear autoregressive lag duration one model, the forecasting phase begins with a sample of process values before t say, Y_1, Y_2, \dots, Y_t , our observed data reflects these values. Our forecasting model forecasts future process values, i.e. $Y_{t+1}, Y_{t+2}, Y_{t+3}, \dots$. Generally speaking, Y_{t+h} is the value of the forecasting process at lead time h ahead of the prediction at lead time t where $h > 1$.

3. Results and Discussions

Monte Carlo simulations were performed to investigate SETAR (p, d) efficiency, where $p, d = 2, 3$, i.e., in the fitting and forecasting of the simulated nonlinear autoregressive model, SETAR (2, 2), SETAR (2, 3), SETAR (3, 2) and SETAR (3, 3) models. On the nonlinear simulated results, the effect of sample size was examined. Under the SETAR (p, d) model, the best order (p) and regime number (d) were calculated. The method was also carried out for Nasarawa State's 10-year monthly Lafia rainfall data up to 2017. For more clarification, the results of the analyses are reported in Tables 1-3 and shown graphically in Figures 1-9.

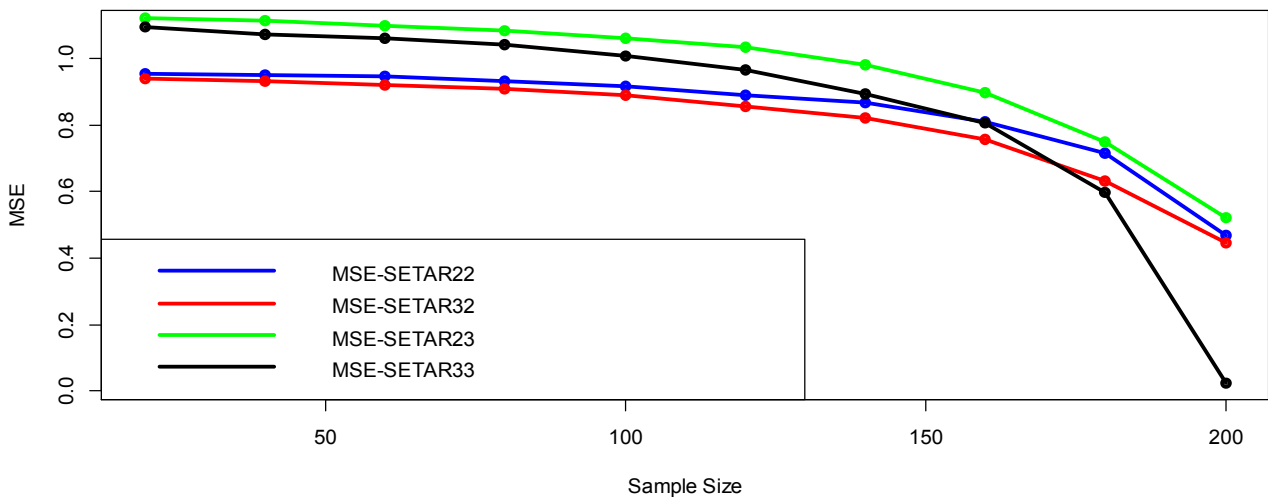


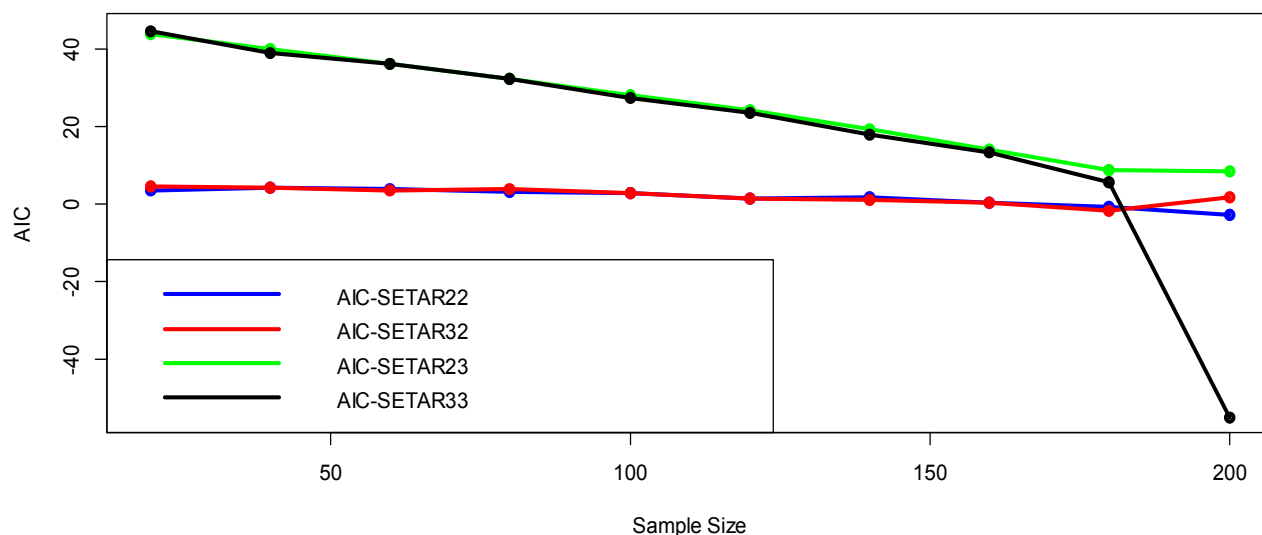
Figure 1. MSE of SETAR (p, d) Model for different Sample Sizes.

The plots of the MSE and AIC for the SETAR (p, d) model for different sample sizes are shown in figures 1 and 2. It can be seen from the graphs that SETAR (3, 2) from sample sizes 20 to 180 is the best model to suit nonlinear results, followed

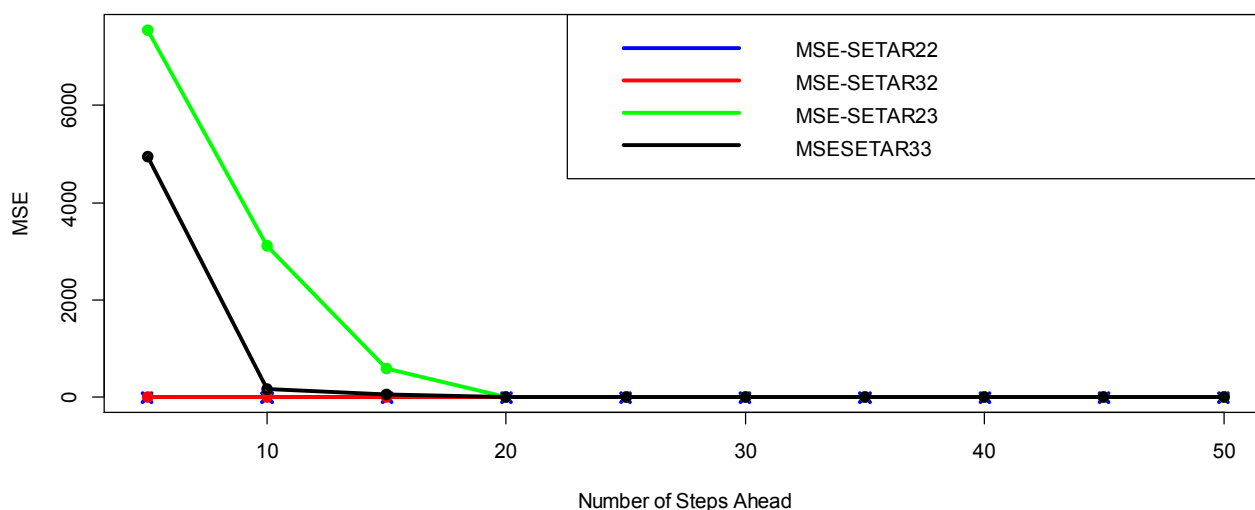
by SETAR (2, 2) on the basis of both MSE and AIC criteria. However, for sample size 180 and above, SETAR (3, 3) outperforms other choices (larger sample sizes). SETAR (2, 3) for all sample sizes is the worst model observed.

Table 1. Results of Relative Performance of SETAR (p,d) Model at Different Sample Sizes.

Sample Size	SETAR (2,2)	SETAR (2,3)	SETAR (3,2)	SETAR (3,3)	SETAR (2,2)	SETAR (2,3)	SETAR (3,2)	SETAR (3,3)
20	0.9529	1.1211	0.9370	1.0884	3.5512	43.7319	4.1406	43.7450
40	0.9488	1.1114	0.9312	1.0739	3.3015	39.8988	3.8982	39.6670
60	0.9378	1.0976	0.9174	1.0568	2.5406	35.6485	2.9950	35.5255
80	0.9273	1.0786	0.9053	1.0318	2.2822	31.3755	2.8977	31.1396
100	0.9156	1.0560	0.8896	1.0012	2.2761	27.3449	2.7791	26.8995
120	0.8925	1.0295	0.8610	0.9631	1.4753	23.6345	1.8456	22.8809
140	0.8589	0.9777	0.8183	0.8948	0.7010	18.8351	0.7177	17.6305
160	0.8175	0.9045	0.7626	0.8028	-0.6712	14.5057	0.4310	13.1912
180	0.7241	0.7448	0.6436	0.5955	-0.2483	8.4707	-1.1195	5.1741
200	0.4700	0.5198	0.2427	0.0676	-2.8632	8.4652	-10.3490	-23.8790

**Figure 2.** AIC of SETAR (p, d) Model for different Sample Sizes.

Therefore, 3rd and 2nd autoregressive and scheme orders are the best autoregressive and scheme orders to be chosen for fitting nonlinear autoregressive time series data with small, moderate and large sample sizes (up to 180) respectively, while those with larger sample sizes (above 180) can be fitted with 3rd autoregressive and scheme orders respectively. The performance of the four models also improves with the minimum value of MSE and AIC as the sample size increases.

**Figure 3.** MSE of Forecast Performances of SETAR (p, d) Model at Sample Size 20.

3.1. Performances of SETAR (p, d) Models on Nonlinear Autoregressive Data

Table 2 and the plots shown in Figures 3-8 showed a comparison of the forecast capacity between the four fitted models and the simulated nonlinear autoregressive data for samples of sizes 20, 100 and 200.

Table 2. Results of Relative Forecast Performance of SETAR (p,d) Model for Different Sample Sizes.

Sample Size	Steps Ahead (d)	SETAR				AIC			
		MSE				AIC			
		(2,2)	(3,2)	(2,3)	(3,3)	(2,2)	(3,2)	(2,3)	(3,3)
20	5	0.6476	0.3823	7555.76	4953.737	0.1302	-7.4221	4.1095	285.8044
	10	0.0004	0.2805	3039.628	173.484	-794.008	-8.1663	-5.027	147.9873
	15	0.0002	0.0028	596.9272	50.6988	-1143.25	-1139.45	-6.4755	47.255
	20	1.23E-05	6.59E-23	6.5361	11.0204	-1325.35	-1409.37	-500.499	-61.2087
	25	8.64E-14	4.25E-32	0.4206	1.534	-1502.38	-1815.57	-1065.69	-110.462
	30	6.41E-33	1.87E-33	0.2092	0.2589	-1883.62	-2280.25	-1085.02	-124.934
	35	6.01E-33	1.83E-33	6.81E-05	0.0116	-2011.42	-2.66E+03	-1.13E+03	-7.79E+02
	40	2.18E-33	1.58E-33	1.52E-31	0.0005	-2.72E+03	-3.04E+03	-1.55E+03	-1.23E+03
	45	3.12E+34	1.39E-33	8.88E-33	6.38E-26	-3.02E+03	-3.07E+03	-1.60E+03	-1.42E+03
	50	1.96E-35	1.14E-33	6.77E-35	4.58E-28	-3.15E+03	-3419.26	-2177.55	-2177.55
100	5	9.25E-01	8.93E-01	3.34E+00	324352.5	-1.08E+01	2.79E+01	4.30E+01	3.03E+01
	10	5.14E-08	8.34E-01	1.11E+00	6407.352	-7.81E+02	5.53E+00	2.28E+01	1.97E+01
	15	9.88E-34	1.75E-35	7.78E-01	492.2123	-1.15E+03	-1.15E+03	1.41E+01	4.83E+00
	20	9.72E-34	1.72E-33	2.89E-01	138.5937	-1.52E+03	-1.55E+03	1.24E-01	-7.37E+02
	25	8.27E-34	1.47E-33	3.06E-02	0.9753	-1.52E+03	-1.91E+03	-8.04E+02	-1.15E+03
	30	2.20E-34	8.77E-34	3.90E-03	0.8231	-2.30E+03	-2.26E+03	-1.06E+03	-1.39E+03
	35	9.07E+35	5.42E-34	3.89E-03	0.755	-2.68E+03	-2.64E+03	-1.14E+03	-1.46E+03
	40	8.58E-35	2.03E-34	4.40E-06	0.04858	-3.12E+03	-3.05E+03	-1.37E+03	-1.72E+03
	45	8.54E-35	1.09E-34	-1.31E+03	0.0122	-3.60E+03	-3.45E+03	-1.86E+03	-2.10E+03
	50	1.63E-35	7.08E-35	-1330.88	0.0003	-3991.13	-3891.98	-2743.7	-2195.3
200	5	9.36E-01	9.82E-01	1.09E+00	31.866	2.41E+01	1.62E+01	5.00E+01	4.98E+01
	10	8.65E-01	9.18E-01	8.74E-01	4.1824	-1.50E+01	-4.08E+01	4.98E+01	3.30E+01
	15	3.00E-33	2.09E-33	3.89E-01	1.2028	-1.15E+03	-1.19E+03	-1.13E+03	1.67E+01
	20	2.76E-33	1.57E-33	3.05E-01	0.9757	-1.54E+03	-1.52E+03	-1.15E+03	-1.48E+03
	25	2.69E-33	7.80E-34	1.97E-02	0.9531	-1.92E+03	-1.91E+03	-1.32E+03	-1.53E+03
	30	1.17E-33	6.26E-34	4.90E-03	0.2659	-2.26E+03	-2.31E+03	-1.43E+03	-1.77E+03
	35	6.24E-34	5.97E-34	1.10E-03	0.1329	-2.65E+03	-2.70E+03	-1.69E+03	-1.82E+03
	40	6.04E-34	2.67E-34	7.00E-04	0.0001	-3.05E+03	-3.08E+03	-1.97E+03	-1.94E+03
	45	4.49E-34	2.00E-34	2.00E-04	4.53E-34	-3.42E+03	-3.47E+03	-2.20E+03	-1.97E+03
	50	1.86E-34	1.03E-34	1.03E-34	6.19E-34	-3777.24	-3744.38	-2705.33	-2577.36

The results in Table 2 show that SETAR (2,2) performs better for sample size 100 while SETAR (3,2) performs better for sample sizes 20 and 200 except for d = 50. Figures 3-8 display the plots of the findings. From the plots, it was observed that SETAR (3, 2) and SETAR (2, 2) have the best forecast from steps 10 and 50 ahead of SETAR (2, 3) and SETAR (3, 3) based on both MSE and AIC. SETAR (2, 3), followed by SETAR (3, 2), are the worst prediction models.

Therefore, the best autoregressive and scheme orders to be chosen at any stage of steps ahead (i.e., from 5 to 50 steps ahead) and sample sizes of 20, 100 and 200 (small, moderate and high) to forecast nonlinear autoregressive time series data are 3rd, 2nd and 2nd, 2nd autoregressive and scheme orders, respectively. When the steps ahead are increased, the forecasting potential of the four models increases.

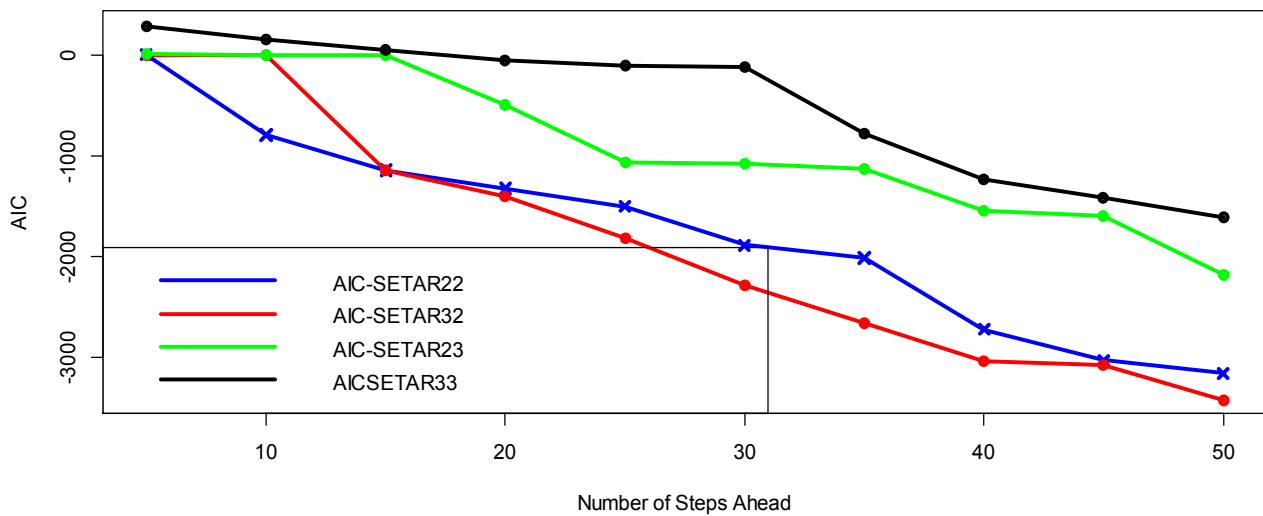


Figure 4. AIC Forecast Performance of SETAR (p, d) Model at Sample Size 20.

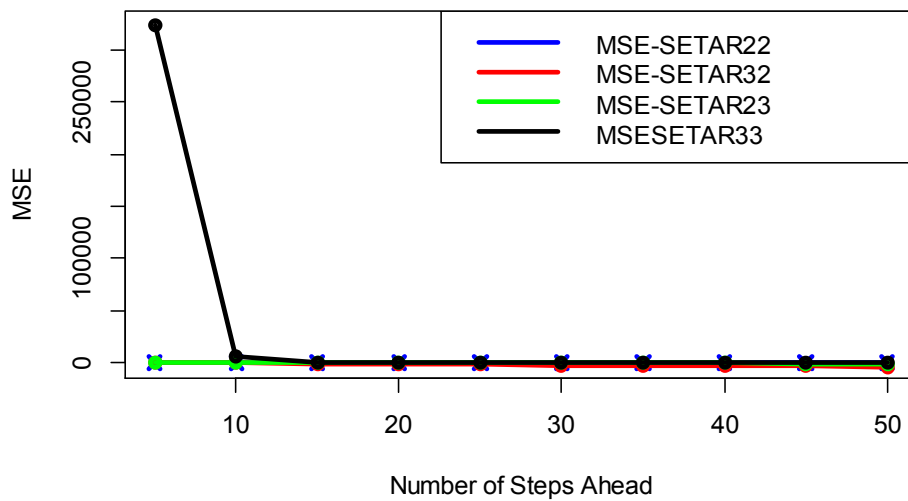


Figure 5. MSE Forecast Performance of SETAR (p, d) Model at Sample Size 100.

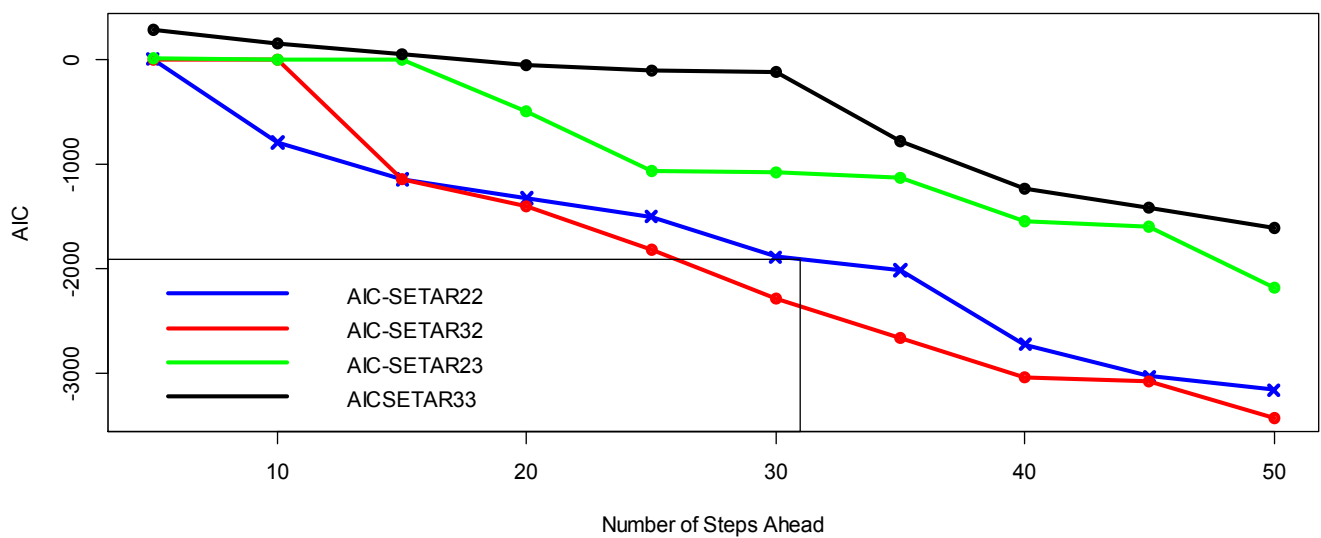


Figure 6. AIC Forecast Performance of SETAR (p, d) Model at Sample Size 100.

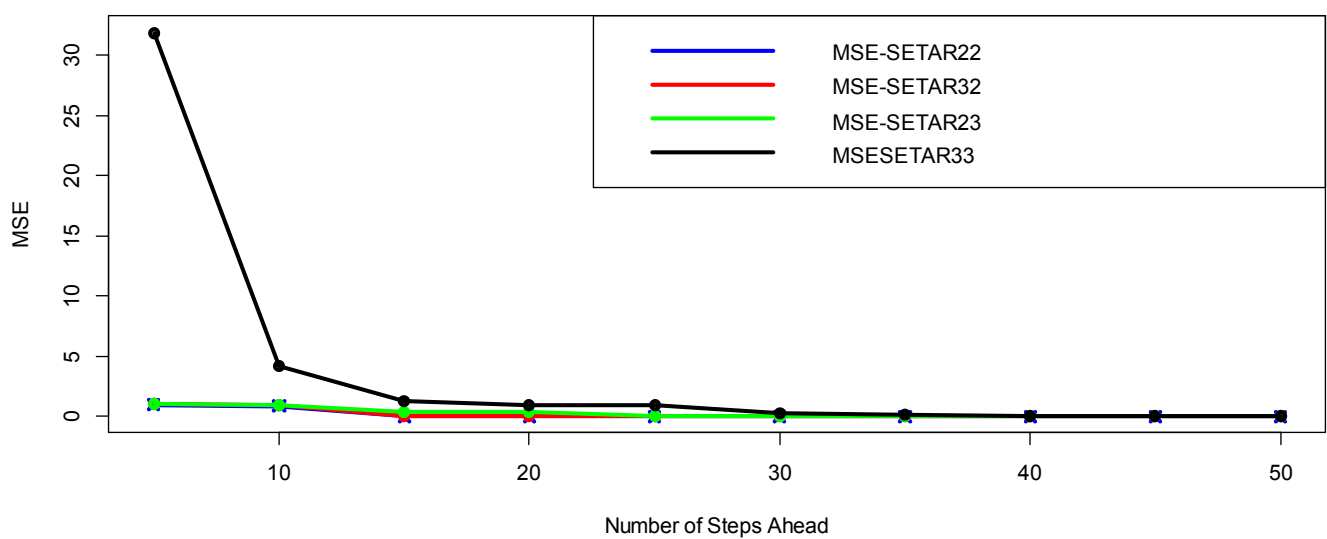


Figure 7. MSE Forecast Performance of SETAR (p, d) Model at Sample Size 200.

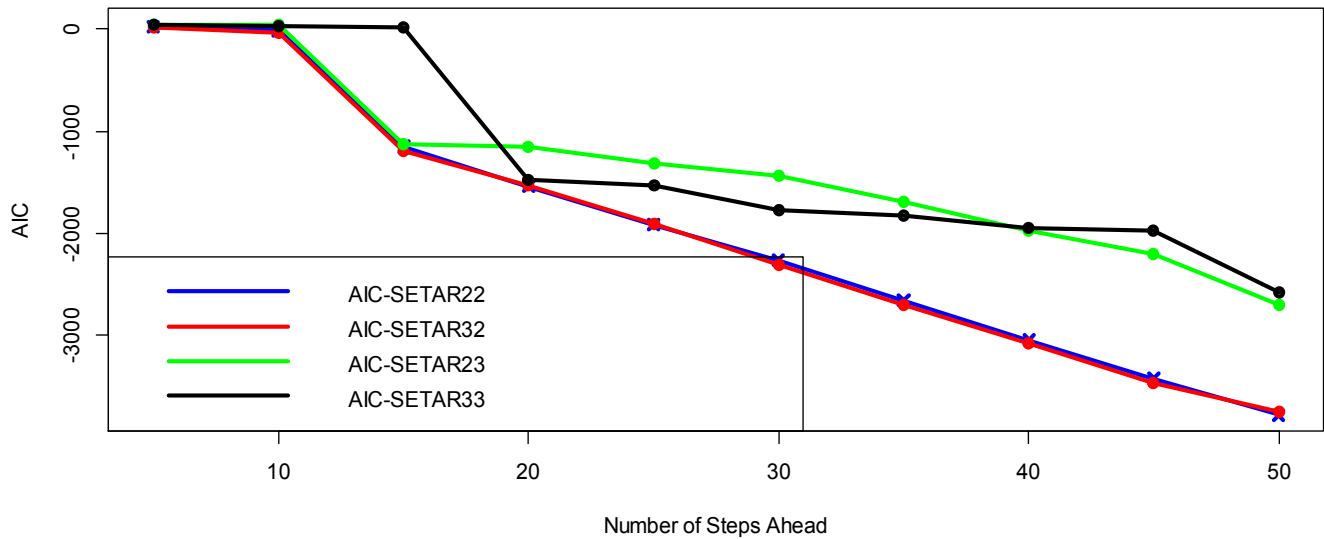


Figure 8. AIC Forecast Performance of SETAR (p, d) Model at Sample Size 200.

3.2. Analysis of Rainfall Data

Before fitting to the nonlinear SETAR models, the Lafia rainfall details, Nasarawa, Nigeria was first tested to determine whether it is linear or not. In measuring nonlinearity using R statistical software, the Tsay F-statistic is used. At 12 orders of auto-regression ($p = 0.4138$), the null nonlinearity hypothesis is not dismissed. Thereafter, to decide the best model as shown in Table 3, the data was

adapted to the class of SETAR models being considered.

The results of the analysis in Table 3 show that SETAR (3, 2) is the best fit for rainfall data, followed by SETAR (2, 2) with the lowest MSE and AIC values of (1320.1870 and 865.8924) and (1392.4540 and 868.8110). It also shows that, with lower standard errors, the parameter estimates of the two best models are important.

Table 3. Results of Fitted mode to Monthly Rainfall Data.

Model		Standard				
SETAR	Regime	Estimate	Error	p-Value	MSE	AIC
(2,2)	1st	2.1707	0.3394	0.0000	1392.4540	868.8110
		-0.3704	0.1012	0.0004		
	2nd	0.0905	0.1617	0.5767		
		-0.7400	0.1281	0.0000		
(3,2)	1st	20.1879	14.8843	0.1778	1320.1870	865.8924
		0.0216	0.2985	0.9423		
		-0.0329	0.1229	0.7894		
	2nd	0.3286	0.0908	0.0005		
		-0.2742	0.1028	0.0008		
		-0.4348	0.1109	0.0002		
(2,3)	1st	-70.7274	66.7048	0.2914	1780.2160	905.1698
		-0.2425	0.1030	0.0203		
	2nd	-0.1802	0.2687	0.5039		
		-0.7971	0.1139	0.0000		
		-0.0516	0.2310	0.8237		
	3rd	-0.1200	0.2644	0.6507		
		-50.6034	61.7894	0.4146		
		-0.1815	0.0961	0.0618		
	1st	-0.3151	0.0797	0.0001		
		-0.1962	0.2536	0.4409		
(3,3)	2nd	-0.7382	0.1396	0.0000	1477.2370	889.1556
		-0.1657	0.3966	0.6770		
		-0.0509	0.2157	0.8139		
	3rd	-0.1174	0.2735	0.6687		
		-0.0396	1.8290	0.9828		

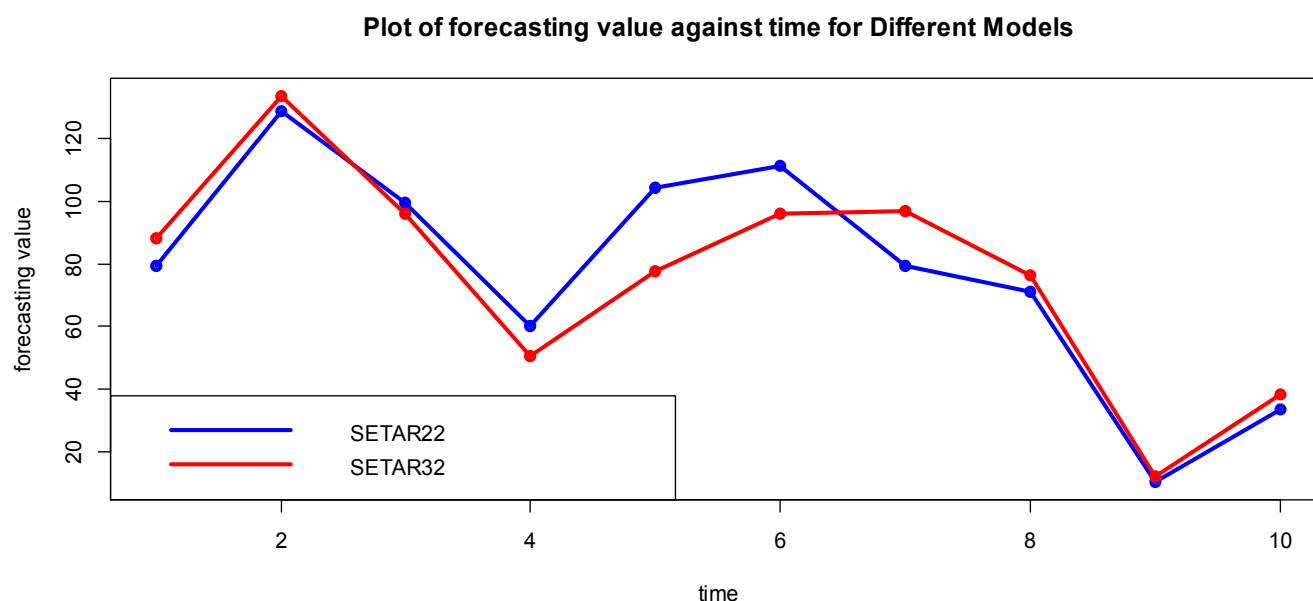


Figure 9. Plot of Forecasting values against time for the best Models for the Lafia rainfall data [SETAR (3,2) and (2,2)].

4. Conclusion

As shown in the study, the results of this research show that SETAR (3, 2) and SETAR (2, 2) are the best fitting models for both simulated and real-life data for small, moderate and large sample sizes. SETAR (3, 2), followed by SETAR (2, 2) at various stages ahead, are also the best forecast models. Although SETAR (3, 3) or higher can be equipped for those with broad sample sizes and higher steps ahead. In addition, SETAR (3, 2) is the best fit for the monthly rainfall data from Lafia, Nasarawa, Nigeria, followed by SETAR (2, 2).

References

- [1] Akeyede, I., Adeleke, B. L. and Yahya, W. B. (2015). On forecast Strength of some linear and nonlinear time series models for stationary data structure, *American Journal of Mathematics and Statistics*, 5 (4): 163-177 DOI: 10.5923/j.ajms.20150504.01.
- [2] Clements, M. P. and Smith, J. (1999). A Monte Carlo study of the forecasting performance of empirical SETAR models, *Journal of Applied Econometrics*, 14, 123-141.
- [3] De Gooijer, J. G. (2001). On threshold moving-average models. *Journal of Time Series Analysis*, 19, 1-18. MR1624163.
- [4] Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive Time series with a unit root, *Journal of American Statistical Association*, 74, 427-431.
- [5] Dufrenot, G., Guegan, D. and Peguin-Feissolle, A. (2005). Long memory dynamics in a SETAR model-applications to stock markets. *Journal of International Financial Markets, Institutions and Money, Elsevier*, 15 (5), 391-406.
- [6] Firat, E. H. (2017). SETAR (Self-exciting Threshold Autoregressive) Non-linear Currency Modelling in EUR/USD, EUR/TRY and USD/TRY Parities. *Mathematics and Statistics* 5 (1): 33-55, DOI: 10.13189/ms.2017.050105.
- [7] Franses, P. H. and van Dijk, D. (2000). *Non-linear time series models in empirical finance*, Cambridge, New York.
- [8] Gibson, D. and Nur, D. (2011). Threshold Autoregressive Models in Finance: A Comparative Approach, Proceedings of the Fourth Annual ASEARC Conference, 17-18 February, University of Western Sydney, Paramatta, Australia.
- [9] Grabowski, D., Staszewska-Bystrova, A. and Winker, P. (2017). Generating prediction bands for path forecasts from SETAR models. *Studies in Nonlinear Dynamics & Econometrics*, 21 (5), 20160066, ISSN (Online) 1558 3708, DOI: <https://doi.org/10.1515/snde-2016-0066>.
- [10] Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis, *Journal of Econometrica* 64, 413-30.
- [11] Keenan, D. M. (1985). A Tukey nonadditivity-type test for time series nonlinearity, *Biometrika* 72, 39-44.
- [12] Tong, H. (1978). On a threshold model, *Pattern Recognition and Signal Processing*, Amsterdam: *Sijthoff & Noordhoff*, 101-41.
- [13] Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion), *Journal of the Royal Statistical Society, Series B, Vol. 42* (3), 245-292.
- [14] Tong, H. (1983). Threshold models in nonlinear time series analysis, *Lecture Notes in Statistics, No. 21, Heidelberg: Springer*, 62-64.
- [15] Tong, H. (1990). *Nonlinear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.
- [16] Tong, H. and Yeung, I. (1991). Threshold autoregressive modelling in continuous time. *Statistica Sinica*, 1 (1991), 411-430.

- [17] Tong, H. (2011). "Threshold models in time series analysis — 30 years on. *Statistics and Its Inference* 2 (2). 1-36, DOI 10.4310/SII.2011.v4.n2.a1.
- [18] Tsay, R. S. (1986). Nonlinearity tests for time series, *Biometrika* 73, 461-466.
- [19] Tsay, R. S. (2010). *Analysis of Financial Time Series*; 3rd Edition, Published by John Wiley & Sons, Inc., Hoboken, New Jersey. Published simultaneously in Canada.
- [20] Watier, L. and Richardson, S. (1995). Modeling of an epidemiological time series by a threshold autoregressive model. *The Statistician* 44 (3): 353-364.