TITLE PAGE

TIME SERIES ANALYSIS ON MEASLES CASES IN NIGERIA

BY

OBAFAIYE ADEGBOYEGA BENEDICT

PGDST/16328022

A RESEARCH WORK SUBMITTED TO THE DEPARTMENT OF STATISTICS, FACULTY OF SCIENCE, UNIVERSITY OF ABUJA, ABUJA

IN PARTIAL FULFILMENT OF THE AWARD OF POST GRADUATE DIPLOMA (PGD) IN STATISTICS

2018

CERTIFICATION

This is to certify that this project work was carried out under my supervision by OBAFAIYE ADEGBOYEGA BENEDICT with Matric Number PGD 16328022 of University of Abuja, Abuja.

....

Dr. Asemota, Omorogbe Joseph Date

Supervisor

....

Dr. Yahaya .H Date

Head of Department

DEDICATION

I dedicate this project work to the glory of almighty God.

ACKNOWLEDGEMENT

My gracious gratitude goes to almighty God for merciful, compassionate, kind and forbearing over my life and family. My unreserved appreciation also goes to my parents, Mrs. Oluwafunmike Lilian Kanjuni and Late Mr. Clement Obafaiye for their enormous contribution throughout my stay in postgraduate.

I also wish to extend my profound gratitude to my supervisor Dr. Asemota, Mr. S.S Ikusika, Mr. F.K Lawal, Mrs. Ayodele, Clementina, Pauline, Silas Kanjuni, I wish you success in your future endeavour Amen.

ABSTRACT

This research work describes a study that used measles disease data collected through Knoema health surveillance system to evaluate univariate time series method namely; autoregressive integrated moving average (ARIMA). The data obtained from 1980 to 2016 were used as modeling data and forecasting samples, respectively. The performances were evaluated based on three metrics: mean absolute error (MAE), mean absolute percentage error (MAPE), and mean square error (MSE). A low normalized BIC of21.817 was recorded. The accuracy of the statistical model in forecasting future measles disease proved its effectiveness in measles disease breakout surveillance. Although the outcome of this research work, has shown that measles outbreak in the nearest future will take a downward trend from 2017 to 2019, as shown in the forecasted output. It was also observed that 40.9% estimate of the proportion of the total variation in the series (measles_1) is explained by the model. The result of this research work has shown that funds for measles can be diverted to other diseases as little fund is required to facilitate measles vaccine and improve measles vaccination in the country.

TABLE OF CONTENT

Tittle page		i
Certification		ii
Dedication		iii
Acknowledgement		iv
Abstract		v
Table of content		vi
CHAPTER ONE: INTRODUCTION		
1.1	Background of the study	1 – 2
1.2	Statement of the problem	2 – 3
1.3	Aim and objectives of the study	3
1.4	Significance of the study	4
1.5	Scope of the study	4
1.6	Definitions of Term	
CHAPTER TWO: LITERATURE REVIEW		
2.1	Review of Related Works	5 - 7
CHAPTER THREE: METHODOLOGY		9
3.1	Introduction	9
3.2	Source of data	9
3.3	Method of analysis	10

3.3.1 Identification of ARIMA (p, d, q) Models 10

3.3.2 Auto-Regressive Components 11

3.3.3 Moving Average Components 11 – 12

3.3.4 Mixed Models 12 – 15

3.3.5 Seasonal Cycles and Trends 15

3.3.6 Forecasting 15

CHAPTER FOUR: DATA PRESENTATION AND DATA ANALYSIS 18

4.1 Data presentation 18 – 19

4.2 Data analysis 19 – 28

CHAPTER FIVE: DISCUSSION OF FINDINGS 29

5.1 Discussion of findings 29 – 32

CHAPTER SIX: SUMMARY, CONCLUSION AND RECOMMENDATION 33

6.1 Summary 33

6.2 Conclusion 33 – 34

6.3 Recommendation 34

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background of the study

Time series analysis is the collection of data at specific intervals over a period of time, with the purpose of identifying trends, cycles, and seasonal variances to aid in the forecasting of a future event. Data is any observed outcome that is measurable. Unlike in statistical sampling, in time series analysis, data must be measured over time at consistent intervals to identify patterns that form trends, cycles, and seasonal variances. These data could be measured every minute, hourly, daily, weekly, monthly, quarterly or yearly. Measurements at random intervals lose the ability to predict future events. In economic forecasting, time series can be used to predict the future economy of a particular country, it can also be useful for sales forecasting, budgetary analysis, yield projections, process and quality control, inventory studies and workload projections. So many areas where time series are **observed** and analyzed is countless. Incidentally, measles cases in Nigeria fall within the category of time series data, and the purpose of time series data are mainly;

i. to understand the underlying stochastics mechanism that gives rise to an observed series.
ii. to forecast or predict future values of the series based on the recorded or observed history of that series.

Measles is a highly contagious disease caused by a virus known as rubeola virus. Before the introduction of measles vaccine in 1963 and widespread vaccination, major epidemics occurred approximately every 2–3 years and measles caused an estimated 2.6 million deaths each year according to W.H.O. The disease remains one of the leading causes of death among young children globally, despite the availability of a safe and effective vaccine. Approximately 110000 people died from measles in 2017 – mostly children under the age of 5 years "World Health Organization". Measles is caused by a virus in the paramyxovirus family and it is normally

passed through direct contact and through the air. The virus infects the respiratory tract, then spreads throughout the body. Measles is a human disease and is not known to occur in animals. Accelerated immunization activities have had a major impact on reducing measles deaths. In 2016, an estimated 90000 people died from measles ' an 84% drop from more than estimated 550 000 in 2000 according to a new report published today by leading organizations. This mark the first time global measles death has fallen below 100,000 per year according to Joint news release CDC/GAVI/UNICEF/WHO.

Measurements at random intervals lose the ability to predict future events. In economic forecasting, time series can be used to predict the future economy of a particular country, it can also be useful for sales forecasting, budgetary analysis, yield projections, process and quality control, inventory studies and workload projections. So many areas where time series are observed and analyzed are countless. Incidentally, measles cases in Nigeria fall within the category of time series data and the purpose of time series data analysis.

1.2 Statement of the problem

Measles cases in Nigeria is becoming alarming and worrisome as quite a number of cases has been recorded in the past, with clear evidence of what might be responsible for the unstable cases recorded. In this research work, univariate time series will be used, to identify the model that best describe the measles cases data in order to predict future cases with a view to aiding planning strategies for eradication of measles in Nigeria.

1.3 Aim of this Study

The broad aim is to fit a univariate time series model to measles cases in Nigeria. The specific objectives are:

i. **To identify the model that best describe the cases of measles in Nigeria;**

ii. **To employ the identified model** to forecast future cases of measles;

iii. **To suggest eradication / control strategies based on the forecasted values**

1.4 Significance of the Study

i. Modelling measles cases would be of great importance to the Nigeria government, development partners, (USAID, W.H.O, DFID, JICA etc), stakeholders in the health sector as well as researchers in the field of time series.

ii. This is particularly important to the sustainable development goals. (SDG) which aims to ensure healthy lives and promote well-being for all at all ages.

iii the study will guide the government in monitoring and evaluation (M&E) as it relates to measles cases in Nigeria. This will in turn guide budget projections for the Ministry of Health in Nigeria.

1.5 scope of the study

i. the research work covered the yearly record of measles cases from 1980 to 2016 in Nigeria.

ii. the study considered a univariate approach based on box – Jenkins procedures.

1.6 Definitions of Terms

Observation: The DV cases at one time period. The observed value can be from a single case or an aggregate observation from numerous cases.

Random shock: The random component of a time series. The shocks are reflected by the residuals (or errors) after an adequate model is identified.

ARIMA (p. d. q): The acronym for an auto-regressive integrated moving average model. The three terms to be estimated in the model are auto-regressive (p), integrated (trend—d), and moving average (q).

Auto-regressive terms (p): The number of terms in the model that describe the dependency among successive observations. Each term has an associated correlation coefficient that describes the magnitude of the dependency. For example, a model with two auto-regressive terms (p=2) is one in which an observation depends on (is predicted by) two previous observations.

Moving average terms (q): The number of terms that describe the persistence of a random shock from one observation to the next. A model with two moving average terms (q=2) is one in which an observation depends on two preceding random shocks.

Lag: The time periods between two observations. For example, lag 1 is between Y_t and Y_{t-1}. Lag 2 is between Y_t and Y_{t-2}. Time series can also be lagged forward, Y_t and Y_{t+1}. Calculating

Differencing: differences among pairs of observations at some lag to make a non-stationary series stationary.

Stationary and non-stationary series: Stationary series vary around a constant mean level, neither decreasing nor increasing systematically over time, with constant variance. Non-stationary series have systematic trends, such as linear, quadratic, and so on. A non-stationary series that can be made stationary by differencing is called "non-stationary in the homogenous sense."

Trend terms (d): The terms needed to make a non-stationary time series stationary. A model with two trend terms (d=2) has to be differenced twice to make it stationary.

The first difference removes linear trend, the second difference removes quadratic trend, and so on.

Autocorrelation: Correlations among sequential scores at different lags. The lag 1 autocorrelation coefficient is similar to correlation between the pairs of scores at adjacent points in time, $r_{Y_t, Y_{t-1}}$ (e.g., the pair at time 1 and time 2, the pair at time 2 and time 3, and so on). The lag 2 autocorrelation coefficient is similar to correlation between the pairs of scores two time periods apart, $r_{Y_t, Y_{t-2}}$ (e.g., the pair at time 1 and time 3, the pair at time 2 and time 4, and so on).

Autocorrelation function (ACF): The pattern of autocorrelations in a time series at numerous lags; the correlation at lag 1, then the correlation at lag 2, and so on.

Partial autocorrelation function (PACF): The pattern of partial autocorrelations in a time series at numerous lags after partial-ing out the effects of autocorrelations at intervening lags.

CHAPTER TWO

LITERATURE REVIEW

2.0 INTRODUCTION

This chapter present the review of the theoretical and empirical literatures as it relates to measles incidence.

2.1 THEORETICAL AND EMPERICAL REVIEW

Measles, one of the most infectious viral diseases of humans affecting over 95% of exposed individuals in the absence of vaccination, is spread by the respiratory route and remains a major cause of mortality in children, particularly in developing countries (van den Ent et al ., 2011).

Umeh and Ahaneku used a descriptive analysis of measles case-based surveillance data that were collected by the WHO and the state MOH between 2007 and 2011 to find out if there was any association between measles immunization coverage and measles outbreak; the inclusion criteria were all those that met the measles diagnostic criteria of clinical confirmation or as confirmed by the laboratory in the absence of measles vaccination. It was discovered that As the proportion of cases with febrile rash who were immunized decreased from 81% in 2007 to 42% in 2011, the laboratory confirmed cases of measles increased from two in 2007 to 53 in 2011.Of the laboratory confirmed cases of measles, five (7%) occurred in children less than 9 months, 48 (64%) occurred in children 9 - 59 months and 22 (29%) occurred in children less than 59 months old. Seventy five percent of all laboratory confirmed cases of measles occurred in rural areas.

In an attempt to look at the epidemiology of measles in South-West Nigeria, Fatiregun. 2014 *et al.* analyzed measles case-based surveillance data from 2007 to 2012 The authors used a descriptive analysis (persons, place, and time) of measles cases and which was confirmed through laboratory and epidemiological link. Fatiregun, 2014 *et al.* predicted expected measles cases in 2015 using additive time series model. Furthermore, in a similar study on trends and patterns of under-fives vaccination in Nigeria, using four National Demographic and Health surveys datasets involving a total of 44,071 (weighted) children from 1990 to 2008; the authors examined child health information including the proportion of those who had some or completed their routine childhood vaccinations, the trends, as well as a pattern of vaccination over 18 years. The authors also selected certain factors and regressed them to obtain predictors of child vaccinations in Nigeria. Considering the importance of timeliness and completeness of reporting on all suspected infectious diseases, a retrospective review of surveillance records was conducted between January 1, 2007 and June 30, 2008. The outcome of several studies has shown that measles outbreaks are associated with factors that include: weak measles case-based surveillance in some areas, lack of awareness about the disease among parents, vaccine stock-out, and lack of adequate cold chain equipment to preserve the vaccine in remote hard-to-reach areas.This was done by review of records of suspected measles from the registers of 23 health facilities in Nigeria.

Fatiregun AA and Odega CC 2016 *et al.* used a capture-recapture method to obtain an estimate of the total number of measles cases required for the study area within the period under review.

Although other primates are also susceptible to measles and develop similar clinical disease, these populations are not large enough to support sustained transmission of the virus and thus humans are the only natural reservoir (Griffin, 2007). Measles has an incubation period of approximately 10-12 days, followed by a 3-4 day prodromal period of fever, anorexia, malaise and one or more of the three C's (cough, coryza, conjunctivitis). During this prodromal period, Koplik's spots (small, bright red spots with a blueish-white Centre) appear on the buccal mucosa. These spots are considered to be pathognomonic for measles diagnosis prior to onset of rash (Moss & Griffin, 2009). The prodrome is followed by a characteristic erythematous maculopapular rash that appears first on the face and behind the ears, spreads to the trunk and extremities, and fades after 3-4 days.

People with measles are infectious for several days before and after the onset of rash. In uncomplicated measles, clinical recovery begins soon after onset of rash, and results in viral clearance and lifelong immunity. As a consequence of the immunosuppressant induced by measles infection, and the direct damage to the respiratory tract (loss of cilia), secondary bacterial, viral and parasitic infections may occur. In developed countries, complications such as otitis media, gastroenteritis, pneumonia, myocarditis and pericarditis occur in approximately 10% of measles cases, with encephalitis occurring in a very small subset of cases. In developing countries, complication rates may reach 80%. Pneumonia and gastroenteritis due to secondary infections are the most common fatal complications, especially in malnourished and immuno compromised children (Duke & Mgone, 2003; Griffin, 2007). Up to 15% of immuno competent adults with measles may also experience pneumonia as a direct result of the measles virus (MV) infection, as opposed to a secondary infection. Childhood blindness associated with keratitis and corneal lesions

and exacerbated by vitamin A deficiency, is a frequent complication of measles, especially in developing countries.

According to Asongo, A.I, Jamala, G.Y and Waindu, C, et al (2013), *explained that t*ime series analysis is a very useful aspect in statistics that is helpful and applicable in all field of human endeavor. Its primary purpose is discovering and measuring the various influences for the observed values and data obtained. These are useful in understanding the past behavioral pattern, evaluating current accomplishment, planning future operation and comparing different time series. The study of the past behavior of any observed data enables us to predict future tendencies, to (i.e. measles) is therefore of great assistance. For it is with the help and analysis of this data that approximately correct time to carryout immunization in the future will be known. In addition, the knowledge of the behavior of the variable enables statistician to iron but inter-year variation, thus, seasonal fluctuation may be reduced by taking effective decisions or plans before time. Generally, from the graph, the period of ten years of study from 1996 - 2005 did experienced that the cases of measles in Federal Medical Center is higher from the last to the first quarter in each successive year. This shows that there are serious cases of measles recorded at the end and the beginning of each year (November, December, January, February and March) in Federal Medical Center, Makurdi. November to March is known to be hot season in Makurdi metropolis. This clearly shows that measles occurs most during hot seasons.

Having reviewed the work done on measles cases in Nigeria, their scope, methodology, findings with a view to pointing out the gap in the literature, which Univariate time series analysis will help to accomplish by sufficiently forecasting, discussing results and making adequate recommendation based on findings.

CHAPTER THREE

RESEARCH METHODOLOGY

3.0 Introduction

This chapter outlines the methods and steps employed in carrying out this research work. It presents the model building strategy espoused in Box-Jekins (Box, Jekins & Reinsel (1994)).

3.1 Method of analysis

Statistical tool that will be used in this research work is the Univariate ARIMA model. Three stages are involved in this method and these are identification, estimation and diagnostic checking.

3.2 Model Identification

The ARIMA (auto-regressive, integrated, moving average) model of a time series is defined by three terms (p, d, q). Identification of a time series is the process of finding integer, usually very small (e.g., 0, 1, or 2), values of p, d, and q that model the patterns in the data. When the value is 0, the element is not needed in the model. The middle element, d, is investigated before p and q. The goal is to determine if the process is stationary and, if not, to make it stationary before determining the values of p and q. Recall that a stationary process has a constant mean and variance over the time period of the study.

In the simplest time series, an observation at a time period simply reflects the random shock at that time period, at, that is: Y_t.

The random shocks are independent with constant mean and variance, and so are the observations. If there is trend in the data, however, the score also reflects that trend as represented by the slope of the process. In this slightly more complex model, the observation at the current time, Y_t, depends on the value of the previous observation, Y_{t-1}, the slope, and the random shock at the current time period: $Y_t = Ø_0 (Y_{t-1}) + \varepsilon_t$.

To see if the process is stationary after linear trend is removed, the first difference scores at lag 1 are plotted against years, as seen in Figure 4.1.2. If the process is now stationary, the line will be basically horizontal with constant variance.

3.2.1 Auto-Regressive Components

The auto-regressive components represent the memory of the process for preceding observations. The value of p is the number of auto-regressive components in an ARIMA (p, d, q) model. The value of p is 0 if there is no relationship between adjacent observations. When the value of p is 1, there is a relationship between observations at lag 1 and the correlation coefficient $Ø_1$ is the magnitude of the relationship. When the value of p is 2, there is a relationship between observations at lag 2 and the correlation coefficient $Ø_2$ is the magnitude of the relationship. Thus p is the number of correlations you need to model the relationship. For example, a model with p = 2, ARIMA (2, 0, 0), is

$$Y_t = Ø_1 Y_{t-1} + Ø_2 Y_{t-1} = \varepsilon_t \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad 1$$

3.2.2 Moving Average Components

The moving average components represent the memory of the process for preceding random shocks. The value q indicates the number of moving average

components in an ARIMA (p, d, q). When q is zero, there are no moving average components. When q is 1, there is a relationship between the cur- rent score and the random shock at lag 1 and the correlation coefficient T_1 represents the magnitude of the relationship. When q is 2, there is a relationship between the current score and the random shock at lag 2, and the correlation coefficient T_2 represents the magnitude of the relationship. Thus, an ARIMA (0, 0, 2) model is

$$Y_t = \varepsilon_t - \emptyset_1\varepsilon_{t-1} - \emptyset_2\varepsilon_{t-2} \quad \ldots\ldots 2$$

3.2.3 Mixed Models

Somewhat rarely, a series has both auto-regressive and moving average components so both types of correlations are required to model the patterns. If both elements are present only at lag 1, the equation is:

$$Y_t = \emptyset_1 Y_{t-1} - \Theta_1\varepsilon_{t-1} + \varepsilon_t \quad \ldots\ldots 3$$

ACFs and PACFs

Models are identified through patterns in their ACFs (autocorrelation functions) and PACFs (partial autocorrelation functions). Both autocorrelations and partial autocorrelations are computed for sequential lags in the series. The first lag has an autocorrelation between Y_{t-1} and Y_t, the second lag has both an autocorrelation and partial autocorrelation between Y_{t-2} and Y_t, and so on. ACFs and PACFs are the functions across all the lags.

The equation for autocorrelation is similar to bivariate r (Equation 3.29) except that the over- all mean Y is subtracted from each Y_t and from each Y_{t+k}, and the denominator is the variance of the whole series.

$$r_k = \frac{\frac{1}{N-K}\sum_{i=1}^{N-K}(Y_1 - \bar{Y})(Y_{t-k} - \bar{Y})}{\frac{1}{N-K}\sum_{t=1}^{N}(Y_t - \bar{Y})^2} \quad \text{..... 4}$$

where N is the number of observations in the whole series, k is the lag, $\bar{Y}$ is the mean of the whole series and the denominator is the variance of the whole series.

The standard error of an autocorrelation is based on the squared autocorrelations from all previous lags. At lag 1, there are no previous autocorrelations, so r_0^2 is set to 0.

$$SE_{rk} = \sqrt{\frac{1 + 2\sum_{t=0}^{k-1} r_t^2}{N}}$$

The equations for computing partial autocorrelations are much more complex, and involve a recursive technique (cf. Dixon, 1992, p. 619). However, the standard error for a partial autocorrelation is simple and the same at all lags:

$$SE_{pr} = \frac{1}{\sqrt{N}}$$

Calculation of the partial autocorrelations after the first few is labor intensive. However, McCleary and Hay (1980) provided equations showing the following relationships between ACF and PACF for the first three lags.

PACF (1) = ACF (1)

$$\text{PACF (2)} = \frac{ACF(2) - (ACF(1))^2}{1 - [ACF(1)]^2}$$

$$PACF(3) = \frac{-2(ACF(1))ACF(2) - [ACF(1)]^2 ACF(3)}{1+2[ACF(1)]^2 ACF(2) - [ACF(2)]^2 - 2[ACF(1)]^2}$$h

If an autocorrelation at some lag is significantly different from zero, the correlation is included in the ARIMA model. Similarly, if a partial autocorrelation at some lag is significantly different from zero, it, too, is included in the ARIMA model. The significance of full and partial autocorrelations is assessed using their standard errors. Although you can look at the autocorrelations and partial autocorrelations numerically, it is standard practice to plot them. The center vertical (or horizontal) line for these plots represents full or partial autocorrelations of zero; then symbols such as * or _ are used to represent the size and direction of the autocorrelation and partial autocorrelation at each lag. You compare these obtained plots with standard, and somewhat idealized, patterns that are shown by various ARIMA models. The ACF and PACF for the first 16 lags of the measles cases recorded are seen in Figure 3 and fig. 4, as produced by SPSS ACF and PACF. The boundary lines around the functions are the 95% confidence bounds. The pattern here is a large, negative autocorrelation at lag 1 and a decaying PACF, suggestive of an ARIMA (1, 1, 1) model, as illustrated in Table 4.1.8. Recall, however, that the series has been differenced, so the ARIMA, model is actually (1, 1, 1). The series apparently has both linear trend and memory for the preceding random shock. That is, the measles cases recorded is generally increasing, however, the increase in one year is influenced by random events in the epidemic breakout from both the current and preceding years. The q value of 1 indicates that, with a differenced series, only the first of the two correlations will be estimated, the correlation coefficient $\Theta_1$²

Pattern of Autocorrelation

The pattern of autocorrelation is modelled in any time-series study, for itself, in preparation for fore- casting, or prior to tests of an intervention. Are there linear or quadratic trends in the data? Does the previous score affect the current one? The previous random shock? How quickly do autocorrelations die out over time? For the example, is the quality of the computer increasing steadily over the time frame? Decreasing? Is the quality of the computer produced in one time frame associated with the quality in the next time frame? How long do the random shocks in the manufacturing processes linger? Section 18.4.1.5 shows how the autocorrelation functions (ACFs) and partial autocorrelation functions (PACFs) are examined to reveal these patterns.

3.2.4 Seasonal Cycles and Trends

Time-series data are also examined for seasonal cycles if such are possible. Are there weekly, quarterly, monthly, or yearly trends in the data? For example, does the observed cases of measles vary systematically over the weeks of the year? The months of a year? As for auto-regressive and moving average components, ACFs and PACFs are examined to reveal seasonal cycles.

3.2.5 Model Estimation

Model estimation is the most straight forward part of modelling procedure. These involve the estimation, the fixed values of p and q of the parameters $\theta = (\propto_1, \propto_2, \ldots, \propto_p; \theta_1, \theta_2, \ldots, \theta_q)$ associated with model (1.2).

Several methods of estimating parameter of a given ARMA model have proposed in the literature. These are:

a. Generalized least square procedure

b. Maximum (Gaussian) likelihood method.

These method have been found to produce similar parameter estimate for fairly large samples sizes. For brevity, we shall concentrate on the operation of Gaussian likelihood procedure.

3.2.6 Guassian Likelihood Estimation Procedure

The procedure is achieved by optimizing a modified Guassian likelihood of the form

$$L_r(\theta) = \frac{1}{2} \log \sigma_{r^2}(\theta)$$

Where $\sigma_{r^2}(\theta) = T^{-1} \sum_{t=1}^{T} e_\theta(t)^2$ and the residual sequence,

$$e_\theta(t) = \sum_{s=0}^{t-1} \emptyset\ (s) y(t-s)$$

$$\emptyset\ (L) = \emptyset\ (L)^{-1} \emptyset(L) = \sum_{j>0} \emptyset\ (j) L^j$$

The Guassian estimator is given by $\theta_G = \arg\min L_T\ (\theta)$

Where $\theta_G = (\emptyset_1, \emptyset_2, \ldots\ldots, \emptyset_p; \theta_1, \theta_2, \ldots, \theta_q)^L$

The problem of evaluating θ_G is accomplished by considering the iterative scheme

$$\theta_{G^{j+1}} = \theta^j G_\Delta\ T(\frac{\theta}{G})^{-1} \sum_{t-1}^{T} \left(\frac{\partial e_{\theta(t)}}{\partial \theta} \right) |\ e_{\theta(t)} \theta = \theta_{G(j)}\ , \mathrm{j} = 0, 1, 2, \ldots\ldots$$

Where $\Delta_r\ (\theta_G) = T^{-1} \sum_{t=1}^{T} \{ \frac{((\partial e_{\theta(t))^L}}{\partial \theta} \frac{(\partial e_{\theta(t)}}{\partial \theta} \}$

The rate of convergence of this iterative scheme depends largely on the choice of $\theta^{(0)}G$, the initial parameter estimates, with which to commence the iteration. In the

light of this, an obvious choice for the starting value is $\theta_{G(0)} = \theta_{LS}$ *where* θ_{LS} is the least square estimates of θ.

3.2.7 Model Diagnostic Checking

Model diagnostic or model criticism is concerned with testing the goodness of fit of a model and if the fit is poor suggesting appropriate modifications. One useful approach for achieving this is through residual analysis. The major concern here is that the are systematically distributed across the series (e.g., they could be negative in the first part of the series and approach zero in the second part) or that they contain some serial dependency, which may suggest that ARIMA model is inadequate. The analysis of ARIMA residual constitute an important test of the model. The estimation procedure assumes that the residual are not auto-correlated and that they are normally distributed.

Suppose ($\widehat{\alpha_j}, j = 1, \ldots p,\ \widehat{\theta_j}\ \ j = (1, \ldots\ldots, q)$*are the* maximum (Guassian) likelihood or generalized least square estimates of ($\alpha_j j = 1, \ldots p, \theta_j\ j = (1, \ldots, q)$ of the tentatively identified model. The residual can be generated recursively as

$$\varepsilon_t = y_t - \alpha_1 y_{t-1} - \alpha_2 y_{t-2} \ldots - \alpha_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots \theta_q \varepsilon_{t-q},\ t = 1,2,\ldots\ T$$

taking pre sample values of y_t *and* ε_t *to be zero* $(i.e. y_t, \varepsilon_t = 0$ for $t \leq 0)$.

If the system parameter $(\widehat{\alpha_j}, s, \theta_j, s, \delta_{e^2})$ are close to the true values ($\alpha_0, \theta_0, \delta_0$) then the innovation series (ε_t) should have nearly the properties of independent identically distributed normal random variables with mean of zero and variance δ^2. *the first diagnostic check is to* inspect the plot of residual over time. It is expected that the plot should suggest a rectangular scatter around a zero horizontal level with no trend what so ever if the model is adequate.

A formal test for determining the adequacy of model is giving by computing the first K autocorrelation of residual series, $r_k(\hat{\varepsilon}), k = 1, 2, \ldots, k.$ for any ARIMA (p,d,q) process. If the fitted model is adequate, then the test statistics is

$Q = T(T+2)\sum_{k=1}^{k} \frac{r_k(\hat{\varepsilon})^2}{T-K}$,

And Q has a chi square distribution as $\chi^2(k-p-q)$ *distribution, see e.g. Ljung* (1986) and Ljung & Box (1979) for detailed particulars.

3.2.8 Forecasting

The main aim and objective for using time series techniques to analyze data is to predict or forecast its future values. In forecasting, we are concerned with the problem of predicting the values $(y_t, t > T+1)$ of ARIMA process in terms of ($y_t, t = 1,2, \ldots, T$). the idea is to utilize observations taken at or before time T to forecast subsequent behaviours of (y_t). Let's show how the minimum mean square error is derived.

3.2.9 minimum Mean Square Error Forecasts

Let an ARMA model be written as

$\sum_{j=0}^{p} \emptyset_j y_{t-j} = \sum_{j=0}^{q} \emptyset_j \varepsilon_{t-j}$, t = 1,2,3.....

Which can be written more compactly as

$(1 - \emptyset_1 B - \emptyset_2 B^2 - \ldots\ldots\ldots - \emptyset_p B^p)\, y_t = (1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots \ldots - \theta_q B^q)\varepsilon_t$

Since B is the back shift operator the model becomes

$y_t - \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} \ldots\ldots + \emptyset_p y_{t-p} = \in_t - \emptyset_1 \in_{t-1} - \emptyset_2 \in_{t-2} \ldots\ldots \emptyset_q B^q) \in_t$

Where

$$y_t = \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} \ldots\ldots + \emptyset_p y_{t-p} = \in_t - \emptyset_1 \in_{t-1} - \emptyset_2 \in_{t-2} \ldots - \theta_q \epsilon_{t-q}$$

Replacing t with t+i

$$y_{t+i} = \emptyset_1 y_{t+i-1} + \emptyset_2 y_{t+i-2} \ldots\ldots + \emptyset_p y_{t+i-p} = \in_{t+i} - \emptyset_1 \in_{t+i-1} - \emptyset_2 \in_{t+i-2} \ldots - \theta_q \epsilon_{t+i-q}$$

And consider one-step ahead forecast, i.e. i = 1

$$y_{t+1} = \emptyset_1 y_{t+1-1} + \emptyset_2 y_{t+1-2} \ldots\ldots + \emptyset_p y_{t+1-p} = \in_{t+1} - \emptyset_1 \in_{t+1-1} - \emptyset_2 \in_{t+1-2} \ldots - \theta_q \epsilon_{t+1-q}$$

Set Y = ($y_t, y_{t-1}, \ldots\ldots, y_1$) and taken the conditional expectation i.e. E(y_{t+1}/Y).

$$E(y_{t+1}/Y) = E\emptyset_1(y_t/Y) + E\,\emptyset_2(y_{t-1}/Y) + \ldots\ldots + \emptyset_p E(y_{t+1-P}/Y) + E(\in_{t+1}/Y) +$$

$$\emptyset_1 E(\in_t/y) + \emptyset_2 E(\in_{t-1}/y) \ldots\ldots \theta_q \epsilon_{t+1-q}/y)$$

We notice that $E(\in_{t+1}/y)$ is zero because it has not yet happened and its been replaced by its conditional expectation which is zero.

CHAPTER FOUR

DATA ANALYSIS

4.0 Introduction

In this chapter, the estimation procedure introduced in chapter three is applied to analyze the cases of measles recorded.

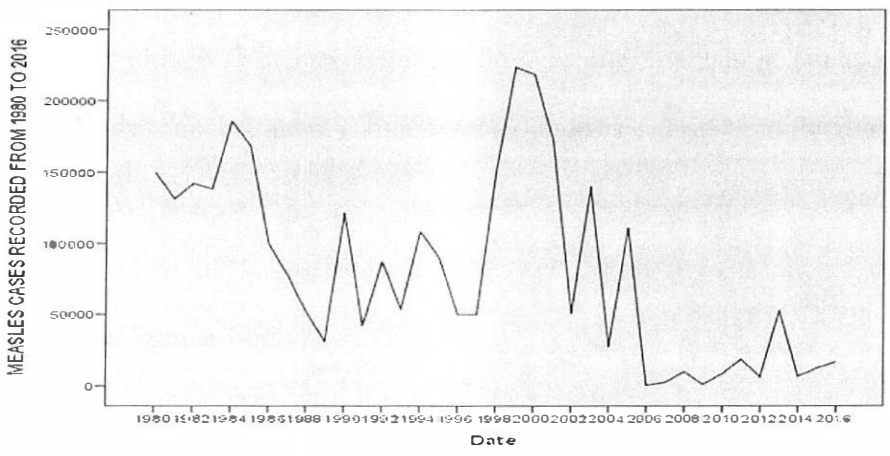
4.1 The Data

In this present research, the yearly totals of measles cases recorded in Nigeria between 1980 to 2016 are properly examined. Data used in this research work are considered to be of public records gotten from knoema website.

No missing data was observed in the data as the analysis of such data could create little or no problem. Time series plot was carried out to illustrate the scope of the data of the measles cases recorded in Nigeria for a period of 37 years. Statistical packages for social sciences (SPSS) was used and the plot is presented in Figure 1 below

Figure 1

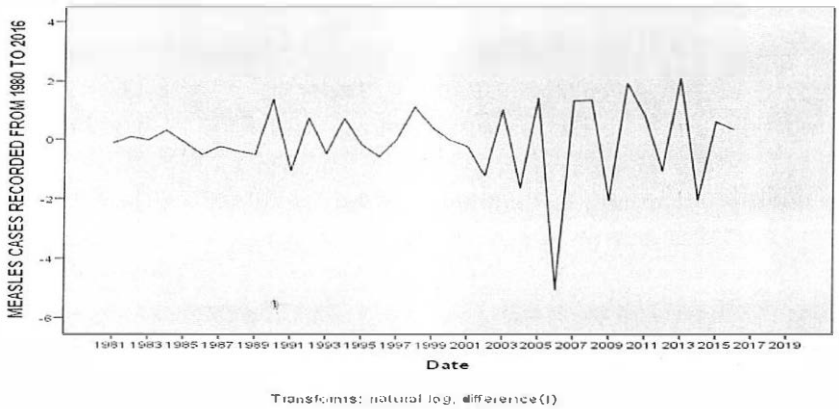
Time series Sequence Plot of Measles cases recorded in Nigeria between 1980 to 2016



Before applying the estimation procedure, it is necessary to study the data to see what behavior they can produce. Figure 1 above helps us to determine if the data (observed measles cases recorded) is stationary or not. It was observed that both the mean and the variance have possible shift over time in the series recorded which indicates that the data is not stationary. Since the mean is changing over time, the trend is removed by differencing once or twice and the variability process may be made stationary by logarithmic transformation. The plot of the differenced data was presented in figure 2

Figure 2

Plot for differenced data using natural logarithmic transformation for measles cases recorded from 1980 to 2016.



From figure 2 above, an examination of the plot above shows that some element of stationarity has been induced in the data. The series now appears stationary with respect to central tendency, therefore second differencing does not appear necessary.

4.2 Identification of Model

Table 1 below shows the plot values for autocorrelations calculated for different lags, standard error, box-Ljung statistics values and probability values using SPSS.

Table 1

Plot values for Autocorrelations calculated

Autocorrelations

Series: MEASLES CASES RECORDED FROM 1980 TO 2016

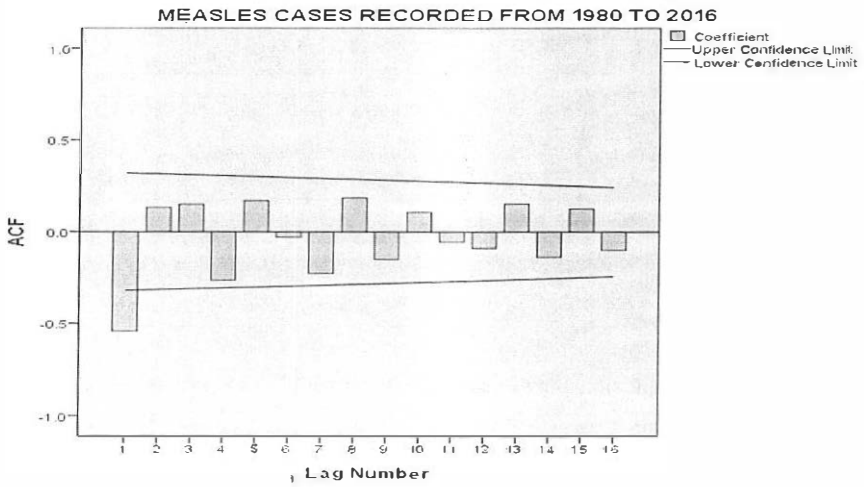
Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	Df	Sig.[b]
1	-.540	.160	11.417	1	.001
2	.132	.158	12.120	2	.002
3	.149	.155	13.042	3	.005
4	-.262	.153	15.983	4	.003
5	.170	.151	17.251	5	.004
6	-.030	.148	17.292	6	.008
7	-.227	.146	19.714	7	.006
8	.186	.143	21.404	8	.006
9	-.149	.140	22.528	9	.007
10	.106	.138	23.121	10	.010
11	-.054	.135	23.279	11	.016
12	-.088	.132	23.717	12	.022
13	.151	.130	25.073	13	.023
14	-.137	.127	26.237	14	.024
15	.125	.124	27.251	15	.027
16	-.098	.121	27.902	16	.032

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

Figure 3

ACF plot for differenced observed values of measles cases recorded



SPSS ACF syntax and output

Table 2

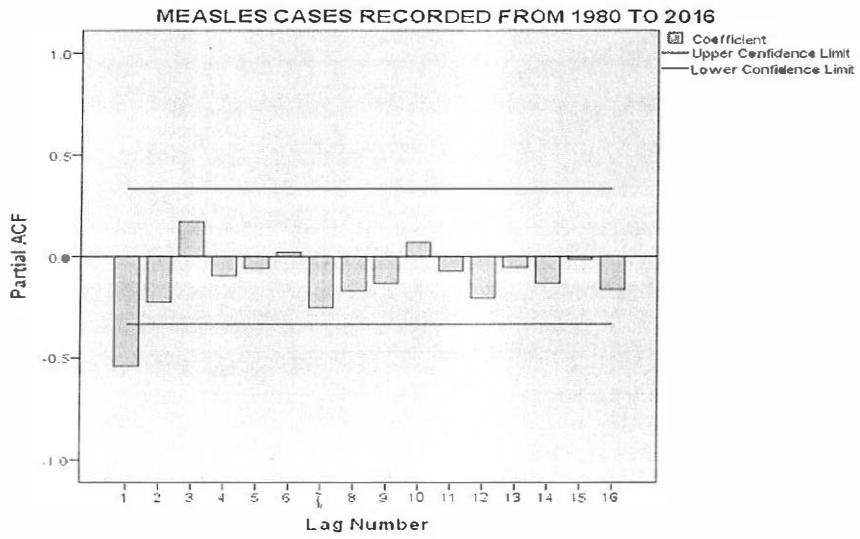
Table 2 shows the plot values for partial autocorrelations for differenced lags and standard error calculated.

Lag	Partial Autocorrelation	Std. Error
1	-.540	.167
2	-.226	.167
3	.171	.167
4	-.095	.167
5	-.058	.167
6	.021	.167
7	-.252	.167
8	-.168	.167
9	-.132	.167
10	.069	.167
11	-.072	.167
12	-.205	.167
13	-.054	.167
14	-.132	.167
15	-.015	.167
16	-.164	.167

Source: SPSS

Figure 4

Partial ACF plot for differenced observed values of measles cases recorded



SPSS ACF syntax and output

Summary of the identified model using transformed natural log, difference (1) value of ACF and PACF plot.

Fig. 3 shows the autocorrelation function for difference 1 in actual value calculated for measles_1 with one negative lag (-0.540) outside the lower confidence limit of the plot which indicate that we have AR = 1 with difference of I = 1. Fig. 4 also shows the partial autocorrelation function for difference in actual value calculated for measles_1 with one negative lag (-0.540) outside the lower confidence limit of the PACF plot which indicate that we have MA = 1 with difference of I = 1. From the autocorrelation and partial autocorrelation plot of lags, the ACF lag plot shows that AR = p = 1, with difference of 1. The Partial ACF plot of lags, also show that MA = q = 1 with difference of I = 1.

Therefore AR = 1, DIFF = I = 1, MA = q = 1

Which indicate that ARIMA (1, 1, 1) process can be used for the model prediction.

Table 3

Model Description

			Model Type
Model ID	MEASLES CASES RECORDED FROM 1980 TO 2016	Model_1	ARIMA(1,1,1)

Source: SPSS

Table 4

Model fit

Fit Statistic	Mean	Minimum	Maximum	Percentile						
				5	10	25	50	75	90	95
Stationary R-squared	.168	.168	.168	.168	.168	.168	.168	.168	.168	.168
R-squared	.409	.409	.409	.409	.409	.409	.409	.409	.409	.409
RMSE	51485.159	51485.159	51485.159	51485.159	51485.159	51485.159	51485.159	51485.159	51485.159	51485.159

MAPE	389.818	389.818	389.818	389.818	389.818	389.818	389.818	389.818	389.818	389.818
MaxAPE	10188.742	10188.742	10188.742	10188.742	10188.742	10188.742	10188.742	10188.742	10188.742	10188.742
MAE	37583.774	37583.774	37583.774	37583.774	37583.774	37583.774	37583.774	37583.774	37583.774	37583.774
MaxAE	134810.644	13481 0.644	134810.644	134810.644	13481 0.644	13481 0.644	13481 0.644	13481 0.644	13481 0.644	13481 0.644
Normalized BIC	21.897	21.897	21.897	21.897	21.897	21.897	21.897	21.897	21.897	21.897

Source: SPSS

4.3 Estimation of the Identified ARIMA Model

Table 4 above shows the model statistics output of the estimated parameter. Efficient parameter estimation were obtained using ARIMA (1,1,1) model for the measles cases recorded from 1980 to 2016, using SPSS

Table 5

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
MEASLES CASES RECORDED FROM 1980 TO 2016-Model_1	0	.168	22.117	17	.180	0

Source: SPSS

Table 6

SUMMARY OF ESTIMATED MODEL

MODEL ESTIMATED	MLE PARAMETER	BIC
ARIMA (1, 1, 1)	ϕ = -0.540 D = 1 Θ = -0.540	21.817
ARIMA (1, 1, 2)	ϕ = -0.540 d = 1 Θ_1 = - 0.540 Θ_2 = - 0.226	23.418
ARIMA (2, 1, 1)	ϕ_1 = -0.252 ϕ_2 = 0.132 d = I Θ_1 = - 0.540	22.319

From this table, by mere examination we can easily detect that the model that best fit the data, which is ARIMA (1, 1, 1).

4.4 Estimation of Parameters

The efficient parameter estimation were obtained using Numerical Algorithm Group (NAG) library subroutine G13 AFF. The iterative scheme converged after three iterations. The model estimated is

$$w_t + 0.542\, w_{t-1} = \hat{\epsilon}_t + 0.226\, \epsilon_{t-1} + 0.132\, \epsilon_{t-2}$$

4.5 Diagnostic Check

Since an appropriate model has been chosen and its parameters estimated, the appropriate diagnostic check was carried out. The diagnostic check is to inspect the plot of residual error is random over time. There was no discrepancy detected in the fitted model. The Ljung-Box Q statistics sig. value shows that the residual error are not random over time.

4.6 Forecasting with the Fitted Model

One of the objective for using time series techniques to analyze data is to predict or forecast its future values. In this research work, we are concerned with the

problem of predicting the values of expected measles cases in Nigeria. In order, to evaluate the forecast accuracy of the identified model, the data for 2017 to 2020 were predicted with the fitted model and the result is presented in table 6

Table 6

Forecast

Model		2017	2018	2019	2020
MEASLES CASES RECORDED FROM 1980 TO 2016-Model_1	Forecast	10162	7856	3685	3018
	UCL	114771	129879	149330	129230
	LCL	-94448	-114168	-141959	-92320

Source: SPSS

CHAPTER FIVE

SUMMARY OF FINDINGS

5.0 Introduction

In this chapter, we shall be discussing the findings obtained in chapter three

5.1 Summary

i. the measles cases recorded in Nigeria from 1980 to 2016 exhibit non-stationary behaviour and some form of transformation was carried out and indeed natural log difference was used and stationarity was attained.

ii. Box-Jenkins approach proved directly on the model by the use of Autocorrelation function (ACF) and Partial autocorrelation function (PACF).

iii. ARIMA (1,1,1) was used as best fit for the model with

$$w_t + 0.542\, w_{t-1} = \hat{\epsilon}_t + 0.226\, \epsilon_{t-1} + 0.132\, \epsilon_{t-2}$$

Summary of findings in this research work was done based on identification of ARIMA (p, d, q) Models using SPSS.

Summary

Table 1 and 2 shows the ACF and PACF calculated for different lag levels (measles_1) with one negative value (-0.540) lag outside the lower confidence limit and upper control confidence limit. Ljung-Box statistics was also calculated. The partial autocorrelation plot also shows that all lags level are within the lower and upper confidence limit except for the first lag of -0.540. Auto regressive integrated moving average ARIMA (1, 1, 1) was used for this research work. Other models like ARIMA (1, 1, 2) and ARIMA (2, 1, 1) was also tested resulting in higher BIC of 23.418 and 22.319 respectively compare to that of ARIMA (1, 1, 1) model with BIC

of 21.817. Table 4 and 5 also show the model summary (measles_1). Minimum average percentage error (MAPE) with the mean = 389.818, the Maximum average percentage error (measles_1) (MaxPE) with the mean = 10188.742, the goodness of fit measures (measles_1) with stationary R squared = 0.168, R squared = 0.409, root mean square error = 51489.159, mean absolute error = 37583.774, normalized BIC of 21.817.

The ARIMA (auto-regressive, integrated, moving average) model of time series in this research work is define by three terms (p, d, q). The process of finding integer, usually very small (e.g., 0, 1, or 2), values of p, d, and q that model the pattern in the data. When the value is zero, the element is not needed in the model. The middle element, d, is investigated before p and q. the goal is to determine if the process is stationary and if not, to make it stationary before determining the value of p and q. it was recall that a stationary process in this research work was based on constant mean and variance over the time period of the study.

ACF and PACF for the first 16 lags of differenced observed value as produced by SPSS. It was observed that lag 1 in autocorrelation has the highest negative value of -0.540 and highest positive value of 0.186 in lag 8. The PACF also recorded highest negative value of -0.540 at lag 1 and the highest positive value 0f 0.171 at lag 3. The boundary lines around the functions are the 95% confidence bounds. The pattern here is a large, negative autocorrelation at lag 1 and a decaying PACF, suggestive of an ARIMA model (1, 0, 1) and because the series has been differenced once, therefore the ARIMA model is actually (1, 1, 1).

Fig. 3 show the autocorrelation function for difference in actual value calculated for measles_1 with one negative lag outside the lower confidence limit of the plot which indicate that we have AR = 1 with difference of I = 1.

From the autocorrelation and partial autocorrelation plot of lags, the ACF lag plot shows that AR = p = 1, with difference of 1. The Partial ACF plot of lags, also show that MA = q = 1 with difference of 1 = 1.

Therefore AR = 1, DIFF = 1 = 1, MA = q = 1

Which indicate that ARIMA (1, 1, 1) process can be used for the model prediction.

Summary of findings was done based on SPSS output.

CHAPTER SIX

CONCLUSSION AND RECOMMENDATION

6.0 Introduction

This chapter contains conclusion and recommendation for this research work.

6.1 Conclusion

From the stationary R square (0.168), we can conclude that we have a good measures that compares the stationary part of the fitted model for measles_1 to a simple mean model. Positive value of stationary r square also indicates that the model under consideration is better than a baseline model. From R square calculated, we can say that only 40.9% estimate of the proportion of the total variation is in the series (measles_1) is explained by the model. We can also conclude from the value of Ljung-Box Q statistics that the residual errors measles_1 are random.

However, the forecasting accuracy based on the selected model ARIMA (1, 1, 1), is assumed to be the best selected model. The research study has found that measles data in Nigeria could be best modelled with ARIMA (1, 1, 1). The study again found out that measles prevalence in the country is expected to decrease if more measures are taken. The result of this research work can be used as a tool to facilitate the introduction of measles vaccine and improve measles vaccination in the country as whole.

5.2 Recommendation

Nigerian Health Service should continue the mass measles vaccination in the region to possibly eradicate the disease and if possible reduce it to the barest minimum.

References

Asongo, A.I, Jamala, G.Y and Waindu, C, *IOSR Journal of Research & Method in Education (IOSR-JRME) Volume 3, Issue 6 published (Nov. –Dec. 2013),*

Epidemiology of Measles in South West Nigeria: Fatiregun AA and Odega CC *et al* (2016).

Descriptive Analysis of Measles cases seen in Tertiary Health Facility "Global Journal of Medicine and Public health". Mohammed Yahaya, Kaoje Aminu Umar, Jiya Fatima Bello, Bello Abubarkar Gwandu, Jimoh Ahmed Kolawole, Raji M.O, Ango U.M, Nakakana Usman Nasir Yusuf Tahir and Ibrahim Bafa Sule.

The Impact of declining vaccination coverage on measles control: Umeh and Ahaneku the Pan Medical Journal 2013

Burden and trend of measles in Nigeria: Rabi Usman, Baffa Sule Ibrahim, Yahaya Mohammed and Patrick Nguku.

WHO. | Measles. WHO [Internet]. World Health Organization; 2017 [cited 2017 Apri 10] available form: http://www.who.int/mediancentre/factsheets/fs286/en/

Ibrahim BS, Gana GJ, Mohammed Y, Bojoga UA, Olufemi AA, et al. 2016 Outbreak of measles in Sokoto State North-western Nigeria, an investigation report 2016. Australas Med J. 9, 32435.10.4066/AMJ.2016.2697

Appendix

OBSERVATION	YEARS			
150000	1980			
131400	1981	146362	32252	260471
142100	1982	133738	29129	238348
138209	1983	132734	28124	237344
185604	1984	134672	30062	239281
168308	1985	161580	56970	266189
100108	1986	170125	65516	274735
78206	1987	122260	17650	226870
52304	1988	81863	-22746	186473
31200	1989	57559	-47050	162169
121305	1990	34539	-70071	139148
42103	1991	80219	-24390	184829
87300	1992	68650	-35960	173259
53500	1993	64154	-40456	168763
108450	1994	61910	-42699	166520
90060	1995	81408	-23202	186017
50008	1996	92314	-12295	196924
50000	1997	60916	-43694	165525
150062	1998	44911	-59698	149521
223430	1999	104999	389	209609
218200	2000	189030	84421	293640
172080	2001	215197	110588	319807
50601	2002	185412	80802	290021
140106	2003	94036	-10574	198646
27309	2004	99260	-5350	203870
110927	2005	67276	-37334	171886
704	2006	72433	-32177	177042
2613	2007	39643	-64967	144252
9960	2008	-3241	-107851	101368
1272	2009	1933	-102676	106543
8491	2010	-349	-104959	104261
18843	2011	516	-104094	105125
6447	2012	9616	-94994	114226
52852	2013	6307	-98303	110917
6855	2014	29223	-75387	133833
12423	2015	20138	-84472	124747
17136	2016	5107	-99503	109717
	2017	10162	-94448	114771
	2018	7856	-114168	129879
	2019	3685	-141959	149330
	2020	3018	-92320	129230

Case Processing Summary

		MEASLES CASES RECORDED FROM 1980 TO 2016
Series or Sequence Length		40
Number of Missing Values in the Plot	Negative or Zero Before Log Transform	3[a]
	User-Missing	0
	System-Missing	3

a. The minimum value is 704.000.

Model Description

Model Name		MOD_7
Series Name	1	MEASLES CASES RECORDED FROM 1980 TO 2016
Transformation		Natural logarithm
Non-Seasonal Differencing		1
Seasonal Differencing		0
Length of Seasonal Period		No periodicity
Maximum Number of Lags		16
Process Assumed for Calculating the Standard Errors of the Autocorrelations		Independence(white noise)[a]
Display and Plot		All lags

Applying the model specifications from MOD_7

a. Not applicable for calculating the standard errors of the partial autocorrelations.

Case Processing Summary

		MEASLES CASES RECORDED FROM 1980 TO 2016
Series Length		40
Number of Missing Values	Negative or Zero Before Log Transform	0
	User-Missing	0
	System-Missing	3[a]
Number of Valid Values		37
Number of Values Lost Due to Differencing		1
Number of Computable First Lags After Differencing		35

a. Some of the missing values are imbedded within the series.