**SARIMA MODEL AND FORECASTING OF ENPLANED PASSENGERS’ TRAFFIC IN MURITALA MOHAMMED INTERNATIONAL AIRPORT**

\*Ajibode, I.A.

Department of Mathematics and Statistics

Federal Polytechnic, Ilaro, Ogun State

\*ilesanmiajibode@yahoo.com

Abstract

*Airlines and Airports generate their revenue from the number of passengers enplane. Enplane passengers refers to persons boarding an airplane. In order to achieve these, it is necessary to put in place a model and forecasting mechanism that will help the administrators of the Airports and Airlines in their planning. Hence, this research work used Seasonal Autoregressive Integrated Moving Average (SARIMA) to model the enplane in Muritala International Airport with monthly data that spans from 2006-2015. R package was used for the analysis. Based on the time plot of the data, ACF and PACF it was discovered that the data is non-stationary and it also exhibit seasonality which necessitate the series to be differenced to attain stationarity. The deseasonalised stationary series was modeled in order to determine the stability of the parameter estimation. Augmented Dickey Fuller test shows that the two differenced data were stationary. The ACF and PACF plots of the ordinary and seasonally differenced series suggested some models for selection but the AICc was used to select the model that really provided the best fit for the time series. From various seasonal models generated with R-console, SARIMA (2,1,1) (0,1,1)12 model was found to best fit enplaned passengers traffic. Ljung-Box test and Shapiro test showed that the residuals of the selected model is independent and normally distributed. The model was used for a short term forecast (2016-2026).*

Keywords:SARIMA, AIC, BIC, AICc, Enplane, Seasonality, Stationary, model.

**1. INTRODUCTION**

There is an increase in patronage of air travel in recent years, and this trend is believed to continue for years ahead (Civil Aviation Authority, 2007). This increase has also affected the number of passengers enplaned in Nigeria. The rise will have direct effect on resource allocation, airport planning, flight schedule and revenue generation.

All over the world, airlines and airports administrators based their planning on passengers’ traffic. Transportation dictionary define enplaned passengers as the total number of revenue passengers boarding aircraft.

Nigeria as one of the developing nations, it is believed that the economic condition will not encourage movement of passengers within and out of the country. But with the predicted upward trend, it is evidenced that Nigeria will also have its own share of passengers traffic in near future.

Looking at the tendency for this rise, there is a need to develop a model that will help in predicting enplane passengers traffic that will aids planning, hence help in estimating revenue generation.

The international airport at Ikeja in Lagos was constructed during [World War II](https://en.wikipedia.org/wiki/World_War_II), though formerly known as Lagos International Airport, it was renamed in 1970s, during construction of the new international terminal, after a former Nigerian military head of state [Murtala Muhammed](https://en.wikipedia.org/wiki/Murtala_Muhammed). The airport is the main base for major local and international carrier airlines, and it serves the entire South Western Nigerian. According to Wikipedia (2016), the airport is the fifth busiest international airport in Africa and third best in Africa as at 2014 with estimated average of 6.3 million passengers per annum.

The continue demand for the services of the airport and based on the average annual passengers traffic, it is important to get a time series model that can best be used for forecast future enplaned passengers’ traffic for the airport. This is very important because of the economic implication of delay in boarding plane and it will also help the management of the airport and the airlines in providing necessary facilities and in estimation of revenue.

Due to the seasonal nature of air passengers’ data, it becomes imperative to utilize good modeling techniques that will assist the airport managers in planning for short and long time periods.

A time series model technique that can be used for seasonality in data set is seasonal autoregressive integrated moving average (SARIMA). This model will assist in developing an adequate model for the ever growing demand for aviation services in Nigeria which in turn will enhances effective air traffic flow management in Muritala Muhammed International Airport.

**2 REVIEW OF RELATED LITERATURE**

Melville (1998), Karlaftis and Papastavrou (1998), Abed, Ba-Fail and Jasimuddin (2001), Ling Lai and Li Lu (2005) affirmed that time-series were widely used in modeling and forecasting of air transport. Grubb and Mason (2001) also support this researcher’s opinion. In the work of Yves and Sandrine (2008) to estimate air traffic loss at Toulouse Blagnac airport, it was concluded that ARIMA models performs better than structural time series model.

In comparison of six different time series model (including SARIMA), it was concluded that the forecasting performance of the models varies widely across series and forecast horizons (Emir and Gabriel 2003). But they still give similar results compared to former works done by the ministry of transport in Canada.

Seasonal ARIMA was employed by Williams and Hoel (2003), Tan et al (2009) to model and predict the international air passenger flow.

In the forecast of inbound tourism demand to Instanbul by Murat (2014). The study was a comparative one between SARIMA and seasonal exponential smoothing models. The researcher affirmed that SARIMA showed best forecast accuracy with lowest deviation (Mean Absolute Percentage Error of 3.42) among the two models.

SARIMA model of (1,1,0) x (1,1,1)12 was proposed by Ette (2013) as adequate for Nigerian monthly air traffic data.

**3. MATERIALS AND METHODS**

Box and Jenkins method called Seasonal Autoregressive Integrated Moving Average (SARIMA) is used in this study. This technique utilizes historical data which involves the Autoregressive (AR) process that takes into account past values, the Moving Average (MA) process which takes into cognizance previous error terms and the integrated part (I). The method is most suitable for seasonal data.

The data is historical monthly data for international passengers’ enplane traffic for Muritala Mohammed International Airport, Lagos, Nigeria, from 2006-2015. The data was extracted from the archive of FAAN official statistics data.

**3.1 SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (SARIMA)**

Box and Jenkins (1970) have generalized a non-seasonal model (ARIMA) to deal with seasonality. The proposed model is called Seasonal ARIMA (SARIMA) model. This model use seasonal differencing of appropriate order to remove seasonality from the series, hence make it non-seasonal.

If d and D are nonnegative integers, then is a multiplicative Seasonal ARIMA model (p, d, q) x (P, D, Q)s process with period s, which can be written as:

(1)

Where:

is white noise with mean zero and variance

(2)

(3)

(4)

(5)

The process is causal if and only if and for. Also, in applications D is rarely more than one and P and Q are always less than three.

It is not reasonable to assume that the seasonal component repeats itself precisely in the same way cycle after cycle. It allows for randomness in the seasonal pattern from one cycle to the next, (Hamilton, 1994).

**3.2 MODEL IDENTIFICATION**

The goodness of fit statistics that are most commonly used for the model selection are AIC, AICc (Corrected Akaike’s Information Critaria) and BIC. The parameters are determined based on likelihood function.

AIC = (6)

This case suggests that an AIC value for each model with the same data set can be calculated, and the “best” model is the one with minimum AIC value, Akaike (1974).

Hurvich and Tsai (1989) showed that the bias in AIC can be approximately eliminated by another non-static penalty term to the AIC, resulting in the corrected AIC, denoted by AICc and defined by the formula:

(7)

The BIC like AIC, uses the maximum likelihood method. BIC is defined as:

(8)

Where,

k = (p + q + P + Q + d + s)

is the sum of square errors, n is the number of observations, p is the order of the non-seasonal part of the (AR) process, q is the order of the non-seasonal part of the (MA) process, P is the order of the seasonal part of the (AR) process, Q is the order of the seasonal part of the (MA) process, d and s represent the differencing order and the seasonal value.

The Shapiro-Wilk tests the null hypothesis that a sample came from a normal distributed population.

The test statistic is:

(9)

The null hypothesis of this test is that the population is normally distributed. Thus, if the p-value is less than the chosen alpha level, then the null hypothesis is rejected. Otherwise, the data is normal, Shapiro and Wilk (1965).

The modified Box-Pierce or Ljung-Box statistic tests for auto-correlation error and it is given by:

(10)

Where: n is the sample size

is the sample autocorrelation at lag k. For significance level α, the critical region of the hypothesis of randomness is where is the α Quartile of the chi-squared distribution with h degrees of freedom.

**3.3 DIAGNOSTIC CHECKING**

The first diagnostic check is to inspect a plot of residuals over time. The Shapiro-Wilk normality test and Ljung-Box test of independence can be applied to the residuals to produce a test statistic which determine whether to reject normality and dependence based on this test.

After a model is identified, the next stage of the SARIMA model building is to estimate the necessary parameters. Maximum likelihood estimation method was used for estimating the parameters with R-package.

**4. RESULTS ANALYSIS**

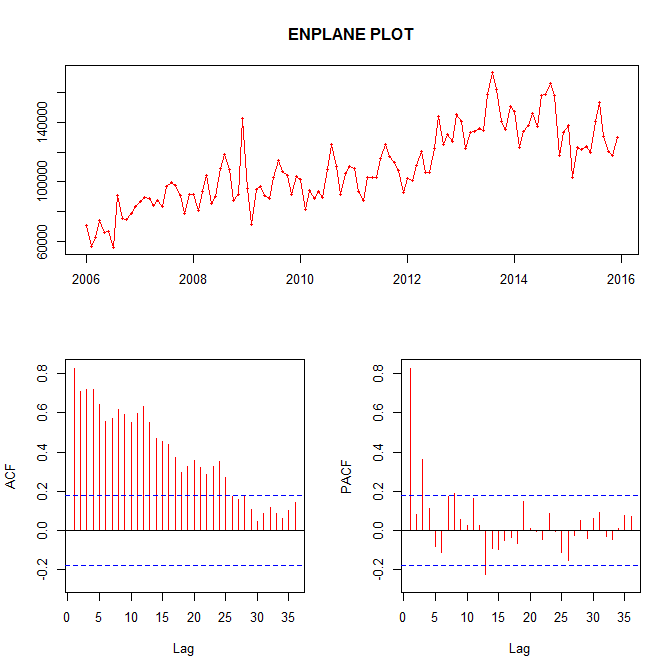
**4.1 DATA ON ENPLANED PASSENGERS’ TRAFFIC**

Table 1 presents the summary statistics result of the data for Enplane data from January 2006 to December 2015. Averagely, 109700 person being enplane monthly at Muritala Muhammed International Airport. The highest number of persons or passengers on monthly basis is 173800. The minimum so far (within the last 10 years) on monthly basis is 55630 persons.

*Table 1: Summary of Enplane Data*

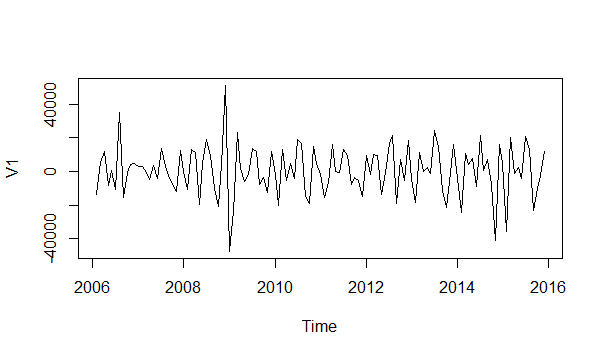
|  |  |  |  |
| --- | --- | --- | --- |
| Minimum | Maximum | Mean | Median |
| 55630 | 173800 | 109700 | 106300 |

From Figure 1 below, it was observed that the pattern of the graph indicates series non-stationarity. There is both upward and downward trend as well as a little seasonal variation which also show that the series is stochastic in nature. The autocorrelation plot indicates significant spikes up to lag 25 but not much of seasonal variation, a downward trend from lag to lag and slight cut off from lag 26 which also indicates an element of non-stationarity. The partial autocorrelation also tails of after lag 2 with a significant spike at lag 13.



*Figure 1: Monthly number of persons being enplane in Muritala Mohammed International Airport*

As a result of this, stationarity was achieved by applying the method of first order and seasonal differencing as shown in Figure 2.



*Figure 2: Plot of 1st seasonal differenced enplane data*

The stationary effect is evidenced in Figure 3 which show the ACF and PACF of the differenced data.



*Figure 3: ACF and PACF of 1st seasonal differenced enplane data*

Since the Dickey-Fuller test statistic is -13.28 and the p-value is 0.0000, we therefore fail to accept H0 and hence conclude that the alternative hypothesis is true; that is, the series is stationary in its mean and variance.

The orders were chosen because of the significant spikes from the ACF and PACF plots. The ACF shows significant spikes in lag1, lag 2, lag 3 and lag 12. The ACF dies down after lag 1 suggesting that p = 2 and q= 1 for the non-seasonal component. Also the PACF shows significant spike in lag 1, lag 2, lag 11, lag 12, lag 24 and lag 45 indicating influence of seasonal components. The PACF dies down at lag lag 1 though a significant spike at lag 2 which was considered spurious and it is neglected. This also suggest

Q=1 would be needed to describe the data coming from a seasonal autoregressive and moving average process.

**4.2 MODEL IDENTIFICATION**

Having made the series stationary, the decision was made on reasonable values of the orders of the Autoregressive (AR(ϕ)), seasonal Autoregressive (SAR(Φ)), ordinary differencing, Moving Average (MA(θ)) and Seasonal Moving Average (SMA(Θ)).

After trying different seasonal ARIMA models of various orders, in order to choose the best model, we look for the model with the least AICc. Brockwell and Davis (1991) in their research suggest that AICc as the primary criterion in selecting the orders of a time series. This was supported by Burnham and Anderson (2004), they noted that AICc converges to AIC as n gets large hence they also affirmed that AICc rather than AIC should generally be employed.

**4.3 PARAMETER ESTIMATION**

After various trial, it was discovered that SARIMA model (2,1,1)(0,1,1)12 gives the minimum AICc. This is observed in table 3 below, with non-seasonal AR (2), MA (1) being significant and seasonal MA(2) being significant.

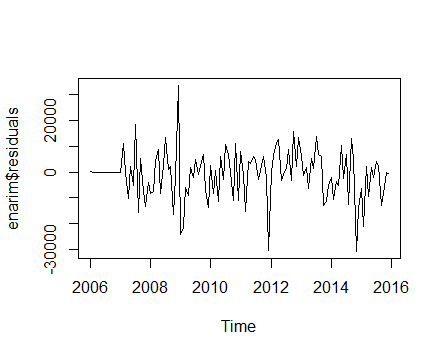
*Table 3: SARIMA (2, 1, 1) (0, 1, 1)12 Fit Result*

|  |  |  |  |
| --- | --- | --- | --- |
| **Coefficients** | **Estimates** | **Standard Error** | **|Z|-value** |
|  | -0.1073 | 0.1876 | 0.5719 |
|  | -0.2573 | 0.1263 | 2.0372\* |
|  | -0.4494 | 0.1863 | 2.4122\* |
|  | -0.8110 | 0.1198 | 6.7696\* |

The model can be express as:

ENPt = + 0.8110 + 0.4494 - 0.3644 + + 0.8927 - 0.8927– 0.15 + 0.15 + 0.2573 - 0.2573 (11)

Also, the diagnostic check carried out in Figure 6 show that there are no significant spikes in ACF and PACF plot of the residual. Also, the residual plot in Figure 5 indicates that the model is stationary.



*Figure 5: Residuals from the fitted SARIMA*

*(2, 1, 1) (0, 1, 1)12 model for the enplane data*

**4.4 DIAGNOSTIC CHECKING AND MODEL VALIDATION**

The Shapiro-Wilk test of normality has a test statistic w=0.96423, leading to a p-value of 0.189583 and normality is not rejected at 1%, 5%, and 10% significant levels. This indicates that the residuals of the chosen model are normally distributed.

Moreover, Bok-Ljung test result gives a chi-square of 37.806 with 48 degrees of freedom leading to p-value of 0.7466. The statistic and large p-value suggests the acceptance of the null hypotheses that all of the autocorrelation functions up to lag 48 are zero. We can conclude that there is no evident for non-zero autocorrelation in the forecast errors up to lag 48 of our fitted model.

The AIC and log-likelihood deal with the fit and parsimony of the model which provides a measure of efficiency and parsimonious prediction. In addition, the seasonal ARIMA

(2, 1, 1) (0, 1, 1)12 model can be confirmed to be good and adequate for the enplane passengers traffic in Muritala Muhammed Airport because it is white noise looking at the ACF and PACF.



*Figure 6: ACF and PACF of Residuals from the fitted SARIMA (2, 1, 1)(0, 1, 1)12 model for the enplane data*

The ultimate aim of building any time series model is forecasting. If this objective is not achieved, the work is incomplete. Forecasts was made for the possible number of enplane passengers for 10 years period. Based on the chosen model, the 10 years is shown in Figure 8. The result shows that there will be downward trend in enplane passengers between November and December of every year but there will be general increase in the number of enplane passengers in Nigeria through Muritala Muhammed International Airport in years to come compared to previous years. This shows that there is need for managements of the airport with airlines to provide necessary facilities that will aids in passengers’ satisfaction in the airport in the coming years.

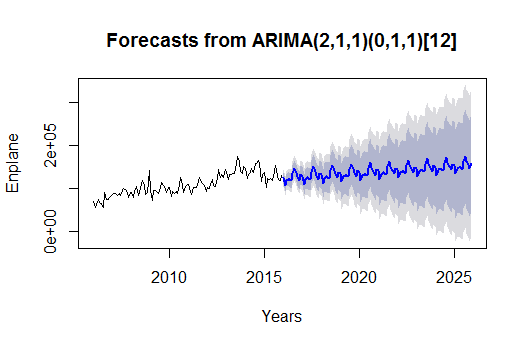
**5. CONCLUSION**

From figure 1, it can be seen that the enplaned passengers’ traffic in Muritala Mohammed international airport exhibit non-stationarity.

This non-stationarity can be attributed to movement of passengers from the country which can be attributed to business transaction, political instability, insecurity, recession in economy and unemployment.

The Seasonal Autoregressive Integrated Moving Average (SARIMA) proposed by Box-Jenkins was employed to analyse the enplaned passengers’ traffic in Muritala International airport, Nigeria from 2006 to 2015. The study is mainly to model and forecast the monthly passengers’ traffic for ten years.

Moreover, several models were developed for the enplaned passengers’ traffic for the international airport but based on minimum corrected akaike information criteria (AICc) value, estimation of necessary parameters and series of diagnostic test were performed. It was observed that SARIMA (2,1,1)(0,1,1)12 model was the best model for forecasting enplaned passengers’ traffic in Muritala Mohammed International Airport.



*Figure 8: Forecasts of the enplane data using the SARIMA (2, 1, 1) (0, 1, 1)12 model.*

**REFERENCES**

Abed, S.Y., Ba-Fail, A.O. & Jasimuddin, S.M. (2001). An econometric analysis of international air travel demand in Saudi Arabia, *Journal of Air Transport Management,* 7, 143-148.

Akaike H. (1974). *A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control, 19, 716-723.*

Box, G.E.P. & Jenkins, G.M. (1970). *Time-Series Analysis, Forecasting and Control*, Holden-Day, San Francisco.

Civil Aviation Authority (2007). A report on passengers’ movement.

Emir, E. & Gabriel, R. (2003). Evaluating Time Series Models In Short and Long Term Forecasting of Canadian Air Passenger Data. Working Paper, Department of Economics, Faculty of Social Sciences, University of Ottawa.

Ette, H.E. (2013). *A Seasonal Time Series Model for Nigerian Monthly Air Traffic Data. IJRRAS, 14(3), 596 - 602.*

Grubb, H. & Mason, A. (2001). Long Lead-Time Forecasting Of UK Air Passengers by Holt-Writers Methods with Damped Trend. *International Journal of Forecasting*, 17, 71 - 82.

Karlaftis, M.G. & Papastavrou, J.D. (1998). Demand Characteristics for Charter Air- Travel, *International Journal of Transport Economics*, XXV (3), 19-35.

Ling Lai & Li Lu, (2005). Impact analysis of September 11 on air travel demand in the USA, *Journal of Air Transport Management*. 11, 455–458.

Melville, J.A. (1998). *An Empirical Model of the Demand for International Air Travel for the Caribbean Region. International Journal of Transport Economics, 25 (3), 313-336.*

Murat Ç. (2014) *Modelling and Forecasting Inbound Tourism Demand to Istanbul: A Comparative Analysis. European Journal of Business and Social Sciences, Vol. 2, No. 12, pp 101-119.*

Tan, M.C., Wong, S.C., Xu, J.M., Guan, Z.R. & Zang, P. (2009). *An Aggregation Approach to Short Term Traffic Flow Prediction. IEEE Transactions on Intelligent Transportation Systems. 10(1), 60-69.*

Tan, M.C.,Wong, S.C., Xu, J.M., Guan, Z.R. & Zang, P. (2009). An Aggregation Approach to Short Term Traffic Flow Prediction. *IEEE Transactions on Intelligent Transportation Systems.* 10(1), 60-69.

Williams, B. M. & Hoel, L. A. (2003) *Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. Journal of Transportation Engineering. 12(9) 664–672.*

Williams, B.M. & Hoel, L.A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of Transportation Engineering* 129, 664–672.

[www.wikipedia.com](http://www.wikipedia.com) retrieved 20/4/2016

Yves, A. & Sandrine, G. (2008). Estimation of Air Traffic Loss at Toulouse Blagnac Airport After September 11, 2001 Attacks. CS-BIGS, 2(1), 1-8.