African Multidisciplinary
Journal of Sciences and Artificial Intelligence

ISSN : 0000-0000
Index: Harvard, Boston, Sydney University, Dimensions, Lens, ResearchGet Scilit, Semantic, Google Scholar, Base etc

https://doi.org/10.58578/AMJSAI.v1i1.3755

Effect of Box-Cox Transformation on a k-th Weighted Moving Average Processes for Time Series

Emwinloghosa Kenneth Guobadia & Kenneth Kevin Uadiale
Federal Medical Centre, Asaba, Nigeria; Federal University of Wukari, Nigeria
kengroove28@gmail.com

Article Info:

Submitted:	Revised:	Accepted:	Published:
Jul 1, 2024	Jul 24, 2024	Jul 27, 2024	Jul 31, 2024

Abstract

In this paper, we examine, if the effect of transformation leads to improvement of model performance in time series modeling. The class of transformations that was considered is the Box-Cox family of transformation on the k-th weighted moving average (k-th WMA) model and autoregressive integrated moving average (ARIMA) model from a given nonstationary economic realization time series data. A real nonstationary economic time series data was used to demonstrate this procedure. The nonstationary time series data can be transformed to stationary data using the process of differencing alongside with Box-Cox transformation. The ARIMA model is fitted to the transformed data using the techniques of Box-Jenkins, where the best ARIMA is selected among the competing ARIMA models using Akaike information corrected criterian (AICc) while the best k-th WMA is selected among the competing models using some evaluation metrics such as root mean square error (RMSE) and mean absolute error (MAE). Finally, the optimal model is selected between ARIMA model and k-th WMA using the RMSE and MAE. Our findings are that the transformed k-th WMA models outperformed the classical ARIMA models for the set of Box-Cox transformation parameters considered for the data used.

Keywords: Box-Cox transformation, k-th WMA model, ARIMA model

INTRODUCTION

An essential aspect of time series modeling is the class of weighted methods (Simple Moving averages, weighted moving averages and exponential smoothing methods) are frequently used in forecasting. The main purpose of each of these methods is to smooth out the random continual change in the time series. These are useful when the time series does not show significant trend, cyclical or seasonal outcomes. That is, the time series is constant. Smoothing methods are generally good for short period forecasts.

These rich classes of model unlike the autoregressive integrated moving average (ARIMA) model do not generally need any form of statistical inference before one can apply them.

There are two different groups; namely averaging methods and exponential smoothing methods and they are extensively used in commerce and industry due to their capacity to remove trend and seasonal outcomes in time series data, Chatfield (2001). The exponential smoothing has been extremely useful for many years. It was first suggested by Holt (1957) and it was proposed to be used for non-seasonal time series exhibiting no trend. This was later extended by Holt (1958) where he offered a procedure that took care of trend. Winters (1965) generalized the method to handle seasonality. Box and Jenkins (1976) mentioned that nonstationary time series does not have natural mean and that the economic forecasting methods used for such time series are the class of exponentially weighted moving average. They however recommend differencing in order to transform nonstationary time series to become stationary. Another possibility is the use of a class of Box-Cox transformation, Box-Cox (1964). Unver et al. (2004) investigated the effects of Box-Cox transformation on estimations of the genetic parameters for egg production traits that do not hold to the assumptions of parametric statistical analysis

Shish and Tsokos (2008) and Tsokos (2010) introduced a class of weighted methods for forecasting nonstationary time series which are called k-th weighted moving average and k-th exponential weighted moving average process. Safi and Dawoud (2013) considered these class of weighted moving average. These weighted methods when applied to original nonstationary series can be transformed into a stationary series using a differencing filter which can also be modeled using ARIMA process due to Box and Jenkins (1976). Shish and Tsokos (2008) and Tsokos (2010), Safi and Dawoud (2013) didn't consider the application of Box-cox transformation using differencing filter.

MATERIALS AND METHODS

The data used in this study represent the daily value of Nigeria Stock Market from 13th March, 2017 to 11th April, 2022, sourced from Nigeria Stock Group.

The statistical procedures used for the analysis of data in this study are those proposed by Box Jenkins (1976) which is called the ARIMA model and k-th WMA model proposed by Shish and Tsoko (2008). The nonstationary data are first of all transformed to stationary data using the process of differencing alongside with Box-Cox transformation. This idea has also been considered by Vijay et al. (2019). As suggested by Franses and De Bruin (2002) a few of such Box-Cox parameters are important for forecasting purposes. The important ones are $\lambda = 0,\ \lambda = 0.5\ and\ \lambda = 1$. The ARIMA model and k-th WMA model are therefore fitted to the transformed data using the techniques of Box-Jenkins for the ARIMA model and techniques of Tsoko (2010) for the k-th WMA model. The best ARIMA model is selected among the competing models using AICc. Similarly, the best k-th WMA is selected among the competing models using RMSE and MAE. Finally, the optimal model between ARIMA model and k-th WMA model is selected using RMSE and MAE.

The Box-Cox transformation

We can transform the series to stationarity using the following Box-Cox transformation defined by:

$$z_t = \begin{cases} log(x_t) - log(x_{t-1}), & (\lambda = 0) \\ x_t^{\lambda} - x_{t-1}^{\lambda}, & (\lambda \neq 0) \end{cases} \quad (1)$$

Then in order to recover back our original series we get:

$$x_t = \begin{cases} x_{t-1} \exp(z_t) \\ \left(z_t + x_{t-1}^{\lambda}\right)^{\frac{1}{\lambda}} \end{cases} \quad (2)$$

Autoregressive Integrated Moving Average model

Given,

$$\emptyset(B)(1-B)^d y_t^\lambda = \theta(B)\varepsilon_t \quad (3)$$

Where d is the differencing parameter and λ is the Box-Cox transformation.

The Criteria for Selecting the forecast with the Best fit Procedure

Model Selection for ARIMA

Once appropriate models are obtained by either through the observation of the autocorrelation function (ACF) plot, where the ARIMA model to be fitted on the series should have the possible smallest parameters, i.e., p and q should be less than or equal to (p, $q \leq 3$) or by direct computation, then the next is to select among them the best model through some information criteria such as Akaike information criterion corrected (AICc) The decision is that ARIMA model with the least AICc is selected as the best model amongst other competing models. The AICc and AIC are obtained as

$$AICc = AIC + \frac{2k^2 + 2k}{n-k-1} \quad (4)$$

where *n* denotes the sample size and *k* denotes the number of parameters.

The k-th Weighted Moving Average Procedure.

For the k-th weighted moving average based on the transformed time series we have

$$w_t = \frac{2}{k(k+1)}\sum_{j=0}^{k-1}(j+1)\, z_{t-k+1-j} \quad (5)$$

Where $z_t = f\left(x_t^\lambda\right)$ defined in equation (1)

We illustrate the method of smoothing times series using the following example for k = 3

$$w_t = \frac{2}{3(3+1)} \sum_{j=0}^{3-1} (j+1)\, z_{t-3+1-j} \qquad (6)$$

$$w_t = \frac{1}{6} \sum_{j=0}^{2} (j+1)\, z_{t-3+1-j} \qquad (7)$$

$$w_t = \frac{1}{6} [z_{t-2} + 2z_{t-1} + 3z_t] \qquad (8)$$

In order to recover z_t, we make use of the following equation

$$z_t = \frac{1}{3} (6w_t - z_{t-2} - 2z_{t-1}) \qquad (9)$$

Evaluation Metrics for the k-th Weighted Moving Average

Once a time series model is developed, the next thing to do is to evaluate the performance of such model on the basis of how well it fits to historical (original) data. This study adopts two metrics which will help in discriminating between the competing models. These metrics are root mean square error (RMSE) and mean absolute error (MAE). They are defined respectively as follows;

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (x_t - \hat{x}_t)^2} \qquad (10)$$

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^{T} |x_t - \hat{x}_t| \qquad (11)$$

where T is the total time series observations, $\hat{x}_t$ is the predicted values, x_t are time series observations.

For the ARIMA (p, d, q), we select the model p, q and the number of differencing parameter d alongside with the parameter λ to achieve stationarity. If d = 1, we test for stationarity using the autocorrelation function (ACF). If stationarity is not achieved, we take d = 2 and check for stationarity again.

Stationarity Test

The stationarity of the time series can be obtained by visualizing the time plot or the autocorrelation function (ACF) plot, while a formal way of ascertaining if there is actually stationarity is by carrying out an Augmented Dickey-Fuller (ADF) test for unit root. The ADF test does not test stationarity explicitly, but test through the presence/absence of a unit root. Thus, if the series tested for stationarity is not stationary, then there is all evidence for transforming the series, and this can be done using differencing or power transformation.

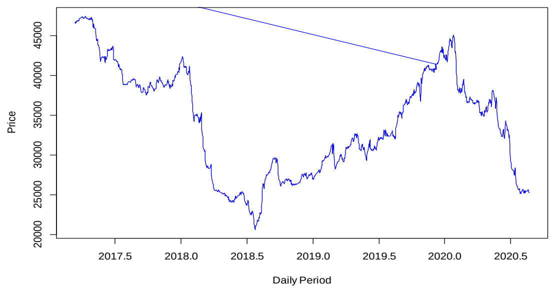

Figure 2: Time series plot for the actual data

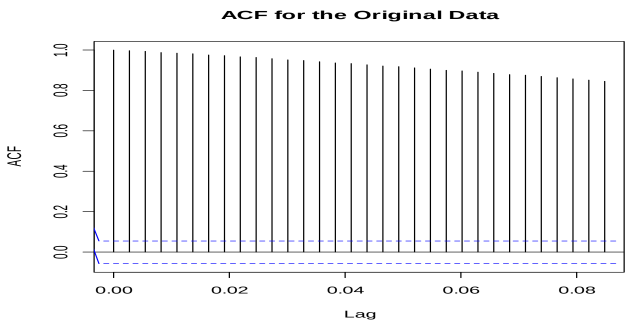

Figure 2: ACF plot for the original data

In Figure 1, a decreasing trend is observed from 13 March 2017 to July 2018, and from August 2018 to January 2020, the price increased intrinsically, and from February 2022, the price later decreases sharply. A sample test of the sample autocorrelation function (SACF) test is conducted for the time series data and the SACF is presented in figure 2. In Figure 2 it is clearly evident that the time series data is non stationary since the SACF refuse to die down quickly.

RESULTS AND DISCUSSION

Table 1 to 3 presents the summary statistics for selection of ARIMA model used in this study. When $\lambda = 0$, ARIMA (0,1,3) has the least Akaike Information Criterion Corrected (AICc = -12337.60) and it is therefore selected as the best model among the competing models for table 1, When $\lambda = 0.5$, ARIMA (0,1,3) has the least Akaike Information Criterion Corrected (AICc = 1082.44) and it is therefore selected as the best model among the competing models for table 2, When $\lambda = 1$, ARIMA (0,1,3)has the least Akaike Information Criterion Corrected (AICc = 15932.09) and it is therefore selected as the best model among the competing models for table 3.

Table 1: Arima Model Selection Table

Table 1: ARIMA Model selection when λ = 0		**Table 2: ARIMA Model selection when λ = 0.5**		**Table 3: ARIMA Model selection when λ = 1**	
ARIMA Model	**AICc**	**ARIMA Model**	**AICc**	**ARIMA Model**	**AICc**
ARIMA (0,1,1)	-12058.82	ARIMA (0,1,1)	1369.35	ARIMA (0,1,1)	16225.52
ARIMA (0,1,2)	-12090.34	ARIMA (0,1,2)	1327.94	ARIMA (0,1,2)	16175.42
ARIMA (0,1,3)	-12337.60	ARIMA (0,1,3)	1082.44	ARIMA (0,1,3)	15932.09
ARIMA (1,1,0)	-12042.12	ARIMA (1,1,0)	1385.98	ARIMA (1,1,0)	16241.48
ARIMA (2,1,0)	-12125.05	ARIMA (2,1,0)	1296.59	ARIMA (2,1,0)	16147.86
ARIMA (3,1,0)	-12233.73	ARIMA (3,1,0)	1191.40	ARIMA (3,1,0)	16044.97
ARIMA (1,1,1)	-12059.47	ARIMA (1,1,1)	1367.86	ARIMA (1,1,1)	16223.28
ARIMA (1,1,2)	-12292.79	ARIMA (1,1,2)	1129.64	ARIMA (1,1,2)	15981.35
ARIMA (2,1,1)	-12316.49	ARIMA (2,1,1)	1109.13	ARIMA (2,1,1)	15963.66

Figure 1 to 3: Presents the pictorial display of the time plot for first differenced data, when $\lambda = 0$, 0.5 and 1, ACF for first differenced data, when $\lambda = 0$, 0.5 and 1, PACF for first differenced data, when $\lambda = 0$, 0.5 and 1 and Time plot for the Box-Cox transformation of 3-WMA , when $\lambda = 0$, 0.5 and 1 the data are all stationary.

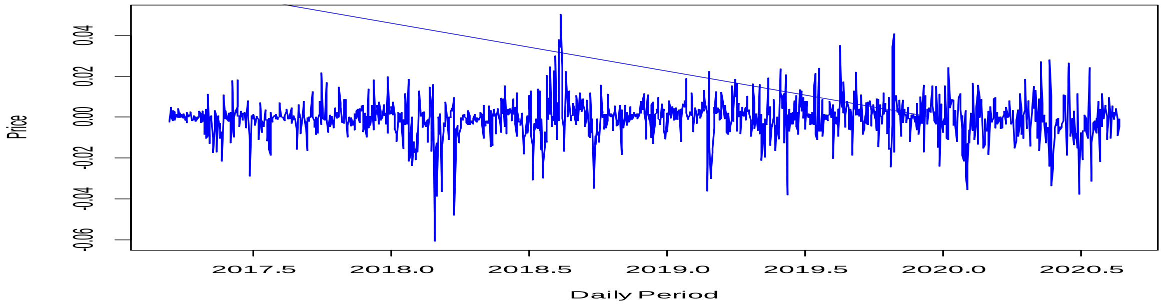

Figure 1 (a): Time series plot for first differenced data, when $\lambda = 0$

Figure 1(a) Time series plot for first differenced data, when $\lambda = 0$, the data are stationary after first differencing.

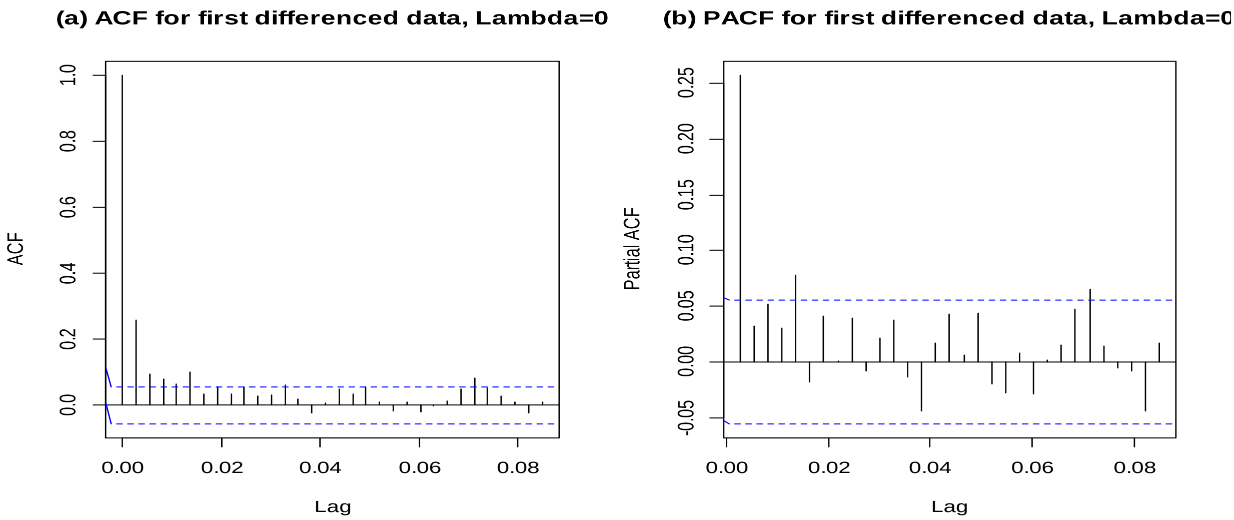

Figure 1(b): (a) ACF plot for first differenced data, $\lambda = 0$; (b) PACF plot for first differenced data, $\lambda = 0$

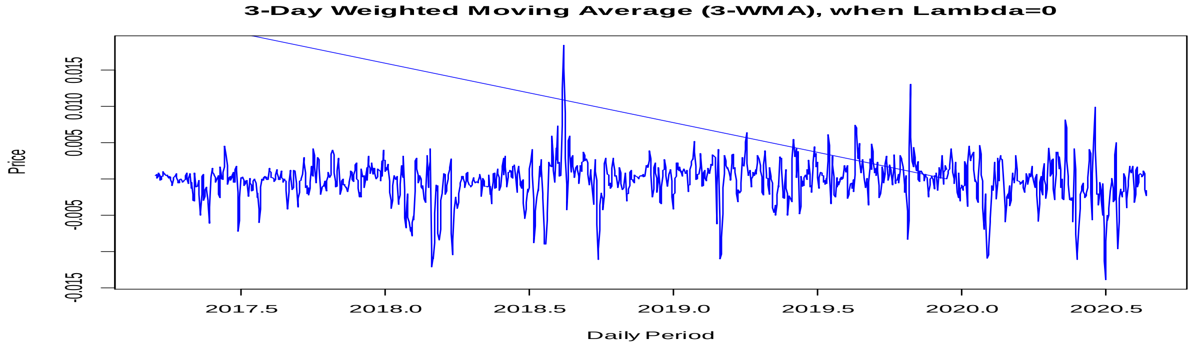

Figure 1(c): Time plot for the Box-Cox transformation of 3-WMA , when $\lambda = 0$

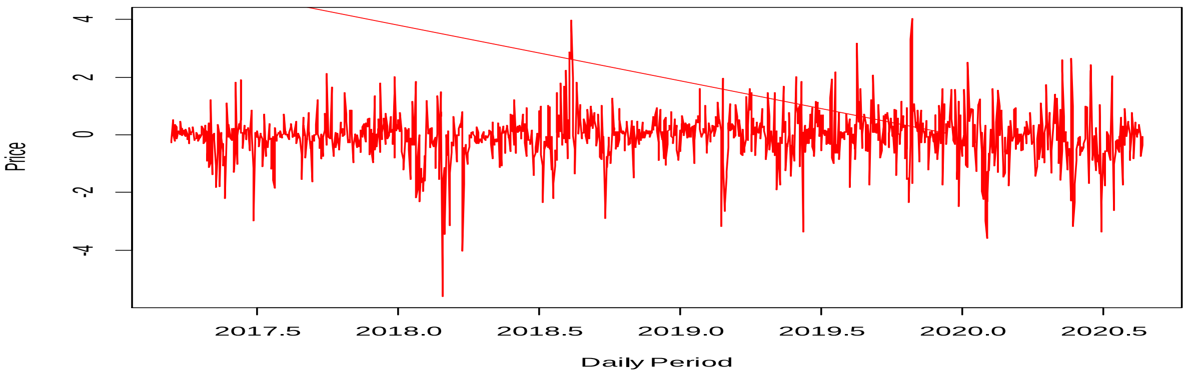

Figure 2(a): Time series plot for first differenced data, when $\lambda = 0.5$

Figure 2(a) Time series plot for first differenced data, when $\lambda = 0.5$, the data are stationary after first differencing

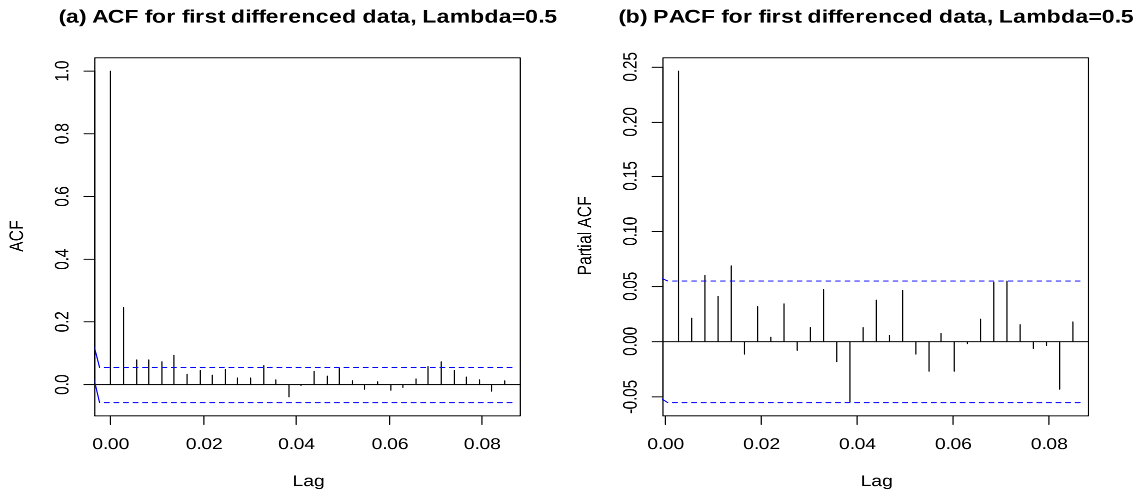

Figure 2(b): (a) ACF plot for first differenced data, $\lambda = 0.5$; (b) PACF plot for first differenced data, $\lambda = 0.5$

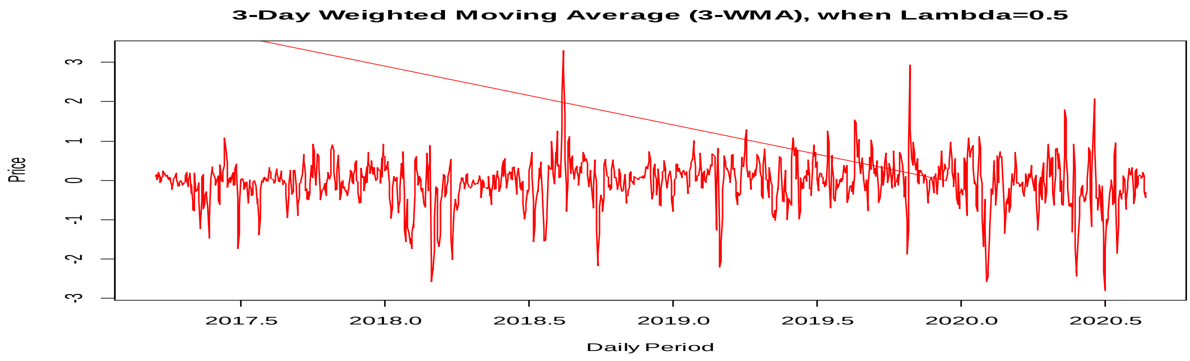

Figure 2(c): Time plot for the Box-Cox transformation of 3-WMA , when $\lambda = 0.5$

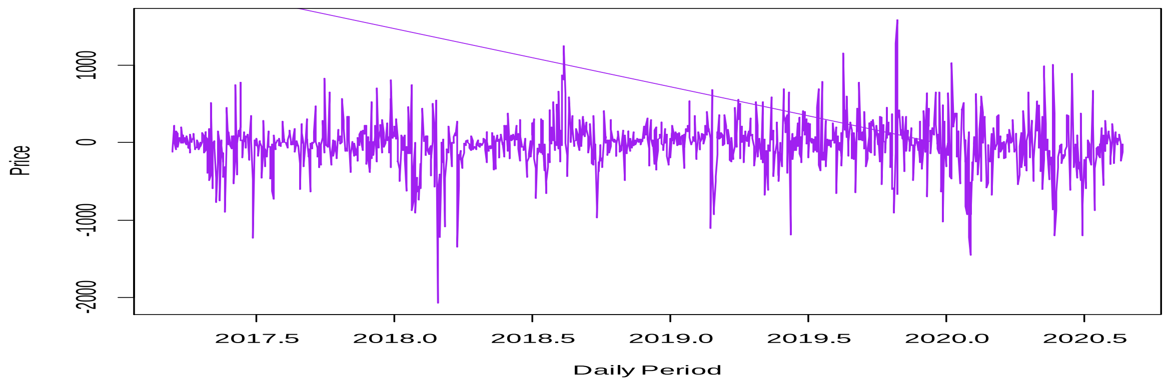

Figure 3(a): Time series plot for first differenced data, when $\lambda = 1$

Figure 3(a) Time series plot for first differenced data, when $\lambda = 1$, the data are stationary after first differencing

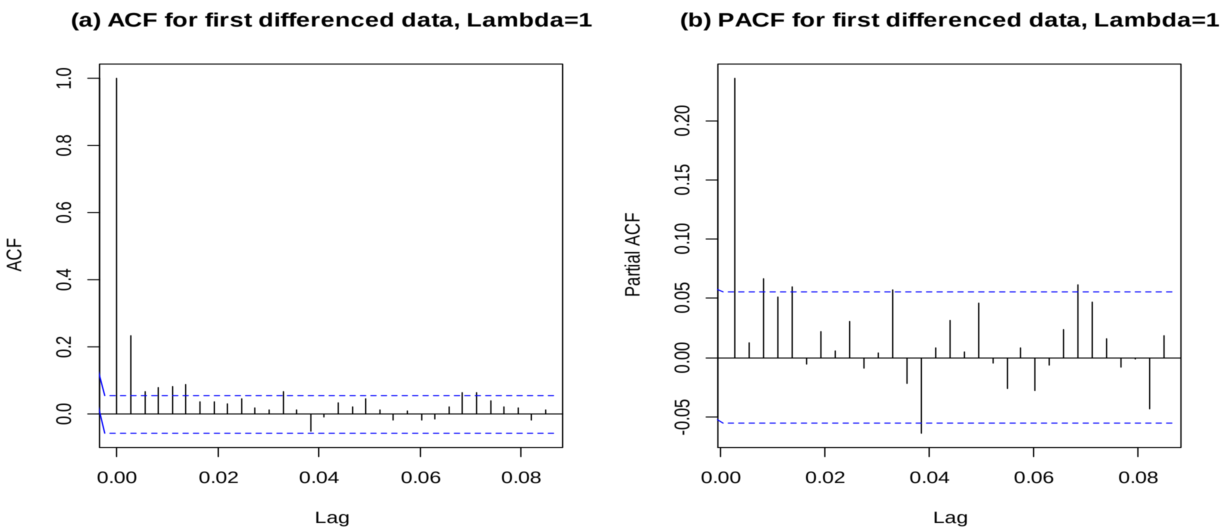

Figure 3(b): (a) ACF plot for first differenced data, $\lambda = 1$; (b) PACF plot for first differenced data, $\lambda = 1$

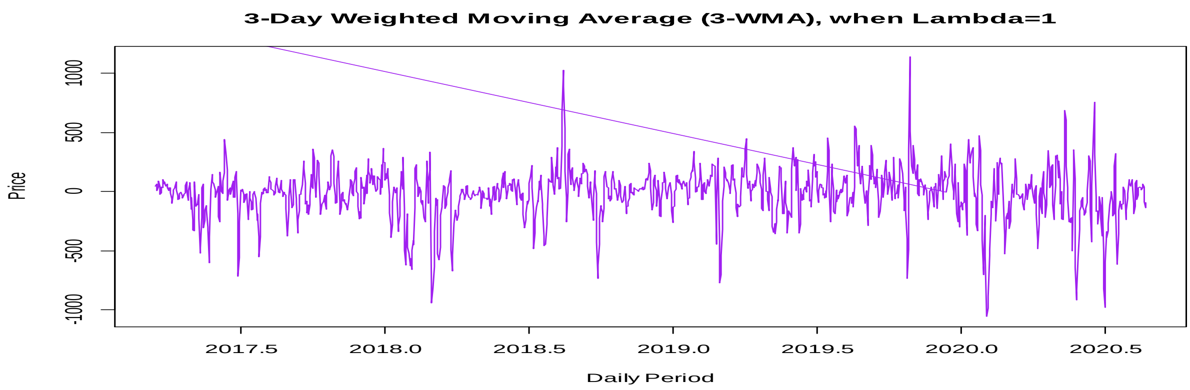

Figure 3(c): Time plot for the Box-Cox transformation of 3-WMA , when $\lambda = 1$

Table 4 to 7 presents the summary statistics for selection of the best model between ARIMA model and K-th WMA model used in this study.

In Table 4, the 3-Day weighted moving average model using Box-Cox transformed data with Box-Cox transformation parameter, $\lambda = 0$, has the least (RMSE and MAE) among the competing models. The 3-WMA model is therefore selected as the best model for $\lambda = 0$, since it has the least RMSE and MAE for the data used.

In Table 5, the 3-Day weighted moving average model using Box-Cox transformed data with Box-Cox transformation parameter, $\lambda = 0.5$, has the least (RMSE and MAE) among the competing models. The 3-WMA model is therefore selected as the best model $\lambda = 0.5$, since it has the least RMSE and MAE for the data used.

In Table 6, the 3-Day weighted moving average model using Box-Cox transformed data with Box-Cox transformation parameter, $\lambda = 1$, has the least (RMSE and MAE) among the competing models. The 3-WMA model is therefore selected as the best model $\lambda = 1$, since it has the least RMSE and MAE for the data used.

In Table 7, the 3-WMA when $\lambda = 0$ has the least (RMSE and MAE). However, it is considered the best model for the daily Nigeria stock price data used.

Table 2: The Comparison Of The Classical ARIMA And 3-Day Weighted Moving Average When $\lambda = 0$

Model	RMSE	MAE
Classical ARIMA	0.00896	0.00588
3-WMA	0.00346	0.00240

Table 3: The Comparison Of The Classical ARIMA And 3-Day Weighted Moving Average When $\lambda = 0.5$

Model	RMSE	MAE
Classical ARIMA	0.81411	0.53578
3-WMA	0.31321	0.21863

Table 4: The Comparison Of The Classical ARIMA And 3-Day Weighted Moving Average When $\lambda = 1$

Model	RMSE	MAE
Classical ARIMA	302.2399	198.3729
3-WMA	115.3396	80.3592

Table 5: Comparing 3-WMA with $\lambda = 0$, $\lambda = 0.5$, and $\lambda = 1$

Model	RMSE	MAE
3-WMA when $\lambda = 0$	0.00346	0.00240
3-WMA when $\lambda = 0.5$	0.31321	0.21863
3-WMA when $\lambda = 1$	115.3396	80.3592

CONCLUSION

The aim of this study is to compare a particular k-th WMA with the classical ARIMA processes for modelling non stationary time series data using a class of Box-Cox Transformation

REFERENCES

Ammeh L.B., Guobadia E.K., Ugoh C.I. (2023), Exponential Smoothing State Space Innovation

Model for Forecasting Export Commodity Price Index in Nigeria. African Journal of Economics and Sustainable Development 6(4), 74-84.DOI: 10.52589/AJESD-XQFAMZMYs

Box G. E. P and Jenkins, G.M., (1976). "Time series analysis: „Forecasting and control,"

Holden-Day, San Francisco

Chatfield, C., Koehler, A.B., Keith Ord., J and Snyder,R.D (2001). A New Look at Models for Exponential Smoothing. Journal of the Royal Statistical Society. Series D, Vol.50, No.2, pp.147-159

Christogonus Ifeanyichukwu Ugoh, Udochukwu Victor Echebiri, Gabriel Olawale Temisan, John Paul, Kenechukwu Iwuchukwu, Emwinloghosa Kenneth Guobadia (2022) On Forecasting Nigeria's GDP: A Comparative Performance of Regression with ARIMA Errors and ARIMA Method, International Journal of Mathematics and Statistics Studies, Vol.10, No.4, pp.48-64

Ekhosuehi, N. (2013). Effect of Box-Cox Transformation on K-th Moving Average processes for, Time series Forecasting Models. Journal of the Nigerian Statistical Association, Vol. 25 pp. 1-11.

Ekhosuehi N, Kenneth GE, Kevin UK. TheWeibull Length Biased Exponential Distribution: Statistical Properties and Applications. Journal of Statistical and Econometric Methods.2020;9(4):15-30.

Emwinloghosa K.G., Pamela O.O., Paschal N.I., Eloho S.O., Agu C. (2023), Modeling and Forecasting Inflation in Nigeria: A Time Series Regression with ARIMA Method. African Journal of Economics and Sustainable Development 6(3), 42-53. DOI:10.52589/AJESD-HFYC2BNW

Emwinloghosa K.G., Pamela O.O., Christogonus I.U. (2023), Impact of Fiscal Policy on Inflation in Nigerian Economy. African Journal of Economics and Sustainable Development 6(4), 37-8.DOI: 10.52589/AJESD-E1RMOYKH

Guobadia Emwinloghosa Kenneth. Statistical Application of Regression techniques in Modeling Road Accidents in Edo State, Nigeria. Sch J Phys Math Stat, 2021 Jan 8(1): 14-18

Guobadia Emwinloghosa Kenneth. Justification of the Nature of Fluctuations in Nigerian Bank Returns: an Empirical Analysis. Sch J Econ Bus Manag, 2021 Jan 8(1): 10-13.

Guobadia Emwinloghosa Kenneth & Ibeakuzie Precious Onyedikachi (2021). Selected Economic Sector Contributionto Nigeria's Gross Domestic Product. Sch Bull, 7(3): 49-59.

Guobadia Emwinloghosa Kenneth & Ibeakuzi Precious Onyedikachi, Short Term Modeling of the NigerianNaira/United States Dollar Exchange Rate Using ARIMA Model. Sch J Phys Math Stat, 2021 Jan 8(1): 8-13.

Guobadia Emwinloghosa Kenneth. A Time Domain Approach to Modeling Nigeria's Gross Domestic Product. SchJ Phys Math Stat, 2021 Jan 8(1): 19-28.

Holt, C.C. (2004) Forecasting Seasonal and Trends by Exponentially Weighted Moving Averages. Int. J. Forecast. 2004, 20, 5–10.

Holt, C.C. (1957) Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages. ONR Memorandum, Vol. 52, Carnegie Institute of Technology, Pittsburgh. Available from the Engineering Library, University of Texas, Austin.

Ibeakuzie Precious Onyedikachi & Guobadia Emwinloghosa Kenneth (2021). Research Ethics Grasp and Enactment; A Casewith University of Benin. Sch Bull, 7(3): 60-71

Shih, S. and Tsokos, C.P. (2008). A weighted Moving Average process for Forecasting. Journal of Modern Applied Statistical Methods.

Tsokos, C.P. (2010), K-th Moving, Weighted and Exponential Moving Average for Time Series Forecasting Models, European Journal of pure and applied Mathematics. Vol.3, No.3 406-416

Winters, P.R. (1960), Forecasting Sales by Exponentially Weighted Moving Averages. Manag. Sci.,6, 324–342.

Ugoh C.I., Abode J.O., OnyiaC.T., Omoruyi P.O., Guobadia E.K. (2023), Evaluating the Determinants of Exchange Rates in Emerging Markets: Evidence from Nigeria and South Africa. African Journal of Economics and Sustainable Development 6(2), 49-63.DOI: 10.52589/AJESD-VB4NTHBE

Ugoh C.I., Igbinosa E.S., Akanno F.C., Omoruyi P.O., Guobadia E.K. (2023),Investigating the Determinants of Inflation in Leading Economies in Africa: A Panel Data Analysis. African Journal of Economics and Sustainable Development6(2), 30-48. DOI:10.52589/AJESD-VGH59XLX

Unver, Y. Akbas, Y. and Oguz, I. 2004. Effect of Box-Cox transformation on genetic parameter estimation in layers. Turk Journal of Veterinary Science, 28, 249-255

Safi, S.K and Dawoud, I.A. (2013). Comparative study on forecasting accuracy among moving average models with simulation and PALTEL stock market data in Palestine. International