Ilorin Journal of Science
Volume 4, Number 2, 2017, pp. 263 – 283 (Printed in Nigeria)
ISSN: 2408 – 4840 © 2017 Faculty of Physical Sciences, University of Ilorin
https://doi.org/10.54908/iljs.2017.04.02.006

ILORIN
JOURNAL OF SCEINCE

ILJS-17-009

Modelling Air Passenger Traffic Flow in Murtala Muhammad International Airport Lagos, Nigeria: A Time Series Approach

Omolohunnu*, O. F. and Yahaya, A.

Department of Statistics, Ahmadu Bello University, Zaria, Nigeria.

Abstract

In this study, Artificial Neural Network (ANN), Seasonal Auto-Regressive Integrated Moving Average (SARIMA) and Holt-Winters Exponential Smoothing (HWES) models are used to model air passenger traffic flow in Murtala Muhammad International Airport (MMIA), Lagos Nigeria. The performances of these proposed models are compared for in-sample and out-of-sample performance by employing static forecast procedure over January 2014 to December 2015 forecast horizon. The best models from the SARIMA, ANN and HWES were selected by employing some performance metrics comprising, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) and residual diagnostics. The selected models forecasting performances were compared using the statistical loss functions, Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) for the measurement of forecast accuracy. Results show that ANN outperforms the other models in the domestic sector, while the HWES had the best performance in the international sector even though it was outperformed by SARIMA in the domestic sector. ANN yielded the best in-sample performance for domestic and international air passenger traffic. It was concluded that the ANN, which represents a class of non-linear time series model is very efficient in mimicking time series pattern and giving good forecast.

Keyword: ANN, SARIMA, Holt-Winters Exponential Smoothing, AIC, BIC, Air Passenger Traffic

1. Introduction

In recent past, aviation in Nigeria has been recognized as a key sector of the economy. It is one of the indices for measuring the development of a country. The importance of this sector to the Nigerian Economy cannot be overemphasized. The International Air Transport Association (IATA) released the airline forecast for 2013-2017 which indicated a 31% increase in passenger number between 2012 and 2017. By 2017 the total passenger numbers are expected to rise to 3.91billion- an increase of 930 million passengers over the 2.98 billion carried out in 2012 according to (IATA).

*Corresponding Author: Omolohunnu, O. F.
Email: funshijojo@yahoo.co.uk

Predicting future air passenger traffic flow is important as it allows air transport authorities to put in place necessary infrastructures and offer airline companies the capacity to match the increasing passenger demand for air transportation. Several studies have been carried out since 1990's that use time series models to forecast air passenger traffic. The forecasting performance of each model varies depending on the country under study, the type of flight considered (domestic or international), the performance measure and the forecasting horizon (Emiray and Rodriguez, 2003). (Doguwa and Alade, 2015) proposed three statistical models in modeling Nigerian external reserves, SARIMA, seasonal autoregressive integrated moving average with an exogenous input (SARIMA-X) and ARDL an autoregressive distributed lag model. Using pseudo-out-of-sample forecasting they concluded that SARIMA model outperformed the other models in three to six months forecast horizon whereas ARDL model performs better in one to two months forecast horizon. (Ajibode, 2016) utilized SARIMA model in forecasting enplane passenger traffic in MMIA, the best SARIMA model was selected based on AIC.

No methodological approach has been found to always dominate another in terms of forecasting performance (Shen *et al.*, 2011). However one model that has been successfully used in its various applications is the Artificial Neural Network (ANN). ANN is a mathematical model inspired by the function of the human brain and its use is mainly motivated by its capability of approximating any Borel-measurable function to any degree of accuracy (Hornik *et al.*, 1989). Though, ANN is a powerful forecasting tool, it could be flawed due to overfitting if not properly implemented.

It is important to examine the strength of ANN over other models to see if it corroborates in this case. Also, with the rate of increase in passenger traffic and increasing influence of Aviation on the Nigerian economy; it is imperative that time series models are accurately selected for proper management in the airlines and effective management of the Nigerian Airspace. Many researchers adopt the more common time series approach in predicting air passenger traffic. In some cases salient characteristics in the time series data might not be captured due to the simplicity of the adopted models. However, there is a need to make a comparison of these methods with a more robust approach in modelling the air passenger traffic so as to improve forecast accuracy. A review of existing literature shows that time series models have been used in forecasting air passenger traffic in various destination, however to the best of our knowledge no study has employed the ANN, SARIMA and Holt-winters Exponential Smoothing in modelling the air passenger traffic in MMIA.

The exponential smoothing model (Additive and Multiplicative) and SARIMA methods have been employed effectively in variety of studies. Though the Holt-winters method is simple to apply, it has the disadvantage of being sensitive to structural breaks which are not accounted for. (Gelper *et al.,* 2010). It does not capture non linearity in data set. (Burger *et al.,* 2001) compared various time series models in forecasting the US tourism demand to Durban, they concluded that exponential smoothing model had more forecast accuracy. According to (Zhang 2003), ARIMA and SARIMA models are among the most widely used in the air passenger forecasting literature. Their popularity comes from the fact that they are based on very few assumptions. Furthermore, they are easy to specify with the Box- Jenkins methodology. The SARIMA model has the draw-back of not capturing shocks and non-linear relationship in the variables, unlike ANN.

In this paper, ANN, SARIMA and Holt-winters Exponential Smoothing models were used to evaluate the accuracy of monthly forecast of air passenger traffic flow in MMIA for domestic and international air passenger traffic. Holt-winters and SARIMA models are simpler, more parsimonious in their application, while ANN involves estimation of more parameters which could affect accuracy of forecast in some cases. The performance of ANN was examined in comparison with the selected models from the SARIMA model and the Holt-winters Exponential Smoothing model. This was achieved by using statistical loss function, MAPE and RMSE criteria for the measurement of forecast accuracy.

2. Materials and Methods

In modelling the air passenger traffic flow in MMIA, this study used the time series data set of monthly air passenger traffic flow for both domestic and international flights between January 2003 and December 2015 obtained from the Nigerian Airspace Management Agency. The Eviews9, R version 3.24 and Zaitun statistical packages were used for the analysis.

2.1 Seasonal Auto-Regressive Integrated Moving Average (SARIMA)

SARIMA from Box-Jenkins depends hugely on the concept of stationarity of the time series data which is achieved by the process of differencing. This methodology is concerned with iteratively building a parsimonious model that accurately represents the past and future patterns of a time series (Louvieris, 2002). The general form of an Auto-Regressive Moving Average (ARMA) model is:

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \ , \tag{1}$$

where ϕ_i's $(i = 1..., p)$ are called the autoregressive parameters, θ_i's $(i = 1..., q)$ are called the moving average parameters and ε's are white noise error terms.

The ARMA model above can be expressed in a more simplified form as:

$$\phi(B)y_t = \theta(B)\varepsilon_t , \tag{2}$$

where B represents the backward shift operator or the lag operator, $\phi(B)$ is the autoregressive polynomial of order p and $\theta(B)$ is the moving average polynomial of order q.

$$B^m y_t = y_{t-m} , \tag{3}$$

$$\phi(B) = 1 - \phi_1 B - ... - \phi_p B^p , \tag{4}$$

$$\theta(B) = 1 + \theta_1 B_1 + ... + \theta_q B^q . \tag{5}$$

The ARIMA model, denoted as ARIMA (*p*, *d*, *q*) contains the additional parameter *d*, which differentiates it from the ARMA (*p*, *q*) model. The parameter *d* specifies the level of non-seasonal differencing that is required to render a non-stationary time series stationary.

The generalized form of the ARIMA (*p*,*d*,*q*) model is written as:

$$\phi(B)(1-B)^d y_t = \theta(B)\varepsilon_t <=> \phi(B)\Delta^d y_t = \theta(B)\varepsilon_t \ d \geq 0, \tag{6}$$

where $\Delta^d = (1-B)^d$ denotes the non-seasonal differencing operator. The SARIMA model is a modified ARIMA(*p*,*d*,*q*) model which is denoted as SARIMA (*p*,*d*,*q*) $(P,D,Q)_s$,; *p* represents the number of parameters in the autoregressive model (AR), *d* the degree of differencing, *q* the number of parameters in MA model, *P* the number of parameters in AR seasonal model, *D* the seasonal differencing degree, *Q* the number of parameters in MA seasonal model, and *s* the period of seasonality. This is expressed as follows:

Seasonal autoregressive terms => $\Phi(B^s) = 1 - \Phi_1 B^s - ... - \Phi_p B^{PS}$, (7)

Seasonal moving average terms => $\Theta(B^s) = 1 + \Theta_1 B^s + ... + \Theta_Q B^{QS}$. (8)

This gives the SARIMA expression in equation (9)

SARIMA $(p,d,q)(P,D,Q)_s$

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D y_t = \theta(B)\Theta(B^S)\varepsilon_t \tag{9}$$

$d, D \geq 0$.

Equation (9) is the generalized SARIMA model. The following steps are involved in employing the Box-Jenkins methodology namely; Identification of SARIMA (p,d,q)(P,D,Q) structure, estimating the model parameters, fitting test on the estimated residuals (diagnostic test) and forecasting future outcomes based on historical data.

In selecting the best SARIMA models for the air passenger traffic, the four steps were taken.

2.2 SARIMA model selection

The Box-Jenkins methodology was employed in selection of the appropriate SARIMA models for international and domestic air passenger traffic.

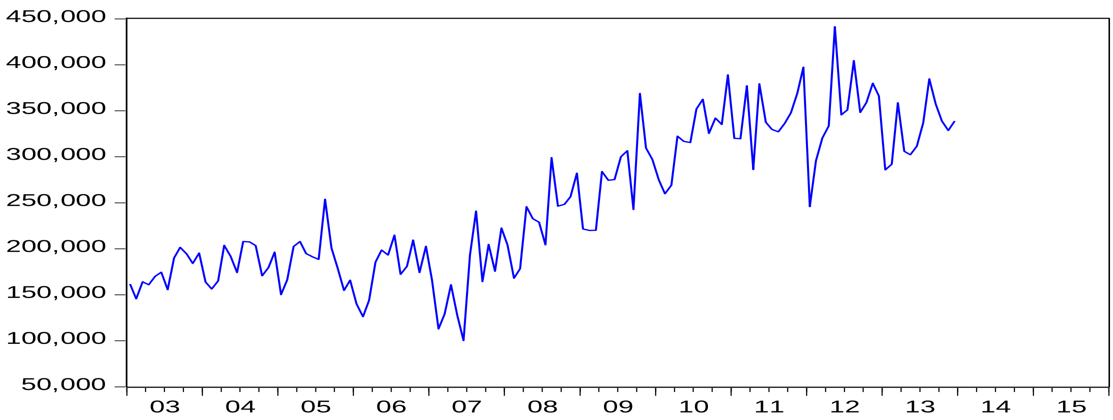

Figure 1: Time plot of Domestic Air passenger Traffic.

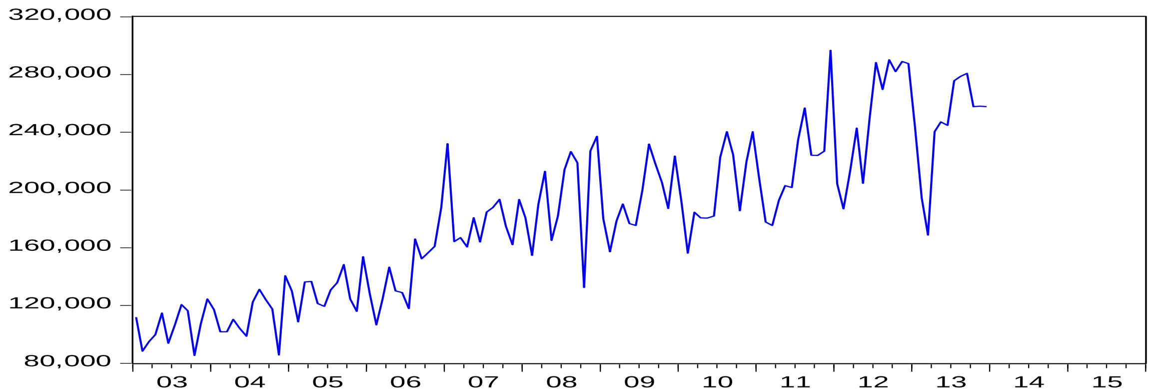

Figure 2: Time plot of international air passenger traffic.

The time plots in figures 1 and 2, reveal non-stationarity, which is common to most time series data. The data were log transformed and attained stationarity at first difference. This was verified using the Augmented Dickey-Fuller (ADF) test, (Engle and Granger, 1987), in this study, the generalized equation is given as:

$$\Delta y_t = \alpha y_{t-1} + \sum_{i=1} \beta \Delta y_{t-1} + \lambda_t + \gamma + \mu_t \qquad (10)$$

$\mu_t \sim$ IID(0, δ^2), where t is time trend $\Delta y_t = y_t - y_{t-1}$, y_t is the natural logarithm at t, α, β, λ_t and γ are the parameters to be estimated and μ_t is the error.

The hypothesis for testing for stationarity using the augmented Dickey Fuller test is given as:

H_0:α = 0 (exhibits unit root or stochastic trend),

H_i:α< 0.

The t-statistic of the parameter $\propto$ is evaluated against a critical value at a level of significance. If the ADF test statistic is greater than the critical value, it is concluded that the series has unit root and is said to be non-stationary. The null hypothesis, H_0: α = 0, is not rejected.

The selection of the best model is based on model selection criteria such as the Akaike Information Criterion (AIC) and Schwarz Bayesian Information Criterion (SBC or BIC) (Hu, 2002). The Ljung-Box portmanteau test was employed for the diagnostic test of the residuals. The null hypothesis tests for absence of residual autocorrelations for h lags. The test statistics is given thus:

$$Q = n(n+2)\sum_{k=1}^{h} \frac{\rho_k^{\ 2}}{n-k}, \qquad (11)$$

$$\rho_k = \frac{\sum_k^n \varepsilon_t \varepsilon_{t-k}}{\sum_{t=1}^{n} \varepsilon_t^{\ 2}}. \qquad (12)$$

ρ_k is the autocorrelation at lag k; $\varepsilon's$ are the residuals of the fitted model; n is the total number of lags considered.

The Ljung-Box rejects the null hypothesis (indicating that the model has significant lack of fit if $Q > \chi^2_{1-\alpha,d}$, where $\chi^2_{1-\alpha,d}$ is the chi-square distribution table value with d degrees of freedom and significance level α. Ljung – Box test was implemented using the statistical Eviews9 package.

Five candidate SARIMA models were considered for domestic and also five SARIMA models for international air passenger traffic. $SARIMA(1,1,1)(0,1,1)_{12}$ and $SARIMA(3,1,1)(2,1,2)_{12}$ were selected using AIC, BIC and RMSE, also considering model parsimony. This was verified by Ljung-Box portmanteau test, for autocorrelation in the residuals of the selected SARIMA models.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.31...	-0.31...	11.922	0.001
		2	-0.24...	-0.38...	19.436	0.000
		3	0.183	-0.05...	23.580	0.000
		4	-0.11...	-0.19...	25.258	0.000
		5	-0.00...	-0.08...	25.259	0.000
		6	0.111	0.001	26.827	0.000
		7	-0.12...	-0.09...	28.850	0.000
		8	0.017	-0.04...	28.888	0.000
		9	0.013	-0.09...	28.909	0.001
		1...	0.114	0.139	30.612	0.001
		1...	0.025	0.137	30.697	0.001
		1...	-0.47...	-0.47...	60.793	0.000
		1...	0.239	-0.16...	68.566	0.000
		1...	0.143	-0.12...	71.376	0.000
		1...	-0.13...	-0.02...	73.739	0.000
		1...	0.137	0.029	76.367	0.000
		1...	0.034	0.137	76.530	0.000
		1...	-0.16...	0.069	80.451	0.000
		1...	0.096	-0.04...	81.768	0.000
		2...	0.045	0.027	82.059	0.000
		2...	-0.12...	-0.05...	84.439	0.000
		2...	0.011	0.016	84.457	0.000
		2...	0.139	0.151	87.352	0.000
		2...	-0.02...	-0.17...	87.479	0.000
		2...	-0.10...	-0.10...	89.034	0.000
		2...	0.011	-0.11...	89.054	0.000
		2...	0.019	-0.00...	89.108	0.000
		2...	-0.03...	0.091	89.321	0.000
		2...	-0.02...	0.010	89.423	0.000
		3...	0.078	0.009	90.405	0.000
		3...	0.022	0.034	90.481	0.000
		3...	-0.07...	-0.04...	91.430	0.000
		3...	0.098	-0.02...	93.032	0.000
		3...	-0.01...	0.063	93.051	0.000
		3...	-0.17...	0.005	98.082	0.000
		3...	0.152	-0.08...	102.08	0.000

Figure 3: ACF and PACF plot of Δ^{12} LDAPT.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.51...	-0.51...	31.932	0.000
		2	0.004	-0.34...	31.934	0.000
		3	0.051	-0.19...	32.261	0.000
		4	-0.05...	-0.17...	32.636	0.000
		5	0.075	-0.04...	33.338	0.000
		6	-0.03...	-0.01...	33.462	0.000
		7	-0.04...	-0.07...	33.766	0.000
		8	0.015	-0.10...	33.794	0.000
		9	0.115	0.091	35.515	0.000
		1...	-0.18...	-0.09...	40.002	0.000
		1...	0.319	0.320	53.565	0.000
		1...	-0.52...	-0.37...	90.892	0.000
		1...	0.268	-0.25...	100.63	0.000
		1...	0.110	-0.08...	102.29	0.000
		1...	-0.13...	-0.02...	104.90	0.000
		1...	0.075	0.035	105.69	0.000
		1...	-0.10...	-0.08...	107.21	0.000
		1...	0.095	-0.02...	108.50	0.000
		1...	-0.02...	-0.05...	108.58	0.000
		2...	0.086	0.138	109.65	0.000
		2...	-0.24...	-0.06...	118.59	0.000
		2...	0.290	0.070	131.06	0.000
		2...	-0.16...	0.123	135.22	0.000
		2...	0.087	-0.13...	136.37	0.000
		2...	-0.00...	-0.05...	136.37	0.000
		2...	-0.13...	-0.03...	139.27	0.000
		2...	0.162	0.044	143.38	0.000
		2...	-0.10...	-0.03...	145.15	0.000
		2...	0.088	-0.03...	146.39	0.000
		3...	-0.08...	0.010	147.41	0.000
		3...	0.057	0.004	147.94	0.000
		3...	-0.10...	0.030	149.89	0.000
		3...	0.175	-0.06...	155.04	0.000
		3...	-0.17...	-0.02...	160.02	0.000
		3...	0.062	0.044	160.68	0.000
		3...	0.030	-0.11...	160.84	0.000

Figure 4: ACF and PACF plot of Δ^{12} LIAPT.

Figures 3 and 4, $\Delta^{12} LDAPT$ and $\Delta^{12} LIAPT$ represent log of seasonally differenced data for domestic and international passenger traffic respectively. The plots show ACF and PACF for both domestic and international air passenger traffic. This gives ample knowledge in identifying the order p, q, P and Q parameters of the proposed SARIMA models by visual inspection. From Figure 3, it is observed that the ACF cuts at q=2 and Q=1 which suggests a moving average parameter of order 2 and a seasonal moving average parameter of order 1. The PACF plot in Figure 3, by visual observation suggests AR parameter of order 2 and a Seasonal AR of order 1. Similarly from Figure 4, by visual inspection, it is observed that the ACF cuts at q=1and Q=1, while the PACF slightly cuts at p=3(AR parameter of order 3) and P = 1

(Seasonal AR parameter of order 1). From this inference the best SARIMA models were selected based on error diagnostics, AIC, BIC, RMSE and also keeping in mind parsimony. Based on this, 5 SARIMA models were considered each for domestic and international air passenger traffic.

2.2 Holt-Winters (HW) exponential smoothing model

The data used in this study consist of trend, seasonal component and white noise. It is however appropriate to apply necessary smoothing technique to model the data used. Exponential Smoothing could be Single Exponential Smoothing, Double Exponential Smoothing or Triple Exponential Smoothing.

The Triple Exponential Smoothing was considered in this study since the time series data has trend, seasonality and expectedly white noise components. There are basically two main HW models namely, additive and multiplicative. In both versions, forecast will depend on the components, level, trend and seasonal co-efficient. The additive version ought to be considered whenever the seasonal pattern of a series has constant amplitude over time (Kalekar, 2004). This is given, equations 16 to 18:

$$y_t = a_t + b_t t + s_t + \varepsilon_t \text{ with } \sum_{t=1}^{t=s} S_t = 0, \tag{13}$$

where a_t represents the level of the series at t, b_t the slope of the series at time t, s_t seasonal coefficients of the series at time t and s the periodicity of the series. ε_t's are error with mean 0 and constant variance.

In the case of the multiplicative seasonal model we have:

$$y_t = (a_t + b_t t)s_t + \varepsilon_t \qquad \sum_{t=1}^{t=s} s_t = s. \tag{14}$$

In this study the multiplicative and additive model were examined, equation (15) is used for forecast.

$$y_{t+h} = (a_t + hb_t)s_{t-12+h}\ , \tag{15}$$

where h represents the forecasting horizon and where a_t, b_t and s_t are estimated with the equations below:

$$a_t = \alpha(\frac{y_t}{s_{t-12}}) + (1-\alpha)(a_{t-1} + b_{t-1}) \tag{16}$$

$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1} \tag{17}$$

$$s_t = \gamma(\frac{y_t}{a_t}) + (1-\gamma)s_{t-12}\,, \qquad (18)$$

where $0 < \alpha < 1, 0 < \beta < 1$, $0 < \gamma < 1$, ∝, β and γ are smoothing parameters for the level, trend and the seasonal component respectively. The closer the parameters to zero indicates less emphasis on recent observation, while the selection of the parameters close to one indicates more emphasis on recent observation. This parameters could be selected subjectively based on adequate knowledge and experience of the time series data. In this paper, the smoothing constants were estimated objectively by minimizing the sum of squared error (SSE) of the forecast. The value that yields the smallest SSE was selected. The values are from zero to one.

2.3 Artificial Neural Network (ANN)

The ANN methodology was inspired in the biological theories on the functions of human brain (McClelland and Rumelhart, 1986). This study employs the Feed Forward Neural Network (FFNN) using the method of back propagation learning algorithm. The ANN has so many advantages over the conventional statistical methods of forecasting, comprising; pattern recognition ability as well as ability to capture non-linearity in the training set.

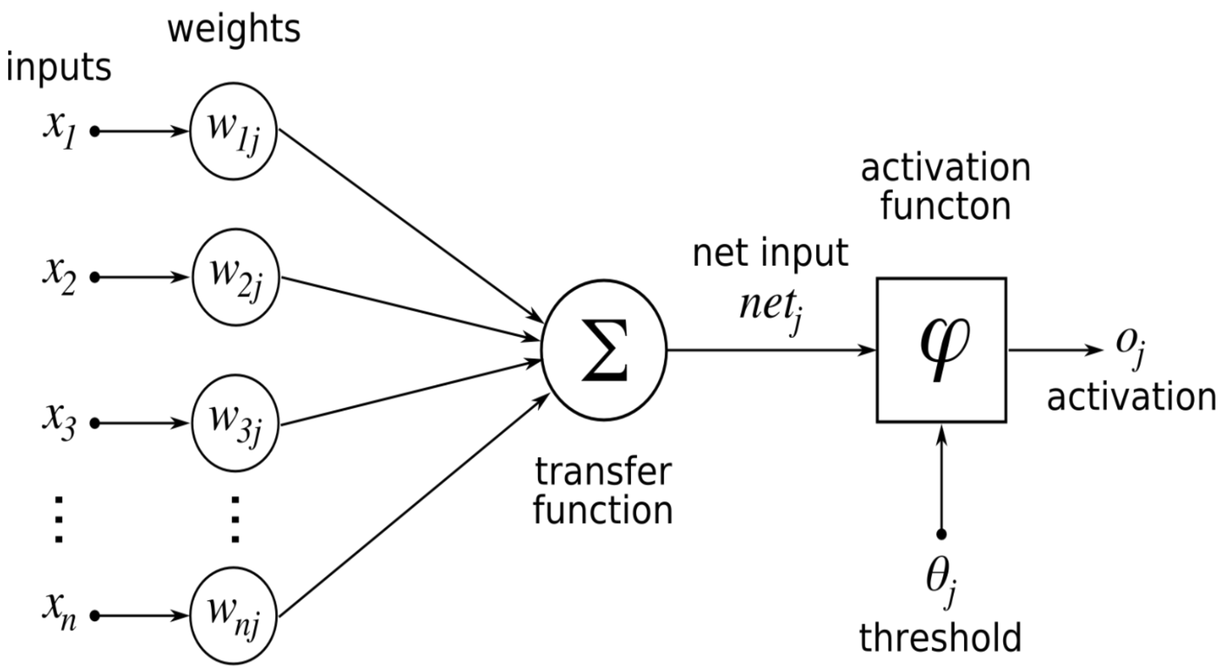

Figure 5: Typical Artificial Neural Network Architecture.

In the architecture of ANN (see Fig. 5), the neurons are connected so that the output from one unit can serve as part of the input to another. Let N be the neurons with outputs $O_j\,(j=1......N)$

Input I to the next neuron is given in equation (19)

$$I_i = W_{i1} * O_1 + W_{i2} * O_2......W_{iN} * O_N = \sum_j W_{ij} * O_j\,. \tag{19}$$

In the back propagation technique, artificial neurons are organized in layers and send their signals forward and then errors are propagated backwards. The network receives inputs by neurons in the input layer and the output of the network is given by the neurons on an output layer. Training a neural network to learn patterns in a given data involves iteratively presenting it with examples of the correct known answers, so as to find the set of weights between the neurons that determine the global minimum of the error function (Kaastra and Boyd, 1996). The back propagation algorithm uses the steepest-descent minimization method for the weight adjustment and threshold coefficient. Implementation of the back propagation algorithm using the sigmoid and bipolar sigmoid functions and weight adjustment are illustrated in the equations below:

$$f_j(\bar{x},\bar{w}) = \sum_{i=0}^{n} x_i w_{ji}\ , \tag{20}$$

where $x_i's$ are the inputs, $w_{ji}'s$ are the respective weights applied. In this case, activation is only dependent on inputs and respective weights. If the output function is equal to the activation, the neuron is called linear. In such a situation we are only trying to make a straight line fit. However a combination of the linear activation function with sigmoid was employed in this study. There are different types of sigmoidal functions such as, the hyperbolic tangent, logistic sigmoid, bipolar sigmoid function and the Elliot transfer function. For this study, we will consider the logistic sigmoid function bounded between 0 and 1 and its linearly transformed version, the bipolar sigmoid function bounded between -1 and 1, the one that models the series best will be adopted.

Equations (21) and (22) represent the sigmoid and bipolar sigmoid functions respectively.

$$y_j(\bar{x},\bar{w}) = \frac{1}{1+e^{f_j(\bar{x},\bar{w})}}\ , \tag{21}$$

$$y_j(\bar{x},\bar{w}) = \frac{2}{1+e^{f_j(\bar{x},\bar{w})}} - 1 \Rightarrow \frac{1-e^{f_j(\bar{x},\bar{w})}}{1+e^{f_j(\bar{x},\bar{w})}}\,. \tag{22}$$

The sigmoid function has non-linear characteristics, monotonically increasing and continuous differentiable. For the sigmoid activation function, $y_j(\bar{x},\bar{w})$ we take the range $[0,1]$, this implies that; as $x_i \to \infty$, $y_j(\bar{x},\bar{w}) = 1$ and also as $x_i \to -\infty$, $y_j(\bar{x},\bar{w}) = 0$. The bipolar sigmoid function takes values within the range $[-1,1]$.

In the training process, the weights are adjusted in such a way that the error is minimized. This is achieved by observing the difference between the actual output and the desired output. The error function of each neuron can be defined thus:

$$E_j(\bar{x},\bar{w},d) = (y_j(\bar{x},\bar{w}) - d_j)^2 \quad . \tag{23}$$

The equation (23) gives the error for the individual neuron. The collective error for the entire network is the sum of the individual error of the neurons. This is indicated below in equation (24):

$$E(\bar{x},\bar{w},\bar{d}) = \sum_j (y_j(\bar{x},\bar{w}) - d_j)^2 \ . \tag{24}$$

In order to adjust the weights appropriately the back propagation will calculate how the errors depend on the outputs, inputs and weights. Thereafter we employ an appropriate method, gradient descendent, to adjust the weight to meet the required target. The gradient descendent method is indicated below:

$$\Delta w_{ji} = -\lambda \frac{\partial E}{\partial w_{ji}}, \tag{25}$$

where, Δw_{ji} is the adjustment of each weight, λ is a constant, which is the learning rate, the success of the neural network convergence depends a lot on the learning rate. This is multiplied by the dependence of the previous weight on the error network, the derivative of E in respect to w_{ji}.The equation (25) will be implemented until the desired weight with minimal error is obtained. In implementing the back propagation algorithm it is important to know how much the error depends on the output. This is expressed below in equation (26):

$$\frac{\partial E}{\partial y_j} = 2(y_j - d_j). \tag{26}$$

It is also important to calculate how much the output depends on the activation, which in turn depends on the weights from equations (20) and (21) or (22) depending on the activation function being used.

$$\frac{\partial y_j}{\partial w_{ji}} = \frac{2e^{-f_j}}{1+e^{-f_j}} - 1 \Rightarrow \frac{(1-e^{-f_j})}{(1+e^{-f_j})}$$

$$= 2(1+e^{-f_j})^2 e^{-f_j}$$

$$= \frac{1}{2}\frac{(4e^{-f_j})}{(1-e^{-f_j})^2}$$

$$= \frac{1}{2}\left[\frac{(1+e^{-f_j})^2-(1-e^{-f_j})^2}{(1+e^{-f_j})^2}\right]$$

$$= \frac{1}{2}\left[\frac{(1+e^{-f_j})^2}{(1+e^{-f_j})^2} - \frac{(1-e^{-f_j})^2}{(1+e^{-f_j})^2}\right]$$

$$= \frac{1}{2}\left[1-\frac{(1-e^{-f_j})^2}{(1+e^{-f_j})^2}\right]$$

$$= \frac{1}{2}\left[1-y_j^{\,2}\right] \tag{27}$$

Equation (27) shows the derivative of the bipolar sigmoid function. The derivative of the logistic sigmoid function is simply given as:

$$\frac{\partial y_j}{\partial w_{ji}} = \frac{\partial\left(\frac{1}{1+e^{f_j}}\right)}{\partial w_{ji}} = y_j(1-y_j)\,, \tag{28}$$

$$\frac{\partial y_j}{\partial w_{ji}} = \frac{\partial y_j}{\partial f_j}\times\frac{\partial f_j}{\partial w_{ji}} = \frac{1}{2}\left[1-y_j^{\,2}\right]x_i\,. \tag{29}$$

From equations (26) and (27) we have the resulting equation (30) below, using the bipolar activation function.

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial y_j}\times\frac{\partial y_j}{\partial w_{ji}} = 2(y_j-d_j)\frac{1}{2}(1-y_j^{\,2})x_i = (y_j-d_j)(1-y_j^{\,2})x_i \tag{30}$$

Therefore:

$$\Delta w_{ji} = -\lambda(y_j-d_j)(1-y_j^{\,2})x_i\,, \tag{31}$$

when applying the logistic sigmoid function the change of weight is given as equation (32):

$$\Delta w_{ji} = -2\lambda(y_j-d_j)y_j(1-y_j)x_i\,. \tag{32}$$

The equations (31) and (32) can be used for ANN with two layers using the bipolar sigmoid function and logistic sigmoid function respectively. However, the iteration is implemented continuously for several layers of the neural network as obtainable in the multi-layer feed forward neural network.

2.4 Performance Evaluation of the Models

In evaluating the forecast performance accuracy in each of the models, the training set of data and a test set were used. The training data set contains data from 2003 to 2013, a total of 132 monthly observations in each case, domestic and international air passenger traffic. The test data set contains data from 2014 to 2015, a total of 24 monthly observations in each case. The static out-of-sample forecast is employed in this study. One step ahead forecast, the forecast for the month of January 2014, is made by using the training data set, this estimate is added to the training data set which in turn is used to make forecast for the month of February 2014. The forecast estimate for the month of February 2014 is added to the new training set so as to obtain forecast for March 2014.

The iteration continues until the estimate for the last month, December 2015. This estimates from the static forecast are compared with the actual test data set, using statistical loss functions. Based on this, the values of the MAPE and RMSE for the respective proposed models are compared so as to select the model that has the best out-of-sample performance for the forecast horizon.

3 Results and Discussion

Table 1: Postulated SARIMA models for Domestic Air Passenger Traffic.

Model	**AIC**	**BIC**	**RMSE**	**Normality Test**		**Serial Correlation**	
				JB Test	P-value	Q-statistic	P-value
$SARIMA(2,1,1)(0,1,1)_{12}$	−1.222	−1.104	0.129	2.269	0.322	26.027	0.762
$SARIMA(2,1,1)(1,1,1)_{12}$	−1.157	−1.004	0.132	1.066	0.587	23.667	0.587
$SARIMA(1,1,1)(1,1,1)_{12}$	−1.123	−0.997	0.135	1.587	0.452	34.654	0.342
$SARIMA(1,1,1)(0,1,1)_{12}$	−1.203	−1.109	0.130	2.102	0.350	32.940	0.470
$SARIMA(1,1,2)(1,1,1)_{12}$	−1.135	−0.984	0.133	1.172	0.557	29.161	0.561

From Table 1, considering the models performance in terms of AIC, BIC and RMSE $SARIMA(2,1,1)(0,1,1)_{12}$ and $SARIMA(1,1,1)(0,1,1)_{12}$ with lower AIC and BIC values performed competitively better than the other SARIMA models. The L-jung Box portmanteau test for residual autocorrelation shows that the residuals are not serially correlated and the Jarque-Bera normality test confirms that the residuals are normal. Both models should be adequate in modelling the domestic air passenger traffic. The $SARIMA(1,1,1)(0,1,1)_{12}$ is the selected model since it is more parsimonious.

Table 2: Postulated SARIMA models for International Air Passenger Traffic.

Model	**AIC**	**BIC**	**RMSE**	**Normality Test**		**Serial Correlation**	
				JB Test	P-value	Q-statistic	P-value
$SARIMA(1,1,2)(1,1,1)_{12}$	−1.558	−1.408	0.108	32.600	0.000	23.989	0.811
$SARIMA(3,1,1)(0,1,2)_{12}$	−1.527	−1.361	0.110	13.923	0.009	27.100	0.618
$SARIMA(2,1,2)(2,1,2)_{12}$	−1.841	−1.596	0.092	1.302	0.522	31.495	0.296
$SARIMA(1,1,2)(0,1,1)_{12}$	−1.553	−1.435	0.109	27.881	0.000	25.362	0.791
$SARIMA(3,1,1)(2,1,2)_{12}$	−1.968	−1.721	0.086	4.352	0.114	16.820	0.952

From Table 2, $SARIMA(3,1,1)(2,1,2)_{12}$ was selected to be the best model for the international air passenger traffic. Compared to the other proposed SARIMA models, $SARIMA(3,1,1)(2,1,2)_{12}$ has the lowest AIC and BIC, the errors are not serially correlated, errors are normally distributed. The $SARIMA(2,1,2)(2,1,2)_{12}$ is the SARIMA model that performs closely to the selected model $SARIMA(3,1,1)(2,1,2)_{12}$ but the selected model performs generally better. The other models fail the Jarque-Bera test for normality.

Table 3 presents the estimated parameters of the seasonal models for both domestic and international air passenger traffic. The diagnostics checks performed indicate that the models should be adequate in modelling the air passenger traffic. The plots of the selected SARIMA models for the air passenger traffic are shown in figures 6 and 7 which indicate good fits. The performances of these models were evaluated in comparison with the Holt-winters Exponential smoothing and the performance of the artificial neural network.

Table 3: Parameter of Estimated SARIMA Models.

Domestic air passenger traffic $SARIMA(1,1,1)(0,1,1)_{12}$			**International Air passenger traffic** $SARIMA(3,1,1)(2,1,2)_{12}$		
Variables	**Estimates**	**P-value**	**Variables**	**Estimates**	**P-value**
$AR(1)$	0.133	0.364	$AR(1)$	-0.390	0.085
$MA(1)$	-0.709	0.000	$AR(2)$	-0.324	0.079
$SMA(12)$	0.906	0.000	$AR(3)$	-0.176	0.242
JB Test	2.102	0.350	$SAR(12)$	-0.195	0.029
Q-Test	32.940	0.470	$SAR(24)$	-0.275	0.007
AIC	-1.203		$MA(1)$	-0.459	0.040
BIC	-1.109		$SMA(12)$	-0.665	0.000
			$SMA(24)$	0.872	0.000
			JB Test	4.352	0.113
			Q-Test	16.820	0.952
			AIC	-1.968	
			BIC	-1.721	

Estimates of the parameters in the Table 4 show that the additive Holt-winters model is best suited for modelling the air passenger traffic for both domestic and the international flight based on AIC and BIC criteria. This is justified, given that the data were log transformed hence making the components additive.

In estimating the ANN model for the air passenger traffic flow, Feed Forward New Network (FFNN) implemented by the back propagation algorithm was used. 4 models, with 3 layers (input layer, hidden layer and output) as expected of most standard neural network architecture were considered. The models with 12 neurons in the input layer, 4 neurons in the hidden layer and one neuron in the output layer (12-4-1), 12 neurons in the input layer, 6 neurons in the hidden layer and 1 neuron in the output layer (12-6-1), 12 neurons in the input layer, 8 neurons in the hidden layer and one neuron in the output layer (12-8-1) and (12-12-1) network were examined.

The Sigmoid and Bipolar Sigmoid activation functions were used in the learning process so as to examine which of these activation functions models the time series data better by observing the mean squared error yielded in each process. In Table 5 it was observed that in the 4 ANN models that were considered, the best results were obtained when the bipolar sigmoid activation function was used. This was deduced by the measure of accuracy. In each of the cases better results were obtained with the increase in number of neurons in the hidden layer, with the model 12-12-1 using bipolar sigmoid function yielding the best results based on the in-sample measure of error. However, the ANN 12-12-1 model for the international air passenger traffic was not as effective, when checking the out-sample performance, 0.32836RMSE compared with the other ANN models with fewer neurons in the hidden layer. ANN 12-4-1 tends to perform averagely better with better out-sample, forecast performance, $RMSE = 0.16185$. For the domestic air passenger traffic, ANN 12-4-1 which has fewer neurons (fewer parameters) and sufficiently good forecast accuracy based on the measure of error would be employed in comparison with the best SARIMA and HWES model. The input neurons are the lagged series, $y_{t-1,} y_{t-2} .... y_{t-12}$ of the time series data, these are weighted randomly and linearly combined and the results are modified by the bipolar sigmoid function to serve as input to the next layer.

Table 4: Estimates of the Holt-Winters exponential smoothing parameters.

Variables		**Model**	α **(Level)**	β **(Trend)**	γ **(Seasonal)**	**BIC**	**AIC**
Domestic		Multiplicative	0.2174	0.0000	0.2032	192.4516	146.3267
Domestic		Additive	0.3300	0.0000	0.0335	166.4400	120.3152
International		Multiplicative	0.3034	0.0253	0.2447	140.8886	94.7637
International		Additive	0.2676	0.0000	0.000	116.9094	70.7846

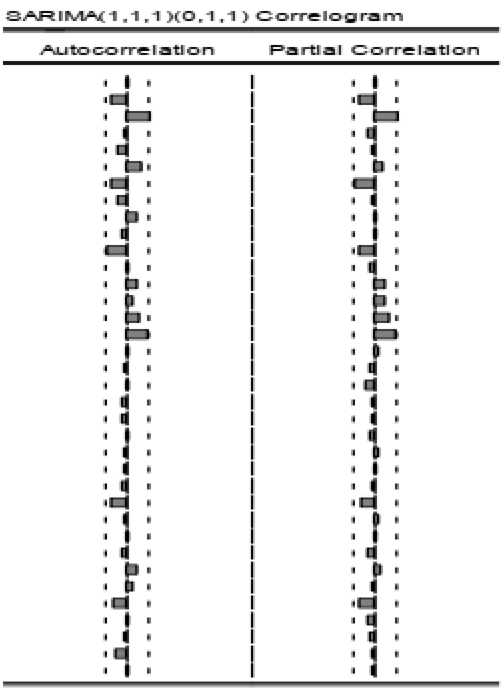

Figure 6: fitted model for Domestic.

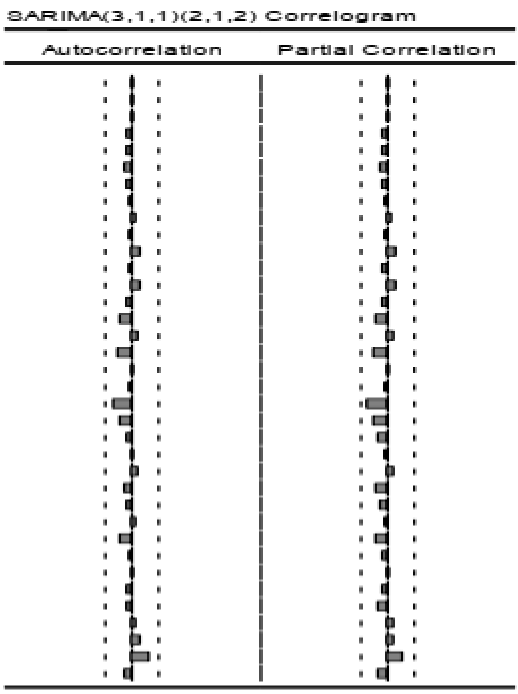

Figure 7: fitted model for international

Table 5: The ANN models for Domestic and International Air passenger Traffic.

DOMESTIC				
Model	**Activation function**	**Iterations**	**MAE**	**MSE**
12-4-1	Sigmoid	20000	0.10098	0.01740
	Bipolar Sigmoid	30000	0.05980	0.00645
12-6-1	Sigmoid	20000	0.10135	0.01754
	Bipolar sigmoid	30000	0.04687	0.00410
12-8-1	Sigmoid	20000	0.10015	0.01733
	Bipolar Sigmoid	30000	0.04020	0.00324
12-12-1	Sigmoid	20000	0.10049	0.01724
	Bipolar Sigmoid	40000	0.01689	0.00079
INTERNATIONAL				
Model	**Activation function**	**Iterations**	**MAE**	**MSE**
12-4-1	Sigmoid	20000	0.08080	0.01154
	Bipolar Sigmoid	20000	0.05694	0.00538
12-6-1	Sigmoid	20000	0.08173	0.01177
	Bipolar sigmoid	20000	0.03962	0.00277
12-8-1	Sigmoid	20000	0.07998	0.01143
	Bipolar Sigmoid	30000	0.02933	0.00160
12-12-1	Sigmoid	20000	0.07994	0.01150
	Bipolar Sigmoid	30000	0.01028	0.00028

The Sigmoid and Bipolar Sigmoid activation functions were used in the learning process so as to examine which of these activation functions models the time series data better by observing the mean squared error yielded in each process. In Table 5 it was observed that in the 4 ANN models that were considered, the best results were obtained when the bipolar sigmoid activation function was used. This was deduced by the measure of accuracy. In each of the cases better results were obtained with the increase in number of neurons in the hidden layer, with the model 12-12-1 using bipolar sigmoid function yielding the best results based on the in-sample measure of error. However, the ANN 12-12-1 model for the international air passenger traffic was not as effective, when checking the out-sample performance, 0.32836RMSE compared with the other ANN models with fewer neurons in the hidden layer.

ANN 12-4-1 tends to perform averagely better with better out-sample, forecast performance, $RMSE = 0.16185$. For the domestic air passenger traffic, ANN 12-4-1 which has fewer neurons (fewer parameters) and sufficiently good forecast accuracy based on the measure of error would be employed in comparison with the best SARIMA and HWES model.

The input neurons are the lagged series, $y_{t-1,} y_{t-2} .... y_{t-12}$ of the time series data, these are weighted randomly and linearly combined and the results are modified by the bipolar sigmoid function to serve as input to the next layer.

3.1 Performance comparison of the models

Table 6: Performance comparisons of models for air passenger traffic.

Domestic air traffic passenger				
Model	**In-Sample**		**Out-Sample**	
	MAE	**RMSE**	**MAPE**	**RMSE**
SARIMA $(1,1,1)(0,1,1)_{12}$	0.0983	0.1304	0.5291	0.0808
Holt-Winters	0.09183	0.12162	0.6000	0.0932
ANN $(12-4-1)$	0.05896	0.00803	0.40918	0.06364
International air passenger traffic				
Model	**In-Sample**		**Out-Sample**	
	MAE	**RMSE**	**MAPE**	**RMSE**
SARIMA $(3,1,1)(2,1,2)_{12}$	0.0632	0.0864	0.9605	0.1453
Holt-Winters	0.07291	0.10081	0.6435	0.0979
ANN $(12-4-1)$	0.05694	0.07335	0.98190	0.16185

Table 6 gives the empirical results of the forecasting performance of the models, ANN, SARIMA and Holt-Winters. All the examined models produced good forecast in the sectors considered, since the MAPE and RMSE are generally low. In modelling the air passenger traffic for both domestic and international flights, the ANN models performed very well in the in-sample forecast. ANN $(12-4-1)$ outperforms the Holt-Winters and SARIMA models for in-sample and out-sample forecast in the domestic sector. It is observed that the values of MAE, MAPE and RMSE are smaller than the other models.

ANN $(12-4-1)$ model for international air passenger traffic, by comparing the MAPE, MAE and RMSE has a better in-sample performance than the other models but has the least out-sample forecast performance. The effectiveness of the ANN models in this study corroborates emphases made on the flexibility and excellent function approximation capability of the ANN in previous studies by (White, 1989). Though the ANN $(12-4-1)$, in modeling the international air passenger traffic, has the least out sample performance, observing the MAE and RMSE for the training set it is evident that the ANN has a great learning and pattern recognition ability. From the time plot of the international air traffic passenger, the exhibition of a more stable characteristics was obtained compared to the time plot for domestic air passenger traffic. This could be the reason for the better performance of the ANN in modelling the domestic air passenger traffic, since the ANN does very well in capturing some hidden salient characteristics and non-linearity in data.

The Holt-Winters Exponential Smoothing model shows better post forecast accuracy, for international air passenger traffic, than SARIMA $(3,1,1)(2,1,2)_{12}$ and ANN, based on the MAE, RMSE and MAPE. In modelling the domestic air passenger traffic, the ANN had the best in-sample and out-of-sample performance, though SARIMA $(1,1,1)(0,1,1)_{12}$ performed competitively with ANN.

4 Conclusion

Three time series models were considered, the seasonal auto-regressive integrated moving average (SARIMA), Holt-Winters Exponential Smoothing model and the Artificial Neural Network. These time series models were tested on the international air passenger and domestic traffic, to evaluate their performances and test the predictive capability of the artificial neural network relative to the other models.

The models were estimated on the training data set which covers the period from January 2003 to December 2013. The test set which covers the period from January 2014 to 2015 December was used to obtain the forecast accuracy measure of these models based on RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error). Empirical results show that all the models provide good forecasts of the air passenger traffic for international and domestic.

Comparing results across the models, it was observed that no model completely outperforms the other in all the sectors. However, the ANN model was found to be very efficient and had the best in-sample performance across the two sectors.

In modelling the domestic air passenger, the ANN model was seen to be significantly dominant, while for the international air passenger traffic the ANN also gave the best in-sample accuracy performance but the least out-sample performance. The Holt-Winters exponential smoothing and SARIMA both yielded good results. The Holt-Winters exponential smoothing out-performed the SARIMA and ANN for the out-of-sample forecast of international air passenger traffic while the SARIMA was more dominant than Holt-Winters in the domestic sector.

Conclusively, for future research, modelling the International air passenger traffic, which appears to have a simpler time series, ANN had the least out-sample forecast accuracy. This implies that complex models may not necessarily yield the best forecast accuracy in modelling simple time series data. This could be due to overfitting which may end up yielding out of sample forecast which is not too accurate. This study has been able to establish the effectiveness of the ANN in modeling and forecasting time series data and also select time series models that gave fairly accurate forecast of air passenger traffic in Murtala Muhammed International Airport Lagos, Nigeria.

Acknowledgement

I am very thankful to God for the provision of sustenance throughout the period of this work. Great appreciation to Dr. A. Yahaya for the immense effort, contribution and being an integral part of this research. I wish to acknowledge the contributions of the lecturers of the Department of Statistics, Ahmadu Bello University (ABU), Zaria. My deepest thanks to my wife, Mrs. Temitope Omolohunnu and our kids Tomisin and Toyosi for their understanding, motivation and support always. I give immense appreciation to my parents, siblings and friends for being there for me. God bless you all.

References

Ajibode, I. A. (2016): SARIMA model and forecasting of enplaned passenger' traffic in Murtala Mohammed International Airport. *Presented at 13th National conference of Academic staff union of polytechnic (ASUP)*. Kano, Nigeria.

Burger, C. J. S. C., Dohnal, M., Kathrada, M. and Law, R. (2001): A practitioners guide to time- series methods for tourism demand forecasting—a case study of Durban, South Africa. *Tourism management*. **22** (4), 403-409.

Doguwa, S. I. and Alade, S. O. (2015): On time series modeling of Nigeria's external reserves. *CBN Journal of Applied Statistics*. **6**(1), 1-28.

Emiray, E. and Rodriguez, G. (2003): *Evaluating Time Series Models in Short and Long-Term Forecasting of Canadian Air Passenger Data*. Department of Economics, University of Ottawa.

Engle, R. F. and Granger, C. W. (1987): Co-integration and error correction: representation, estimation, and testing. *Econometrica*. **55** (2), 251-276.

Gelper, S., Fried, R. and Croux, C. (2010): Robust forecasting with exponential and Holt–winters smoothing. *Journal of forecasting*. **29** (3), 285-300.

Hornik, K., Stinchcombe, M. and White, H. (1989): Multilayer feedforward networks are universal approximators. *Neural networks*. **2**(5), 359-366.

Hu, L. (2002): Estimation of a censored dynamic panel data model. *Econometrica*. **70** (6), 2499-2517.

IATA (2013): www.iata.org/pressroom/Pr/pages/2013-12-10-01.

Kaastra, I. and Boyd, M. (1996): Designing a neural network for forecasting financial and economic time series. *Neurocomputing*. **10** (3), 215-236.

Kalekar, P. S. (2004): Time series forecasting using holt-winters exponential smoothing. *Kanwal Rekhi School of Information Technology*.

Louvieris, P. (2002): Forecasting international tourism demand for Greece: A contingency approach. *Journal of Travel & Tourism Marketing*. **13**(1-2), 21-40.

McClelland, J. L. and Rumelhart, D. E. (1986, January): A distributed model of human learning and memory. In *Parallel distributed processing* (pp. 170-215). Mit Press.

Shen, S., Li, G. and Song, H. (2011): Combination forecasts of international tourism demand. *Annals of Tourism Research.* **38** (1), 72–89.

White, H. (1989): Neural-network learning and statistics. *AI expert*. **4** (12), 48-52.

Zhang, G. (2003): Time series forecasting using a hybrid arima and neural network model. *Neurocomputing* 50: 159-175.