

Handling Seasonal Autoregressive Integrated Moving Average Model with Correlated Residuals

I. A. Iwok

Department of Mathematics/Statistics, University of Port-Harcourt, Port-Harcourt, Nigeria

Abstract This work proposed a method of handling a seasonal autoregressive integrated moving average (SARIMA) model with correlated residuals. SARIMA model was fitted to a rainfall data and the residuals a_t obtained were found to be correlated. This indicated the inadequacy of the model. The correlated residual series were modelled separately using Box and Jenkins autoregressive moving average (ARMA) methods. The resulting residual model was added to the SARIMA component to obtain a new model. The new model was re-fitted to the data and the residuals ε_t obtained were found to be uncorrelated; confirming the adequacy of the new model.

Keywords ARIMA models, SARIMA model, White noise process, Autocorrelation function and Partial autocorrelation function

1. Introduction/Review

In time series modelling, one of the crucial steps taken to ensure model adequacy is by considering the behaviour of the residuals. Such a step is referred to as diagnostic checks. If the model is adequate, the residuals are expected to be uncorrelated. In more precise terms, the residuals are expected to follow a white noise process. If, however, the residuals are found to be correlated, an alternative model is employed to fit the data.

According to [2], no model form ever represents the truth absolutely. Thus, given sufficient data, statistical tests can discredit models that could nevertheless be entirely adequate for the data at hand. Tests can fail to indicate serious departures from assumptions because these tests are insensitive to the types of discrepancies that occur. The best policy therefore, is that our diagnostic checks must be sensitive to discrepancies that are likely to happen. However, in a case where the model is inadequate after subjecting it to a diagnostic check; an alternative model should be tried [2].

[7] presented theoretical basis for modelling univariate traffic condition data streams as seasonal autoregressive integrated moving average processes. The method employed the Wold decomposition theorem with the assertion that a one week lagged first seasonal difference applied to discrete interval traffic condition data will yield a weakly stationary transformation. The assertions and findings contradicted the

statement of [4] which says that extending simple ARIMA models to include seasonal and other effects in practice does not have a substantial impact on the results. [7] showed that a first seasonal difference taken at a one week lag is the key to a proper application of ARIMA modelling to time indexed traffic volumes.

[6] established a seasonal ARIMA model that was able to capture and explain the patterns and the determinants of German tourism demand in Croatia. Autocorrelation and partial autocorrelation of the residuals were used as diagnostic checks of the model; and SARIMA (0,0,0)(1,1,3)₄ was found to fit the data adequately.

[3] showed that the class of seasonal ARIMA models with a seasonal moving average factor can be parsimoniously generalised to model time series with heteroskedastic seasonal frequency components. [3] used frequency specific models to decompose the factor by associating one moving average coefficient with a proper subset of the several frequencies 1, 2, 3, 4, 5 and 6 cycles per year and a second coefficient with the complementary subset. A generalization of Akaike information criteria was presented to determine these subjects. The properties of seasonal adjustment filters and adjustments obtained from the new models were examined as forecasts.

[5] investigated the use of double seasonal ARIMA model for forecasting load demand. The mean absolute percentage error (MAPE) was used as the measure of forecasting accuracy. Least squares method was used to estimate the coefficients in a double SARIMA model. Comparing the forecasting performances by using k-step ahead forecasts and one-step ahead forecasts, it was found that MAPE for the one-step ahead out-sample forecasts from any horizon ranging from one-week lead time to one month lead time

* Corresponding author:

ibywok@yahoo.com (I. A. Iwok)

Published online at <http://journal.sapub.org/ajms>

Copyright © 2017 Scientific & Academic Publishing. All Rights Reserved

were all less than 1%. Indeed, a double seasonal ARIMA model with one-step ahead forecast was proposed and was shown to be appropriate model for forecasting the two-seasonal cycles Malaysia load demand time series.

The above reviewed works on seasonal models dwell so much on direct applications. This work, however, seeks to address a method of handling a seasonal autoregressive integrated moving average (SARIMA) model where the fitted model fails to be adequate and the residuals are correlated.

2. Methodology

2.1. 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.

2.2. Backward Shift Operator

This is defined as $B^j Y_t = Y_{t-j}$.

2.3. Difference Operator

This is defined as $\nabla = 1 - B$.

2.4. The Seasonal Autoregressive Integrated Moving Average (SARIMA) Model

The SARIMA model denoted: $SARIMA(p, d, q) \times (P, D, Q)_s$ can be expressed as

$$\phi_p(B)\phi_p(B^s)\nabla^d\nabla_s^D Y_t = \theta_q(B)\theta_q(B^s)a_t \quad (1)$$

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$, $\Phi(B) = 1 - \Phi_{1,s} B^s - \Phi_{2,s} B^{2s} - \dots - \Phi_{p,s} B^{ps}$, $\nabla^d = (1 - B)^d$, $\nabla_s^D = (1 - B^s)^D$, $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$,

$$\Theta(B^s) = 1 - \Theta_{1,s} B^s - \Theta_{2,s} B^{2s} - \dots - \Theta_{Q,s} B^{Qs}$$

Y_t is the time series at period t , a_t is the (residuals) white noise process, B represents the backshift operator, S is the season, p is the order of non-seasonal autoregressive components,

P is the order of seasonal autoregressive components, d is the order of non-seasonal differencing,

D is the order of seasonal differencing, q is the order of non-seasonal moving average component,

Q is the order of seasonal moving average component.

3. The Residual ARMA Method

Assuming the a_t in (1) does not follow a white noise process; this approach proposes that the residuals \hat{a}_t be modelled separately by ARMA methods and the result be added to the SARIMA part. That is,

$$\phi_p(B)a_t = \theta_q(B)\varepsilon_t \quad (2)$$

where

ε_t is the (residuals) white noise process obtained from modelling a_t separately using the ARMA methods. Equation (2) implies that

$$a_t = \frac{\theta_{q(a_t)}(B)}{\phi_{p(a_t)}(B)} \varepsilon_t \quad (3)$$

where,

$\theta_{q(a_t)}(B)$ is the moving average operator from the a_t series.

$\phi_{p(a_t)}(B)$ is the autoregressive operator from the a_t series.

Putting (3) in (1), we obtain

$$\varepsilon_t = \frac{\phi_p(B)\phi_p(B^s)\nabla^d\nabla_s^D\phi_{p(a_t)}(B)}{\theta_q(B)\theta_q(B^s)\theta_{q(a_t)}(B)} Y_t.$$

The residuals obtained by this approach are found to be uncorrelated and follow a white noise process as will be demonstrated in the next section.

4. Illustration and Result

The above proposed method was illustrated using monthly rainfall series (between 1995 and 2014) from Nigeria Meteorological Agency, Markurdi. The analysis was carried out using the Minitab software. The series is found to indicate seasonality as shown in the raw data, autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of figures 1, 2 and 3 below. In these figures, the series indicates significant spikes at lags 12, 24, 36, Thus the season, $s = 12$. Figure 1 clearly shows that the series is not stationary; and differencing transformation was applied to obtain stationarity (see figure 4). The ACF plot of the stationary (DY_t) series in figure 5 decays exponentially to zero at the seasonal lags 12, 24, 36, ... ; and alternate at the non-seasonal lags. However, the PACF of DY_t (see figure 6) cut off after lag 12 at the seasonal lags and after lag 1 at the non-seasonal lags. The behaviours of the ACF and PACF suggest that the appropriate model for the series is SARIMA (1,1,0)(1,0,0)₁₂. Applying equation (1), the model can be expressed as:

$$(1 - \phi_1 B)(1 - \Phi_{1,12} B^{12})(1 - B)Y_t = a_t$$

which gives

$$Y_t = Y_{t-1} + \phi_1 Y_{t-1} + \phi_1 Y_{t-2} + \Phi_{1,12} Y_{t-12} - \Phi_{1,12} Y_{t-13} - \phi_1 \Phi_{1,12} Y_{t-12} + \phi_1 \Phi_{1,12} Y_{t-14} + a_t \quad (4)$$

Substituting the SARIMA coefficient provided by the Minitab software in table 1, we have:

$$Y_t = Y_{t-1} - 0.5333Y_{t-1} - 0.5333Y_{t-2} + 0.5833Y_{t-12} - 0.5833Y_{t-13} + 0.3111Y_{t-12} - 0.3111Y_{t-14} + a_t \quad (5)$$

The residual autocorrelation function of the above model is displayed in figure 7. As clearly seen, there are some significant spikes which are out of the two boundary lines. This is an indication that the residuals a_t are correlated and do not follow white noise process. Hence, the model (5) is

inadequate.

Table 1. SARIMA Coefficients

Final Estimates of Parameters					
Type		Coef	SE Coef	T	P
AR	1	-0.5333	0.0795	-6.71	0.000
SAR	12	0.5833	0.0793	7.36	0.000

4.1. The Residual ARMA Approach

In this approach, the correlated residuals a_t are subjected to the Box and Jenkins (1976) ARMA methods. The autocorrelation function (ACF) of the a_t series displays a sinusoidal pattern (see figure 7) while its partial autocorrelation function (PACF) cut off after lag 1 (see figure 8). Thus the identified model is ARMA (1, 0) which can be expressed as

$$(1 - \phi_1 B)a_t = \varepsilon_t \Rightarrow a_t = \phi_1 a_{t-1} + \varepsilon_t \quad (6)$$

As described in the methodology, equation (6) is substituted in equation (4) to give:

$$Y_t = Y_{t-1} + \phi_1 Y_{t-1} + \phi_1 Y_{t-2} + \Phi_{1,12} Y_{t-12} - \Phi_{1,12} Y_{t-13} \\ - \phi_1 \Phi_{1,12} Y_{t-12} + \phi_1 \Phi_{1,12} Y_{t-14} + \phi_1 a_{t-1} + \varepsilon_t \quad (7)$$

which yields

$$Y_t = Y_{t-1} + 0.231Y_{t-1} - 0.3412Y_{t-2} + 0.2562Y_{t-12} \\ + 0.1223Y_{t-13} + 0.6131Y_{t-12} \\ - 0.5214Y_{t-14} + 0.3742a_{t-1} + \varepsilon_t \quad (8)$$

The autocorrelation function of the residual ε_t is displayed in figure 9. Clearly, there is no significant autocorrelation of the residuals since all the autocorrelations are within the limits $\pm 2/\sqrt{n}$ (shown by the two lines in figure 9; n is the sample size). Hence, the fitted model (8) is adequate.

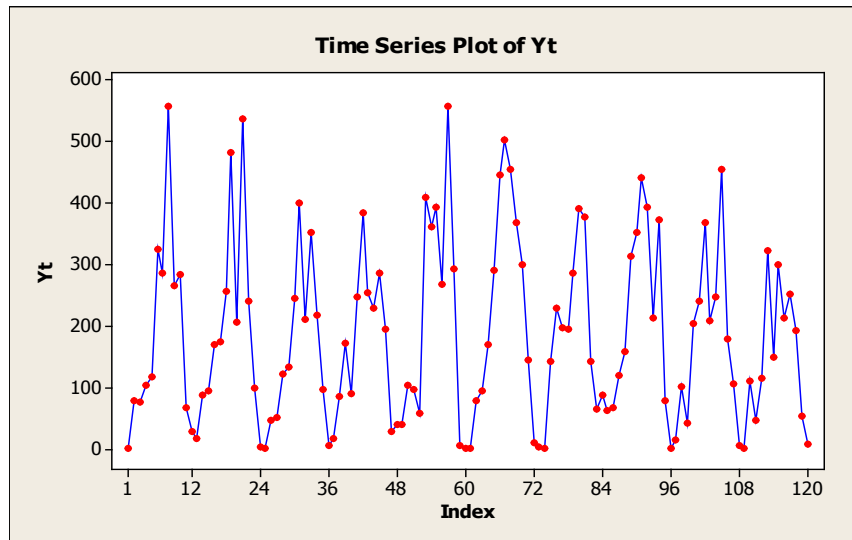


Figure 1. Raw data plot of the original series (Y_t)

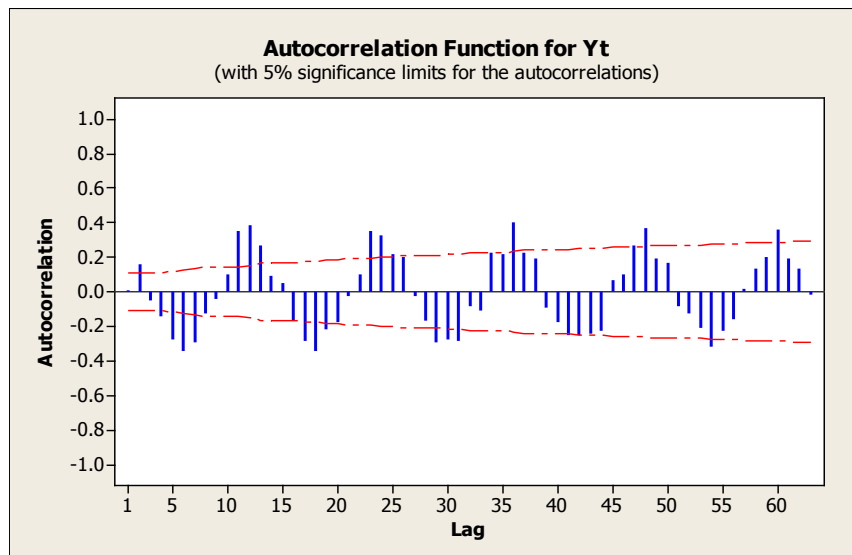


Figure 2. Autocorrelation plot of the original series (Y_t)

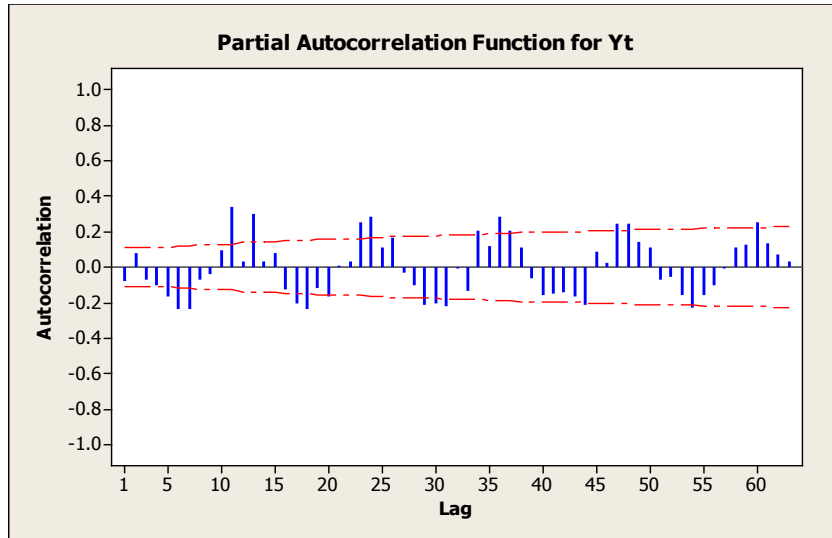


Figure 3. Partial Autocorrelation plot of the original series (Y_t)

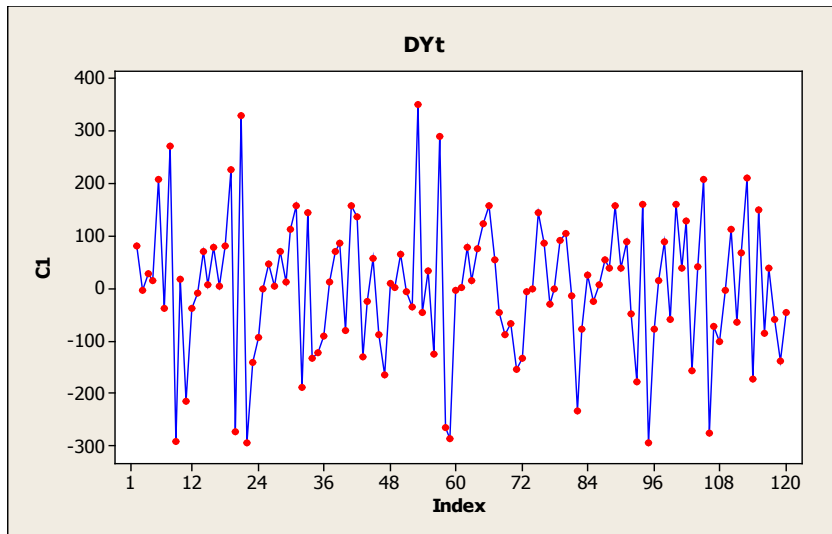


Figure 4. Plot of the differenced series (DY_t)

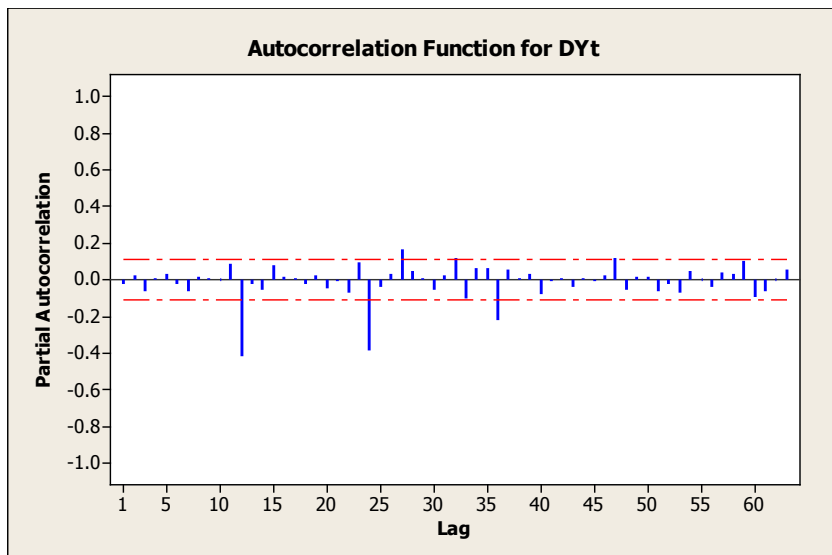


Figure 5. Autocorrelation plot of the differenced series (DY_t)

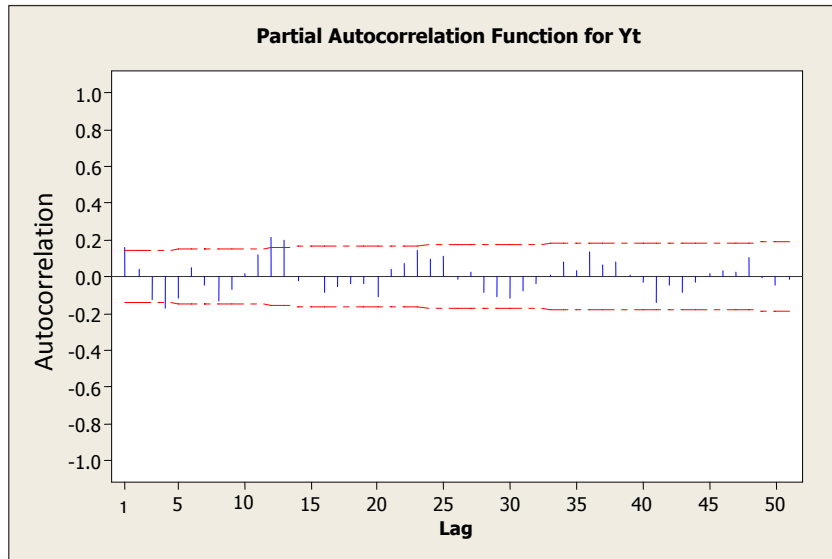


Figure 6. Partial Autocorrelation plot of the differenced series (DY_t)

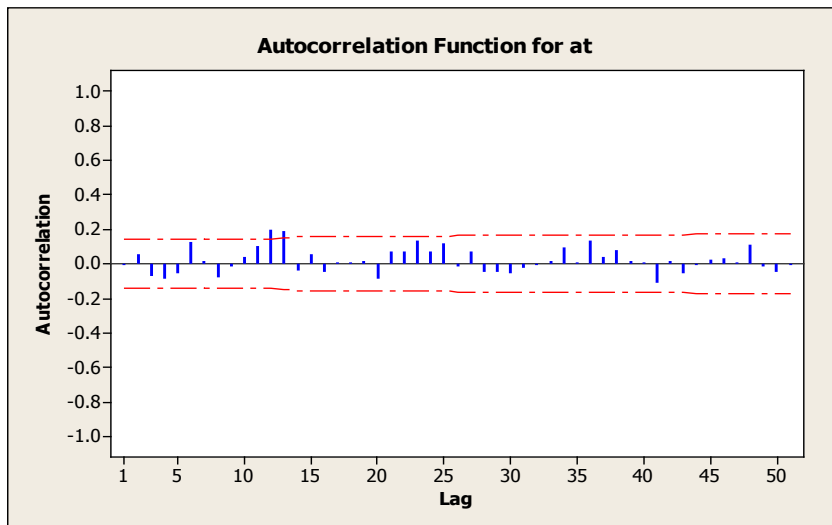


Figure 7. Autocorrelation plot of the residual series (a_t)

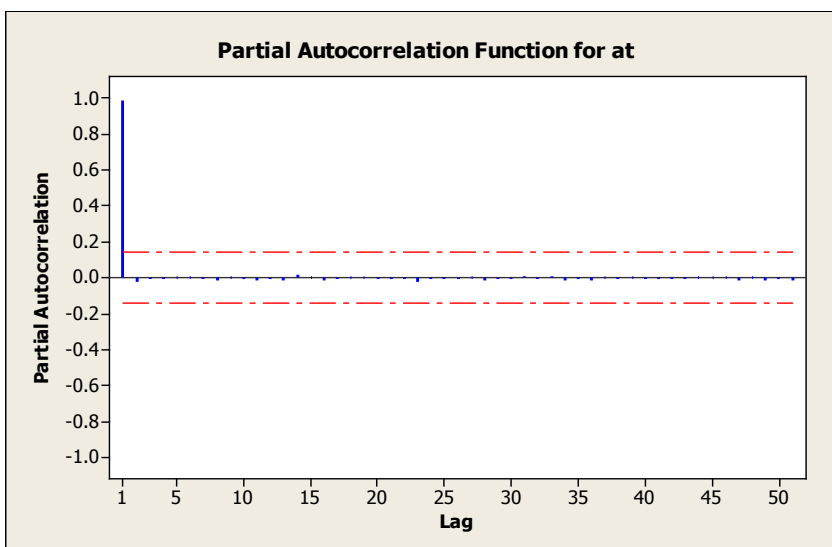


Figure 8. Partial Autocorrelation plot of the residual series (a_t)

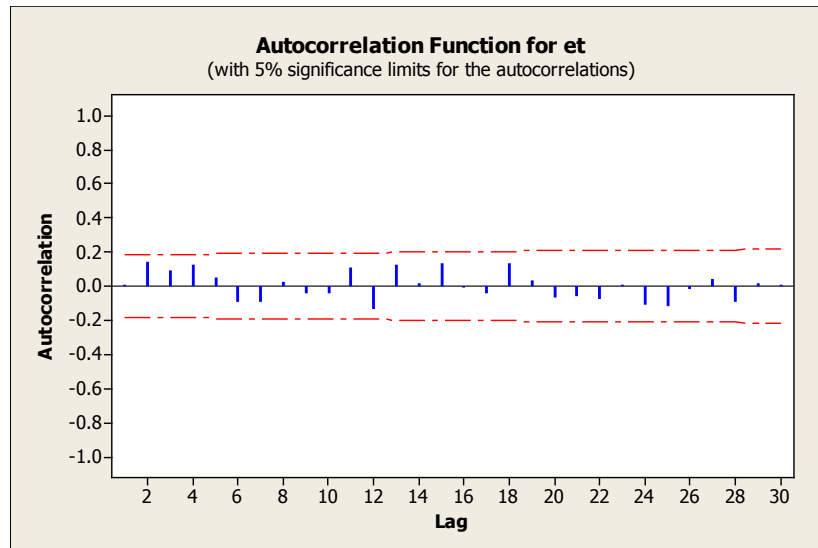


Figure 9. Autocorrelation plot of the residual series (ε_t)

5. Discussion and Conclusions

Usually, in time series modelling; if the fitted model is inadequate, it is replaced by an entirely new model that best fit the data. However, it may be difficult to search for the appropriate model through the usual process and sometimes may be time consuming. The proposed method in this work is advantageous in the sense that it is cheap, simple and less time consuming. Rather than trying or searching for alternative models {as proposed by [2]}, this method utilizes the residuals obtained from the first fitted model to build the final model that best fit the data. As illustrated above, the method has yielded dependable results. It is therefore believed that this proposal has offered another possibility of modelling time series processes.

REFERENCES

- [1] Box, G.E.P. and Jenkins, G.M. (1976). Time Series Analysis; forecasting and control. Holdenday Inc., San Francisco, California. ISBN: 0-8162-1104-3.
- [2] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (2008). Time Series Analysis; forecasting and control. Fourth Edition. A John Wiley & Sons, Inc., Hoboken, New Jersey. ISBN: 978-0-470-27284-8 (cloth). QA 280.B67 2008. 519.5'5-dc22.
- [3] John, A.D., David, F.F., Turkur, S.M., Kellie, C.W. and Donald, E.K. (2007). New ARIMA Models for seasonal time series and their application to seasonal adjustment and forecasting. Journal of Applied Statistics. Vol. 6, No. 7, pp. 33-39.
- [4] Kirby, H., Dougherty, M. and Watson, S. (1997). "Should we use neural networks or statistical models for short term motorway traffic forecasting?" International Journal of Forecasting; 13(1), 43-50.
- [5] Norizon, M., Maizah, H.A., Zuhaimy, I. and Suhartono (2010). Double SARIMA model for forecasting load demand. MATIMATIKA, 2010, vol. 26, No. 2, pp. 217-231.
- [6] Tea, B. and Maja, M. (2015). Modelling international tourism demand using seasonal ARIMA models. Tourism and Hospitality Management Vol. 21, No. 1, pp. 19-31.
- [7] Williams, B.M. and Hoel, L.A. (2003). Modelling and forecasting vehicular traffic flow as seasonal ARIMA process: Theoretical basis and empirical results. Journal of Transportation Engineering © ASCE. Vol. 6, No. 664. DOI: 10.1061/(ASCE) 0733-947X (2003) 129: 6 (664).