International Journal of Mathematics and Statistics Studies
Vol.9, No.1, pp.16-28, 2021
Published by ECRTD-UK
Print ISSN: 2053-2229 (Print), Online ISSN: 2053-2210 (Online)

COMPARATIVE PERFORMANCE BETWEEN THE BOX - JENKINS AND TIME SERIES REGRESSION MODELS

Iwok, Iberedem A.[1] **and Udoh, Grace M.**[2]

[1]Department of Mathematics/Statistics, University of Port Harcourt, Nigeria
[2]Department of Statistics, Akwa Ibom State Polytechnic, Ikot Osurua, Nigeria
Corresponding e-mail: ibywok@gmail.com

ABSTRACT: *This work compared the performance of the Box-Jenkins and time series regression models. The two methods were theoretically presented. Data was also collected for fitting the models. Three test measures were used for the comparative analysis. The results showed that the time series regression model performs better than the Box-Jenkins model.*

KEYWORDS: autocorrelation function, partial autocorrelation function, white noise process, time series regression, Arima models.

INTRODUCTION

In recent times, time series modelling and forecasting have proven to have fundamental importance in practical domains. Due to this, researches are intensified and geared towards developing and improving methodologies that can give accuracy and efficiency of the modelling process. Two of such methodologies are the Box-Jenkins and the Time Series Regression Methods. The aim of time series analysis is to use the past observations to develop and appropriate model that describes the underlying structure of the series for the prediction of future. This can be achieved by using any of the aforementioned methods.

Hejase and Assi (2012) established a weather model for the United Arab Emirate (UAE) using 10 years weather data. The research employed various models like classical empirical models, artificial neural network models, and time series regression models with autoregressive integrated moving average (ARIMA) models. The work used time series regression to model the mean daily and monthly global solar radiation for the city of Al-Ain. The Analysis was shown to yield accurate average long-term prediction performance of solar radiation in Al-Ain. The low corresponding values of mean bias error (MBE), mean absolute bias error (MABE), mean absolute percentage error (MAPE), and root-mean-square error (RMSE) confirmed the adequacy of the obtained model for long-term prediction of GSR data in Al-Ain, UAE.

Bhaskaram *et al* (2013) investigated the short term associations between exposures such as air pollution, weather variables or pollen, and health outcomes such as mortality, myocardial infection or disease-specific hospital admissions. Typically, for both exposure and outcome, data were available at regular time intervals and the aim was to explore the short-term associations between them. General features of time series data were outlined and real focus was made on time series regression that differ from other regression methods. Short term fluctuations were modelled in the presence of seasonal and long-term patterns, dealing with time varying confounding factors and lagged associations between exposure and outcome. The result showed that time series regression is no different from regression techniques used in other areas.

Imai *et al* (2015) presented a time series regression approach to obtain potential solutions for five issues arising in changes in immune population, strong autocorrelations, a wide range of plausible lag structures and association patterns, seasonality adjustments and large over dispersion. The approach was illustrated with cholera, rainfall, influenza and temperature data sets. Modifications were made to standard time series regression practice using sums of past cases as proxies for the immune population and logarithm of lagged disease counts to control autocorrelation. The results showed that time series regression may be used to investigate the dependence of infectious disease on weather but are likely to require modifying to allow for certain features.

Tebb *et al* (2015) provided an introduction to time series methodologies that is oriented toward issues within psychological research. This was accomplished by first introducing the basic characteristics of time series data. Various time series regression models were explicated to achieve a wide range of goals. The paper described how regressive techniques and autoregressive integrated moving average models can be combined in a dynamic regression model that can simultaneously explain and forecast a time series variable. Thus, the paper seeks to provide an integrated resource for psychological researchers interested in analysing time series data.

In this work, however, we shall compare the performance of the time series regression model with the more popular Box and Jenkins model.

METHODOLOGY

Time Series Regression

The time series regression technique uses the general pth-order polynomial trend model of the form:

$$X_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_p t^p + \varepsilon_t \qquad (1)$$

The above expression (1) is a function of time ; where $t \in T$, $t = \pm 1, \pm 2, \pm 3, \ldots$ and ε_t is a white noise process.

If there is no trend in the time series; then there is no long run growth or decline in the time series over time; hence

$$X_t = \beta_0$$

Equation (1) can be estimated by

$$\widehat{X_t} = \widehat{\beta_0} + \widehat{\beta_1} t + \widehat{\beta_2} t^2 + \cdots + \widehat{\beta_p} t^p$$

and

$$\widehat{\varepsilon_t} = X_t - \widehat{X_t}$$

Linear Trend

This indicates a straight line long run growth or decline over time and is represented by:

$$X_t = \beta_0 + \beta_1 t + +\varepsilon_t \qquad (2)$$

The increasing or decreasing nature of the time varying function is indicated by the sign of β_1 which is either greater than or less than 0.

Other Trends

If $p = 2$, we have the quadratic trend given by

$$X_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t \qquad (3)$$

This gives a quadratic or curve linear long run change over time.

Other trends can also be observed in time series for $p \geq 3$ with one or more reversals in curvature.

Any time series X_t showing any of these trends can be modelled using the appropriate polynomial function of time t.

Box and Jenkins Method

The Box and Jenkins method use models such as Autoregressive (AR), Moving Average (MA), Mixed Autoregressive-Moving Average (ARMA) models. For a non-stationary series, the time series has to be made stationary by some transformation methods and the model is identified by examining the behaviour of the Autocorrelation function (ACF) and Partial Autocorrelation function (PACF).

Autoregressive (AR) Model

An autoregressive process model of order p denoted by $AR(p)$ is given as:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t \qquad (4)$$

where,

$$x_t = X_t - \mu$$

and $\phi_1, \phi_2, \dots, \phi_p$ are the unknown parameters to be estimated in the model.
Expression (4) can be represented as:

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right) x_t = \varepsilon_t$$

$$\Longrightarrow \phi_p(B) = \varepsilon_t$$

where,

$$\phi_p(B) = \left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)$$

Moving Average (MA) Models

A moving average model of order q denoted $MA(q)$ is expresed as:

$$x_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \qquad (5)$$

where,

$$x_t = X_t - \mu$$

and $\{\varepsilon_t\}$ is a zero mean white noise process with constant variance.
Expression (5) can be written:

$$x_t = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right) \varepsilon_t$$

$$\Longrightarrow x_t = \theta_q(B) \varepsilon_t$$

where,

$$\phi_q(B) = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right)$$

Mixed Autoregressive – Moving Average Process (Model)

A combination of $AR(p)$ and $MA(q)$ results in the *mixed autoregressive – moving average time series model* of order (p, q) denoted as $ARMA(p, q)$.

This is represented as:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (6)$$

$$\Rightarrow \left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right) x_t = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right) \varepsilon_t$$

$$\Rightarrow \phi_p(B) x_t = \theta_q(B) \varepsilon_t$$

Autoregressive Integrated Moving Average (ARIMA) Models

For a non-stationary time series to be stationary, it has to be differenced d times. Thus, given a non-stationary series X_t, we can fit the dth difference:

$$x_t = \nabla^d X_t$$

to a stationary ARMA model; where $\nabla = 1 - B$ and $B^m X_t = X_{t-m}$. The result gives rise to an ARIMA model denoted as $ARIMA(p, d, q)$; where p is the order of autoregression (AR), d is the degree of differencing and q is the order of moving average.

In general, the $ARIMA(p, d, q)$ model is expressed as:

$$\phi(B) \nabla^d X_t = \phi(B)(1 - B)^d X_t = \varphi(B) X_t = \theta(B) \varepsilon_t \qquad (7)$$

where,

$$\varphi(B) = \phi(B) \nabla^d = \phi(B)(1 - B)^d = \left(1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_{p+d} B^{p+d}\right)$$

$$\Rightarrow \left(1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_{p+d} B^{p+d}\right) X_t = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right) \varepsilon_t$$

$$\Rightarrow X_t = \sum_{j=1}^{p+d} \varphi_j X_{t-j} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \qquad (8)$$

White Noise Process

A process $\{\varepsilon_t\}$ is said to be a white noise process with mean 0 and variance σ_ε^2 written $\{\varepsilon_t\} \sim WN(0, \sigma_\varepsilon^2)$, if it is a sequence of uncorrelated random variables from a fixed distribution. A well fitted model is expected to follow a white noise process.

Autocovarance

The autocovariance at lag k denoted γ_k is defined by:

$$\gamma_k = Cov(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)]$$

Autocorrelation

The autocorrelation at lag k denoted by ρ_k is defined as:

$$\rho_k = \frac{Cov(X_t, X_{t+k})}{\sqrt{[Var(X_t)][Var(X_{t+k})]}} = \frac{E[(X_t - \mu)(X_{t+k} - \mu)]}{\sqrt{\{E[(X_t - \mu)^2]\}\{E[(X_{t+k} - \mu)^2]\}}}$$

Autocorrelation Function(ACF)

The autocorrelation function $\{\rho_k\}$ of $\{X_t\}$ is the plot of ρ_k against the lag k. For a white noise process, the autocorrelation function of the residual is expected to lie within $\pm 2/\sqrt{N}$.

Partial Autocorrelation

Is the conditional correlation between X_t and X_{t+k} after their mutual linear dependency on the intervening variables $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$ has been removed.

Partial Autocorrelation Function(PACF)

The partial autocorrelation function $\{\phi_{kk}\}$ of $\{X_t\}$ is a plot of the partial autocorrelations against the lag k.

Model Evaluation

After fitting the Time Series Regression and the Box-Jenkins models and confirming the adequacy of the models; a comparative study between the two approaches shall be based on the following statistics:

(i) The Mean Square Error (MSE)

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(\mathrm{X}_i - \hat{X}_i\right)^2$$

(ii) The mean absolute error (MAE)

$$MAE = \frac{1}{N}\sum_{i=1}^{N} \left| \mathrm{X}_i - \hat{X}_i \right|$$

(iii) The mean absolute percentage error (MAPE)

$$MAPE = \left[\frac{1}{N}\sum_{i-1}^{N}\left|\frac{\mathrm{X}_i - \hat{X}_i}{\mathrm{X}_i}\right|\right] \times 100$$

Data Analysis

The data used in this work is the average daily sales per month in Dollars from Wimpey Supermarket in Port Harcourt, Nigeria. The data is displayed in the Appendix of this work.

Regression Method (RM)

```
The regression equation is
Xt = 2206 + 7.30 t


Predictor      Coef    SE Coef        T       P
Constant    2206.20       7.86   280.69   0.000
t            7.2983     0.3705    19.70   0.000


S = 23.0903   R-Sq = 91.9%   R-Sq(adj) = 91.7%
```

Time Series Plot of Residual and Its Autocorrelation Function of the RM

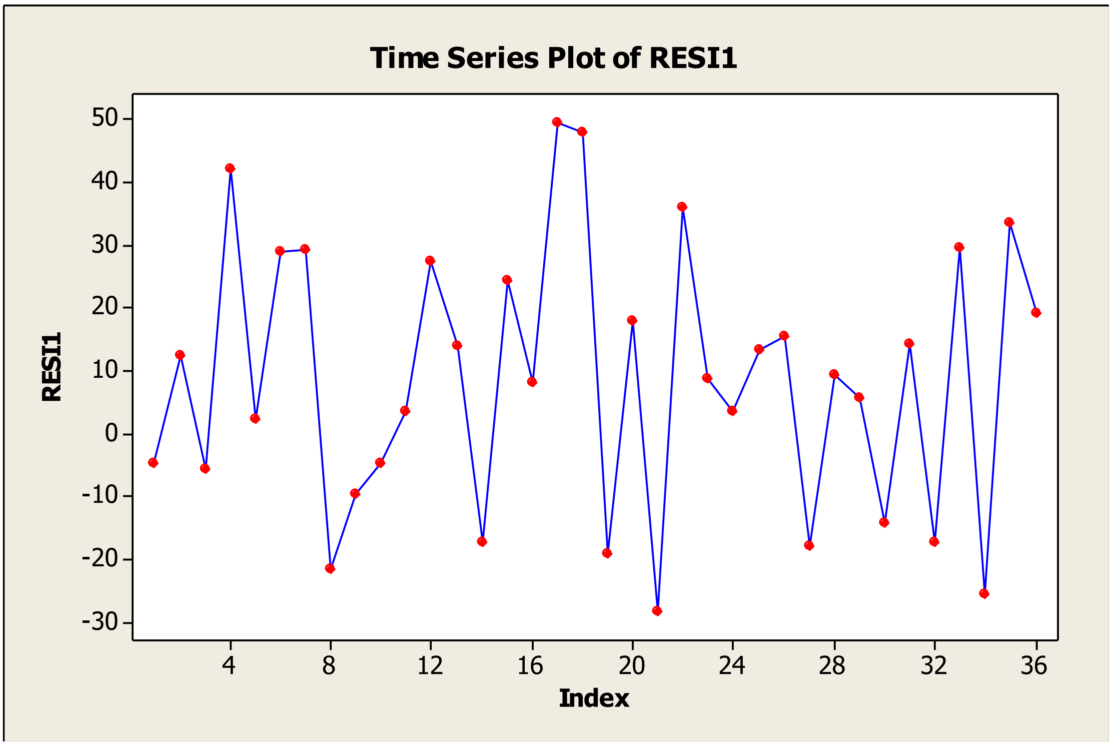

Figure 1: Time Series Plot of the Residual from the RM

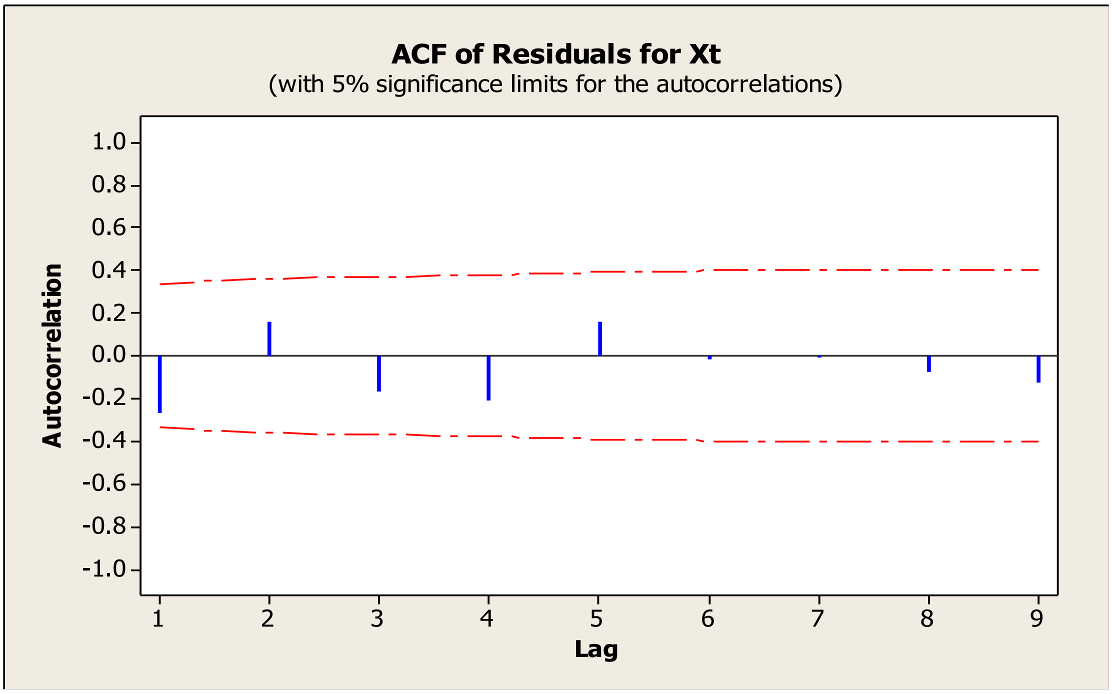

Figure 2: Autocorrelation function of the Residual from RM

The plot of the residual shows that the residual follows a white noise process and the residual autocorrelation function contains no spike. Hence, the fitted model is adequate.

Box and Jenkins Method (BJM)

Raw Data Plot

The raw data plot of the sales is shown in figure 3 below:

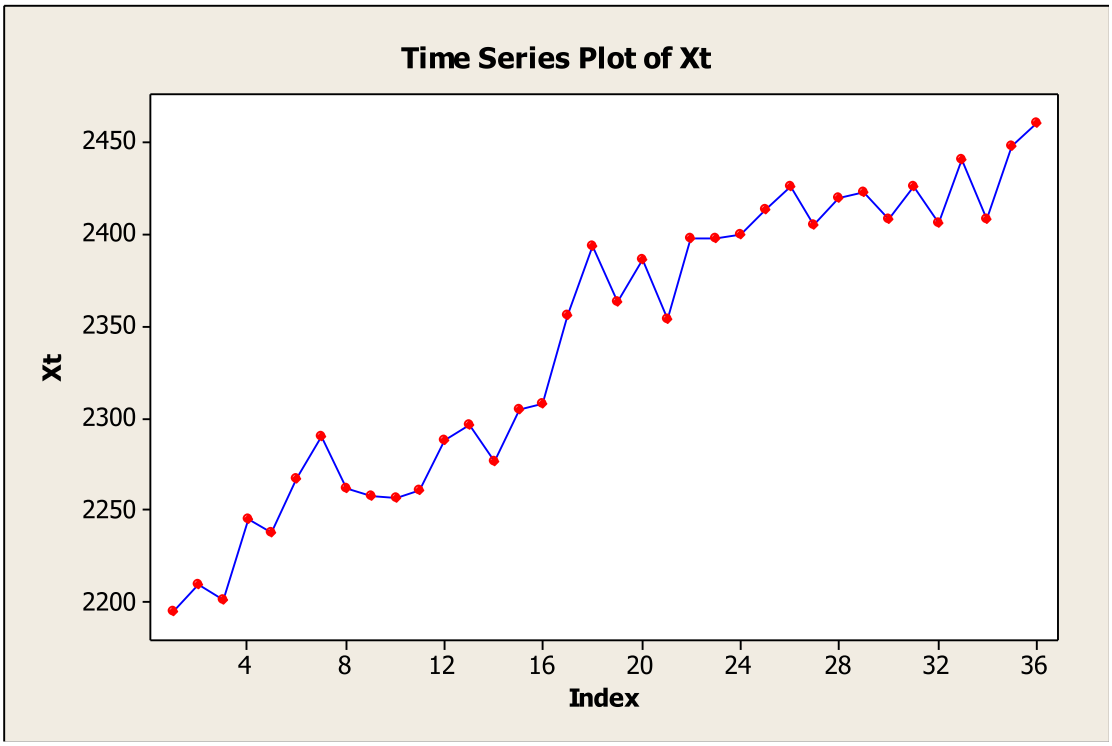

Figure 3: Time series plot of the average daily sales per month in Dollars

Stationary plot of the Data

Figure 3 clearly indicates that the series is non-stationary and requires differencing of the form: $DX_t = x_t = X_t - X_{t-1}$ to obtain stationarity. The result of the stationary series is as shown in the following plot.

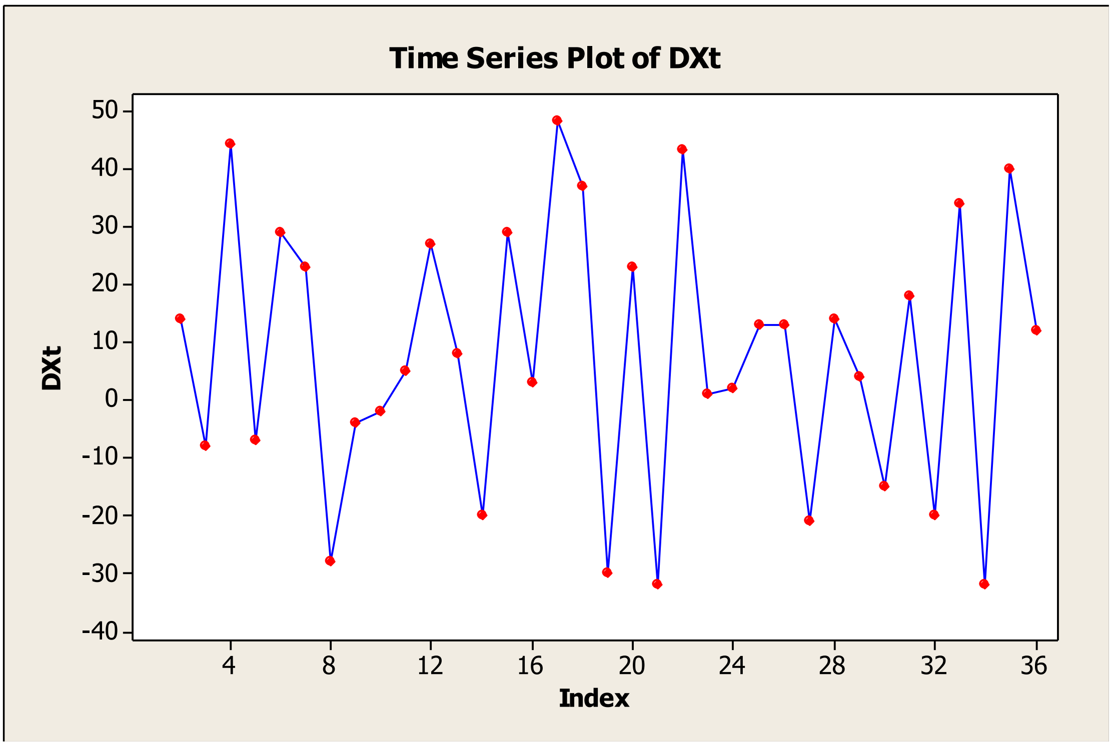

Figure 4: Differenced Series Plot of X_t

By mere inspection, the differenced series is stationary and the Box-Jenkins method can now be applied.

Model Identification

Autocorrelation and Partial Autocorrelation

As shown in figures 5 and 6 below, the autocorrelation function of DX_t declines exponentially to 0 while the partial autocorrelation function cut off after lag 1. This suggest a tentative model of ARIMA(1, 1, 0) which can be expressed as:

$$x_t = \phi_1 x_{t-1} + \varepsilon_t \qquad (9)$$

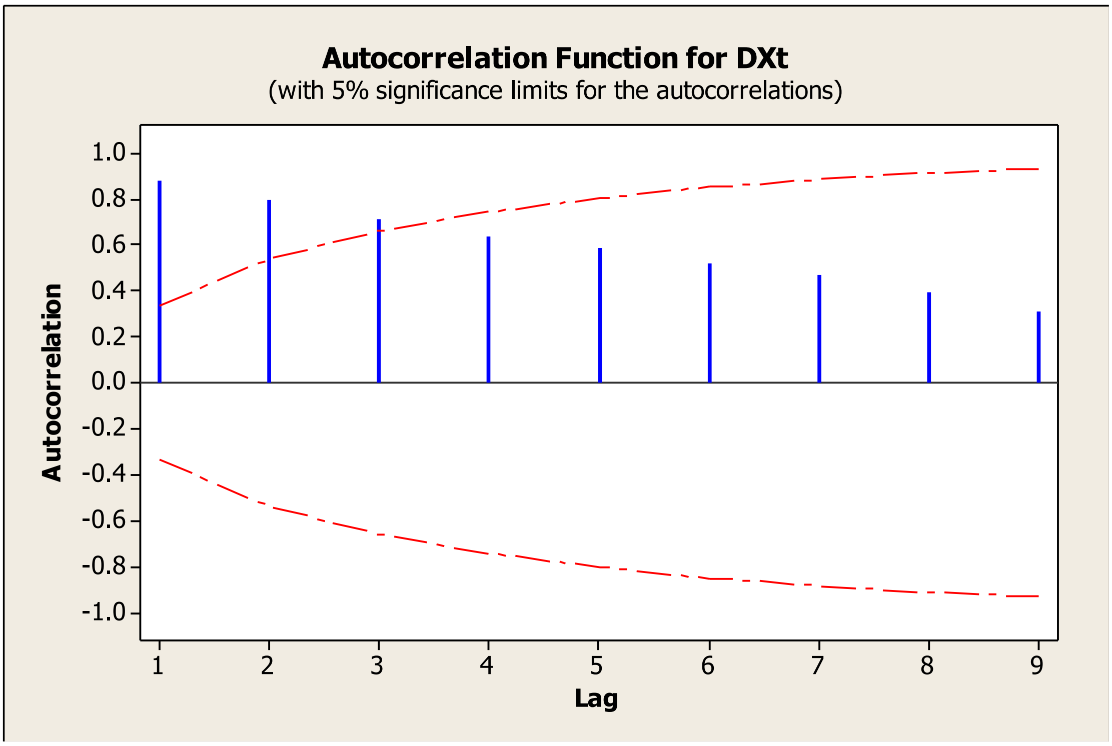

Figure 5: Autocorrelation Function Plot for DX_t

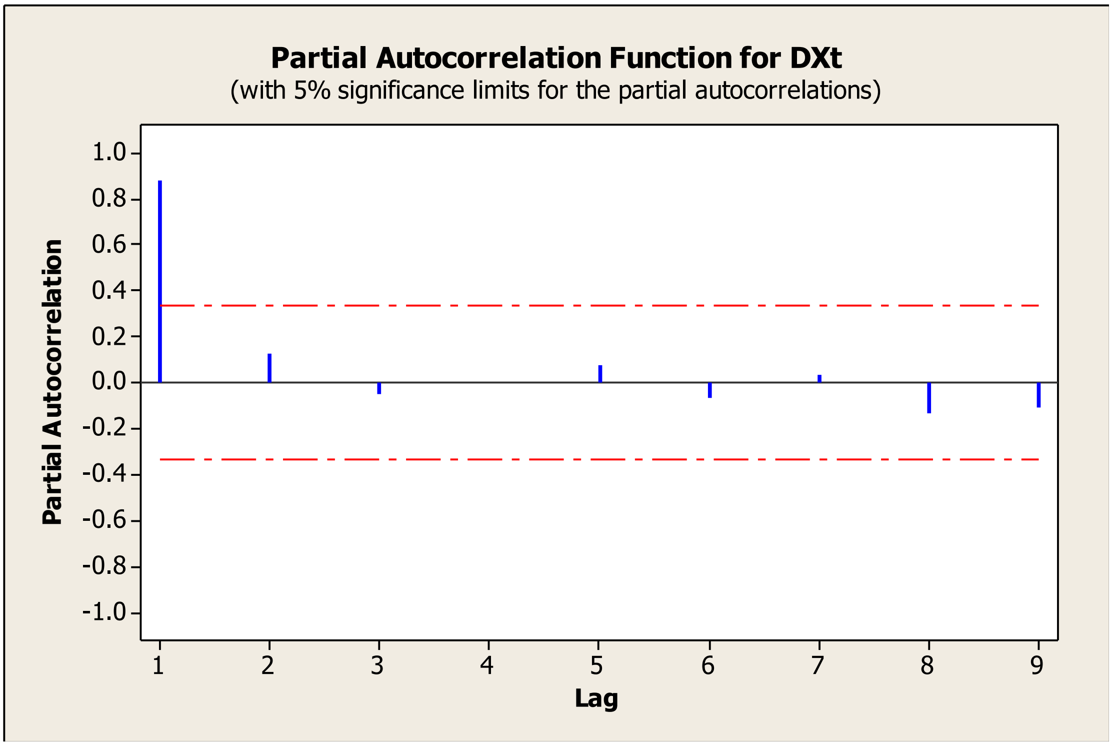

Figure 6: Partial Autocorrelation Function Plot for DX_t

Residual Analysis

Time Series Plot of the Residual

The residual obtained from the fitted [ARIMA(1,1,0)] model is shown below (figure 7). The plot shows that the residual follows a white noise process, indicating a good fit.

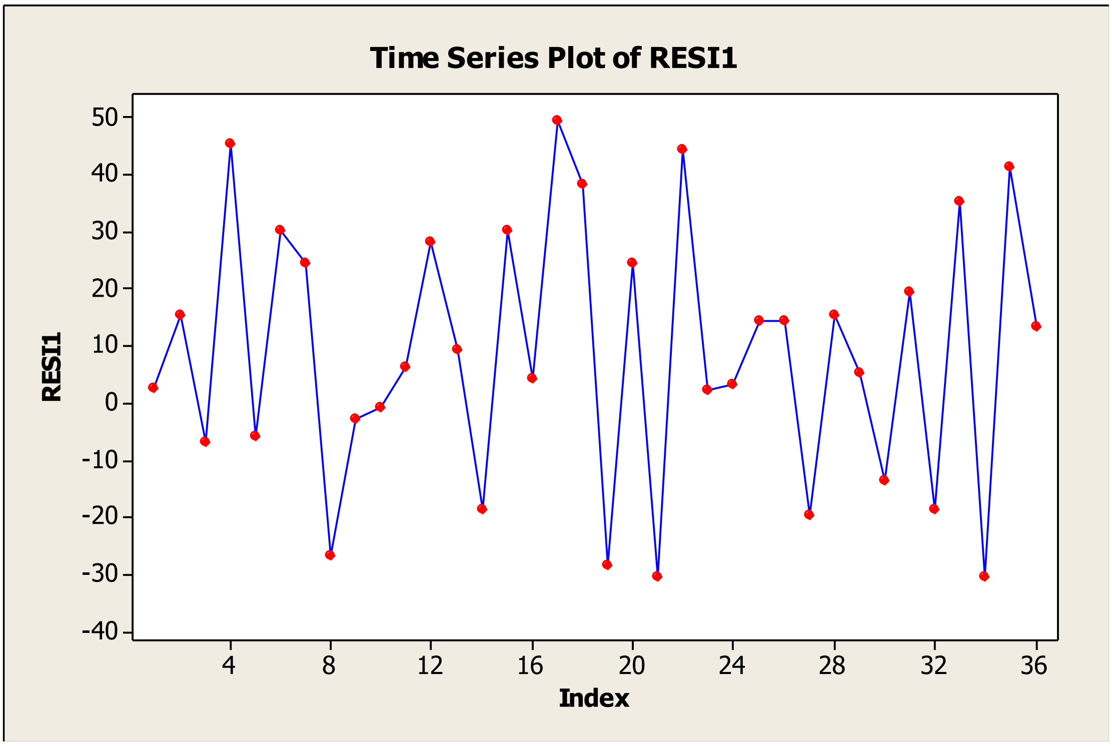

Figure 7: Time Series Plot of the Residual from BJM

Autocorrelation Function of the Residual

Figure 8 shows no spike in the autocorrelation function plot. This indicates that the fitted model is adequate.

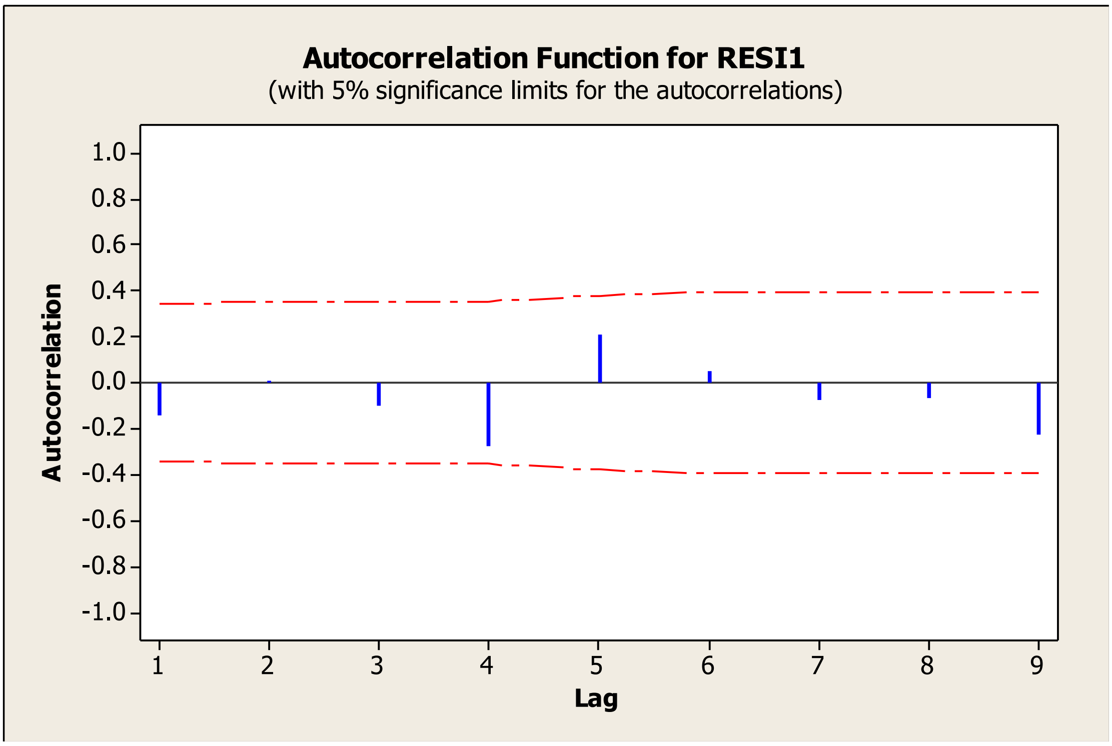

Figure 8: Autocorrelation Function of the Residual from BJM

Conclusion: Since in BJM, both the residual plot and its autocorrelation function fulfils the assumption of Model adequacy; we conclude that the fitted model in expression (9) is adequate. The fitted model is:

$$x_t = 0.34x_{t-1} + \varepsilon_t$$

Comparative Performances of the Estimated Models

By our analysis, the two competing models (RM and BJM) have been found to be adequate. The next step involves fishing out the most preferred models. This will be achieved by subjecting the two models to the model evaluation test described in the methodology. The evaluation involves comparing the values of the different types of errors incurred by the models. The results are tabulated in 1 table below.

Table 1: Error Comparison Table between RM and BJM Models

Model	**MSE**	**MAE**	**MAPE**
Time Series Regression	5.124	3.152	7.492
Box-Jenkins Method	7.428	5.451	7.723

As seen in the above table, the time series regression model (RM) incurs less error than the Box-Jenkins model (BJM) in all the test measures. Hence, the time series regression model is the most preferred.

CONCLUSION

Despite the newness and complexity of the Box and Jenkins method of analysing time series, the old time series regression models still exhibit superiority over the modern Box-Jenkins models. This is clearly shown by the result of the above test measures.

REFERENCES

Bhaskaram, K., Gasparrini, A., Hajat, S., Smeeth, L. and Armstrong, B. (2013). Time Series Regression Studies in Environmental Epidemiology. International Journal of Epidemiology 2013; 42: 1187-1195. Doi:10.1093|ije|dyt092 .

Imai, C., Armstrong B., Chalabi, Z., Mangtani, P. and Hashizume, N. (2015). Time Series Regression Model for Infectious Disease and Weather. Environmental Research 142 (2015) : 319-327. www.elsevier.com|locate|envres

Hejase, H.A.N. and Assi, A.H. (2012). Time Series Regression model for Prediction of Mean Daily Global Solar Radiation in Al-Ain, UAE. Open Access-Hindawi Journal. Volume 10 |Article ID 412471 | https://doi.org/10.5402/2012/412471 .

Tebb, A.T., Tay, L., Wang, W. and Huang, O. (2015). Time Series Analysis for Psychological Research: Examining and Forecasting Change. Front. Psychol., 09 June 2015 | https://doi.org/10.3389/fpsyg.2015.00727 .

APPENDIX : Average Daily Sales per month in Dollars
Source: Wimpey Supermarket, Port Harcourt, Nigeria.

S/N		S/N		S/N		S/N		S/N		S/N	
1	2195	*7*	2290	*13*	2296	*19*	2363	*25*	2413	*31*	2426
2	2209	*8*	2262	*14*	2276	*20*	2386	*26*	2426	*32*	2406
3	2201	*9*	2258	*15*	2305	*21*	2354	*27*	2405	*33*	2440
4	2245	*10*	2256	*16*	2308	*22*	2397	*28*	2419	*34*	2408
5	2238	*11*	2261	*17*	2356	*23*	2398	*29*	2423	*35*	2448
6	2267	*12*	2288	*18*	2393	*24*	2400	*30*	2408	*36*	2460