A SEASONAL TIME SERIES MODEL FOR NIGERIAN MONTHLY AIR TRAFFIC DATA

Ette Harrison Etuk
Department of Mathematics/Computer Science, Rivers State University of Science and Technology, Nigeria

ABSTRACT
Nigerian Monthly Air Traffic Data (NAP) is analysed as a time series. The non-seasonal difference of its seasonal (i.e. 12-month) difference (DSDNAP) is observed to show some seasonality. The autocorrelation function of DSDNAP reveals a 12-month seasonality, the involvement of a seasonal moving average component of order 1 and the product of two autoregressive components: one non-seasonal and the other seasonal, both of order one. Therefore, a $(1, 1, 0)x(1, 1, 1)_{12}$ is proposed and fitted to the data. This model has been demonstrated to be adequate.

Keywords: *Air Traffic Data, Seasonal Time Series, ARIMA models, Nigeria.*

1. INTRODUCTION

A time series may be defined as data collected sequentially in time, the time points often equally spaced. A property of such a series is that neighbouring values are correlated. This correlation is called *autocorrelation.* Put as a function of the lag separating the correlated values, it is called *autocorrelation function* (ACF). The graph of the ACF is called the *correlogram.*

A *stationary* time series refers to a time series with a constant mean, a constant variance and autocorrelation that is a function of the lag separating the correlated values. A stationary time series $\{X_t\}$ is said to follow an *autoregressive moving average model of order p and q*, denoted by ARMA(p, q) if it satisfies the following difference equation

$$X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - ... - \alpha_p X_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + ... + \beta_q \varepsilon_{t-q} \quad (1)$$

where $\{\varepsilon_t\}$ is called a *white noise process* and defined as a sequence of uncorrelated zero mean random variables with constant variance. The model (1) may be alternatively put as

$$A(L)X_t = B(L)\varepsilon_t \quad (2)$$

where

$$A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - ... - \alpha_p L^p$$

and

$$B(L) = 1 + \beta_1 L + \beta_2 L^2 + ... + \beta_q L^q$$

and L is the backward shift operator defined by $L^k X_t = X_{t-k}$. Besides stationarity, another necessary property for a time series is *invertibility*, which may be defined as the situation whereby the model is associated with a unique autocorrelation structure (Priestley[1]). For the model above to be stationary the equation A(L) = 0 must have roots all outside the unit circle and for it to be invertible, the equation B(L) = 0 must have all roots outside the unit circle.

If p = 0, the model (1) or (2) becomes a *moving average model of order q*, designated MA(q). If, however, q = 0, the model (1) or (2) an *autoregressive model of order p*, designated AR(p). An autoregressive model of order p may be more specifically written as

$$X_t - \alpha_{p1} X_{t-1} - \alpha_{p2} X_{t-2} - ... - \alpha_{pp} X_{t-p} = \varepsilon_t \quad (3)$$

The sequence of the last coefficients $\{\alpha_{ii}\}$ is called the *partial autocorrelation* (PACF) of $\{X_t\}$. The PACF of an AR(p) cuts off at lag p, whereas that of an MA model dies off slowly. The ACF of an MA(q) model cuts off at lag q but the PACF dies off slowly. AR and MA models have certain duality relationships. For instance, a finite-order model of one type is equivalent to an infinite-order model of the other type.

Most real-life time series exhibit non-stationary behaviour. Box and Jenkins ([2]) proposed that such a series could be made stationary after differencing of an appropriate order. Let the minimum order of differencing applied to a time series $\{X_t\}$ to render it stationary be d. This d^{th} difference of X_t denoted by $\nabla^d X_t$ and defined by $\nabla^d X_t = (1 - L)^d X_t$ is used in lieu of X_t in model (1) or (2). Then the resultant model is called an *autoregressive integrated moving average model of orders p, d and q* , designated ARIMA(p,d,q) in $\{X_t\}$.

Seasonality refers to a tendency for a time series to fluctuate periodically. Many economic time series are seasonal, fluctuating according to natural "seasons". Assuming that s is the period of seasonality, a time series $\{X_t\}$ is said to follow a multiplicative $(p, d, q)x(P, D, Q)_s$ seasonal ARIMA model if

$$A(L)\Phi(L^s)\nabla^d\nabla^D_s X_t = B(L)\Theta(L^s)\varepsilon_t \quad (4)$$

where $\Phi(L) = 1 + \phi_1 L + \phi_2 L^2 + ... + \phi_P L^P$ and $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_Q L^Q$ and the ϕ's and θ's are constants such that the roots of $\Phi(L) = 0$ and $\Theta(L) = 0$ are all outside the unit circle for stationarity and invertiblity respectively.

The purpose of this work is to fit a seasonal ARIMA model to Nigerian Air Traffic Data (NAP). Earlier works on Air Traffic Modelling exist. For instance, Box and Jenkins ([1]) modelled monthly totals (in thousands of passengers) from 1949 to 1960 as $(0, 1, 1)x(0, 1, 1)_{12}$. He used this data to introduce the concept of a seasonal ARIMA model. In the Nigerian scene, Ogbudinkpa ([3]) used the traditional approach of breaking the series into the traditional components: secular trend, seasonal movement and irregular movement, to analyse Nigerian Air Traffic Data. Etuk *et al* ([4]) fitted to the data an ARIMA(7, 1, 0) model.

2. MATERIALS AND METHODS

The data for this work are ninety six monthly totals of Nigerian Air Traffic Data in nearest thousands covering 2004 to 2011 retrievable from the Federal Airways Authority of Nigeria (FAAN) website www.faannigeria/documents/statistics.

2.1. Determination of orders p, d, q, s, P, D and Q

Seasonality often becomes apparent from the time plot of a time series. In that case, s is the observed period. Seasonality of period s is suggestive if the correlogram shows a significant spike at the lag s, with neighbouring spikes non-significant. A negative spike at lag s also is indicative of the involvement of a seasonal MA component of lag 1, that is, Q=1. On the other hand, a positive spike at lag s indicates the involvement of a seasonal AR component of lag 1, that is, P=1. To avoid undue model complexity, it has been advised that $d + D < 3$. The orders p and q are the lags of the cutting off of the ACF and PACF respectively. Moreover, knowledge of the autocorrelation structure of the hypothesized model provides further basis for order determination. Box and Jenkins ([2]) and Madsen ([5]) are a few of authors that have written extensively on seasonal time series models.

2.2. Model Estimation

Once orders have been determined, the model could be estimated. Involvement of items of a white noise process in the model to be estimated often entails the application of non-linear optimization techniques for model estimation. Usually an initial estimate is made. Then by an iterative process improvement is made on the estimates until convergence to an optimal estimate is achieved. The criterion of optimization could be the least error sum of squares, the maximum likelihood, the maximum entropy, etc. For pure AR or pure MA models, there exist linear optimization techniques (see for example, Box and Jenkins, [1]; Oyetunji, [6]). There are attempts to adopt linear optimization techniques for mixed ARMA models (see for example Etuk ([7], [8]). In this work we are using Eviews software which employs the least sum of squares technique.

2.3. Diagnostic Checking

Sequel to model estimation goodness-of-fit of the model to the observations must be ascertained. It involves some residual analysis. Under the assumption of model adequacy, the residuals should be uncorrelated, have mean zero and follow a normal distribution.

3. RESULTS AND DISCUSSION

The time plot of NAP in Figure 1 shows a depression in 2008 and a generally increasing trend. Seasonality is not obvious. Seasonal (i.e. 12-month) differencing yields SDNAP which apart from a peak in 2009 shows no trend on the overall. Seasonality is not evident still (see Figure 2). Non-seasonal differencing of SDNAP yields DSDNAP which shows no trend either. Even though its time plot in Figure 3 does not reveal a seasonal nature its correlogram in Figure 4 reveals seasonality of order 12, involvement of a seasonal MA component and a seasonal AR component. Moreover the spikes at lags 1, 12 and 13 in the PACF suggest the involvement of the product of two autoregressive components: one seasonal and the other non-seasonal. The model $(1, 1, 0)x(1, 1, 1)_{12}$ is hereby proposed. That is,

$$DSDNAP = \alpha_1 DSDNAP_{t-1} + \alpha_{12} DSDNAP_{t-12} + \alpha_{13} DSDNAP_{t-13} + \varepsilon_{t-12} + \varepsilon_t \quad (5)$$

The model (5) estimated as summarized in Table 1 is given by

$$DSDNAP + 0.3123 DSDNAP_{t-1} + 0.5074 DSDNAP_{t-12} + 0.2528 DSDNAP_{t-13} + 0.8507 \varepsilon_{t-12} = \varepsilon_t \quad (6)$$

It is noteworthy that all the coefficients of the model (6) are statistically significant, each being more than twice its standard error. Moreover as much as 69% of the variation in DSDNAP is explained by the model. Figure 5 shows a very close agreement between the model and the data. Figure 6 is a histogram of the residuals. The probability curve shows a nearly normal distribution of mean zero (apart from the outlier of value close to 500). The correlogram of the residuals in Figure 7 is such that no correlation is statistically significant. All these show that the model is adequate.

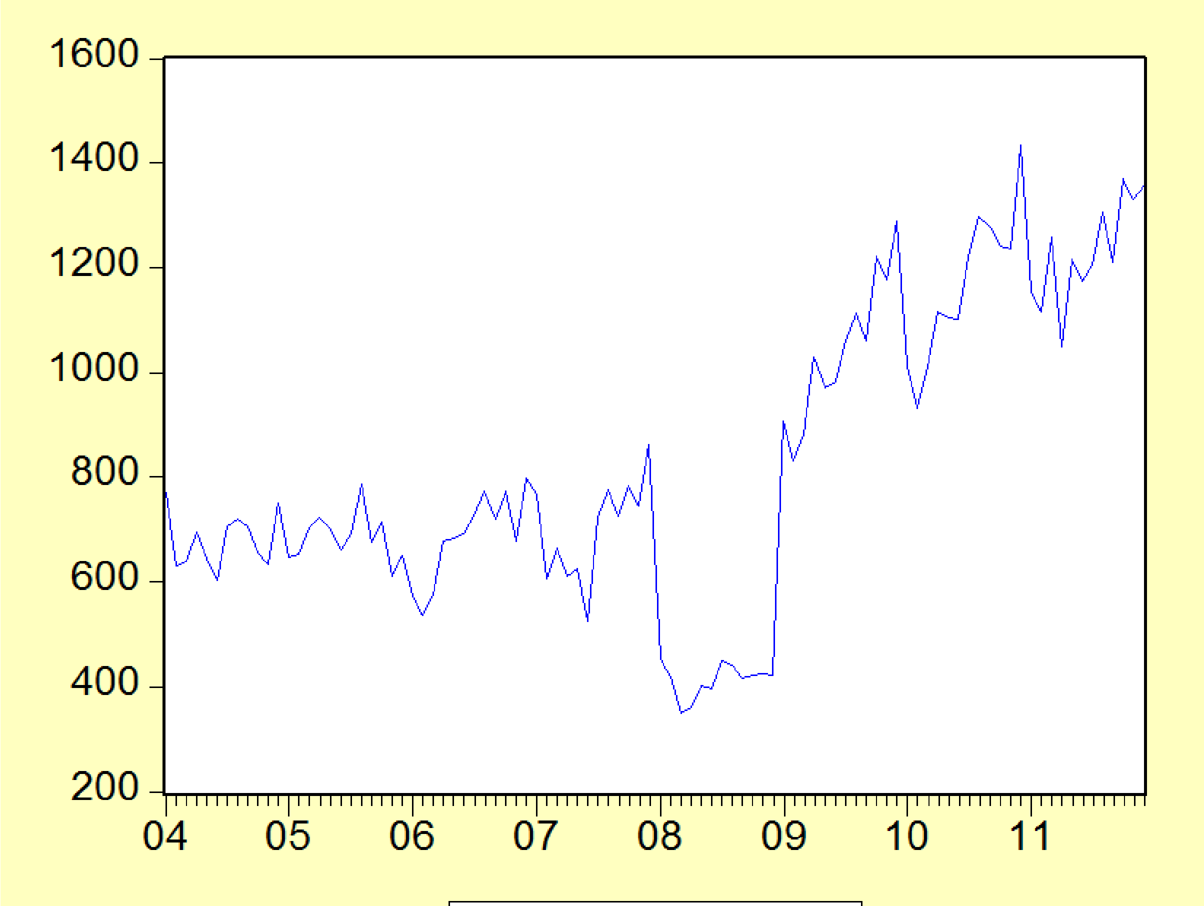
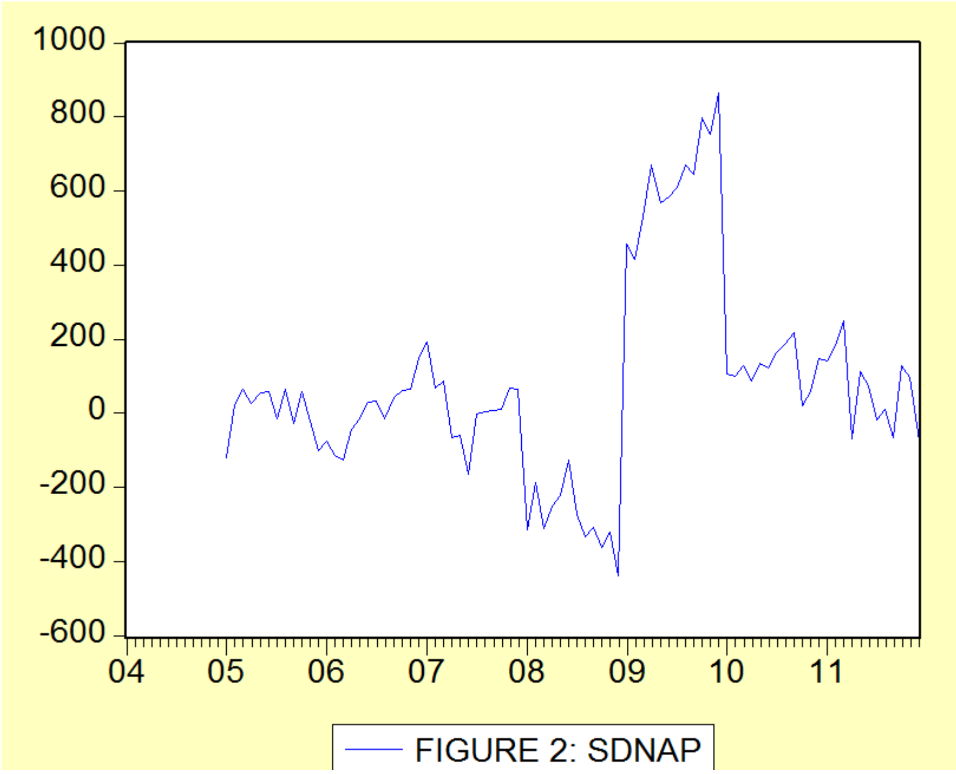
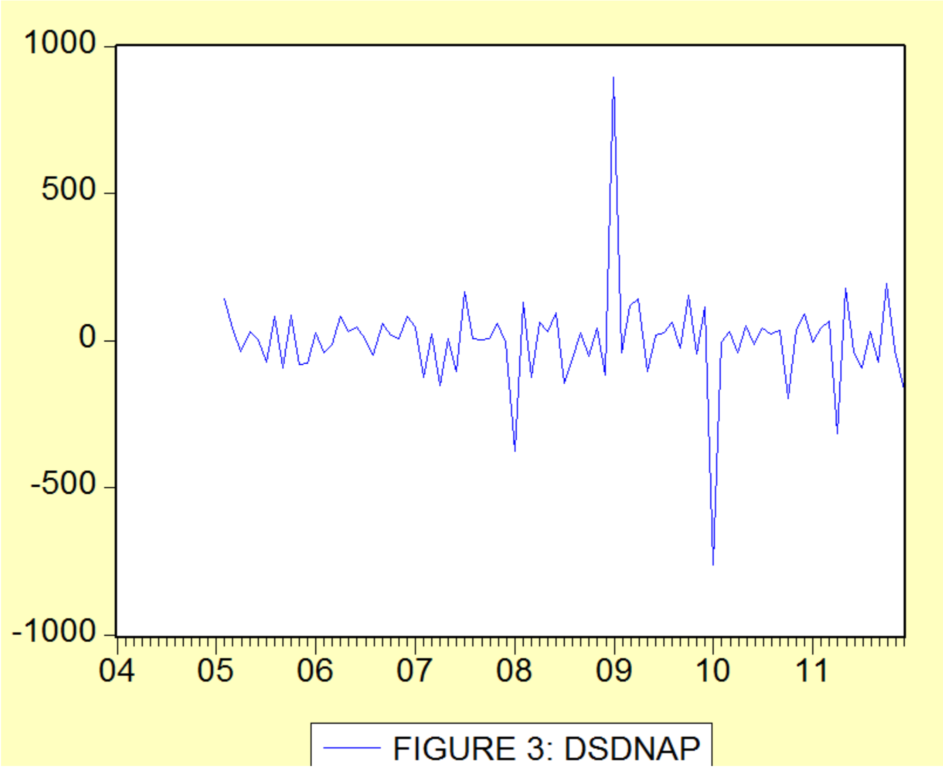
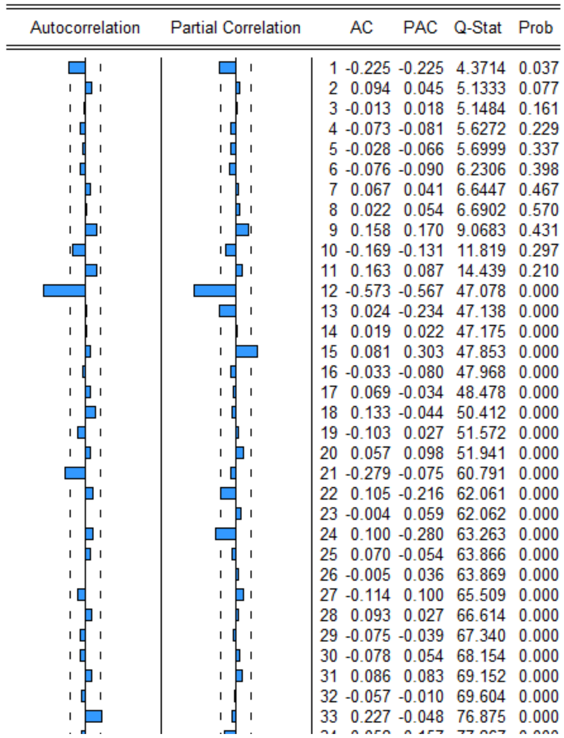


FIGURE 1: NAP


1000
800
600
400
200
0
-200
-400
-600
04
05
06
07
08
09
10
11
FIGURE 2: SDNAP


1000
500
0
-500
-1000
04
05
06
07
08
09
10
11
FIGURE 3: DSDNAP



	AC	PAC	Q-Stat	Prob
1	-0.225	-0.225	4.3714	0.037
2	0.094	0.045	5.1333	0.077
3	-0.013	0.018	5.1484	0.161
4	-0.073	-0.081	5.6272	0.229
5	-0.028	-0.066	5.6999	0.337
6	-0.076	-0.090	6.2306	0.398
7	0.067	0.041	6.6447	0.467
8	0.022	0.054	6.6902	0.570
9	0.158	0.170	9.0683	0.431
10	-0.169	-0.131	11.819	0.297
11	0.163	0.087	14.439	0.210
12	-0.573	-0.567	47.078	0.000
13	0.024	-0.234	47.138	0.000
14	0.019	0.022	47.175	0.000
15	0.081	0.303	47.853	0.000
16	-0.033	-0.080	47.968	0.000
17	0.069	-0.034	48.478	0.000
18	0.133	-0.044	50.412	0.000
19	-0.103	0.027	51.572	0.000
20	0.057	0.098	51.941	0.000
21	-0.279	-0.075	60.791	0.000
22	0.105	-0.216	62.061	0.000
23	-0.004	0.059	62.062	0.000
24	0.100	-0.280	63.263	0.000
25	0.070	-0.054	63.866	0.000
26	-0.005	0.036	63.869	0.000
27	-0.114	0.100	65.509	0.000
28	0.093	0.027	66.614	0.000
29	-0.075	-0.039	67.340	0.000
30	-0.078	0.054	68.154	0.000
31	0.086	0.083	69.152	0.000
32	-0.057	-0.010	69.604	0.000
33	0.227	-0.048	76.875	0.000

FIGURE 4: CORRELOGRAM OF DSDNAP

TABLE 1: MODEL ESTIMATON

Dependent Variable: DSDNAP
Method: Least Squares
Date: 10/13/12 Time: 20:30
Sample(adjusted): 2006:03 2011:12
Included observations: 70 after adjusting endpoints
Convergence achieved after 8 iterations
Backcast: 2005:03 2006:02

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.312758	0.116583	-2.682701	0.0092
AR(12)	-0.507429	0.117786	-4.308066	0.0001
AR(13)	-0.252777	0.129115	-1.957773	0.0545
MA(12)	-0.850673	0.042652	-19.94446	0.0000

R-squared	0.693390	Mean dependent var	0.557143
Adjusted R-squared	0.679453	S.D. dependent var	173.9601
S.E. of regression	98.49074	Akaike info criterion	12.07325
Sum squared resid	640228.1	Schwarz criterion	12.20173
Log likelihood	-418.5637	F-statistic	49.75236
Durbin-Watson stat	1.977239	Prob(F-statistic)	0.000000

Inverted AR Roots	.92 -.25i	.92+.25i	.68 -.67i	.68+.67i
	.26 -.92i	.26+.92i	-.23 -.92i	-.23+.92i
	-.50	-.65 -.68i	-.65+.68i	-.89 -.25i
	-.89+.25i			
Inverted MA Roots	.99	.85+.49i	.85 -.49i	.49+.85i
	.49 -.85i	.00 -.99i	-.00+.99i	-.49 -.85i
	-.49+.85i	-.85 -.49i	-.85+.49i	-.99

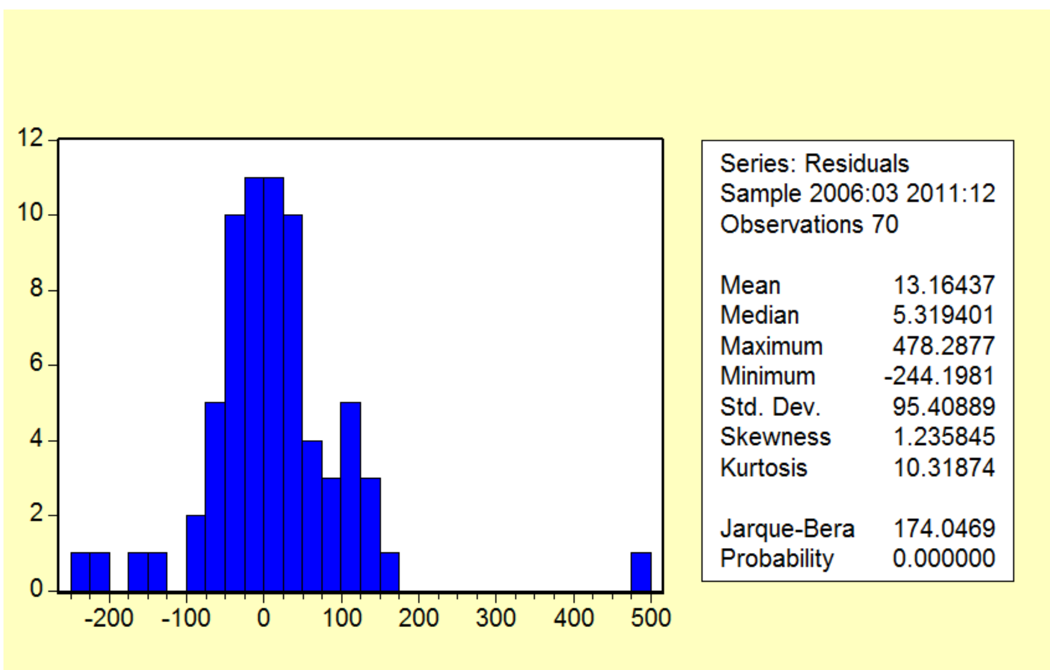
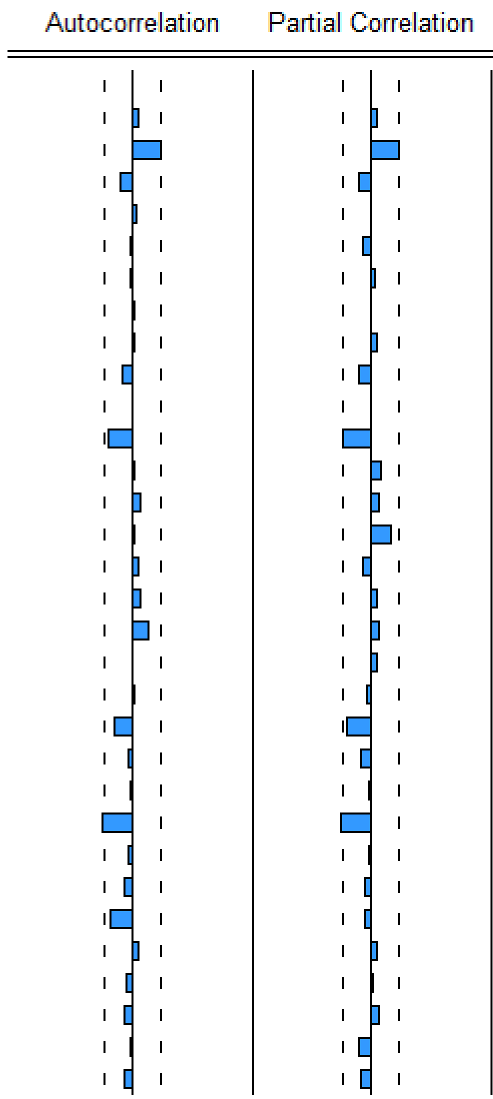


FIGURE 6: HISTOGRAM OF THE RESIDUALS



	AC	PAC	Q-Stat	Prob
1	-0.008	-0.008	0.0048	
2	0.043	0.043	0.1416	
3	0.244	0.245	4.6241	
4	-0.108	-0.111	5.5184	
5	0.027	0.005	5.5760	0.018
6	-0.014	-0.068	5.5913	0.061
7	-0.022	0.034	5.6283	0.131
8	0.019	0.003	5.6579	0.226
9	0.022	0.049	5.6984	0.337
10	-0.091	-0.113	6.3911	0.381
11	0.007	0.005	6.3949	0.494
12	-0.216	-0.249	10.466	0.234
13	0.014	0.095	10.483	0.313
14	0.075	0.068	10.995	0.358
15	0.022	0.176	11.038	0.440
16	0.046	-0.069	11.232	0.509
17	0.058	0.052	11.555	0.564
18	0.139	0.073	13.422	0.494
19	-0.006	0.046	13.425	0.570
20	0.014	-0.039	13.446	0.640
21	-0.150	-0.201	15.759	0.541
22	-0.039	-0.089	15.915	0.598
23	-0.023	-0.013	15.973	0.659
24	-0.259	-0.254	23.298	0.274
25	-0.040	-0.018	23.476	0.319
26	-0.080	-0.054	24.205	0.337
27	-0.189	-0.056	28.408	0.201
28	0.053	0.052	28.742	0.230
29	-0.054	0.017	29.102	0.260
30	-0.076	0.062	29.828	0.275
31	-0.013	-0.109	29.848	0.321
32	-0.066	-0.093	30.434	0.343

FIGURE 7: CORRELOGRAM OF THE RESIDUALS

4. CONCLUSION

It may be concluded that NAP follows the multiplicative seasonal $(1, 1, 0)x(1, 1, 1)_{12}$ model. This model has been shown to be adequate.

5. REFERENCES

[1] M. B. Priestley, "Spectral Analysis and Time Series". Academic Press, London (1981).

[2] G. E. P. Box and G. M. Jenkins, "Time Series Analysis, Forecasting and Control". Holden-Day, San Francisco (1976).

[3] R. N. Ogbudinkpa, Time Series Analysis of Nigeria's Airways Passenger Traffic 1959-1974, *Jour. Of Nigerian Statistical Association*, **1**, Number 3, 13-29 (1983).

[4] E. H. Etuk, C. E. Onwukwe, I. E. Egong and J. N. Ezeorah, An ARIMA model for Nigerian Aviation traffic Data, *International Journal of Natural and Applied* Sciences, 2, Number 1, 81-85 (2007).

[5] H. Madsen, "Time Series Analysis", Chapman/CRC, London (2008).

[6] O. B. Oyetunji, Inverse Autocorrelations and Moving Average Time Series Modelling. *Jour. Of Official Statistics*, 1, 315-322 (1985).

[7] E. H. Etuk, On the Selection of Autoregressive Moving Average Models. Ph. D. Thesis, Department of Statistics, University of Ibadan, Nigeria (1987).

[8] E. H. Etuk, An Autoregressive Integrated Moving Average (ARIMA) Simulation Model: A Case Study. *Discovery and* Innovation, **10**, Numbers 1 & 2, 23-26 (1998).