International Journal of Mathematics and Statistics Studies
Vol.4, No.6, pp.25-37, December 2016
Published by European Centre for Research Training and Development UK (www.eajournals.org)

AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODEL FOR THE MAJOR AIRLINE DISASTERS IN THE WORLD FROM 1960 THROUGH 2013

Nasiru, M.O., Department of Statistics, University of Abuja, Nigeria
Email: nasiru.olakorede@uniabuja.edu.ng
Ipinyomi R.A., Department of Statistics, University of Ilorin Nigeria
Email: raipinyomi@unilorin.edu.ng
Yahaya H.U., Department of Statistics, University of Abuja, Nigeria
yahagumau@gmail.com
Zubair M.A., Department of Statistics, University of Abuja, Nigeria
Email: anono2009@yahoo.com

ABSTRACT: *This research fit a univariate time series model to the major Airline Disasters in the world from 1960 through 2013. The Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) model was estimated and the best fitting ARIMA model was used to obtain the post-sample forecasts for five years. The fitted model was ARIMA (0,1,1) with Akaike Information Criteria (AIC) of 323.14, Normalized Bayesian Information Criteria (BIC) of 327.04, Stationary* R^2 *of 0.348.This model was further validated by Ljung-Box test with no significant Autocorrelation between the residuals at different lag times and subsequently by white noise of residuals from the diagnostic checks performed which clearly portray randomness of the standard error of the residuals, no significant spike in the residual plots of ACF and PACF. The forecasts value indicates that Airline Disasters will increase slightly with almost equal number of cases for the next five years (2014-2018).*

KEYWORDS: ARIMA, Time Series, Box- Jenkins, Ljung-Box, Stationarity, Unit Root, Airline Disasters, Forecast

INTRODUCTION

In the airline industry, Airline accidents are paramount issue since they have a significant impact on the demand for air travel, it significantly affects the finances of the Airlines, as a results of random occurrences of Airline accidents, Airline passengers switch for a safer Airline or choose an alternative mode of transportation. This issue has been examined among others by Rose, Nancy L. (1992), Borenstein and Zimmerman (1988), Bosch et al. (1998)

Literature has shown that previous research on Airline accidents focused on the effect of fatal accident on the equity values of Airlines, for example, Borenstein and Zimmerman (1998), Mitchell and Maloney (1989), Bosch et al (1998). They also examined the impact of fatal accidents on Air travel demand. Their results showed that fatal accident have a significant negative effect on the stock value of the Airlines involved in such accidents. Other results showed that fatal accidents have no significant effect on the equity values of other Airlines. Mitchell and Maloney (1989), in their research, found out that equity value of the Airline involved in a fatal accidents falls only if the accidents is the fault of the company. Their findings further showed

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

that; if the financial market is perfect, the stock value of the Airline should have already considered the expected responses of the demand. Their findings may suggest that fatal accidents have little impact on the total demand for Air travel. Borenstein and Zimmerman (1988), further examined the impact of fatal accidents on the demand for Air of crash Airlines remained largely unaffected by the fatal accidents prior to deregulation.

Airline accidents data can be viewed as a count data which has been primarily categorized as cross-sectional time series data, and panel count data. Over the past decades, Poisson and Negative Binomial (NB) models have been used widely to analyse cross-sectional and time series count data, and random effect and fixed effect Poisson and NB models have been used to analyse panel count data. However, in recent time, literature suggests that although the underlying distributional assumptions of these models are appropriate for cross-sectional count data, they are not capable of taking into account the effect of serial correlation often found in pure time series count data. Real-valued time series models, such as the autoregressive integrated moving average (ARIMA) model, introduced by Box and Jenkins have been used in many applications over the last few decades. However, when modeling non-negative integer-valued data where the dataset is relatively low (less than 30 Observations) such as traffic accidents over time, Box and Jenkins models may be inappropriate. This is mainly due to the normality assumptions of error in the ARIMA model. Over the last few years, a new class of time series models known as integer-valued autoregressive (INAR) Poisson models has been studied by many authors. This case of model is particularly applicable to the analysis of time series count data as these models hold the properties of Poisson regression and able to deal with serial correlation, and therefore offers an alternative to the real-valued time series models.

Mohammed A. Quddus (2008), introduced the class of INAR models for the time series analysis of traffic accidents in Great Britain. He compared the performance of the INAR models with the class of Box and Jenkins real-valued models, his result suggest that the performance of these two classes of models is quite similar in terms of coefficients estimates and goodness of fit for the case of aggregated time series traffic accidents data. This is because the mean of the counts is high in which case the normal approximations and the ARIMA models may be satisfactory.

Mohammed A. Quddus (2008), in his work, developed accident prediction models of a highly aggregated time series process of annual road traffic fatalities in Great Britain. He employed a range of econometrics models such as ARIMA, NB, and INAR Poisson models. He investigated the performance of the fitted models. His result implied that the best accident prediction model for the aggregated time series count data was achieved when ARIMA model was used. This is due to the fact that this model is able to take into account both serial correlation and non-stationarity normally found in a time series dataset.

This research will contribute to the literature by fitting a univariate time series ARIMA models for the number of cases of Airline disasters in the world from 1960 through 2013 and use the best fitted model to make forecast for five years. This long period was chosen because the number of cases is a count data which the number of observations must be large enough (at least 50-100 observations) to meet-up the normality assumption of ARIMA model. Although, Airline accidents can be viewed with rare events, however, in this research observations were collected

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

annually from 1960 through 2013 in an ordinal basis and the interval over which the data was taken remain the same over time to generate the time series data used for this research. The objectives of this research are: (i) to evaluate the pattern and duration of the airline disasters in the world from 1960 through 2013 (ii) to fit a univariate time series ARIMA model for Airline disasters and (iii) use the fitted model to make five years forecast

Methodology and Model Specification

The model used in this study is the ARIMA proposed by Box and Jenkins (1976). The preliminary test for stationarity and seasonality of the data was conducted in which differences (d) as well as transformation were taken. After the stationarity of the series was attained, ACF and PACF of the stationary series are employed to select the order p and q of the ARIMA model. At this stage, different candidates' model manifested and their parameters were estimated using the maximum likelihood method. Based on the model diagnostic tests and parsimony we obtained the best fitting ARIMA model. The Mathematical model for Auto Regressive of order p as well as that of Moving Average of order q are given respectively as

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + \epsilon_t \qquad (2.1)$$

and $$y_t = \in_t - \theta_1 \in_{t-1} - \theta_2 \in_{t-2} - \cdots - \theta_q \in_{t-q} \qquad (2.2)$$

The ARMA process of order (p,q) is written as

$$y_t - \Phi_1 y_{t-1} - \Phi_2 y_{t-2} - \cdots - \Phi_p y_{t-p} = \in_t - \Phi_1 \in_{t-1} - \Phi_2 \in_{t-2} - \cdots - \Phi_q \in_{t-q} \qquad (2.3)$$

Method of Estimation: ARIMA Methodology

The Box-Jenkins model building techniques consists of the following four steps:

Step 1: Preliminary Transformation: If the data display characteristics violating the stationarity assumption, then it may be necessary to make a transformation so as to produce a series compatible with the assumption of stationarity. After appropriate transformation, if the sample autocorrelation function appears to be nonstationary, differencing may be carried out.

Step 2: Identification: If $y_t$is the stationary series obtained in step 1, the problem at theidentification stage is to find the most satisfactory ARMA (p,q) model to representy$_t$

Box – Jenkins(1976) determined the integer parameters (p,q) that govern the underlying processy$_t$by examiningthe autocorrelations function (ACF) and partial autocorrelations (PACF) of the stationary series, y_t. (Salau, M.O.(1998) explained that it is better to entertain more than one structure for further analysis because the evidence examined at this stage does not point clearly in the direction of a single model.Salau, M.O. (1998) stated that this decision can be justified on the ground that the objective of the identification phase is not to rigidly select a single correct model but to narrow down the choice of possible models that will then besubjected to further examination.

Step 3: Estimation of the model: This deals with estimation of the tentative ARIMA model identified in step 2. The estimation of the model parameters can be done by the conditional least squares and maximum likelihood.

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

Step 4: Diagnostic checking: Having chosen a particular ARIMA model, and having estimated its parameters, the adequacy of the model is checked by analyzing the residuals. If the residuals are white noise; we accept the model, else we go to step 1 again and start over.

ANALYSIS AND RESULTS

Time Series Graph of the Raw Data

Time series plots which display observations on the y-axis against equally spaced time intervals on the x-axis used to evaluate patterns and behaviors in data over time for major Airlines disasters in the world is displayed in the Figure 1 below. The data used for this research was sourced from www.airdisasters.co.uk from 1960 through 2013.

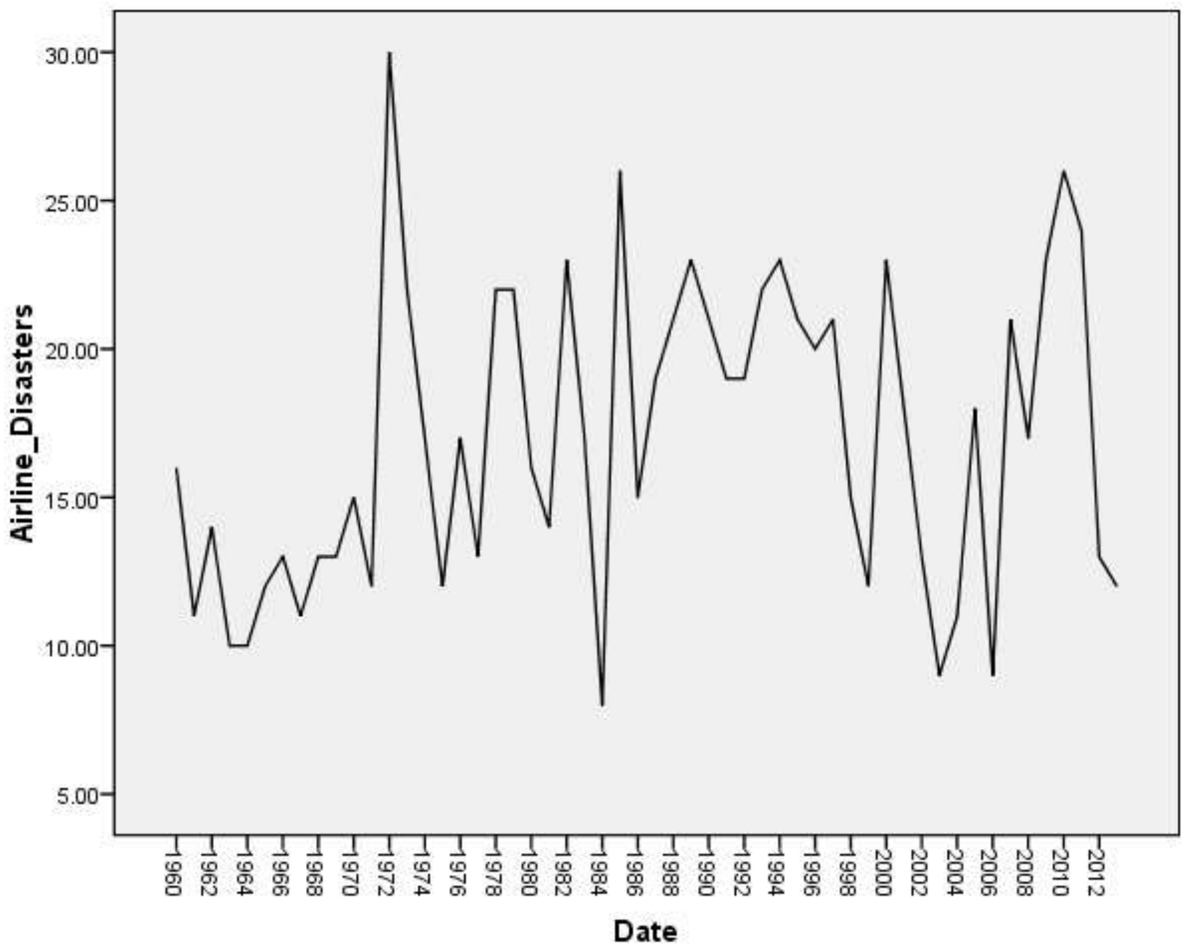


Figure1: Time Series Graph of Airline Disasters 1960 - 2013

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

Table1: Unit Root and Stationarity Tests of Airline Disasters

Test type	Test Statistics	Lag Order	p-value
KPSS	0.3959	1	0.01
PPT	-40.6613	3	0.0789

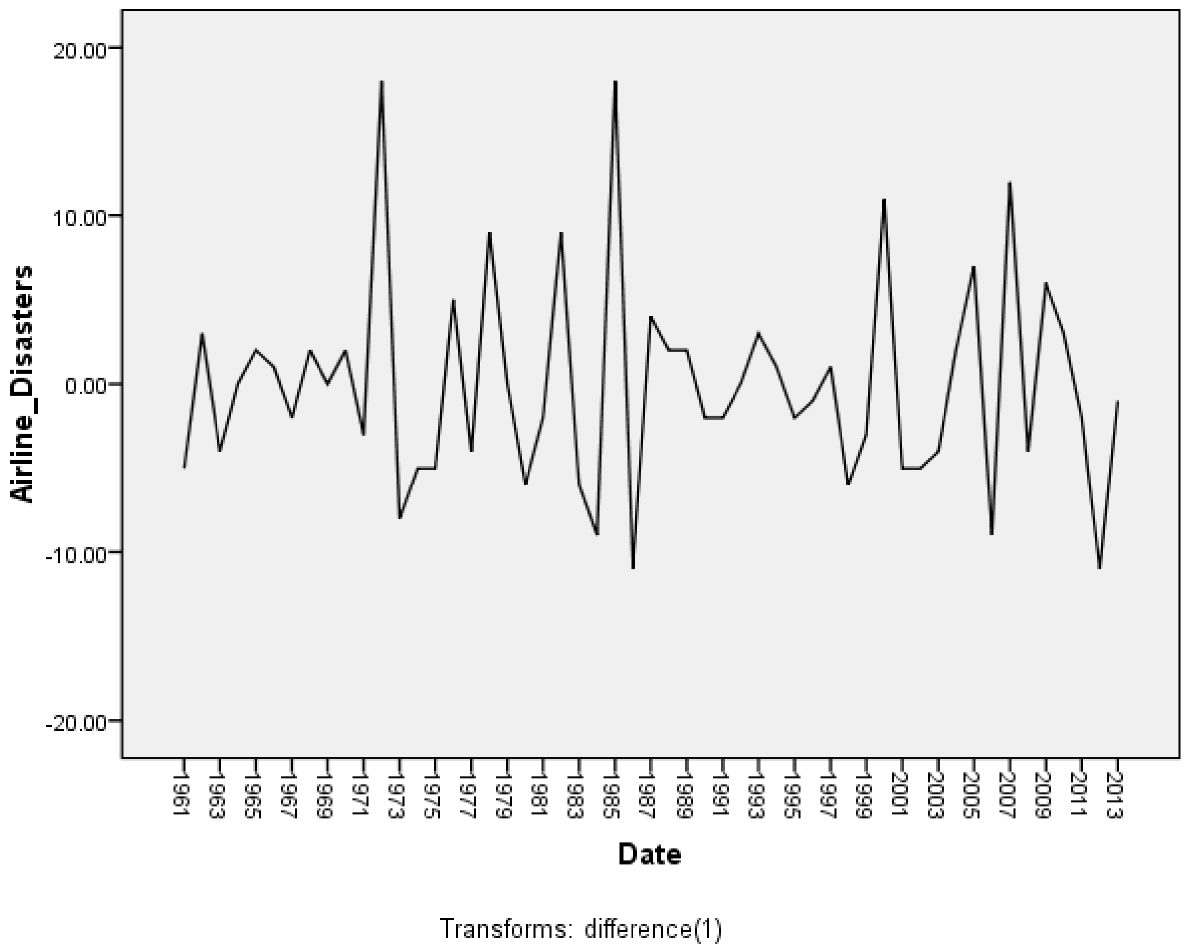


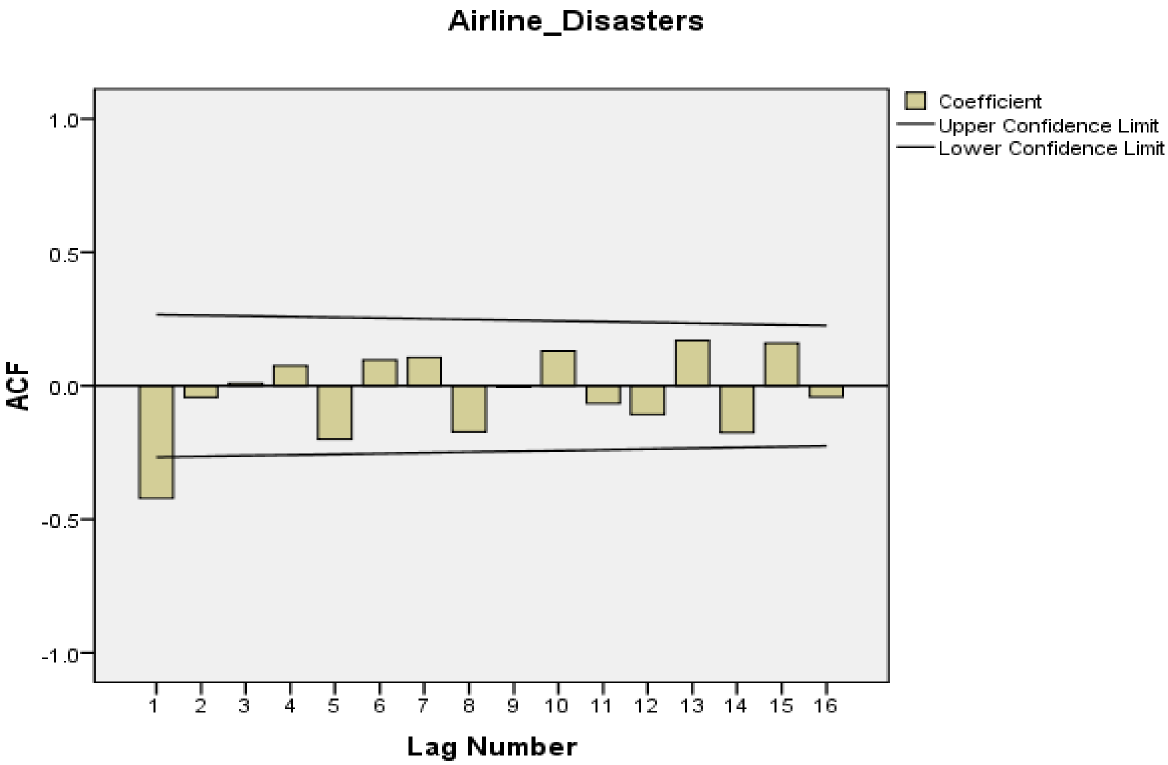
Fig 2: FIRST Order Differenceof Major Airline Disasters

Tab2 : Unit Root and Stationarity Tests for the Differenced Major Airline Disasters

Test type	Test Statistic	Lag order	p-value
KPSS	0.0487	1	0.100
PPT	-63.1537	3	0.01

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

International Journal of Mathematics and Statistics Studies
Vol.4, No.6, pp.25-37, December 2016
Published by European Centre for Research Training and Development UK (www.eajournals.org)



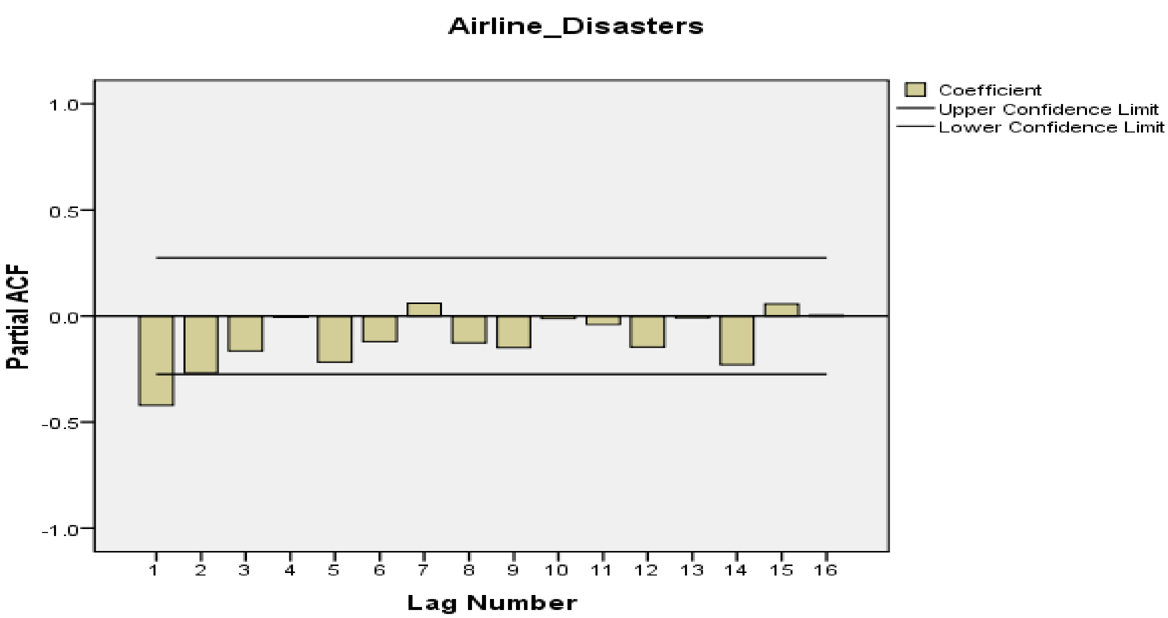


Fig 3; Plots Of ACF and PACF Of Major Airline Disasters

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

Tab 3: ARIMA MODELS RESULTS

ARIMA Structures	Parameter Estimates	p-value	S.E	Stationary R^2	Normalized BIC	AIC
ARIMA(0,1,1)	MA1=0.835	<0.0001	0.134	0.348	327.04	323.14
ARIMA(1,1,1)	AR1=0.243 MA1=0.995	0.122 0<0.0001	0.154 0.650	0.346	330.10	324.25
ARIMA(2,1,1)	AR1=0.230 AR2=0.067 MA1=0.998	0.153 0.672 <0.0001	0.158 0.157 1.693	0.325	334.02	326.22

Tab 4: ARIMA (0,1,1) RESULTS

Stationary R^2	Normalized BIC	AIC		
0.348	327.04	323.14		
Coefficient	Estimate	S.E	t-value	p-value
MA 1	0.835	0.134	9.959	<0.0001
Constant	0.092	0.132	0.696	0.490

Ljung-Box Test of ARIMA (0,1,1)

Test Type	Q-statistic	Df	p-value
Ljung-Box	8.772	17	0.947

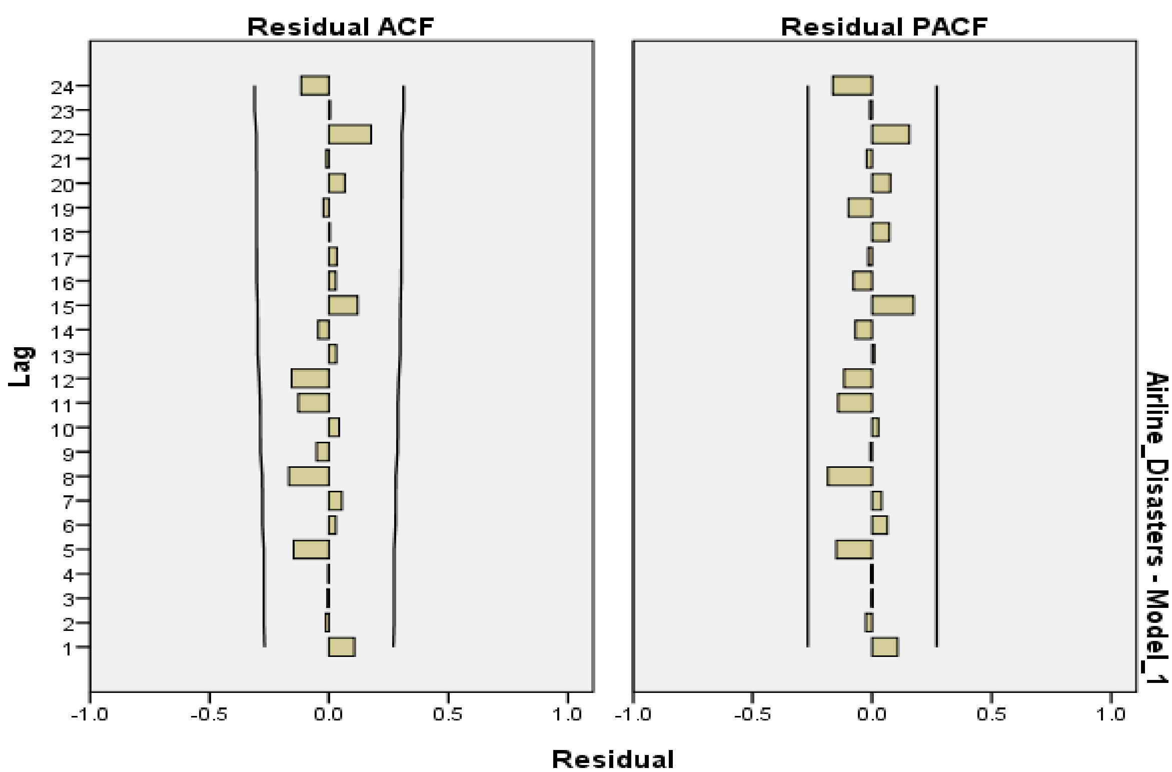


Figure 4: Plot of Residual ACF and PACF of Major Airline Disasters

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

Table 5: Forecasts results with the Fitted ARIMA (0,1,1)Model

Year	LCL	FORECAST	UCL
2014	7.50	17.93	28.36
2015	7.45	18.02	28.60
2016	7.40	18.11	28.83
2017	7.35	18.21	29.06
2018	7.31	18.30	29.29

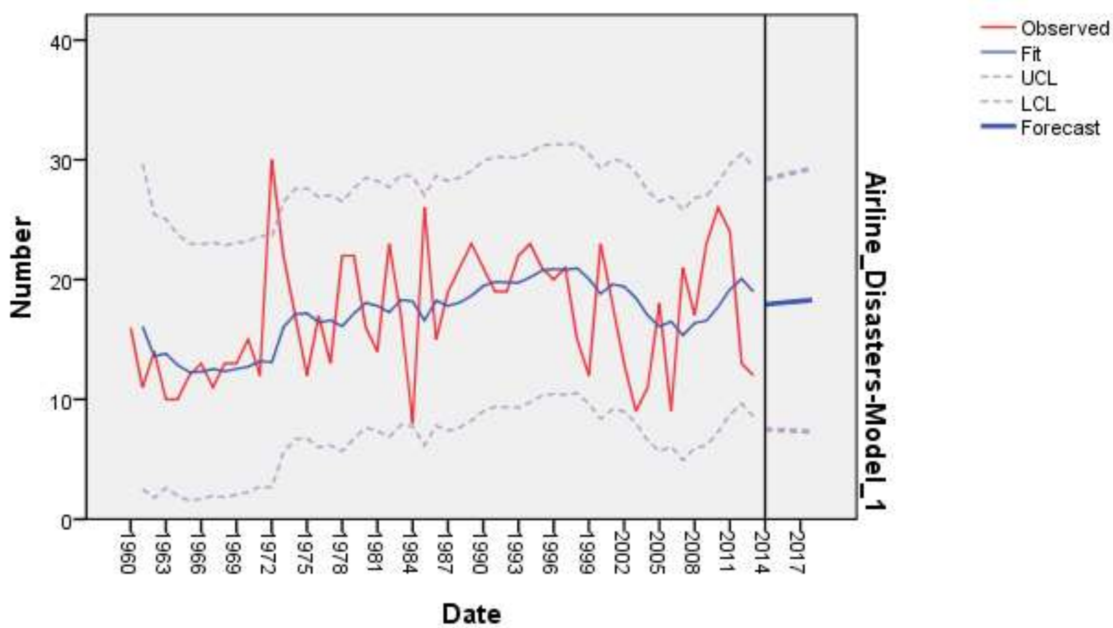


Figure 5: The plot of the observed and forecast value of Airline Disasters

DISCUSSION OF RESULTS

The time series plots of the raw data in Figure 1 indicates clearly that the occurrence of major Airline Disasters in the world from 1960 through 2013 was not constant but rather varied from one year to the other with no systematically visible pattern, structural breaks, outliers, and no identifiable trend components in the time series data or non-monotonous (that is consistently increasing or decreasing). This behavior clearly revealed that non- stationarity was inherent in the data. The unit root tests provide a more formal approach in determining whether the series is stationary or not such as Kwiatkowski-Phillips-Schmidt-Shin (KPSS) and Phillips-Perron Unit Root Tests (PPT), thesetestswere carried out as shown in Table1. We employed the unit root testing procedures of Hamilton, J.D. (1994). The following hypotheses were tested:

For KPSS

H_0: the series is stationary or has no unit root

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

Vs
H_1: the series is not stationary or has a unit root

For PPT

H_0: the series is not stationary or has a unit root
Vs
H_1: the series is stationary or has no unit root

The decision rules involves accepting H_0 where the p-value is greater than critical value of 0.05, and fails to reject if otherwise. KPSS test statistic has a p-value less than the critical value of 0.05 as presented in Table1, rejects the null hypothesis. The PPT has a p-value greater than 0.05 fails to reject the null hypothesis. It is clear from the time series plot of Airline Disasters and the unit root test that the series has to be transformed or differenced to stabilize or stationarize the data before its capability is assessed or improvements are initiated.

The time series of the first differenced Figure2is stationary because the mean and variance were constant over time which means that the mean is exactly zero which confers a stationary series. The unit root test which is a formal test of stationary was performed as shown in Table2. Table2depicts the KPSS and the PPT for the first order differenced of the series, The KPSS test statistic has a p-value greater than the critical value of 0.05 do not reject the null hypothesis of having a level stationary series. Philips-Perron Test on the other hand has its p-values rejects the null hypothesis of a unit root at 5% significance level, since its p-values is less than 0.05. It therefore can be concluded that the time series plot of the first differenced indicates that the stationarity ofthe series was achieved at first differenced.

Figures3 comprises the plots of ACF and PACF. If the PACF display a sharp cutoff while the ACF decay more slowly (i.e., has significant spikes at higher lags), we say that the series display an AR signature, however, if the ACF display a sharp cutoff while the PACF decay more slowly, we say that the series display an MA signature. The lag at which the ACF cut off is the indicated number of MA terms. Based on Figure3 plot for major Airline disasters,, it can be seen that there is a slow decay in the PACF, but has a cut-off at lag 1, and lag2 suggesting AR(1) and AR(2) respectively, with a single negative significant spike around the ACF which displays a sharp cutoff at lag 1. This pattern is typical to a Moving Average (MA) process of order one. Hence a number of possible models were identified, these models are: ARIMA (1,1,1), ARIMA (2,1,1), and ARIMA (0,1,1). We proceeded to further statistically analyzed these three possible models and the results were summarized in table3

Based on the parameters estimates in Table3 of major Airline disasters, the estimate of all the AR models were found to be statistically insignificant because their p-values were all greater than 0.05 Therefore the null hypothesis (H_o) of parameter are or equal zero is not rejected resulting in their removal from the model. The estimates of the MA model on the other hand, was found to be statistically significant because it p-value is less than 0.05 significance level. In addition, comparing the ARIMA (0,1,1) and the ARIMA (1,1,1), ARIMA (2,1,1) models above in terms of the Stationary $R^{2,}$ BIC, AIC respectively, clearly prefers ARIMA (0,1,1) model since It has highest Stationary $R^{2,}$ smallest BIC, and smallest AIC. The summary of the parameter estimates of ARIMA (0,1,1) was stated in tab4. In conclusion, based on the parameter estimates in the Tab3 above, we

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

chose ARIMA(0,1,1) as the best model for the Airline Disasters in the world. The model is thus given as:

$\nabla^1 Y_t=0.092-0.835\varepsilon_{t-1}-\varepsilon_t \Rightarrow Y_t- Y_{t-1}=0.092-0.835\ \varepsilon_{t-1}-\varepsilon_t$ (5.1)

This model is a special case of ARIMA model, which is called an Integrated Moving Average (IMA) Model.

This model was diagnosed by Ljung-Box test and the p-value was quite large (greater than the usually chosen critical level (0.05), the test is not significant and therefore we do not reject the null hypothesis, thus the residuals appear to be uncorrelated. This indicates that the residuals of the fitted ARIMA (0,1,1) model is a white noise, and for that reason, the model fit the series quietly well, the parameter of the model are significant and the residuals are uncorrelated.

The plots Fig4 comprises of the time plot of the residuals, ACF plot of the residuals and the PACF plot of the residuals respectively. The time plots of the residuals clearly showed that the residuals appear to be randomly scattered, no evidence exists that the error terms are correlated with one another as well as no evidence of existence of an outlier. The residuals or errors are therefore conceived of an independently identically distributed sequence with a constant variance and a zero mean. The ACF and the PACF plots of the residuals shows no evidence of a significant spike (the spikes are within the confidence limits) indicating that the residuals seems to be uncorrelated. Therefore, the ARIMA (0,1,1) model appears to fit well so we can use this model to make forecasts. . This also shows that the residuals of ARIMA (0,1,1) model is a white noise process. Thus the residual plots corroborate the conclusion of the Ljung-Box test.

FORECASTING WITH THE FITTED MODEL

Thus, in time series modeling, researchers are motivated by the desire to produce a forecast with minimum error as possible. In this section, we assessed the forecasting performance of Box-Jenkins models. Literature has shownthat the Box-Jenkins method give better forecasts than the traditional econometric methods.

Forecasting the Major Airline disasters in the world using a univariate Time Series Model, we computed one-step ahead forecasts for the fitted model, i.e. ARIMA (0,1,1). These forecasts and their 95% confidence interval i.e. Lower confidence limit (LCL) and upper confident limit (ULC) for five years (i.e. 2014 – 2018) were summarized in Table5, while Fig5 depicts the observed and forecast plot of Major Airline disasters in the world. The values of this forecasts shows that occurrences of airline disasters will slightly increase for the next five years. The forecast equation used is given as:

$\hat{Z}_t = 0.092+Z_{t-1}-0.835\varepsilon_{t-1}$ (6.1)

CONCLUSIONS

This research fit a univariate time series model to the major Airline Disasters in the world from 1960 through 2013.Theevaluation ofpattern revealed that occurrences of Airline Disasters were not constant but rather varied from one year to the other with no systematically visible pattern. The

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) model was estimated and the best fitted ARIMA model was used to obtain the post-sample forecasts for five years. The fitted model was ARIMA (0,1,1) with Normalized Bayesian Information Criteria (BIC) of 327.04, Akaike Information Criteria (AIC) of 323.14, and Stationary R^2 of 0.348. .the model is given as:

$$\nabla^1 Y_t=0.092-0.835\varepsilon_{t-1}-\varepsilon_t \Rightarrow Y_t- Y_{t-1}=0.092-0.835\ \varepsilon_{t-1}-\varepsilon_t \quad (7.1)$$

This model was further validated by Ljung-Box test with no significant Autocorrelation between the residuals at different lag times and subsequently by white noise of residuals from the diagnostic checks performed which clearly portray randomness of the standard error of the residuals, no significant spike in the residual plots of ACF and PACF.

The fitted model was used to obtain the post-sample forecast for five years. Weassessed the forecasting performance of Box-Jenkins models. We computed one-step ahead forecasts for the fitted mode, i.e. ARIMA (0,1,1). These forecasts and their 95% confidence interval i.e. Lower confidence limit (LCL) and upper confident limit (ULC) for five years (i.e. 2014 – 2018) indicates that Airline Disasters will increase slightly with almost equal number of cases for the next five years (2014-2018).

REFERENCES

Borenstein, Severin; and Zimmerman, Martin B. (1988) "*Market Incentives for Safe Commercial Airline Operation.*" American Economic Review 78:913-35

Bosch et al (1998). *The Competitive Impact of Air Crashes: Stock Market Evidence.*" Journal ofLaw and Economics: Vol. 41: No. 2, Article 8.

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

Hamilton, J. D. (1994), *Time Series Analysis. Princeton University Press, New Jersey, USA.*

Mitchell, Mark L., and Michel T. Maloney (1989),*The Role of Market Forces inPromoting Air Travel Safety:* Journal of Law and Economics (Impact Factor 0.89)

Mohammed A. Quddus (2008). *Time Series Count Data Models: An Empirical Application to Traffic Accidents.* Journal of Accident Analysis and Prevention. 40 (5),pp. 1732-1741

Rose, Nancy L. (1992). *Economic analysis of Airline Safety:* Journal of Economic Perspectives,Vol.6 issue 2, pages 75-94

Salau M.O (1998). *ARIMA modeling of Nigeria's crude oil Export,* AMSE, modeling, measurement and control. Vol.18.No.1,1-20

www.airdisasters.co.uk (2002). *Major Airline Disasters involving commercial passengers Airline 1920-2002.*

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)