INTERNATIONAL JOURNAL OF NATURAL AND APPLIED SCIENCES (IJNAS), VOL. 2, NO. 1(2007); P. 81 – 85, 4 FIGS.

An ARIMA model for Nigerian aviation traffic data

E. H. Etuk*[1], C. E. Onwukwe,[2] I. E. Egong[2] and J. N. Ezeorah[2]

ABSTRACT

The Nigerian aviation data has been analyzed by traditional time series methods. These methods are getting out of fashion. However, by the use of the more modern enhanced Box-Jenkins techniques we fit an adequate ARIMA model to the data set.

INTRODUCTION

Ogbudinkpa (1983) has analyzed Nigerian aviation traffic data (1959-1974) as a time series. He used the traditional approach which involves the identification and unscrambling of the component movements: trend, seasonality, cyclical variation and the irregular movement.

Contemporary techniques of time series analysis which, of course, are an improvement on the traditional ones employ models of the autoregressive moving average (ARMA) family. A stationary time series $\{X_t\}$ is said to follow an ARMA model of order (p, q) if

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

where $\{\varepsilon_t\}$ is a white noise process with variance σ^2 and the α_i's and the β_j's are constants, such that

$$1 + \alpha_1 z + \ldots + \alpha_p z^p \neq 0, |z| \leq 1 \qquad (2)$$

and

$$1 + \beta_1 z + \ldots + \beta_q z^q \neq 0, |z| \leq 1 \qquad (3)$$

Box and Jenkins (1976) have shown that differencing a certain kind of non-stationary time series $\{X_t\}$ to a suitable order makes it stationary. Let the d^{th} order difference of $\{X_t\}$ be denoted by $\nabla^d X_t$. Box and Jenkins (1976) defined this as an autoregressive integrated moving average model of order (p, d, q), designated ARIMA(p, d, q), of X_t. ARIMA models have been successfully applied to model real-life series (e.g. Ozaki, 1977, Etuk(1987), Oyetunji (1985), Priestley (1981).

In the sequel, we shall model the Nigerian Airways passengers' traffic data (1959-1974) using an ARIMA model.

ARIMA MODELLING.

Let $\{X_t\}$ be a stationary time series. Then $\gamma_k = E[(X_t - \mu)(X_{t-k} - \mu)]$, k = 0, 1, 2, … (where $\mu = E(X_t)$) is called its autocovariance of lag k. Then $\rho_k = \gamma_k/\gamma_0$, k = 1, 2, 3, … is known as its autocorrelation of lag k. Both parameters γ_k and ρ_k may be estimated, respectively, by $c_k = \sum(X_t - \bar{X})(X_{t-k} - \bar{X})/N$ and $r_k = c_k/c_0$, k = 1, 2, … with $\bar{X} = \sum X_t/N$, given the realization $X_1, X_2, \ldots, X_N$ of $\{X_t\}$.

Assuming no β_i in equation (1) is significant, the autoregressive model of order p (AR(p))

$$X_t + \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} = \varepsilon_t \qquad (4)$$

results. Put more specifically, equation(4) can be written in the form

$$X_t + \alpha_{p1} X_{t-1} + \ldots + \alpha_{pp} X_{t-p} = \varepsilon_t \qquad (5)$$

The sequence $\{\alpha_{kk}\}$ referred to as the partial autocorrelation function (PACF) of $\{X_t\}$ is known to truncate at lag p if $\{X_t\}$ follows an AR(p). For such a series $\{\rho_k\}$ is known to attenuate. If none of the α_i's in equation (1) is significant, then

$$X_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \ldots + \beta_q \varepsilon_{t-q} \qquad (6)$$

is said to be a moving average model of order q (MA(q)). For such a model the autocorrelation function (ACF) $\{\rho_k\}$ cuts off at lag q and the PACF dies out exponentially. The sample values of these functions are then used as a preliminary guide for model identification. For instance, if $\{r_k\}$truncates at lag w, an MA(w) is suggestive.

Cleveland(1972) has defined the inverse autocorrelation of lag k of a time series $\{X_t\}$ as $\gamma i_k = \int_{-\pi}^{\pi} e^{iwk} fi(w)\, dw = \gamma i_{-k}$, k = 0, 1, 2, … where fi(w) is the reciprocal of the spectral density function f(w). The inverse autocorrelation function (IACF) is defined as

$$\rho i_k = \gamma i_k / \gamma i_0, \; k = 0, 1, 2, \ldots$$

*Corresponding author
Manuscript received by the Editor September 6, 2006; revised manuscript accepted April 2, 2007.
[1]Department of Mathematics, Rivers State University of Science and Technology, Port Harcourt, Nigeria.
[2]Department of Mathematics /Statistics & Computer Science, University of Calabar, Calabar, Nigeria.

For the model equation (4), the IACF is defined by

$$\rho i_k = \begin{cases} (\alpha_k + \sum \alpha_j \alpha_{j+k})/(1 + \sum\alpha_j^2), & k = \pm 1, \pm 2, \dots, \pm p \\ 0, & |k| > p \end{cases} \quad (7)$$

and

$$\gamma i_0 = (1 + \sum\alpha_j^2)\,\sigma^2.$$

Corresponding to the two ways of estimating the spectral density function, namely, by window estimates and by an AR fit, Cleveland (1972) has suggested these two approaches for the estimation of ρi_k. The autoregressive approach consists of fitting an AR(p) with a sufficiently high value of p for a good fit and using the parameter estimates in equation (7).

The moving average (MA) and autoregressive (AR) models have some dual relationships. These include the fact that the ACF of one behaves like the IACF of the other. The dual of the AR PACF in the inverse MA model is called the inverse partial autocorrelation function (IPACF). The use of the sample values of all the four: ACF, IACF, PACF and IPACF for the initial model identification has been advocated (e.g. Hipel *et al.* (1977), Etuk(1988)).

A variety of methods have been proposed to estimate an ARMA model once preliminary identification has been made (e.g. Durbin (1959), Walker (1962), Hannan(1969), Akaike (1973), Box and Jenkins (1976), Oyetunji (1985), etc.). Most of these methods are based on the maximum likelihood principle, the differences arising from the choice of initial estimates.

It has been shown that the fitting of AR models by the least squares approach may be done by the use of the following Levinson-Durbin (1960) algorithm:

$$\left.\begin{aligned} \phi_k &= \alpha_{k+1,k+1} = (r_{k+1} - \sum\alpha_{kj}\, r_{k+1-j}) \\ \alpha_{k+1,j} &= \alpha_{kj} - \phi_{k+1}\, \alpha_{k,k+1-j}, \quad j = 1, 2, \dots, k \end{aligned}\right\} \quad (8)$$

where the form equation (5) is used for the AR(p) model.

To enhance model selection, automatic order determination criteria like AIC (Akaike, 1974), BIC (Akaike, 1977), SIC (Schwarz, 1978), etc. are used. Each of them selects the model corresponding with its minimum within a chosen order range. AIC is defined by

AIC = - ln (maximum likelihood) + 2 (number of parameters)

which for equation (1) is given by $AIC(p+q) = N \ln \sigma^2_{p+q} + 2(p+q)$

where σ^2_{p+q} is the maximum likelihood estimate of the residual variance after fitting an ARMA (p, q) model. By considering purely autoregressive behaviour Etuk (1987) has shown that AIC is the best such criterion for full order modelling.

THE NIGERIAN AIRWAYS PASSENGERS' TRAFFIC DATA (1959 – 1974) (Ogbudinkpa, 1983, pp. 14).

The series has a trend as evident from the time-plot (Ogbudinkpa, 1983, pp. 17). First order differencing removes the trend; the ACF and IPACF of first order differences die out fast (Figs. 1 and 4) to confirm stationarity. The PACF and IACF (Figs. 2 and 3) suggest an AR(7) fit; the IACF was estimated by fitting an AR(30) model and using equation (7) and the IPACF by fitting MA models of orders 1 to 30 using IACF in lieu of ACF in equation(8).

Using AIC, the AR(7)

$$X_t + 0.541X_{t-1} + 0.206X_{t-2} + 0.146X_{t-3} + 0.149X_{t-4} - 0.001X_{t-5} + 0.287X_{t-6} + 0.204X_{t-7} = \varepsilon_t, \quad \sigma^2 = 4845313 \quad (9)$$

was fitted to the first order differences.

Applying Box-Pierce (1970) diagnostic checks to the residuals of the model (9), the value of the test statistic is

$(n-p) \sum h_k^2 = 26.971$

where n is the number of values in the differenced series and h_k the autocorrelation of lag k of the residuals. This is non-significant. Hence, the model is adequate.

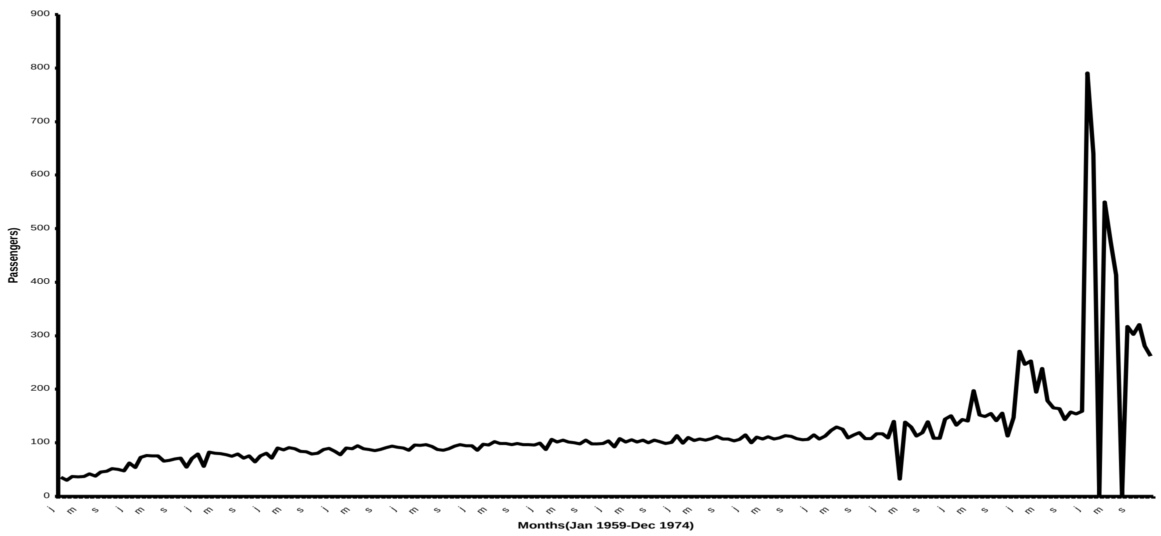

Fig .1 Time plot of Nigeria Airways Passengers Data

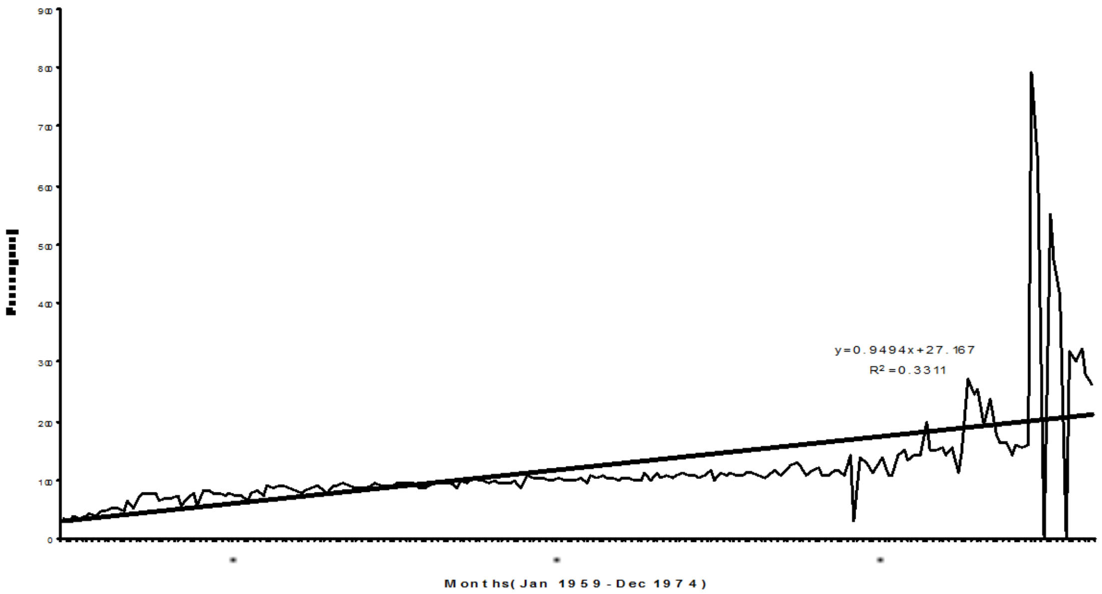

Fig .2 Time plot of Nigeria Airways Passengers Data With Trend

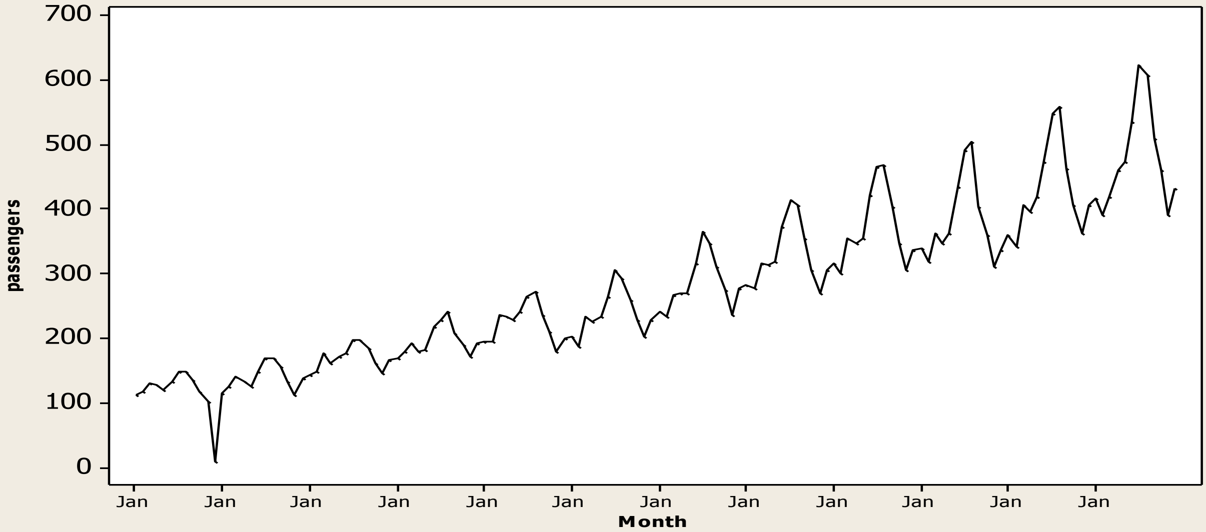

Fig. 3 Time Series Plot of Passengers

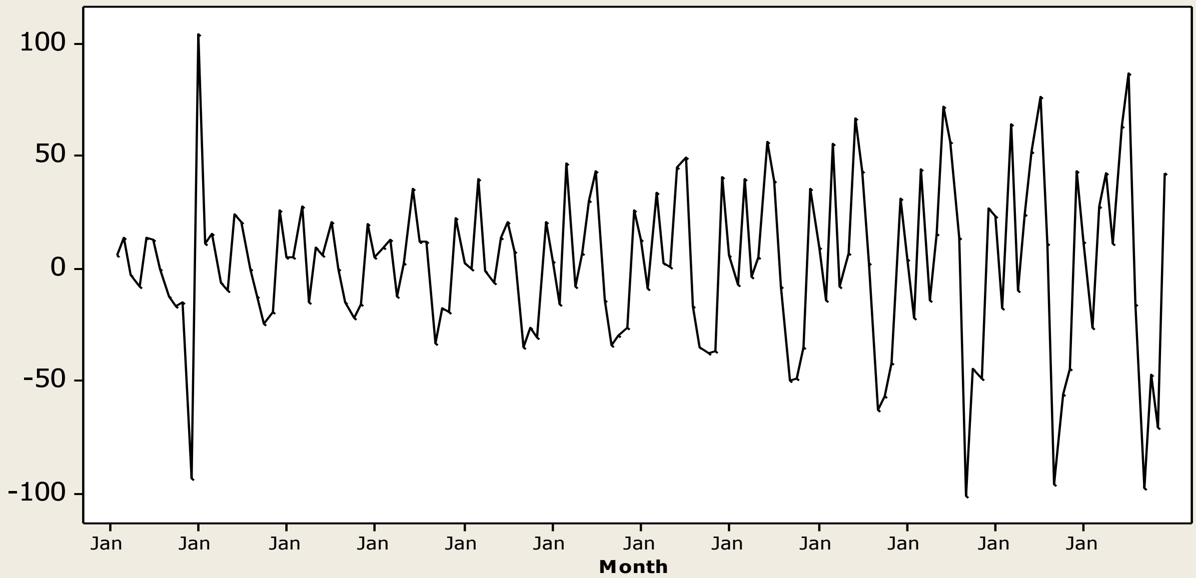

Fig. 4 First Difference of the Original Data

CONCLUSION

The model fitted is an ARIMA (7,1,0). That is, a current value of the first differences depends on the last seven values. Box and Jenkins (1976) have given various procedures for forecasting on the basis of such a model. ARMA models belong to the domain of the general linear process. Non-linear models have now been proposed (e.g. Prietley, 1978) and have been demonstrated to explain the variation in many time series better than linear ones. This is expected for a linear model is often an oversimplification. Perhaps, a non-linear model can better approximate the Nigerian Airways passengers' traffic data.

REFERENCES

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika,* 60: 255 – 265.

Akaike, H. (1974). A new look at the statistical model identification. *I. E. E. E. Trans. Automatic Control, AC-19*: 716 – 722.

Akaike, H. (1977). An entropy maximization principle. In: *Proc. Symp. On Applied Statistics* (Krishnaiah P. ed.). North-Holland, Amsterdam.

Box, G. E. P. and Jenkins, G. M. (1976). *Time series analysis, fortecasting and control*. Holden-Day, San Francisco.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive –integrated moving average time series models. *J. Amer. Statist. Assoc.* 65: 1509 – 1526.

Cleveland, W. S. (1972). The inverse autocorrelations of a time series and their applications. *Technometrics,* 14: 277 – 293.

Durbin, J. (1959). Efficient estimation of parameters in moving average models. *Biometrika,* 46: 306 – 316.

Etuk, E. H. (1987). On the selection of autoregressive models. Ph. D. Thesis, Dept. of Statistics, University of Ibadan, Nigeria.

Etuk, E. H. (1988). On autogressive model identification. *Jour. Of Official Statistics*, 4(2):113 – 124.

Hannan, E. J. (1969). The estimation of mixed moving average autoregressive systems. *Biometrika,* 56: 579 – 593.

Hipel, K. W., Mcleod, A. L., and Lennox, W. C. (1977). Advances in Box-Jenkins modelling1. *Water Resources Research*, 13, : 567 – 575.

Ogbudinkpa, R. N. (1983). Time series analysis of Nigerian Airways passenger traffic 1959 – 1974. *Jour. Of Nigerian Statistical Association.* 1(3) :13 – 29.

Oyetunji, O. B. (1985). Inverse autocorrelations and moving average time series modelling. *Jour. Of Official Statistics*, 1: 315 – 322.

Ozaki, T. (1977). On the order determination of ARIMA models. *Appl. Statist*. 26: 290 – 301.

Priestley, M. B. (1978). Non-linear models in time series analysis. *The Statistician*, 27: 159 – 176.

Priestley, M. B. (1981). *Spectral analysis and time series*. Academic Press, London.

Schwarz, G. (1978). Estimating the Dimension of a model. *Annals of Statistics,* 6 : 461 – 464.

Walker, A. M. (1962). Large sample estimation of parameters for autoregressive processes with moving average residuals. *Biometrika*, 49: 117 – 131.