



Science Forum (Journal of Pure and Applied Sciences)

Journal Homepage: <https://atbu.edu.ng/science-forum/>



Compare the Performances and Identification of Optimal Lag Order on Selected Linear and Non-Linear Time Series Models of Different Orders

T. A. BATURE¹, K. E. LASISI², A. ABDULKADIR², M. O. ADENOMON³

¹Dept. of Mathematics, Nigerian Army University Bui, Borno State, Nigeria.

²Dept. of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi

³Department of Statistics, Nasarawa State University Keffi, Nigeria.

ABSTRACT

This study sets out to evaluate and compare the performance of Autoregressive (AR) models against modified versions of several smooth transition autoregressive models, including Inverse Smooth Transition Autoregressive (ISTAR), Exponential Smooth Transition Autoregressive (ESTAR), and Trigonometric Smooth Transition Autoregressive (TSTAR) models. Specifically, we examine how these models perform across varying sample sizes and different orders. A comprehensive numerical simulation was conducted to assess the relative efficiency of both linear and nonlinear models, with sample sizes ranging from 20 to 250, for models of the first, second, and third orders. The evaluation of each model's performance is based on widely used statistical selection criteria, such as the Bayesian Information Criterion (BIC), Hannan-Quinn Information Criterion (HQC), and Akaike Information Criterion (AIC). For each order, the model with the lowest value for a given criterion is considered the most suitable. The findings indicate that, for second-order models, AR models consistently outperform their smooth transition counterparts, as evidenced by lower AIC and BIC values. However, for third-order models, the TSTAR model emerges as the most effective, outperforming other models based on HQC and AIC. These results highlight the strengths and weaknesses of both linear and nonlinear models at different complexity levels, providing useful insights into model selection for time series analysis.

ARTICLE INFO

Received 12 August 2024 Received in

Revised Form 21 August 2024

Accepted 2nd September 2024

Published 2nd December 2024

Available online December 2024

KEYWORDS

Smooth Transition Autoregressive, Sample Sizes, Orders and functions, Linear and Non-Linear Models.

1.0 Introduction

By examining historical occurrences that have been tracked and compiled at regular intervals, time series models are used to forecast future events. A time series model typically includes the mean component and the conditional variance component. (Box and pierce 1970). Time series is an order in which data points are measured consecutively across time. In fact, time series data are made up of a variety of numbers and times. A sequence of clearly defined procedures is followed to implement autoregressive, moving average, exponential, seasonal, and autoregressive moving average modelling. The initial stage involves the identification of the model. The identification step entails defining the appropriate structure. (Autoregressive (AR) models, Smooth Transition Autoregressive, Exponential Smooth Autoregressive model and Trigonometric model) if under distributions the model is stationary or non-stationary. Plots of the autocorrelation function and partial autocorrelation function are sometimes used to identify. Occasionally, an automated iterative approach that fits a large number of possible model structures and orders them is used to perform identification. We then use a goodness-of-fit statistic to find the best model.

Regressive in nature The persistence, or autocorrelation, of a time series is captured by mathematical models like moving average and smooth transition. Econometrics, hydrology, engineering, and many other professions frequently use these models. There are several justifications for employing trigonometric, exponential smooth autoregressive, smooth transition autoregressive, and autoregressive (AR) models in data analysis. By illuminating the fundamental mechanisms that give rise to the series' persistence, modelling can offer important new perspectives on the physical system. These models are also useful for predicting, from past values, the behavior of an economic data set or a time series. To evaluate the possible importance of other system variables, one might use this forecast as a point of reference. They are commonly employed to

forecast economic and industrial time series. For simulation, can also utilise exponential smooth autoregressive model, trigonometric model, smooth transition autoregressive (STAR), and iberregressive (AR) models. These approaches enable the creation of synthetic series that resemble an observed series' persistence structure. When establishing confidence ranges for statistics and estimated econometric quantities, simulations are especially helpful.

Autoregressive, autoregressive moving average, moving average, exponential, and seasonal modelling abide by a precise set of guidelines. The first step is to identify the model. Finding the right structure for the model—whether stationary or non-stationary under various distributions—requires identification. This can be accomplished with a variety of models, like Autoregressive (AR) models, Smooth Transition Autoregressive models, Exponential Smooth Autoregressive models, and Trigonometric models. The autocorrelation function and partial autocorrelation function graphs can be examined to aid with identification. A common method for achieving identification is an automated iterative procedure that fits different model structures and ordering. We then use a goodness-of-fit statistic to find the best model.

2.0 Literature Review

The autoregressive (AR) model is commonly used as the classical model for the mean part. A prominent topic of discussion when studying time series data is stationarity, which describes how a certain variable behaves over time. Stationarity consists of three parts. The series' mean remains constant, suggesting that there is no innate tendency for the series' mean to change over time. Furthermore, it is presumable that the series' variance won't change with time. Most people agree that the autocorrelation pattern holds true for the whole series. Many non-linear time series models have been proposed in the last two decades. These include, among many others, the bilinear model by Granger and Andersen (1978), the threshold AR model by Fan, J.Q. (2019), the amplitude dependent exponential AR (EX PAR) model by Maulana, A., & Slamet, I. (2020), and the random

coefficient of AR model by Ratnasingam, S., & Ning, W (2020).

Time series modelling employs a number of techniques that presuppose linear relationships between variables. But occasionally, data fluctuations don't exhibit basic linearity, making them challenging to interpret and forecast with accuracy. Hence, for such data, it would not be practicable to expect a single, linear model to capture these distinct behaviours. Linear relationships and various combinations of them are frequently insufficient for accurately characterising the behaviour of such data. Time series analysis reveals a multitude of nonlinear features, including cycles, thresholds, bursts, chaos, heteroscedasticity, asymmetries, and various combinations of these. Tong (1990), Granger and Teräsvirta (2019), Franse, and van Dijk (2020), Tsay(2020) and Kim and Nelson (2019) have discussed the different models that can be formulated in these forms.

The key assumption in analyzing time series data is stationarity. The term "stationarity" refers to the continuity across time of the statistical characteristics influencing the process's behaviour. Essentially, there is statistical equilibrium in the process. A process Y_t is regarded as perfectly steady if and only if the combined distribution of Y_t matches that of Y_{t-k} for any values of t and k .

$t = 1, 2, \dots, k$. In other words, the Y 's are (marginally) identically distributed (Tsay and Kung-Sik, (2018). It follows that $E(Y_t) = E(Y_{t-k})$ for every t and k in order to maintain the mean function's constant across time. Furthermore, $Var(Y_t) = Var(Y_{t-k})$ for all t and k so that the variance is also constant over time. A stationary time series' basic tenet is that it displays white noise, which means that the model's error term has to have a normal distribution with a zero mean and a particular variance σ^2 .

Imagine a collection of time-ordered data points, denoted as $\{y_1, y_2 \dots \dots y_n\}$, that come from seeing a certain phenomenon observed. Observations taken over time, either continuously or at predetermined intervals,

make up time series data. (Kim, 2022). One of the essential aspects in model fitting, particularly when working with time series data and econometrics, is understanding the most effective methods to employ. AR models, Smooth Transition Autoregressive, Exponential Smooth Autoregressive model and Trigonometric model. These models' central tenet is the requirement for stationarity, which states that the input data must show stationarity.

Regressive in nature Mathematical models that depict a time series' persistence, or autocorrelation, include moving average and smooth transition models. The models find widespread use in the domains of hydrology, engineering, econometrics, and related topics. There are multiple justifications for using Autoregressive (AR) models, Smooth Transition Autoregressive (STAR) models, Exponential Smooth Autoregressive models, and Trigonometric models to data. By offering insights into the underlying physical mechanism that gives the series persistence, modelling helps improve understanding of the physical system. The models are also useful for predicting, from past values, the behaviour of an economic data set or a time series. This prediction can be used as a guide to determine the possible importance of other system factors. They find widespread application in the prediction of industrial and economic time series. Autoregressive (AR) models, Smooth Transition Autoregressive (STAR), Exponential Smooth Autoregressive model, and Trigonometric model can be used for simulation. This involves generating synthetic series that have the same persistence structure as an observed series. Simulations are particularly valuable for determining confidence ranges for statistical measures and estimated values in econometrics.

These studies suggest that, in trying to decide by classical methods whether economic data are stationary or integrated, it would be useful to perform tests of the null hypothesis of stationarity as well as tests of the null hypothesis of a unit root. This paper provides a straight forward test of the null hypothesis of stationarity against the alternative of a unit root at different

order of autoregressive and moving average and various sample sizes. There have been surprisingly few previous attempts to test the null hypothesis of stationarity. Park and Mahdi, E. (2020) consider a test statistic which is essentially the F statistic for 'superfluous' deterministic trend variables; this statistic should be close to zero under the stationary null but not under the alternative of a unit root. Zhang, J and Zhou, B (2022) considers the Dickey-Fuller test statistics, but estimates both trend-stationary and difference-stationary models and then uses the bootstrap to evaluate the distribution of these statistics.

3.0 Material and Methods

In this section, Simulation studies will be conducted to investigate the performances of Tests of stationarity under different orders of Autoregressive with Modified version models.

$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t$, where e_t is white noise process with zero mean and variance σ^2 . $\phi_1, \phi_2, \dots, \phi_p$ are autoregressive parameters. The mathematical model for existing and proposed models are

$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t$ is an existing model for Autoregressive model, where e_t is white noise process with zero mean and variance σ^2 .

ISTAR_(p) is $Y_t = \frac{\phi_1 Y_{t-1} + Y_{t-2} + \dots + Y_{t-p} + e_t}{e^{\phi_1 Y_{t-1} + Y_{t-2} + \dots + Y_{t-p} + e_t}}$ is called Inverse Smooth Transition Autoregressive model. Where e_t is white noise process with zero mean and variance (σ^2), Y_t is variable of interest at time (t), $\phi_1, \phi_2, \dots, \phi_p$ are the coefficient that define the unit root.

ESTAR_(p) is $Y_t = e^{\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t}$ is called Exponential Smooth Transition Autoregressive model.

TSTAR_(p) is $Y_t = \text{Sin}\phi_1 Y_{t-1} + \text{Sin}\phi_2 Y_{t-2} + \dots + \text{Sin}\phi_p Y_{t-p} + e_t$ is called Trigonometric Smooth Transition Autoregressive model.

selection rule

We calculate the Mean Absolute Percentage Error (BIC), Mean Square Error (HQC), and Akaike Information Criteria (AIC) for the given sample sizes. (n)=20, 50, 70, 100, 120, 150, 180, 200, 230 and 250 For each model, we determine the model with the lowest values for these criteria, which we consider to be the optimal model. In this study order 1,2 and 3 of each models are considered in respective case and situation.

4.0 Results and Discussion

This section includes simulation studies for the Autoregressive models and suggested models, including ISTAR, ESTAR and TSTAR at various orders under stationary conditions, evaluate and contrast the results, and look at how each model's sample size at various orders affected the simulated data for each model.

For the simulation study, the parameters of the linear and nonlinear Autoregressive functions in models 1-4 are fixed as $\phi_1 = 0.5, \phi_2 = 0.3$ and $\phi_3 = 0.1$. The sample sizes simulated from each of model cases are;

20,50,70,100,120,150, 180, 200, 230 and 250. The simulation study was conducted 1000 times for various versions of the model functions, using a certain sample size.

Table 1 The fitted models' performances for the first order on the Basis of BIC criteria. $AR(1) = Y_{ti} = 0.5Y_{t-1} + e_t$, $ISTAR = \frac{0.5Y_{t-1} + e_t}{e^{0.5Y_{t-1} + e_t}}$, $ESTAR(1) Y_{ti} = e^{0.5Y_{t-1} + e_t}$, $TSTAR(1) = \sin 0.5Y_{t-1} + e_t$

Sample Size(n)	BIC			
	AR	ISTAR	ESTAR	TSTAR
20	10.0674	11.1344	12.159	11.1802
50	10.0647	11.1272	12.1383	11.1409
70	10.0421	11.0122	11.0484	11.0981
100	10.964	10.9966	10.9768	11.0980
120	10.9481	10.9905	10.9676	11.0707
150	10.9018	10.9764	10.9863	10.065
180	10.8956	10.9661	10.9147	10.9639
200	10.8767	10.9428	10.9208	10.946
230	10.8748	10.9142	10.8939	10.9079
250	10.8716	10.8344	10.8915	10.8878

Table 2 The fitted models' performances for the first order on the Basis of HQC criteria. $AR(1) = Y_{ti} = 0.5Y_{t-1} + e_t$, $ISTAR = \frac{0.5Y_{t-1} + e_t}{e^{0.5Y_{t-1} + e_t}}$, $ESTAR(1) Y_{ti} = e^{0.5Y_{t-1} + e_t}$, $TSTAR(1) = \sin 0.5Y_{t-1} + e_t$

HQC			
AR	ISTAR	ESTAR	TSTAR
11.142	11.871	12.358	11.3786
11.059	11.287	12.311	11.1587
11.042	11.168	12.108	11.1492
11.020	11.166	11.0865	11.1454
10.935	11.139	11.0642	11.141
10.891	11.096	11.0553	11.0710
10.888	11.051	11.9663	11.0464
10.887	10.947	11.9421	11.0459
10.866	10.907	11.9208	11.9079
10.731	10.839	10.9147	11.8878

Table 3 The fitted models' performances for the second order on the Basis of AIC criteria. $AR(2): Y_{ti} = 0.5Y_{t-1} + 0.3Y_{t-2} + e_t$, $ISTAR = \frac{0.5Y_{t-1} + 0.3Y_{t-2} + e_t}{e^{0.5Y_{t-1} + 0.3Y_{t-2} + e_t}}$, $ESTAR(2): Y_{ti} = e^{0.5Y_{t-1} + 0.3Y_{t-2} + e_t}$

$TSTAR(2): Y_{ti} = \sin 0.5Y_{t-1} + \sin 0.3Y_{t-2} + e_t$

Sample Size(n)	AIC			
	AR	ISTAR	ESTAR	TSTAR
20	10.064	11.1314	11.156	12.1772
50	10.061	11.1242	11.1353	12.1379
70	10.039	11.0092	11.0454	12.0951
100	9.961	10.9936	10.9738	11.077
120	9.9451	10.9875	10.9646	11.0677
150	9.8988	10.9734	10.9833	11.062
180	9.8926	10.9631	10.9117	10.9609
200	9.8737	10.9398	10.9178	10.943
230	9.8718	10.9112	10.8909	10.9049
250	9.8686	10.8314	10.8691	10.8848

Table 4 The fitted models' performances for the second order on the Basis of BIC criteria. AR(2): $Y_{ti} = 0.5Y_{t-1} + 0.3Y_{t-2} + e_t$, ISTAR = $\frac{0.5Y_{t-1} + 0.3Y_{t-2} + e_t}{e^{0.5Y_{t-1} + 0.3Y_{t-2} + e_t}}$, ESTAR(2): $Y_{ti} = e^{0.5Y_{t-1} + 0.3Y_{t-2} + e_t}$, TSTAR(2): $Y_{ti} = \sin 0.5Y_{t-1} + \sin 0.3Y_{t-2} + e_t$

BIC			
AR	ISTAR	ESTAR	TSTAR
12.0584	11.0254	14.15	11.1712
12.0557	11.0182	14.1293	11.1319
12.0331	11.0032	13.0394	11.0891
11.9555	10.9376	12.9678	11.0713
10.9391	10.9215	12.9586	10.0617
10.8928	10.8674	11.9773	10.056
10.8866	10.8571	11.9257	10.9549
10.8677	10.8338	11.9118	10.937
10.8658	10.8052	10.8849	10.8989
10.8626	10.8254	10.8631	10.8788

Table5: The fitted models' performances for the third order on the Basis of HQC criteria. AR(3): $Y_{ti} = 0.5Y_{t-1} + 0.3Y_{t-2} + 0.1Y_{t-3} + e_t$, ISTAR(3): $Y_{ti} = \frac{0.5Y_{t-1} + 0.3Y_{t-2} + 0.1Y_{t-3} + e_t}{e^{0.5Y_{t-1} + 0.3Y_{t-2} + 0.1Y_{t-3} + e_t}}$, ESTAR(3): $Y_{ti} = e^{0.5Y_{t-1} + 0.3Y_{t-2} + 0.1Y_{t-3} + e_t}$, TSTAR(3): $Y_{ti} = \sin 0.5Y_{t-1} + \sin 0.3Y_{t-2} + \sin 0.1Y_{t-3} + e_t$.

Sample Size(n)	HQC			
	AR	ISTAR	ESTAR	TSTAR
20	12.1039	11.8321	12.3198	10.0396
50	12.0206	11.2487	12.2722	10.0197
70	11.0033	11.1293	12.069	10.0102
100	10.9818	11.1274	11.0475	10.0064
120	10.8964	11.1001	11.0252	10.0042
150	10.8526	11.0577	10.0163	10.032
180	10.8499	11.0124	10.9273	10.0074
200	10.8485	10.9083	10.9031	10.0069
230	10.827	10.8687	10.8818	10.0065
250	10.6927	10.8006	10.8757	10.0044

Table 6 The fitted models' performances for the third order on the Basis of AIC criteria. $AR_{(3)}: Y_{ti} = 0.5Y_{t-1} + 0.3Y_{t-2} + 0.1Y_{t-3} + e_t$, $ISTAR_{(3)}: Y_{ti} = \frac{0.5Y_{t-1} + 0.3Y_{t-2} + 0.1Y_{t-3} + e_t}{e^{0.5Y_{t-1} + 0.3Y_{t-2} + 0.1Y_{t-3} + e_t}}$, $ESTAR_{(3)}: Y_{ti} = e^{0.5Y_{t-1} + 0.3Y_{t-2} + 0.1Y_{t-3} + e_t}$, $TSTAR_{(3)}: Y_{ti} = \sin 0.5Y_{t-1} + \sin 0.3Y_{t-2} + \sin 0.1Y_{t-3} + e_t$.

AIC			
AR	ISTAR	ESTAR	TSTAR
10.0264	11.0934	12.118	10.0092
10.0237	11.0862	12.0973	10.0089
10.0011	10.9712	12.0074	10.0071
10.923	10.9556	11.9358	9.9069
10.9071	10.9495	11.9266	9.8965
10.8608	10.9354	11.9453	9.8040
10.8546	10.9251	11.8737	9.8229
10.8357	10.9018	11.8708	9.8050
10.8338	10.8732	11.8529	9.8025
10.8306	10.7934	11.8311	9.8018

5.0 Conclusion

In this study, we conducted a thorough comparison of the efficiency and performance of Autoregressive (AR) models with several modified versions of smooth transition autoregressive models, namely the Inverse Smooth Transition Autoregressive (ISTAR), Exponential Smooth Transition Autoregressive (ESTAR), and Trigonometric Smooth Transition Autoregressive (TSTAR) models. The primary aim was to investigate how these models behave under different conditions, particularly varying sample sizes and model orders, to determine their relative effectiveness.

An extensive simulation study was performed to evaluate each model across a broad range of sample sizes, from 20 to 250, and for models of first, second, and third orders. The evaluation process utilized standard statistical selection criteria, including the Bayesian Information Criterion (BIC), Hannan-Quinn Information

Criterion (HQC), and Akaike Information Criterion (AIC). These criteria were employed to identify the optimal model for each order, with the model exhibiting the lowest values for any of these criteria being considered the best-performing model for that particular order.

The results of the simulation, as summarized in Tables 1 to 6, reveal several important findings. For first-order models, AR models consistently provided the best fit, as indicated by the lowest BIC and HQC values. Similarly, for second-order models, AR models once again outperformed the smooth transition models, showing the lowest AIC and BIC values. However, for third-order models, the Trigonometric Smooth Transition Autoregressive (TSTAR) model emerged as the superior choice, outperforming other models based on HQC and AIC values.

These findings underscore the versatility and effectiveness of AR models, particularly in lower orders, while also highlighting the potential of

TSTAR models for more complex, higher-order situations. The results provide valuable insights for model selection in time series analysis, showing how different criteria can point to varying optimal models depending on the order and complexity of the data.

REFERENCES

Akaike, H. (1973). Maximum likelihood identification of Gaussian auto-regressive moving-average models." *Biometrika*, 60, 255–266

Alok, M. S. (2020). *Diagnostic Checking for Linearity in Time Series Models* (Master's thesis).

Amelot, L. M. M., SubadarAgathe, U., & Sunecher, Y. (2021). Time series modelling, NARX neural network and hybrid KPCA–SVR approach to forecast the foreign exchange market in Mauritius. *African Journal of Economic and Management Studies*, 12(1), 18-54.

Barnett, D.F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøstheim and A.H. Würtz (eds.), *Nonlinear Econometric Modeling*, Cambridge: Cambridge University Press.

Barrientos, A. F., & Canale, A. (2021). A Bayesian goodness-of-fit test for regression. *Computational Statistics & Data Analysis*, 155, 107104.

Boero, G. and Marrocu, E. (2018). The performance of SETAR models: a regime conditional evaluation of point, interval and density forecasts, Working Paper.

Boero, G. and Marrocu, E. (2022). The performance of non-linear exchange rate models: a forecasting comparison, *Journal of Forecasting*, Vol. 21(7), 513-542.

Box, G.E.P. and Pierce, D.A. (1970) Distribution of the residual autocorrelation in autoregressive integrated moving average time series models. *Journal of the American Statistical Association* 65, 1509-1526.

Chan, K. S., and Tong, H. (2015). On Estimating Thresholds in Autoregressive Models. *Journal of Time Series Analysis* 7: 178–90.

Chen, R., and R. Tsay (2023). Nonlinear additive ARX models, *Journal of the American Statistical Association*, 88(423), 955–967.

Chen, S. X., Li, J., & Zhong, P. S. (2019). Two-sample and ANOVA tests for high dimensional means. *The Annals of Statistics*, 47(3), 1443-1474.

Cheng, C., Guan, Y., Chen, G., & Gan, M. (2020). Parameters Estimation of RBF-AR Model Based on EM and Variable Projection Algorithm. In *2020 Chinese Control And Decision Conference (CCDC)* (pp. 1678-1683). IEEE.

Horváth, L., Miller, C., & Rice, G. (2020). A new class of change point test statistics of Rényi type. *Journal of Business & Economic Statistics*, 38(3), 570-579.

Mahdi, E. (2020). Mixed Portmanteau Tests for Simultaneous Linear and Nonlinear Dependency in Time Series. *arXiv preprint arXiv:2008.08176*.

Terasvirta, T., (2020). The Performance of RaHQCy test and white test of Terasvirta Test in detecting Nonlinearity.

Terasvirta, T., Tstheim, D., & Granger, C. W. (2022). *Modelling nonlinear economic time series*. University Press.

Zhang, J. T., Guo, J., & Zhou, B. (2022). Testing equality of several distributions in separable metric spaces: A maximum mean discrepancy-based approach. *Journal of Econometrics*.