

# Forecast Comparison of Seasonal Autoregressive Integrated Moving Average (SARIMA) and Self Exciting Threshold Autoregressive (SETAR) Models

Akintunde Mutairu Oyewale<sup>1,\*</sup>, Olalude Gbenga Adelekan<sup>1</sup>, Oseghale Osezuwa Innocent<sup>2</sup>

<sup>1</sup>Department of Statistics, School of Applied Sciences, The Federal Polytechnic, Ede, Nigeria

<sup>2</sup>Department of Statistics, College of Natural Sciences, Joseph Ayo Babalola University, Ikeji-Arakeji, Nigeria

## Email address:

waleakintunde2004@gmail.com (A. M. Oyewale), olaludelekan@yahoo.com (O. G. Adelekan),

innocentoseghale@hotmail.com (O. O. Innocent)

\*Corresponding author

## To cite this article:

Akintunde Mutairu Oyewale, Olalude Gbenga Adelekan, Oseghale Osezuwa Innocent. Forecast Comparison of Seasonal Autoregressive Integrated Moving Average (SARIMA) and Self exciting Threshold Autoregressive (SETAR) Models. *American Journal of Theoretical and Applied Statistics*. Vol. 6, No. 6, 2017, pp. 278-283. doi: 10.11648/j.ajtas.20170606.13

**Received:** August 13, 2017; **Accepted:** August 31, 2017; **Published:** November 21, 2017

---

**Abstract:** Financial and Economic time series literatures have shown that financial and economic time series data exhibit non-linearity in their behavior. In order to be mindful of such behavior as applied to Nigeria inflation rates, this study therefore, applies a two stages non-linear self-exciting threshold autoregressive model (SETAR) to Nigeria inflation rates. The results obtained for both in-sample and out-of-sample forecast performances for SETAR model were compared with results of linear seasonal autoregressive integrated moving average (SARIMA). On the basis of in-sample forecast performance of linear SARIMA and non-linear SETAR, using performance measure indices like MAE and RMSE, the results obtained indicated that non-linear SETAR model performed better than linear SARIMA. So also for the out-of-sample forecast performance using multi-step ahead forecast performance, the results also indicated that non-linear SETAR out performed linear SARIMA.

**Keywords:** SETAR Model, SARIMA Model, Inflation Rates, In-Sample, Out-of-Sample, Forecast Performance

---

## 1. Introduction

It is obvious that linear time series models are too restrictive to capture economically interesting asymmetries and empirically observed nonlinear dynamics. This has over the years being the research focus of academia, investors and practitioners. Models that can capture nonlinear dynamics have also been the research attraction of time series and econometrician statisticians so as to obtain different field specific forecast models such as Bilinear models, Random Coefficient Models, State Dependent Models SETAR models, GARCH models and soon. Vast result has been concentrated on describing sunspot and Canadian lynx data example of this include but not limited to [11], Granger and Terasvirta [12], [9], [8], and references therein.

Of interest to time series experts and econometricians is the behavior of regime switching models whereby the parameters of models are made to change depending on the

occurrence of a particular event, episode or policy (e.g. recessions or expansions, periods of low/high stock market valuations, low/high interest rates etc) but are otherwise constant within regimes. Popular models that can be categorized within this group are the well known Markov switching models popularized by Hamilton's early work [4] and which model parameter change via the use of an unobservable discrete time Markov process. This class of models in which parameter changes are triggered by an unobservable binary variable has been used extensively as an intuitive way of capturing policy shifts in Macroeconomic models as well as numerous other contexts such as forecasting economic growth and dating business cycles. In Leeper and Zha [6], Farmer, [8], [7], for instance the authors use such models to introduce the concept of monetary policy switches and regime specific Taylor rules. Other particularly fruitful areas of application of such regime switching specifications has involved the dating of Business Cycles, the

modelling of time variation in expected returns among numerous others (see Hamilton [4], [8] etc.). An alternative, parsimonious and dynamically very rich way of modelling regime switching behaviour in economic data is to take an explicit stand on what might be triggering such switches and adopt a piecewise linear setting in which regime switches are triggered by an observed variable crossing an unknown threshold. Such models have been proposed by Howell Tong in the mid 70s and have gone through an important revival following their adoption by Economists and Econometricians during the 80s and 90s following the methodological work of Bruce Hansen (see also Hansen (2011) and references therein for a historical overview), [15], [12], [14] and others. When each regime is described by an autoregressive process and the threshold variable causing the regime change is also a lagged value of the variable being modeled we have the well known Self Exciting Threshold Auto-Regressive class of models (SETAR) extensively studied in the early work of [11] and others [12-14]. In general however the threshold principle may apply to a wider range of linear univariate or multivariate models and need not be solely conned to autoregressive functional forms. Similarly the threshold variable triggering regime switches may or may not be one of the variables included in the linear part of the model. Despite their simplicity, such models have been shown to be able to capture a very diverse set of dynamics and asymmetries particularly relevant to economic data. Important examples include the modelling of phenomena such as costly arbitrage whereby arbitrage occurs solely after the spread in prices exceeds a threshold due for instance to transport costs [6], [14], and [16]. It is the intention of this study to compare the forecast performance of SETAR and SARIMA models.

## 2. Mathematical Specification of the Models

The generalization of well-known Box-jenkins ARIMA model to accommodate data with characteristics of both season and non-seasonal is called seasonal autoregressive integrated moving average (SARIMA) model. ARIMA model is the combination of Autoregressive (AR) and Moving average (MA) models. The autoregressive part takes care of the past information of the series under consideration while the moving average aspect takes care of the past error of the series [4] and [11]. The ARIMA model with its order is usually written as ARIMA ( $p, d, q$ ) where  $p, d$  and  $q$  are integers greater than or at least equal to zero and refers to the order of the autoregressive, integrated and moving average parts respectively. Parameter  $p$  refers to number of autoregressive lags; the second parameter  $d$  means the other of integration that reflects the stationarity of data and  $q$  refers to the lag in moving average.

The ARIMA model is written as

$$\phi(L)(1-L)^d y_t = \theta(L)\varepsilon_t, \{\varepsilon_t\} \sim WN(0, \sigma^2)$$

Where  $\varepsilon_t$ , follows a white noise (WN) process. The autoregressive operator and moving average operator are as defined below;

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$$

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$$

$\phi(L) \neq 0$  for  $|\phi| < 1$ , the process  $\{y_t\}$  is stationary if and only if  $d = 0$ , if this happens it reduces to ARMA ( $p, q$ ).

The generalization of ARMA model to SARIMA model occurs when series contains seasonal and non-seasonal features. This behavior rendered ARIMA model inefficient to be applied to the series with such behavior, because it might not be able to adequately account for the seasonal pattern of the series and if use may lead to wrong order selection for non-season components. The SARIMA model is denoted by ARIMA ( $p, d, q$ ) ( $P, D, Q$ )<sub>s</sub>. The lag form of which was given by (Halim and Bisomo, 2008) is given below:

$$\phi(L)\Phi(L^s)(1-L)^d(1-L^s)^D y_t = \theta(L)\Theta(L^s)\varepsilon_t$$

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$$

$$\Phi(L^s) = 1 - \Phi_1 L^s - \Phi_2 L^{2s} - \dots - \Phi_p L^{ps}$$

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$$

$$\Theta(L^s) = 1 - \Theta_1 L^s - \Theta_2 L^{2s} - \dots - \Theta_q L^{qs}$$

Having established the stationarity condition of data, the suitable model can thereafter be determined. The Order of the model which AR, MA, SAR and SMA terms follow can be determined using ACF and PACF plot. ACF and PACF has spikes at ks and cut off after lag ks at season level. The numbers of significant spikes suggest the order of the model as shown in the table below.

$$\phi(L)\Phi(L^s)(1-L)^d(1-L^s)^D y_t = \theta(L)\Theta(L^s)\varepsilon_t$$

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$$

$$\Phi(L^s) = 1 - \Phi_1 L^s - \Phi_2 L^{2s} - \dots - \Phi_p L^{ps}$$

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$$

$$\Theta(L^s) = 1 - \Theta_1 L^s - \Theta_2 L^{2s} - \dots - \Theta_q L^{qs}$$

Having established the stationarity condition of data, the suitable model can there after be determined. The Order of the model which AR, MA, SAR and SMA terms follow can be determined using ACF and PACF plot. ACF and

PACF has spikes at ks and cut off after lags at season level. The numbers of significant spikes suggest the order of the model as shown in the table below.

2.1. Character Exhibited by ACF and PACF for Pure Seasonal ARMA(P,Q)s

Table 1. ACF AND PACF FOR PURE SEASONAL ARMA(P,Q)s.

	AR(P)s	MA(Q)s	ARMA(P,Q)s
ACF	Tailsoffatlag ks k=1,2,3,...	Cutoffafterlag Qs	Tailsoffatlag ks
PACF	Cutoffafterlag Ps	Tailsoffatlag ks k=1,2,3,...	Tailsoffatlag ks

On the other hand the class of the Threshold Autoregressive (TAR) model commonly use in literature is self-exciting threshold autoregressive (SETAR) model proposed by [11] and [12] (Tong 1978, 1983). For instance Tong and Yeung [12] for beach water pollution, [15] for futures markets, [12] for epidemiological applications, [16] for sea surface temperatures, Montgomery et al. For the examination of statistical properties and forecasting ability of SETAR model check [11], [13], [16], [12] and [14]. The two regime version of the SETAR model of order p as given by [17] is as follows;

$$\begin{aligned}
 y_t = & \phi_0^{(1)} + \sum_{j=1}^{p_1} \phi_j^{(1)} Y_{t-j} + \varepsilon_t^{(1)}, \text{ if } y_{t-d} \leq r_1 \\
 & \phi_0^{(2)} + \sum_{j=1}^{p_2} \phi_j^{(2)} Y_{t-j} + \varepsilon_t^{(2)}, \text{ if } r_1 y_{t-d} \leq r_2 \\
 & \dots \\
 & \dots \\
 & \phi_0^{(k)} + \sum_{j=1}^{p_k} \phi_j^{(k)} Y_{t-j} + \varepsilon_t^{(k)}, \text{ if } r_{h-1} < y_{t-d}
 \end{aligned}
 \tag{1}$$

$\phi_j^{(1)}$  and  $\phi_j^{(2)}$  are the coefficient of lower and upper regime respectively which needs to be determined, r is the threshold value,  $p^{(1)}$  and  $p^{(2)}$  are the order of linear AR model in low and high regime respectively. The orders of the AR model in both regimes are the same.  $y_{t-d}$  is the threshold variable that governs the transition between low and high regimes.

2.2. Forecast from the Model Used in the Study

I. SARIMAMODEL

The forecast equations used in the study is patterned towards cryer and Chan 2008.

$$y_t - y_{t-1} = \Phi(y_{t-1,2} - y_{t-1,3}) + \varepsilon_t - \theta\varepsilon_{t-1} - \Theta\varepsilon_{t-1,2} + \theta\Theta\varepsilon_{t-1,3}$$

$$y_t = y_{t-1} + \Phi y_{t-1,2} - \Phi y_{t-1,3} + \varepsilon_t - \theta\varepsilon_{t-1} - \Theta\varepsilon_{t-1,2} + \theta\Theta\varepsilon_{t-1,3}$$

The one step ahead forecast from the origin is given by

$$\hat{y}_{t+1} = y_t + \Phi y_{t-1,1} - \Phi y_{t-1,2} - \theta\varepsilon_{t-1} - \Theta\varepsilon_{t-1,1} + \theta\Theta\varepsilon_{t-1,2}$$

While the next step is

$$\hat{y}_{t+2} = \hat{y}_{t+1} + \Phi y_{t-1,0} - \Phi y_{t-1,1} - \theta\varepsilon_{t-1} - \Theta\varepsilon_{t-1,0} + \theta\Theta\varepsilon_{t-1,1}$$

And soon. The noise terms are  $\varepsilon_{1,3}, \varepsilon_{1,2}, \varepsilon_{1,1}, \varepsilon_{1,0}, \dots, \varepsilon_i$  as residuals will fit into forecasts for lead times.

II. SETARMODEL

The optimal forecast one step ahead forecast is given by Franses and Van Dijk, [16] as follows:

$$\hat{y}_{t+1|r} = E[y_{r+1} | \Omega_t] = E[F(x_t : \phi)]$$

Where  $\hat{y}_{t+1}$  represents the forecast value for (t+1), and  $\Omega_t$  is the time history of the series upto time t.  $F(x_t : \phi)$  is the non-linear function that represent the SETAR model. The next optimal forecast is:

$$\hat{y}_{t+2|r} = E[y_{r+1} | \Omega_t] = E[F(x_{t+1} : \phi) | \Omega_t]$$

Conclusively, the optimal h-step ahead forecast can be obtained as:

$$\hat{y}_{t+h|r} = E[y_{r+h} | \Omega_t] = F(x_{t+h-1} : \phi)$$

2.3 Forecast Comparism

The forecast used in the study is compared using performance measure indices. It should be noted that a model with minimum forecast errors compare to other competing models. To determine a model that gives the best forecasting accuracy, the accuracy of each model can be checked to determine the performance of the models for both in-sample and out-of-sample forecast. The accuracy of the models are compared using performance measure indices such as mean square error (MSE), mean absolute error (MAE), mean absolute precision error (MAPE) and Theil's u inequality coefficient (TIC).

I. MEAN SQUARE ERROR (MSE)

$$MSE = N^{-1} \sum_{t=1}^T (Y_t - \hat{Y}_t)^2$$

II. MEAN ABSOLUTE ERROR (MAE)

$$MAE = N^{-1} \sum_{t=1}^T |Y_t - \hat{Y}_t|$$

III. MEAN ABSOLUTE PRECISION ERROR (MAPE)

$$MAPE = N^{-1} \sum_{t=1}^T \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100$$

IV. THEIL'S UNEQUALITY COEFFICIENT

$$U = \frac{\sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t^s - Y_t^a)^2}}{\sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t^s)^2 + \frac{1}{n} \sum_{t=1}^n (Y_t^a)^2}}$$

V. BIAS

$$U^M = \frac{(\bar{Y}^s - \bar{Y}^a)^2}{\frac{1}{n} \sum_{t=1}^n (Y_t^s - Y_t^a)^2}$$

VI. VARIANCE

$$U^S = \frac{(\sigma_s - \sigma_a)^2}{\frac{1}{n} \sum_{t=1}^n (Y_t^s - Y_t^a)^2}$$

VII. COVARIANCE

$$U^C = \frac{2(1-\rho)\sigma_s\sigma_a}{\frac{1}{n} \sum_{t=1}^n (Y_t^s - Y_t^a)^2}$$

### 3. Empirical Illustration with Nigerian Inflation Rates Data

The inflation rate data of Nigeria was obtained from the Consumer Price Index of Nigeria Bureau of Statistics from 1993 to 2013. The data covers. All methodologies used were applied to the data. We used E-view and R code for the data analysis.

#### 3.1. Descriptive Statistics

To assess the distributional properties of the series under study, various descriptive statistics are shown below in the figure 1 below:

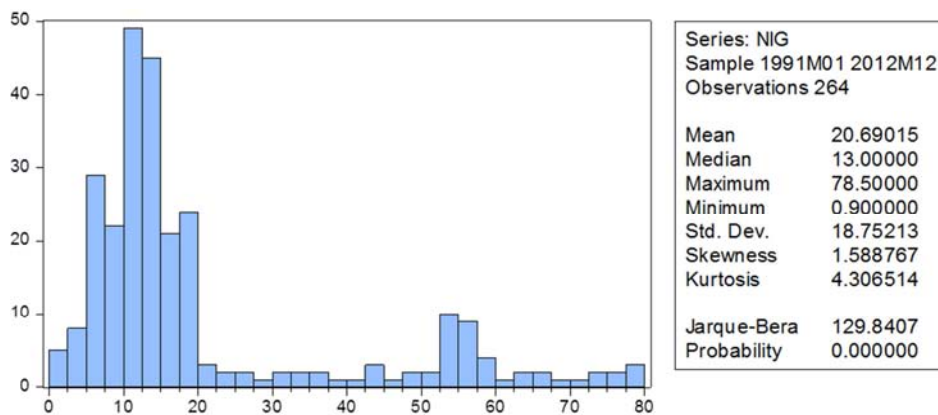


Figure 1. Histogram and descriptive statistics of the series.

The standard deviation is very high, indicating high level of fluctuations in the series, so also the presence of positive skewness with long right tail, showing that the series is non-symmetric. The histogram of the series shows that it is non-normal (leptokurtic).

#### 3.2. Stationarity Test and Model Identification

Two measures were used to determine the stationarity status of the series under study, they are graphical and unit root methods. In the graphical method stationarity was not achieved at level and at the first difference (figures 2 and 3), however, this was achieved at second difference (as shown in figures 4). So also unit root conducted using Augmented Dickey fuller's test revealed that at level and first difference the series was not stationary (Tables 2 and 3), but at second difference stationarity condition was achieved (as shown in table 4):

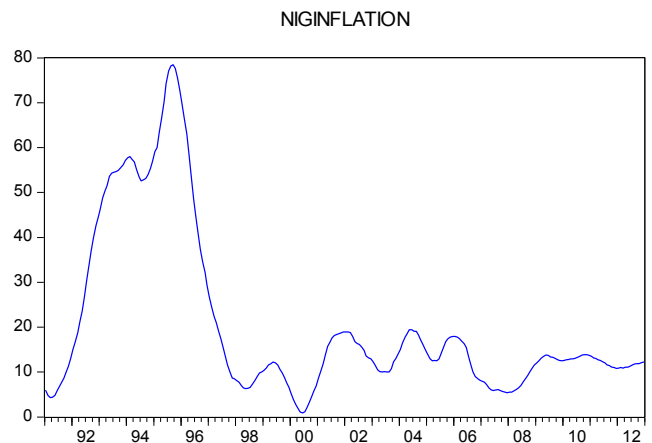


Figure 2. Line Graph of Level of Nigeria's Inflation Rates.

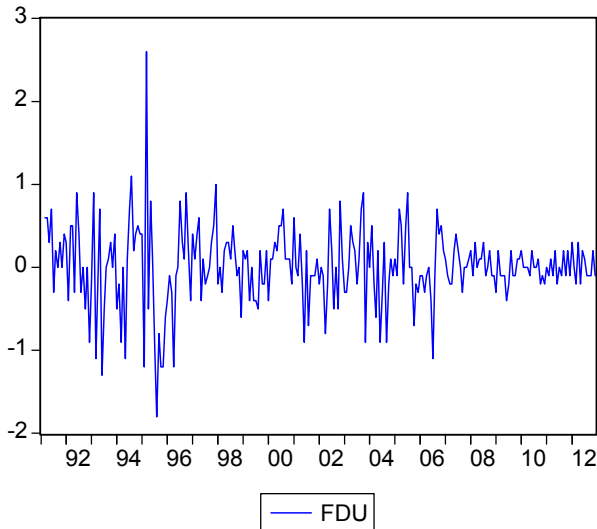


Figure 3. Line Graph of First Difference of Nigeria's Inflation Rates.

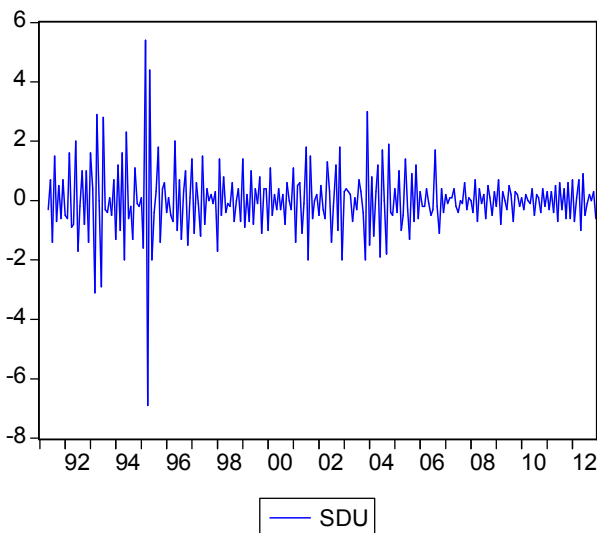


Figure 4. Line Graph of Second Difference of Nigeria's Inflation Rates.

Table 2. Unit root test output for the level of the series.

Series	ADF-Test statistic	Critical value (5%)	Mackinnon prob •
NIG. INFLATION RATE	-2.6645	-2.8729	0.0817

Table 3. Unit root test output for the first difference of the series.

Series	ADF-Test statistic	Critical value (5%)	Mackinnon prob •
NIG. INFLATION RATE	-2.5667	-2.8729	0.1014

Table 4. Unit root test output for the second difference of the series.

Series	ADF-Test statistic	Critical value (5%)	Mackinnon prob •
NIG. INFLATION RATE	-8.1872	-2.8729	0.0000

Table 5. Unit root test output for the second difference of the series.

Series	ADF-Test statistic	Critical value (5%)	Mackinnon prob •
NIG. INFLATIONRATE	-8.1872	-2.8729	0.0000

FORECAST PERFORMANCE MEASURES INDICES

Using all indices as discussed in the mathematical preliminaries, the following tables were generated from the analysis. Both in-sample and out-of sample forecast performances of the series for both season autoregressive integrated moving average (SARIMA) and self exciting transition autoregressive (SETAR) models. The results obtained show that in each case (in-and out-of sample) SETAR model actually out-performed SARIMA model.

Table 6. In-sample forecast performances.

INDICES	SARIMA	SETAR
RMSE	18.75116	1.92342
MAE	14.15307	1.52261
MAPE	120.60480	22.52000
THEIL'SU	0.38616	0.12650
BIAS	0.000003	0.000002
VAR.	0.97294	0.98213
COV.	0.0106	0.00533

Table 7. Out-of sample forecast performances.

INDICES	SARIMA	SETAR
RMSE	1.14573	0.32123
MAE	0.81662	0.21800
MAPE	57.63100	9.8312
THEIL'SU	0.20880	0.00146
BIAS	0.000001	0.000001
VAR.	0.98176	0.99234
COV.	0.018241	0.00001

4. Conclusion

This paper focuses on building a model for Nigerian inflation rates using a season autoregressive integrated moving average (SARIMA) model and Self exciting transition autoregressive (SETAR) model. We used monthly data of Nigerian inflation rates from 1991 to 2012. Stationarity of the series was established using both graphical Augmented-Dickey fuller test, at level and first difference the series was not stationary, but at second difference it was stationary. We thereafter use performance measure indices to evaluate both the in-sample and out-of sample forecast performances which show that SETAR model has a better performances in term of both in and out of sample forecast performance.

References

- [1] Anderson, T. W. (2003). Estimation for Autoregressive Moving Average Models in the Time and Frequency Domains. The Ann. Of Stat. Vol 5 No. 5, 842-865.
- [2] Anderson, T. W. (1977). The Statistical Analysis of Time Series, New York and London Wiley.
- [3] Bonilla, C. R. Romero, M. J., and Hinich, M. J., (2005). Episodic non-linearities in the Latin American Stock Market indices, Applied Economics Letters, vol. 13 pp 195-199.

- [4] Hamilton J. D., (1989). A new approach to the economic analysis of non-stationary time series and the business cycle. *Econometrica*, 57(2): 357–384.
- [5] Lo, M. and Piger, J., (2005). “Is the Response of Output to Monetary Policy Asymmetric? Evidence from a Regime-Switching Coefficients Model,” *Journal of Money, Credit and Banking* 37, 865-887.
- [6] Ocal, N. and Osborn, D. R., (2000). Business cycle nonlinearities in UK consumption and production, *Journal of Applied Econometrics* 15, 27-44.
- [7] Panagiotidis, T. and Pelloni, G. (2003) Testing for nonlinearity in labour markets: the case of Germany and the UK, *Journal of Policy Modeling*, 25, 275–86.
- [8] Preez, J. & Witt, S. F. (2003). Univariate versus Multivariate Time Series Forecasting: An Application to International Tourism Demand. *International Journal of Forecasting*, 19, 435-451.
- [9] Pesaran, M. H. and S. M. Potter, 1997. A Floor and Ceiling Model of US Output. *Journal of Economic Dynamics and Control* 21: 661-695.
- [10] Teräsvirta, T. and Anderson, H., (1992). Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics*, 7: S119–S136.
- [11] Tong, H. (1990) *Non-Linear Time Series: A Dynamical System Approach*. London: Oxford University Press.
- [12] Tong, H., & Lim, K. S. (1980). Threshold autoregressive, limit cycles and cyclical data. *Journal of the Royal Statistical Society Series B*, 42(3), 245–292.
- [13] Tong, H., (1983). *Threshold Models in Non-Linear Time Series Analysis*. Springer-Verlag, New York.
- [14] Tsay, R. S. (1989). Testing and Modeling Threshold Autoregressive Processes. *Journal of the American Statistical Association* 84: 231–240.
- [15] Tsay, R. S. (2005). *Analysis of Financial Time Series 2ED*. John Wiley & Sons, Inc., Hoboken, New Jersey.
- [16] Van Dijk, Teräsvirta, T. and Franses P., (2002). Smooth transition autoregressive models – a survey of recent developments. *Econometric Reviews*, 21(1): 1–47.
- [17] Yaser, A. S., Moody, J. and Weigend, A. (1996). Introduction to financial forecasting. *Applied intelligence*, 6(3), 205-213.
- [18] Zhou, Z. J., & Hu, C. H. (2008). An effective hybrid approach based on grey and ARMA for forecasting gyro drift. *Chaos, Solutions and Fractals*, 35, 525–529.