Forecasting Maternal Mortality Ratio in Juba Teaching Hospital

Madalina Kaku Daniel Lado

I56/67332/2013

School of Mathematics

College of Biological and Physical Sciences

A research project submitted in the partial fulfilment of the requirement for the degree of Masters of Science in Social Statistics

University of Nairobi

July 2015

Declaration

Student Declaration

This project is my original work and has not been presented for any award of a degree in any other University.

Madalina Kaku Daniel Lado

I56/67332/2013

Signature..... Date.....

Supervisors Declaration

This project has been submitted for examination with our approval as supervisors.

Signature..... Date.....

DR. Nelson O. Onyango

School of Mathematics

Signature..... Date.....

Dr. Japheth O. Awiti

School of Economics

Dedication

I dedicate this work to my lovely husband Lazarus Pierentino Lugoi and our children, when God gives us. In addition, I dedicated this work to the sons and daughters of my brothers and sisters, especially to Emma Keji, Gabriel Amule, George Lwoki, Jele, and Isaac Wonge, who are struggling with their education, and to the entire family of late Loro Lwoki.

Acknowledgements

I want to thank the Almighty God for giving me the gift of life and the energy to undertake this master's degree. I am deeply indebted to my supervisors Dr. Nelson O. Onyango and Dr. Japheth O. Awiti for their support, advice, and encouragement throughout my research work. Their rich knowledge, experience and skills in mathematics and research work helped immensely in adding value to this paper. Special mention goes to my sponsors DAAD for financing my studies. I also want to thank the Administration of Juba Teaching Hospital and the Department of Gynecology for providing me with the data for this study. I am equally indebted to Prof. Weke, the Director of School of Mathematics, for having recognized me in his administration and all M.Sc lecturers for playing a key role during my studies. Last but not the least, I am obliged to thank all my colleagues for the teamwork and support, especially Arthur Yambayamba and Abdirahman Omar and all who might have contributed indirectly to the success of this study.

God Bless you all.

Abstract

In an effort to reduce global maternal mortality, all countries that gathered at the United Nation Millennium Summit in 2000 agreed to incorporate maternal mortality as MDG-5. This was intended to improve maternal health by reducing maternal mortality ratio by three quarters by 2015. South Sudan is one of the United Nation countries with the highest mortality rate compared to other Countries worldwide. This study was conducted to model and forecast maternal mortality ratio (MMR) at the Juba Teaching Hospital (JTH) using the ARIMA time series model for the period of January 2008 to December 2014. Within the study period, there were 135 maternal deaths and about 29,711 deliveries, which accounts to MMR of 454 per 100,000 live births. The ARIMA (3, 0, 1) model adequately fitted Maternal Mortality Ratio data and was able to forecast monthly Maternal Mortality ratios at the facility for the period of January 2015 to December 2015.

Contents

List of Tables

List of Figures

Acronym	
ACF	Autocorrelation Function
ADF	Augmented Dickey-Fuller
AIC	Akaike Information Criterion
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
BIC	Bayesian Information Criterion
CMLE	Conditional Maximum Likelihood Estimation
AIC_c	Corrected Akaike Information Criterion
D.F	Degree of Freedom
HQ	Hannan-Quinn Criterion
ICU	Intensive Care Unit
JTH	Juba Teaching Hospital
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
MDG	Millennium Development Goals
ME	Mean Error
MLE	Maximum Likelihood Estimation
MMR	Maternal Mortality Ratio
MSE	Mean Square Error
PACF	Partial Autocorrelation Function
RMF	Real Medicine Foundation
RMSE	Root Mean Square Error
UN	United Nation
ONS	Office for National Statistics
WHO	World Health Organization

Chapter 1

INTRODUCTION

Maternal mortality is defined as the death of a woman while she is pregnant or within 42 days after delivery from any cause related to or aggravated by the pregnancy or its management but not from accidental or incidental causes (Conde-Agudelo and Belizán, 2000).

Maternal Mortality Ratio (MMR) is usually expressed as the number of maternal deaths per 100,000 live births. It is influenced by a number of factors which include socio-cultural, socio-demographic and socio-economic factors, access to health care such as antenatal care and also her nutrition both in childhood and adulthood. Most of the maternal deaths are thus, preventable and this could be achieved through; adequate nutrition, proper health care, including access to family planning, and due presence of a skilled birth attendant during delivery and emergency obstetric care (Asia, 2013).

1.1 Background.

Maternal mortality is an essential indicator of maternal health in both developed and developing countries (Høj *et al* ., 2003).

Women in developing countries on an average have more pregnancies than their counterparts in developed countries; hence, their lifetime risk of death due to pregnancy is higher.

A womans lifetime risk of maternal death, the probability that a 15-year-old woman will eventually die from a maternal cause is, 1 in 3700 in developing countries, versus 1 in 160 in

developing countries (WHO, 2014).

In an effort to reduce worldwide maternal mortality, all the countries that gathered at the United Nation Millennium Summit in 2000, agreed to put maternal mortality as one of the eight Millennium Development Goals (MDGs). In particular, maternal mortality is placed as MDG 5 with the aim to improve maternal health, by reducing the maternal mortality ratio (MMR) by three quarters by 2015 MDG (Report, 2014).

Worldwide, maternal death has declined by 45% from 380 per 100,000 live births in 1990 to 210 per 100,000 live births in 2013, showing an average annual decline of 2.6% (UN, 2014).

The report further indicates that African countries have reduced their MMR from an average of 870 deaths per 100,000 live births in 1990 to 460 deaths per 100,000 live births in 2013, an average reduction of 47%.

Attainment of the MDG target of reducing maternal mortality by three-quarters will require accelerated efforts and stronger political backing for women and children. Improving maternal health is another key to achieving MDG 4 of reducing child mortality (UN, 2014; Asia, 2013).

The maternal mortality ratio for most of Sub-Sahara Africa countries are above the MDG target. However, the average regional figures show that maternal mortality has been decreasing; from 990 per 100,000 live births in 1990, 830 per 100,000 live births in 2000 to 510 per 100,000 live births in 2013 (UN, 2014).

According to WHO (2014), South Sudan has a higher maternal mortality ratio of 730 per 100,000 live births compared to the rest of the countries in the region; for example Kenya has 488 per 100,000 live births in the same report. This implies that MMR in South Sudan is higher when compared to worldwide, continental and regional estimates.

South Sudan with an estimated population of 9.6 million has only one referral hospital, Juba Teaching Hospital (JTH), a 580-bed facility located in Juba City. Due to lack of proper functioning primary health care facilities upcountry, many South Sudanese have nowhere to go to but this national referral hospital. Furthermore, military and police hospitals, if any, are non-functional country wide, hence soldiers and officers share the same limited facilities with civilians at JTH (RMF,2013).

1.2 Problem Statement.

MMR in South Sudan is among the highest in the world and more so in Sub-Sahara Africa. Being a new country, it has low levels of income in most families, low education levels among women, lack of access to appropriate reproductive health care, poor health infrastructure, inadequate medical supplies and insufficient human resource in existing health facilities. Recognizing that these factors could be some of the main factors leading to high MMR; Juba teaching hospital trained its medical staff in emergency obstetric care, increased its surgical supplies to handle complications, ensured better reproductive health services for adolescence and improved family planning care. Despite the new measures, MMR remained high.

Women and girls are a driving force in our economies and when women are healthy, they play a crucial role in the development of communities and nations in general.The attention and care given to women before, during and after pregnancy; inside and outside the health system reflects the relative value a society accords to women.Reducing MMR requires strengthening of health care system. This process takes time and must be fueled by public commitments; sustained by maternal death review and forecasting. One valuble entry point is research on pattern of MMR and forecasting.

Most of the researchers in the reviewed literature used odd ratio and descriptive methods in their analysis which are limited to their data and cannot be used to extrapolate any conclusion about a full population or used to forecast the future trends. This study accounted for this limitation by adopting an ARIMA model to explain the relation between the values from the past and use the model to predict the future values of the variable.

1.3 Research Questions.

1. What is the pattern of maternal mortality ratio (MMR) in Juba Teaching Hospital?
2. How will the MMR in the Juba Teaching Hospital be in the following twelve months after December 2014?

1.4 Objectives.

1.4.1 Main Objective.

To model and forecast maternal mortality ratio at Juba Teaching Hospital using the ARIMA time series model.

1.4.2 Specific Objectives.

1. To construct an appropriate ARIMA model depicting the pattern of MMR.
2. To forecast MMR in Juba Teaching Hospital for January 2015 to December 2015.

1.5 Justification of the Study.

There is no existing literature on forecasting maternal mortality ratio specifically in Juba teaching hospital and also the study accounts for the limitation of methods of analysis used by the existing literature. The study determines the pattern of MMR at the Juba teaching hospital thus providing vital information regarding prevalence of MMR in the health facility. The finding from the study provides important information to policy makers and other stakeholders on future trends of MMR on JTH and thus will assist them in formulation of policies and appropriate strategies for intervention to reduce MMR as well as evaluating and monitoring maternal health policies; making it a core national concern.

Furthermore, the project will provide a baseline for further research on MMR and related topics in the country in general.

Chapter 2

LITERATURE REVIEW

2.1 Introduction.

This chapter gives a brief summary of related literature. It starts with an empirical literature followed by an overview of the literature. The surveyed recent empirical literature is organized chronologically.

Osoro *et al.*(2014) investigated the antecedent factors predisposing women to maternal death in Kisii Level 5 Hospital using a descriptive retrospective study design, for the period ranging from January 01, 2009 to June 30, 2010. They found that out of 72, 42 maternal deaths were as a result of direct obstetric complications which included haemorrhage, post-partum sepsis, pre-eclampsia and abortion. Post partum hemorrhage was the most common complication which contributed to maternal deaths. Their findings also shows that out of 72 maternal deaths, 33 were as a result of indirect causes with peritonitis, heart disease, HIV/AIDS, anaemia, and convulsive disorder. Delayed access to transport, lack of money for user fees, and hospital distance were challenges that led to delay in accessing care. They concluded that lack of access to quality healthcare facility, poor health seeking behaviour and poor socio-economic factors were the main causes of maternal mortality. A study conducted by Savadogo *et al.*(2014) on Maternal Mortality Risk Factors in regional Hospital in Burkina Faso using univariate analysis revealed that the most significant risk factors that led to increased maternal Mortality were: age (women older than 35 and younger than 19 years),

distance from hospital (greater than 10km), multiple deliveries, few antenatal care visits (less than 3 visits), obstetrical maternal mortality direct causes and emergency reference.

Using an ARIMA model and quarterly data for a period of 2000 to 2010; Sarpong (2013) examined maternal mortality ratios at the Okomfo Anokye Teaching Hospital in Kumasi (Ghana) and found that ARIMA (1,0,2) model was the best to predict MMR for the period of 20 quarters since ARIMA (2,0,1) model was insignificant. The calculated MMR was 967.7 Per 100,000 live births. Kitui *et al.*(2013) investigated the factors influencing the place of delivery for women in Kenya. Using multivariate methods of analysis, they found that living in urban areas, being wealthy, more educated, using antenatal care services optimally, region, ethnicity, type of facilities used and lower parity strongly predicted where women delivered. They reported that women most commonly cited distance and/or lack of transport as reasons for not delivering in a health facility and concluded that physical access to health facilities through distance and/or lack of transport, and economic considerations were important barriers for women to delivering in a health facility in Kenya.

Mojekwu and Ibekwe (2012) carried out a study on maternal mortality in Nigeria using simultaneous stepwise multiple regressions. Their findings indicated that delivery by a skilled health professional and educational level of women had more effect in reducing the maternal mortality ratio than other factors. Fawole *et al.*(2012) examined the risk factors for maternal mortality in institutional births in Nigeria. Using stratified multi-stage cluster sampling strategy, they found that within the study period there were 79 maternal deaths and 8,526 live births which accounted for the maternal mortality ratio of 927 per 100,000 live births. Almost about one-fifth (20.5%) of women had no antenatal care while 79.5% had at least one antenatal visit during pregnancy. Four-fifths (80.5%) of all deliveries were normal deliveries. Elective and emergency caesarean section rates were 3.1% and 11.5% respectively. Lack of antenatal care, parity, level of education, and mode of delivery were significantly associated with maternal mortality. Low maternal education, high parity, emergency caesarean delivery, independently predicted maternal mortality. Socio-economic factors have much impact in increasing MMR as high-income countries have lower ratios than low-income countries (Bhutta *et al.*, 2012). They conducted a study on reducing maternal, newborn, and infant

mortality globally and found an estimate of about 500 deaths per 100,000 live births in low-income compared to 4 per 100,000 live births in high income countries. The health related factors that escalate MMR included hypertensive diseases, sepsis/infections, obstructed labor, and abortion-related issues.

Koch *et al.* (2012) used autoregressive models (ARIMA) to assess the main factors related to maternal mortality reduction in Chile. Their findings revealed that the most significant factor for the increase in MMR was abortion. The analysis further revealed that women's education level among other factors; significantly contributed to the decrease in MMR in Chile during the study period. The calculated MMR was 102.3 per 100,000 live births. Ahmed *et al.*(2011) estimated maternal mortality at the Sub-national level in Bangladesh, using an empirical Bayesian prediction method to provide a model based method of estimating maternal mortality ratios in all 64 districts. They found MMR to be ranging from 158 per 100,000 live births to 782 per 100,000 live births. The highest MMR was especially in Sylhet (678 per 100,000 live births) followed by Habiganj (654 per 100,000 live births). According to the study carried out by Hogan *et al.*(2010) on maternal mortality for 181 Countries, using robust analytical methods. The researchers found that there were 342,900 maternal deaths worldwide in 2008, down from 526,300 in 1980. The global MMR decreased from 422 per 100,000 live births in 1980 to 320 per 100,000 live births in 1990 and to 251 per 100,000 live births in 2008. The global maternal deaths show a rate of decline from 1.8% between 1980 and 1990 to 1.4% from 1990 to 2008. There results also shows that HIV epidemic had not contributed to substantial increases in maternal mortality in eastern and southern Africa. Baby (2010) investigated institutional factors affecting Maternal Mortality in Faidpur Medical College Hospital in Bangladesh.Using a cross sectional analysis, he estimated MMR as 2,010.5 per 100,000 live births and the most common contributing factors were delayed or non-attendance by senior doctors, unavailability of intensive care unit (ICU) as well as lack of infrastructures of the health centre. Høj *et al.* (2003) used a prospective survey to assess demographic and obstetric risk factors for pregnancy-related death in a multi ethnic rural population in Guinea-Bissau. Their results indicated that maternal mortality ratio increased with increasing distance from the regional hospital, multiple pregnancy and stillborn fetus. They also found that women living in the region of Gabu had higher mortality than those

living in Biombo. They concluded that screening approach of antenatal care is of limited value for the purpose of reducing maternal mortality. Adopting a descriptive analysis, Cham (2003) analysed the socio-cultural, economic and health service factors contributing to maternal deaths in rural Gambia. He found that bad experience with the health care system, the delay in reaching an appropriate medical facility, lack of transportation or prolonged transportation, seeking care at more than one medical facility and delay in receiving prompt and appropriate care after reaching the hospital are significant causes of Maternal mortality. Health service factors were the most frequently identified contributing factors to maternal deaths and therefore, concluded that improving the quality of and accessibility to emergency obstetrical care services would significantly contribute to the reduction of maternal deaths in the area.

Conde-Agudelo and Belizán (2000) conducted a cross sectional study in Latin America and the Caribbean to investigated the impact of inter-pregnancy interval on maternal morbidity and mortality. Using Crude and adjusted odds ratios they found that women with inter-pregnancy intervals of 5 months or less had higher risks for maternal death, third trimester bleeding , premature rupture of membranes, puerperal endometritis, and anaemia while those with inter-pregnancy intervals longer than 59 months had significantly increased risks of pre-eclampsia and eclampsia; both intervals compared with women with inter-pregnancy intervals of 18 to 23 months. They concluded that inter-pregnancy intervals less than 6 months and longer than 59 months were associated with an increased risk of adverse maternal outcomes.

2.2 Overview of the literature.

Some of the factors identified by the researchers in the literature review as the major contributes of maternal mortality ratio includes: age, distance from hospital, multiple deliveries, few antenatal care visits , obstetrical maternal mortality direct causes, mode of delivery, abortion-related issues, seeking care at more than one medical facility, and emergency reference. In conclusion, the main causes of maternal mortality are lack of access to quality healthcare, poor health seeking behaviour and poor socio-economic factors and therefore,

improving the quality of and accessibility to emergency obstetrical care services would significantly contribute to the reduction of maternal deaths. Reducing MMR requires strengthening of health care system; a process which takes time and requires public commitments as well as maternal death review and forecasting. One valuable entry point is research on pattern of MMR and forecasting. Most of the researchers in the reviewed literature used odd ratio and descriptive methods in their analysis which are limited to their data and cannot be used to extrapolate any conclusion about a full population or used to forecast the future trends. This study accounted for this limitation by adopting an ARIMA model to explain the relation between the values from the past and and use the model to predict the future values of the variable.

Chapter 3

METHODOLOGY

3.1 Introduction.

This chapter describes the theoretical background of the models applied in this study, formulations and the methods of analysis to the data available for the fulfilment of the objectives of the study. It focuses on the Box-Jenkins methodology for constructing ARIMA models and carrying out forecasting.

3.2 Data Source and Type.

This study aimed at modelling and forecasting the patterns of maternal mortality ratios at the Juba Teaching Hospital (JHT). The analysis was based on secondary data available at the Department of Statistics, Directorate of Obstetrics and Gynaecology of JTH. The data collected include monthly records of maternal deaths and monthly records of live birth deliveries for the period January 2008 to December 2014. From the data, we computed maternal mortality ratios using the following expression:

$$MaternalMortalityRatio = (Maternaldeaths \div livebirths) \times 100,000 \qquad (3.1)$$

We calculation of MMR on monthly and yearly basis for the study period. The data had only one variable under study, which was maternal mortality ratio; related to its past values and

current and past error terms. Therefore univariate time series models were purely stochastic processes and do not have explanatory variables. The data points for the study were 84 monthly observations under the seven years duration.

Table 3.1: **Data Structure and Source**

Date months	**Maternal deaths**	**Live births**	**MMR per 100,000 live births**
Jan 2008	1	166	602
⋮	⋮	⋮	⋮
Dec 2014	3	482	622

3.3 The Concept of Time Series.

The special feature of time-series data is that successive observations are usually not independent and so the analysis must take account of the order in which the observations are collected. The series is a time dependent sequence X_t, where t belongs to the set of integers and denotes the time steps. If the time series can be expressed as a known function, $X_t = f(t)$, then it is said to be a deterministic time series. If it is expressed as $X_t = Y(t)$, where Y is a random variable then X_t is a stochastic time series. The main objectives of time-series analysis are; description, modeling, Forecasting and control. Time series analysis aims to decomposes the variation in a time series into several components of trend, periodic and stochastic (Chatfield, 2000; Sarpong, 2013).

3.3.1 Stationary Time Series.

A time series is stationary if its mean, variance and auto-covariance are time invariant; thus, they do not change over time. It remains constant or unchanged. The time series is made stationary, so that we can study its behaviour over the period of consideration and also for the purpose of forecasting.

3.3.2 Non-Stationary Time Series.

A time series is said to be non-stationary, if its mean, variance and auto-covariance changes over the time periods. Thus it is not constant throughout the time.

3.4 The Unit Root Test for Stationarity.

The unit root test is done by Augmented Dickey Fuller (ADF) test which is the same as Dickey-Fuller test, but it is carried out in the context of the model. The advantage of the Augmented Dickey-Fuller test for unit root is that it can accommodate higher order autoregressive process in ε_t (Dickey and Fuller, 1979).

Considering the model below is an $AR(1)$

$$X_t = \rho X_{t-1} + \varepsilon_t \tag{3.2}$$

where ε_t are independent random errors with zero mean and constant variance σ_ε^2. And X_t is a random walk, whereby the value of the series tomorrow is X_{t+1} with its unpredicted change of ε_{t+1} and its value today is X_t .

The Zero Mean.

Now by substituting the values of $t = 1, 2, \ldots, T$ into equation (3.2) and if ρ=1 becomes:

$$X_1 = X_0 + \varepsilon_1$$

$$X_2 = X_1 + \varepsilon_2$$

substituting X_1, we gets X_2

$$X_2 = X_0 + \varepsilon_1 + \varepsilon_2)$$

$$X_3 = X_0 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3$$

$$\vdots$$

$$X_t = X_0 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \ldots + \varepsilon_t$$

which is equal to

$$X_t = X_0 + \sum \varepsilon_t$$

Therefore taking expectation on both sides of the equation

$$E(X_t) = E(X_0 + \sum \varepsilon_t) = X_0$$

$$var(X_t) = E(X_0 + \sum \varepsilon_t)^2 = E(\sum \varepsilon_t)^2 = t\sigma^2$$

This implies, that the mean of X_t is equal to a constants but if t increases, its variance also increases indefinitely such a condition is stationary.

Dickey-Fuller Test.

From equation (3.2)

$$X_t = \rho X_{t-1} + \varepsilon_t (-1 \leq \rho \leq 1)$$

If ρ=1

$$X_t = X_{t-1} + \varepsilon_t$$

$$X_t - X_{t-1} = \varepsilon_t$$

Now using the lag operator L $(1 - L)X_t = \varepsilon_t$, if $(1 - L)$= 0 , we get L=1.

The hypotheses for Dickey- Fuller test are now as follows:

$H_0 : \rho = 1$ (the series is non-stationary and has unit root)

$H_1 :\mid \rho \mid \leq 1$ (the series is stationary and has no unit root)

If we subtract X_{t-1} from both sides of equation (3.2):

$$X_t - X_{t-1} = \rho X_{t-1} - X_{t-1} + \varepsilon_t$$

$$\Delta X_t = (\rho - 1)X_{t-1} + \varepsilon_t$$

let $(\rho - 1)$ equal to γ then it becomes

$$\Delta X_t = \gamma X_{t-1} + \varepsilon_t \tag{3.3}$$

where $\gamma = \rho - 1$

$H_0 : \gamma = 0$ (the series is non-stationary and has unit root)

$H_1 : \gamma \neq 0$ (the series is stationary and has no unit root).

Decision on Null and Alternative hypotheses.

If, we fail to reject the null hypothesis, we conclude that the series is non-stationary and has unit root. But if we reject the null hypothesis then we conclude that the series is stationary and has no unit root.

Decision on computed P-value.

If the computed P-value is greater than 5%, we fail to reject the null hypothesis and we conclude that the series is non-stationary and has unit root. But if the computed P-value is less than 5% we reject the null hypothesis and conclude that the series is stationary and has no unit root.

Decision on critical value and test statistic.

If the absolute value of ADF test statistic is less than the critical value, we fail to reject the null hypothesis and conclude that the series is non-stationary and has unit root. But if the absolute value of ADF test statistic is greater than the critical value we reject the null hypothesis and conclude that the series is stationary and has no unit root. The critical value of ADF test were tabulated by the statisticians' professor David Dickey and professor Wayne Fuller, the Dickey-Fuller critical values are more of negative values that why we are taking absolute value (Hill, *et al.*, 2011).

3.5 Box and Jenkins ARIMA Methodology.

The ARIMA model was introduced by Box and Jenkins (also known as Box-Jenkins mode) in 1960. It is an extrapolation method for forecasting and, like any other such method, it requires only the historical time series data on the variable under forecasting. ARIMA models are the most important for classification of models for forecasting a time series data. Usually the ARIMA model is represented as ARIMA (p, d, q) where p is the number of autoregressive terms, d is the number of non- seasonal differences, and q is the number of lagged forecast errors in the prediction equation (Shrivastav and Ekata, 2012).

The use of ARIMA models is also known as the Box-Jenkins approach; following the work of Box and Jenkins (Box G. and Jenkins, G. 1976) as cited by Atkinson (2005) and according

to (ONS, 2008) .

Box and Jenkins (1976) propose a four-step iterative approach to modeling as follows:

1. Model identification.
2. Model parameter estimation.
3. Model checking (goodness of fit).
4. The forecasting.

The four iterative steps are not straightforward, but are embodied in a continuous path depending on the set of data under study (Sarpong,2013).

In practice, most of the time series are non-stationary. In order to fit a stationary model, it is necessary to remove non-stationary sources of variation by the mean and variance of the original data by differencing the series. Differencing is widely used for econometric data. Such a model is called an Integrated model, because the stationary model that is fitted to the difference data has to be integrated to provide a model for the original non-stationary data (Chatfield, 2004).

It is an extrapolation method for forecasting and, like any other such method, it requires only the historical time series data on the variable under forecasting. ARIMA models are the most important for classification of models for forecasting a time series data.

3.5.1 Assumptions of ARIMA.

The Box-Jenkins model assumes that the time series is stationary; and stationary series has the following assumptions:

1. Normality of Distributions of Residuals ($\varepsilon_t \sim N(\mu, \sigma^2)$).

   The normality of residuals is evaluated in time-series analysis by the normalized plot of residuals for the model before evaluating an intervention. They are independent and normally distributed, with mean zero and homogeneity of variance. It implies that the correlation in the series of observations has been removed to be adequate for the model.

2. Homogeneity of Variance and Zero Mean of Residuals .

   The homogeneity of variance and zero mean is done by the plots of standardized residuals versus predicted values to assess homogeneity of variance over time. But if there is heterogeneity of variance in the series, one can apply a logarithmic transformation to remove these from the series.

3. Independence of Residuals ($\varepsilon_t \sim iidN(0, \sigma^2)$) .

   Once the model is developed and residuals are computed, there should be no remaining autocorrelations or partial autocorrelations at various lags in the ACFs and PACFs; to show that the residual are a white noise process.

3.5.2 Model Specifications.

3.5.2.1 Autocorrelation Function (ACF) and Correlogram.

The autocorrelations at lag $1, 2, 3, \ldots, k$ make up the autocorrelation function (ACF). The plot of ACF against the lag is called a correlogram and helps us visualize the ACF easily. It also behaves a standard tool in exploring a time series before forecasting and it provides a useful check for seasonality, cycles, and other time series patterns. The autocorrelation at lag k, ACF (k), is the (linear) Pearson correlation between observations k time periods (lags) apart. If the ACF (k) differs significantly from zero, the serial dependence among the observations must be included in the ARIMA model (Box and Jenkins, 1976).

ACF is denoted by ρ_k ;

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{cov(X_t, X_{t-k})}{var(X_t)} \tag{3.4}$$

where $-1 \leq \rho_k \leq 1$

To compute this, we must first compute the sample covariance at lag k,$\hat{\gamma_k}$ and the sample variance $\hat{\gamma_0}$ which are defined as

$$\hat{\gamma_k} = \frac{\sum(X_t - \bar{X})(X_t - \bar{X})}{n}$$

$$\hat{\gamma_0} = \frac{\sum(X_t - \bar{X})^2}{n}$$

where n is the sample size and $\bar{X}$ is the sample mean. Therefore, the sample autocorrelation function at lag$k$ is $\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$. A plot of ACF was used in this study to determine the p order of AR model by observing the number of significant spikes in the correlogram.

3.5.2.2 Partial Autocorrelations Function (PACF).

Like the ACF (k), the partial autocorrelation at lag k, or PACF (k), measures the correlation among observations k legs apart. We use the partial autocorrelation function in this study to determine q orders of MA model respectively, by observing the number of significant spikes. Denoted by ϕ_{kk}, PACF is defined as;

$$\phi_{kk} = corr[X_t - E^*(X_t|X_{t-1,\ldots,X_{t-k+1}}), X_{t-k}] \tag{3.5}$$

Where

$$E^*(X_t|X_{t-1,\ldots,X_{t-k+1}})$$

is the minimum mean-squared error predictor of X_t by $X_{t-1}, \ldots, X_{t-k+1}$. In summary, the properties of ACF and PACF of an ARIMA model are illustrated in the table (3.2)

Table 3.2: **Properties of ACF and PACF**

	AR(p)	**MA(q)**	**ARMA (p,q)**
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tail off

Source: Box and Jenkins, 1976

3.5.2.3 Lag.

Lag is a difference in time between an observation and a previous observation. X_{t-k} lags X_t by k periods.

3.5.3 Model Formulation.

3.5.3.1 Autoregressive (AR) Model: AR (p).

AR models are based on the relationship between the current variables of the series X_t , and it previous lagged values of the series plus the current error term. That is, AR (p) is the autoregressive model of order p is of the form:

$$X_t = C + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \ldots + \beta_p X_{t-p} + \varepsilon_t \tag{3.6}$$

In general, AR (p) is of

$$X_t = C + \sum_{i=1}^{p} \beta_i X_{t-i} + \varepsilon_t$$

Where c is the constant, X_t is actual values, and β_i, $(i = 1, 2, 3, \ldots, p)$ are the model parameters and p is the order of the model ε_t is a white noise process which is independent and identically normally distributed $(i.i.d)$ random variables with mean zero and constant variance σ_ε^2 .

3.5.3.2 Moving Average (MA) Model: MA (q).

MA models account for the possibility of a relationship between a variable and the residuals from previous periods. It is the alternative to the autoregressive representation in which the X_t on the left-hand side of the equation are assumed to be a linear combination of the moving average model of order q, assuming that the white noise ε_t on the right-hand side of the defining equation are combination of linear form of the observed data; that is:

$$X_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{3.7}$$

In general, $MA(q)$ is of the form:

$$X_t = \mu + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

where μ is the constant value, θ_j at $(j = 1, 2, 3, \ldots, q)$ are the model parameters and ε_t is white noise series which is independent and identically normally distributed $(i.i.d)$ random variables with mean zero and constant variance σ_ε^2 .

3.5.3.3 Autoregressive Moving Average ARMA (p, q) Model.

This is a general development of autoregressive, moving average, or sometime called mixed autoregressive moving average (ARMA), models for stationary time series. The ARMA (p, q) model, if it is stationary is defined as:

$$X_t = \mu + \sum_{i=1}^{p} \beta_i X_{t-i} + \sum_{j-1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{3.8}$$

where,X_t, at $t = 1, 2, \ldots, n$ is the series being modelled (maternal mortality ratio in the case of this study)

$p =$ is the number of AR Parameters

$\beta_i =$ is the i^th AR parameter

$q =$ is the number of MA parameters

$\theta_j =$ is the j^th MA parameters

$\varepsilon_t =$ is the residual series

The important assumptions involved in such models are that (ε_t) has Zero mean with terms which are uncorrelated and form an independently identically distributed random variable. i.e $\varepsilon_t \sim iidN(0, \sigma^2)$ (Mujumdar and Kumar, 1990; Shumway and Stoffer, 2006).

3.5.3.4 Autoregrssive Integrated Moving Average ARIMA (p, d, q) Model.

The ARIMA model was introduced by Box and Jenkins (also known as Box-Jenkins mode) in 1960. It is an extrapolation method for forecasting and, like any other such method, it requires only the historical time series data of underlying variable for forecasting. ARIMA models are the most important for classification of models for forecasting a time series data. Usually the ARIMA model is represented as ARIMA (p, d, q) where p is the number of autoregressive terms, d is the number of non- seasonal differences, and q is the number of lagged forecast errors in the prediction equation (Shrivastav and Ekata, 2012). If the variable under the study is stationary at level, $I(0)$ or at first difference $I(1)$ determines the order of integration, which is called as ARIMA model.

3.5.4 Model Fitting or Estimation.

The general structure form of an ARIMA model is expressed as follows:

$$\bigtriangledown^d X_t = \mu + \beta_1 \bigtriangledown^d X_{t-1} + \ldots + \beta_p \bigtriangledown^d X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon t - q \tag{3.9}$$

where,

$$\bigtriangledown X_t = X_t - X_{t-1} = (1 - L)X_t$$

This can be written in the general form:

$$\beta(L) \bigtriangledown^d X_t = \bar{\mu} + \theta(L)\varepsilon_t$$

where $\bigtriangledown = 1 - L$

$$\beta(L) = 1 - \beta_1 L - \ldots - \beta_p L^p, \beta_p \neq 0$$

$$\theta(L) = 1 - \theta_1 - \ldots - \theta_q L^q, \theta_q \neq 0$$

where $\beta(L)$ and $\theta(L)$ are polynomials in the lag operator, and have no common root,L is defined such that $L^n X_t = X_{t-n}$

Therefore, the model that will be fitted for this study was

$$X_t = \mu + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \ldots + \beta_p X_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + {}_t \tag{3.10}$$

ε_t = a white noise process.

X_t = the observation under study (maternal mortality ratio)

AR = Autoregressive polynomial $\beta()$, and of the

MA = Moving average polynomial $\theta()$.

3.5.5 Estimation of Model Parameters.

The main approaches to fitting Box-Jenkins models are non-linear least squares and maximum likelihood estimation. Maximum likelihood estimation is generally the most preferred technique. The parameters for this study were estimated using the conditional maximum likelihood estimation for time series. For ARIMA models, L is a function of β_i =

$(\beta_1, \beta_2, \beta_3, \ldots, \beta_n)$, $\theta_i = (\theta_1, \theta_2, \theta_3, \ldots, \theta_n)$, and σ_ε^2 given the observations $X_1, X_2, \ldots, X_n$. The conditional maximum likelihood estimators are defined as those values of the parameters which maximize the likelihood function (Greene, 2012).

The CMLE approach of estimation of parameters of an ARMA (p, q) model is hereby outlined;

$$X_t = \beta_1 X_{t-1} + \beta_2 X_{t-1} + \ldots + \beta_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \tag{3.11}$$

Where $\varepsilon_t \sim$ iidN(0,σ_ε^2) white noise, the joint probability density of $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ is as follows

$$p(\varepsilon|\beta, \mu, \theta, \sigma_\varepsilon^2) = (2\pi\sigma_\varepsilon^2) - \frac{n}{2} exp[-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^{n} \varepsilon_t^2] \tag{3.12}$$

Rearranging equation (3.10) becomes

$$\varepsilon_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + X_t - \beta_1 X_{t-1} - \ldots - \beta_p X_{t-p} \tag{3.13}$$

Therefore, we can write down the likelihood function of the parameters $(\beta, \mu, \theta, \sigma_\varepsilon^2)$. Considering the series of $X = (X_1, X_2, X_3, \ldots, X_n)$ and assume the initial conditions $X_* = (X_{1-p}, \ldots, X_{-1}, X_0)$ and $\varepsilon_* = (\varepsilon_{1-q}, \ldots, \varepsilon_{-1}, \varepsilon_0)$. The conditional log-likelihood function is given by

$$\ln L(\beta, \mu, \theta, \sigma_\varepsilon^2) = -\frac{n}{2} \ln 2\pi\sigma_\varepsilon^2 - \frac{S_*(\beta, \mu, \theta)}{2\sigma_\varepsilon^2} \tag{3.14}$$

Where

$$S_*(\beta, \mu, \theta) = \sum_{t=1}^{n} \varepsilon_t^2(\beta, \mu, \theta|X_*, \varepsilon_*, X) \tag{3.15}$$

is the conditional sum of squares function. The quantities of $\hat{\beta}$,$\hat{\mu}$ and $\hat{\theta}$ which maximize equation (3.13) are called the conditional maximum likelihood estimators. $\ln L(\beta, \mu, \theta, \sigma_\varepsilon^2)$ Involves the data only through $S_*(\beta, \mu, \theta)$, these estimators are the same as the conditional least squares estimators found from minimizing the conditional sum of squares functions $S_*(\beta, \mu, \theta)$,which contain the parameter σ_ε^2. The conditional sum of squares function in equation (3.14) becomes

$$S_*(\beta, \mu, \theta) = \sum_{t=p+1}^{n} \varepsilon_t^2(\beta, \mu, \theta|X) \tag{3.16}$$

After obtaining the parameter estimates of $\hat{\beta}$,$\hat{\mu}$ and $\hat{\theta}$,the estimate $\hat{\sigma}_{\varepsilon}^2$ of σ_{ε}^2 is calculated from

$$\hat{\sigma}_{\varepsilon}^2 = \frac{S_*(\beta, \mu, \theta)}{d.f} \tag{3.17}$$

Where by the degrees of freedom $(d.f)$ equals the number of terms used in the sum of $S_*(\beta, \mu, \theta)$ minus the number of the parameters estimated. If equation (3.16) is used to calculate the sum of squares (Wei, 1990), $d.f = (n-p) - (p+q+1) = n - (2p+q+1)$

3.5.6 Model Selection Criteria.

There are several model selection criteria like, Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (AIC_c), Hannan-Quinn criterion (HQ) and Bayesian Information Criterion (BIC). Based on these methods of model selection criteria, this study used AIC , AIC_c and BIC, the model with the least AIC value will be selected. Using R-software for plotting the graphs and analysis of the data set. The mathematical formulations for the model selection criteria used in this study are as follows:

1. Akaike Information Criterion (AIC):

$$AIC = \ln(\hat{\sigma}_{\varepsilon}^2) + \frac{2k}{n} \tag{3.18}$$

2. Bayesian Information Criterion (BIC):

$$BIC = \ln(\hat{\sigma}_{\varepsilon}^2) + \frac{k}{n}\ln(n) \tag{3.19}$$

where $k = (p + d + q)$ number of estimated parameters,n is number of observations used for estimation and $\hat{\sigma}_{\varepsilon}^2$ is the estimated variance (Dobre and Alexandru, 2008;Greene, 2012)

3.6 Model Diagnostics (Goodness of Fit).

After determining the appropriate lag length, the model diagnostics using Box-Jenkins model is the same as model validation for nonlinear least squares fitting. We use diagnostic test to differentiate whether a time series appears to be non auto correlated.

3.6.1 Diagnostic Test for Residuals.

Box- Jenkins model is considered to be a good model for the data if the residuals should satisfy these assumptions: the residuals should be white noise (or independent when their distributions are normal), should be drawn from a fixed distribution with a homogeneous mean and variance.

1. The Box-Pierce (1970) Q- statistic.

   The Box-Pierce test statistic is used for the sum of squares of the residual autocorrelations. If this test statistic exceeds some critical $t - value$ (found in a table), then the model is declared to be inadequate.

   $$Q = n \sum_{k=1}^{k} \hat{\rho}_k^2 \tag{3.20}$$

   H_0: the residuals are independent and identically distributed

   H_1: the residuals are not independent and indectically distributed

2. Ljung Box (1979) statistic.

   $$Q' = n(n+2) \sum_{k=1}^{k} \frac{\hat{\rho}_k^2}{n-k} \tag{3.21}$$

   the same hypothesis implies on the Box-Pierce test statistic.

3.6.2 Diagnostic Test for the Parameters of the Model.

Model testing was conducted on the basis of the P- value or calculating t test, if the t statistic is greater than 1.96, then we reject the null hypothesis and concluded that the model parameters are statistically significant.

$$t = \frac{\hat{\beta}_i}{s.e(\hat{\beta}_i)} \tag{3.22}$$

$$var(\hat{\beta}_i) = \frac{1 - (\hat{\beta}_i)}{n}$$

$$\hat{\beta}_i \pm 2\sqrt{var(\hat{\beta}_i)}$$

H_0 : all parameters coefficints of β and $\theta = 0$

H_1 : all parameters coefficients of β and $\theta \neq 0$

3.7 Forecasting.

Forecasting is concerned with the process used to predict the unknown. That is to say, it is used to predict the unknown future (time series forecasting), but sometimes we make predictions about people, firms or other objects that are a cross-section forecasting. The field includes the study and application of judgment as well as quantitative I (Statistical) methods (Afzal, et-al., 2002).

Time series forecasting thus can be termed as the act of predicting the future by understanding the past. Therefore, one of the most popular and frequently used stochastic time series models is the Autoregressive Integrated Moving Average (ARIMA) model.

Forecasting for this study was done for the period January 2015 to December 2015, using the estimated model. From equation (3.10) , the following are the forecasting equations: (Nielsen, 2005)

$$X_t = \mu + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \ldots + \beta_p X_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

$$X_{t+1} = \mu + \beta_1 X_t + \beta_2 X_{t-1} + \ldots + \beta_P X_{t-p+1} + \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q+1} + \varepsilon_{t+1}$$

$$X_{t+2} = \mu + \beta_1 X_{t+1} + \beta_2 X_{t-1} + \ldots + \beta_P X_{t-p+2} + \theta_1 \varepsilon_{t+1} \theta_2 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q+2} + \varepsilon_{t+2}$$

$$\vdots$$

$$X_{t+12} = \mu + \beta_1 X_{t+11} + \beta_2 X_{t+10} + \ldots + \beta_P X_{t-p+12} + \theta_1 \varepsilon_{t+11} \theta_2 \varepsilon_{t+10} + \ldots + \theta_q \varepsilon_{t-q+12} + \varepsilon_{t+12}$$

Measuring Accuracy of Forecast.

In most of the forecasting situation, accuracy is treated as the overriding criterion for selecting a forecasting method. The word accuracy is referring to goodness of fit, how good the forecasting model is able to reproduce the data for future from the known observations. To the consumers of the forecasts, it is the accuracy of the future forecast that is most essential. E.g if X_t is the actual observation for time period t and F_t is the forecast for the same period, then the error is defined as (Hyndman and Koehler, 2006) Thus, a scaled error is defined as

$$\pi_i = \frac{\varepsilon_i}{\frac{1}{n-1}\sum_{t=2}^{n} |X_t - X_{t-1}|}$$

The Mean Absolute Scaled Error is simply

$$MASE = E|\pi_i|$$

When $MASE < 1$, the proposed method gives, on average, smaller errors than the one-step errors from the naïve method.

Various measures have been proposed for assessing the predictive accuracy of forecasting models. The most important of these measures are designed to evaluate expost forecasts (Greene,2012;Gor,2009)

1. Mean absolute percentage error (MAPE).

   The mean absolute percentage error (MAPE) is also known as mean absolute percentage deviation (MAPD), is a measure of accuracy of a method for constructing fitted time series values in Statistics , specially in trend estimation. It usually ecpresses accuracy as a percentage, and is defined by the formula:

   $$MAPE = \frac{1}{n}\sum_{t=1}^{n} \mid \frac{(A_t - F_t)}{A_t} \mid \times 100 \tag{3.23}$$

   Where A_t is the actual value and F_t is the forecast value.

2. Root mean squared error (RMSE) and mean absolute error (MAE) is also known as mean absolute deviation (MAD).

   These two measures are based on the residuals from the forecasts . And it is calculated using the following formulas

   $$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(e_i)^2} \tag{3.24}$$

   $$MAD = \frac{1}{n}\sum_{i=1}^{n}|e_i| \tag{3.25}$$

   Where n is the number of period being forecasted.

3. The mean square error (MSE).

   It is common for two forecasting models to be ranked differently depending on the

accuracy measures used. The MSE gives more weight to large errors because they are squared,in these case it is not often used in most practice.

$$MSE = \frac{\sum_{i=1}^{n} |{e_i}^2|}{N-1} \tag{3.26}$$

$$MAD = \frac{\sum_{i=1}^{n} |e_i|}{n} \tag{3.27}$$

4. Mean error (ME) and mean forecast error (MFE) .

   These two measures are computed only for the last half of the data. The forecasting models are evaluated by dividing the data in to two parts. The first part is used to fit the forecasting model and the second part of the data is used to test the model and is called the forecasting sample.Therefore, ME is denoted by e_i

$$ME = A_t - F_t \tag{3.28}$$

$$MFE = \frac{\sum_{i=1}^{n} (e_i)}{n} \tag{3.29}$$

   Notes that if $MFE > 0$, model tends to under - forecast and $MFE < 0$, model tends to over - forecast.

Accuracy in forecasting model is really irrelevant while, accuracy in the forecasting sample is more important because the pattern of the data often changes over time. The forecasting sample is used to evaluate how well the model tracks such changes (Gor *et al.* 2009).

Chapter 4

DATA ANALYSIS AND RESULTS

4.1 Summary of Calculated MMR for the Study Period.

Tables 4.1 and table 1 in the appendix show the calculated MMR on yearly and monthly basis respectively.

Table 4.1: **Calculated Annual MMR**

Year	**Maternal deaths**	**Live births**	**MMR per 100,000 live births**
2008	17	3115	546
2009	29	3315	875
2010	9	3925	229
2011	23	4088	563
2012	10	5120	195
2013	25	4972	503
2014	22	5176	425
Total	135	29711	3336

The results in table 4.1 shows the highest MMR calculated for the hospital was 875 per

100,000 live births, in 2009 while the lowest MMR calculated was 195 per 100,000 live births, recorded in 2012. The average yearly MMR recorded within the study period was 476.6 per 100,000 live births.The table 1 in the appendix shows the monthly calculated MMR for the hospital from the period of January 2008 to December 2014.

From the calculated values, the highest MMR was 1,835 per 100,000 live births recorded in the month of April 2011 while the lowest MMR was 0 which was spread throughout the study period. In particular, the year 2010 had MMR equal to 0 for four consecutive months; that is from February to May. The maternal death within the study period was 135 and the live births were 29,711 contributing to total MMR of 454.4 per 100,000 live births. The average monthly MMR recorded within the study period was 500.7 per 100,000 live births for the period of January 2008 to December 2014.

4.2 Patterns of Maternal Mortality Ratios(MMR).

Figure 4.1, shows that the monthly MMRs recorded during the study period has no trend and the general behaviour was rather irregular.This is also confirmed in the decomposition of the time series data as shown in the Figure 4.2.

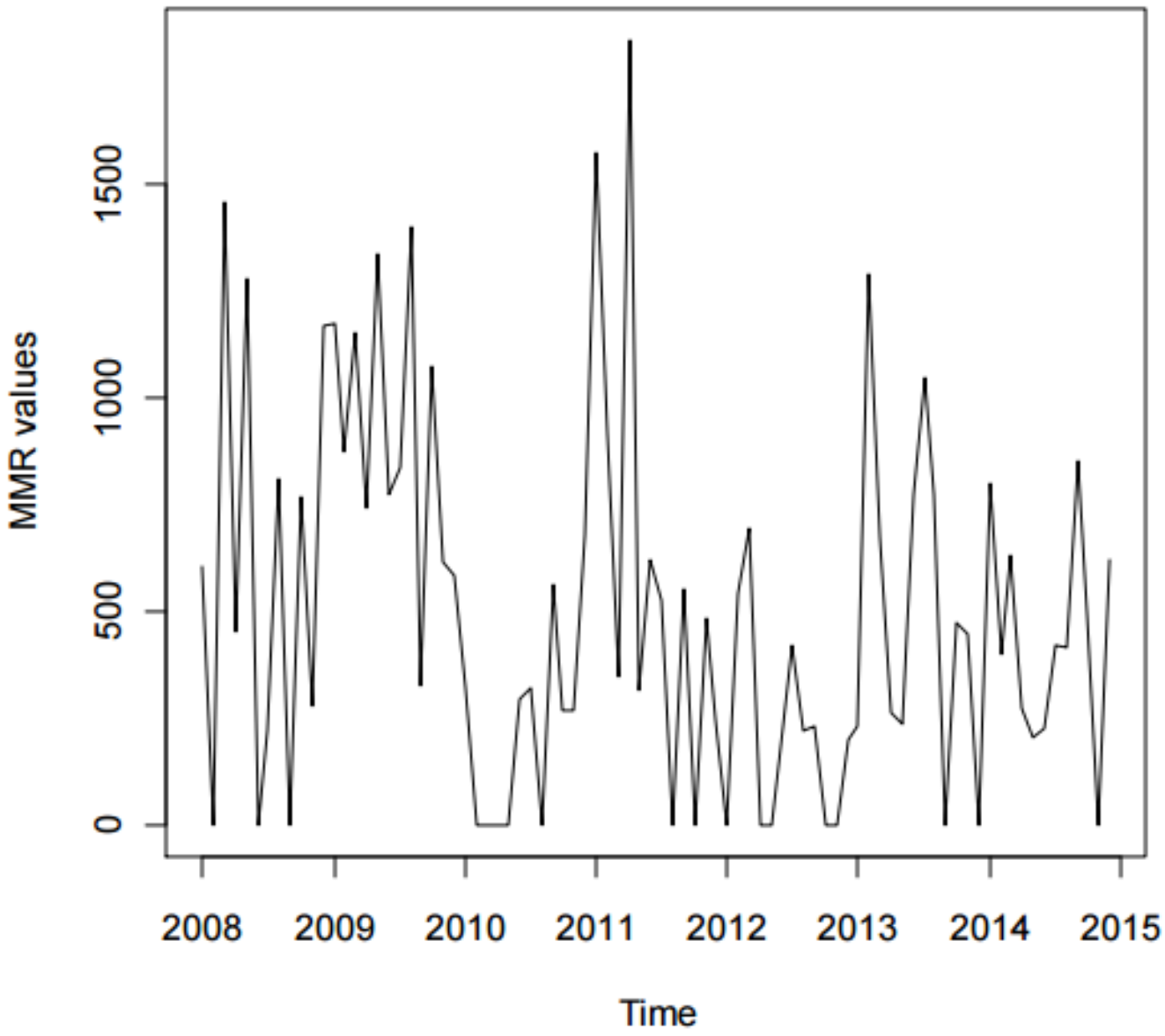

Figure 4.1: Plot of Calculated MMR Patterns from January 2008 to December 2014.

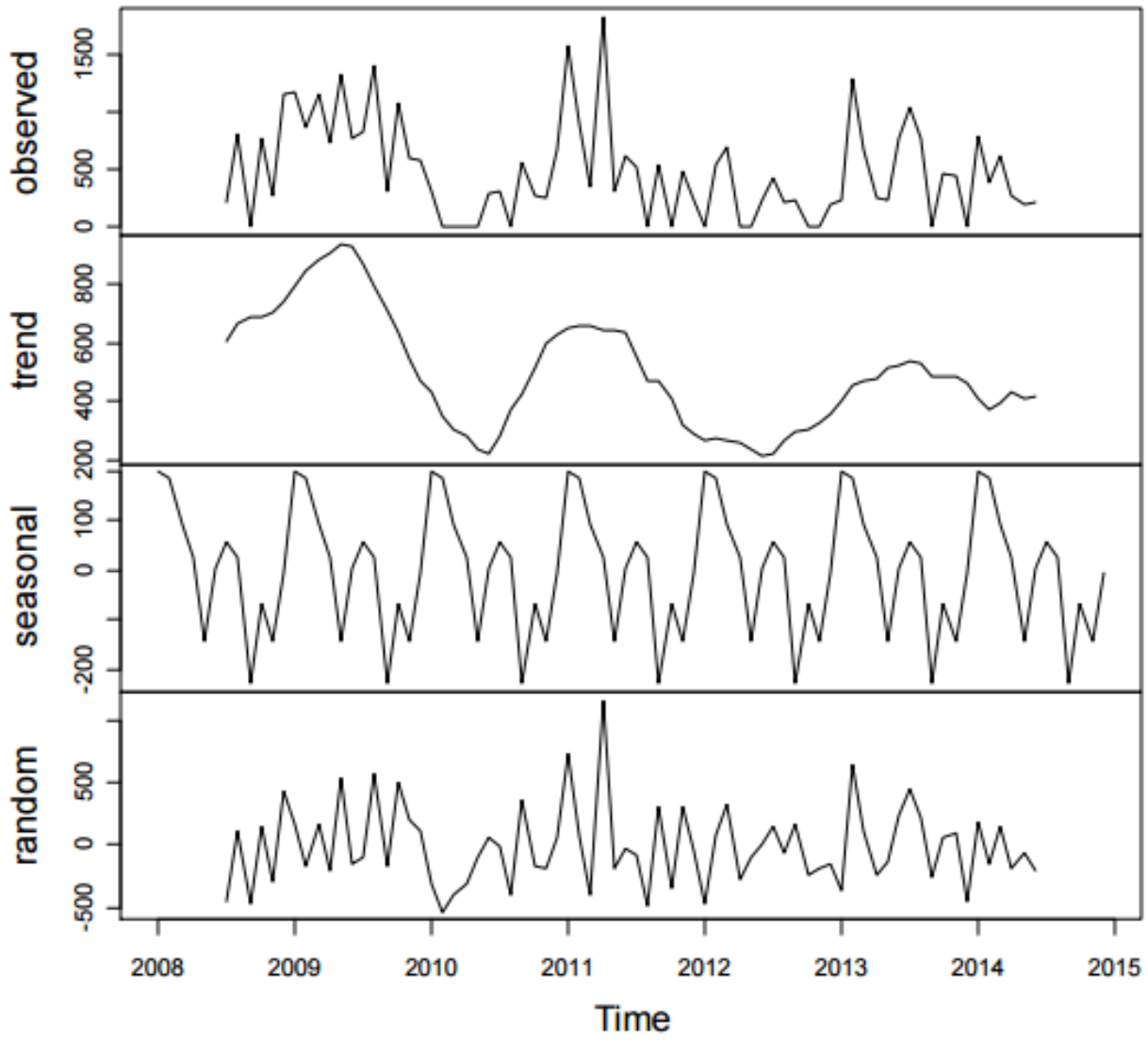

Figure 4.2: Decomposition of the Time Series Data.

4.3 Test for Stationarity of Time Series Data.

Table 4.2: **Augmented Dickey-Fuller Test**

DF Statistic	**lag order**	**P-value**
-7.9934	0	0.01
-4.7007	1	0.01

The results for test of stationarity of time series data is as shown in table 4.2. The calculated Augmented Dickey-Fuller test (ADF), Dickey-Fuller is -7.9934, at the lag order of 0 and the p-value is 0.01 and also ADF value was -4.7007, at the lag order of 1 with the p-value of 0.01. Based on the p-values, the null hypothesis is rejected and the conclusion is that the series of observation is stationary and has no unit root. The data is stationary with zero order differences or integrated of order zero $I(0)$.

4.4 Model Identification.

In identification of an appropriate ARIMA model, we make use of the original plot of the time series data as shown in figure 4.1, ACF and PACF in figure 4.3.

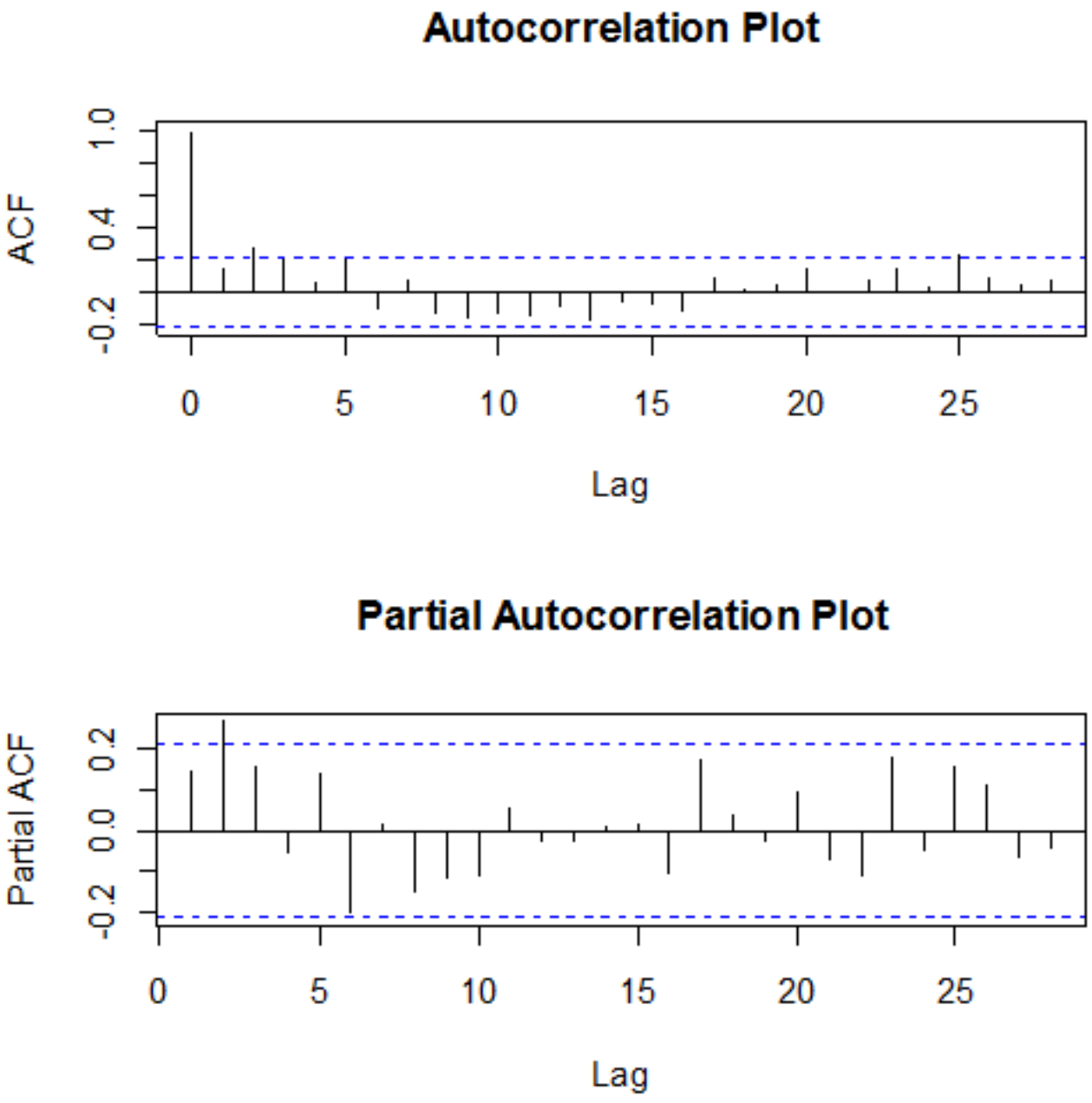

Figure 4.3: Plot ACF and PACF of Monthly MMR.

The ACF plot shows three (3) significant spikes exceeding 95% confidence interval, whereas the PACF plot of the series appears to have one (1) significant spike that exceeds 95% confidence interval. Therefore an AR (3) and MA (1) model is identified based on the data.

4.5 Model Estimation and Evaluation.

The process for choosing the best model depends on choosing the one with minimum AIC, AIC_c and BIC. The models constructed are presented in table 4.3 with their corresponding AIC, AIC_c and BIC.

Table 4.3: **Model Fit Based on AIC and BIC for the Suggested ARIMA**

Model	**AIC**	AIC_c	**BIC**
ARIMA (3,0,1)	**1073.75**	**1075.04**	**1087.41**
ARIMA (2,0,1)	1083.04	1083.95	1094.42
ARIMA (1,0,1)	1084.91	1085.51	1094.02
ARIMA (2,0,0)	1083.02	1083.62	1092.13
ARIMA (1,0,0)	1087.81	1088.16	1094.64
ARIMA (0,0,1)	1088.24	1088.59	1095.07
ARIMA (3,0,0)	1082.15	1083.06	1093.52
ARIMA (0,0,2)	1086.66	1085.26	1093.77
ARIMA (1,0,2)	1082.1	1083	1093.48

From table 4.3 ARIMA (3, 0, 1) was chosen as the appropriate model that fit the data well. This is because it has the minimum AIC, AIC_c and BIC among the different ARIMA models constructed.

Table 4.4: **Estimated Parameters for ARIMA (3, 0, 1)**

Variable	**Coefficient**	**Standard Error**	**t-value**	**p-value**
Intercept	505.106	88.5867	5.7018	<.0001
β_1	-0.7741	0.1049	-7.3794	<.0001
β_2	0.3375	0.1325	2.5472	<.0001
β_3	0.4609	0.1106	4.1673	<.0001
θ_1	0.9944	0.2350	4.2315	0.0000
		σ^2 =141644	log likelihood = -530.87	

Using the method of maximum likelihood estimation, the estimated parameters of ARIMA (3,0,1) model with their corresponding standard error, t-value and the p-values are shown in the Table 4.4. The coefficients of the model are significantly different from zero (based on comparison of absolute values to t-test critical value of 1.96). The p-values and the x^2 -test value of 18.5057 for the estimates also show that all the parameters are statistically significant at 0.05 level of significance. Therefore, we conclude that ARIMA (3, 0, 1) model satisfies the stability condition of good fit. The results from table 4.5 show the accuracy of the model fitted, based on MASE value of 0.6075, since it is less than 1, the model gives on average smaller error.

Table 4.5: **In-Sample Measures of Error for ARIMA (3,0,1) Model**

Accuracy measurements	**Corresponding values**
ME	2.4456
RMSE	376.3564
MAE	282.9222
MPE	2.2133
MAPE	190.2704
MASE	0.6075

4.6 Goodness of Model Fit.

In time series modeling, the choice of a better model fit to the data is directly related to whether the residual analysis is done well. One of the assumptions of ARIMA model is that, for a better model, the residuals must follow a white noise procedure. That is, the residuals are independent identically distributed with mean zero, constant variance, and the residuals are uncorrelated.

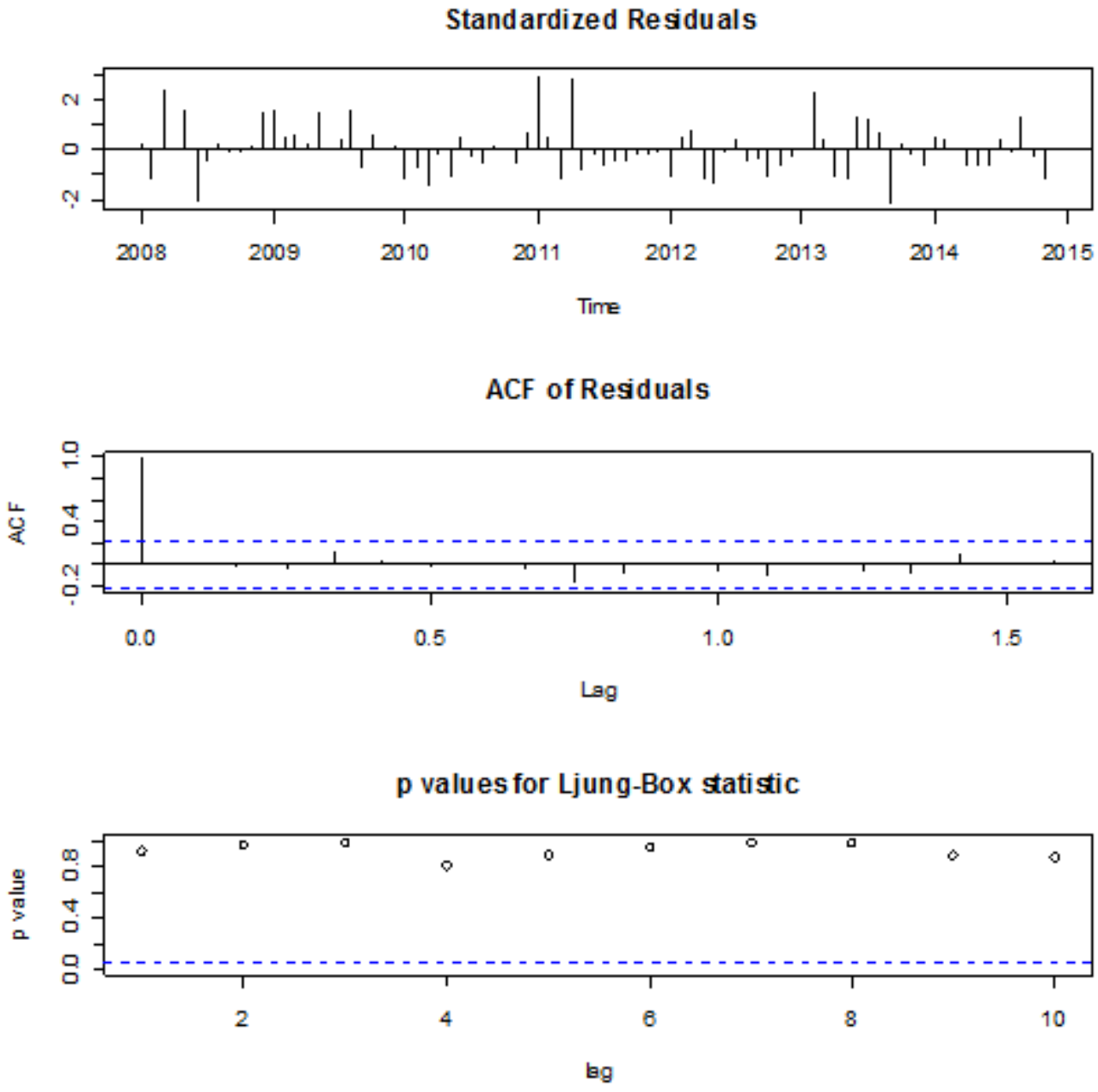

Figure 4.4: Diagnostic and Residual Plots of ARIMA (3,0,1) Model.

Figure 4.4 showing the standardized residual reveals that the residuals of the model have zero mean and constant variance and ACF plot of the residuals and the probability plot of the residuals respectively shows the residuals appear to be random and uncorrelated. The ACF plot shows no evidence of a significant spike. Finally, the results from Box-Pierce statistic clearly exceed 5% at lag orders 10 based on p-value of 0.9935; we fail to reject the null hypothesis and conclude that, there is no significant departure from white noise for the residuals. Thus, the selected model satisfies all the model assumptions and therefore, ARIMA (3, 0, 1) is a white noise process we can use this model to make forecasts. The constructed model is of thus:

$$X_t = 505.106 - 0.7741X_{t-1} + 0.3375X_{t-2} + 0.4609X_{t-3} + 0.9944\varepsilon_{t-1} \tag{4.1}$$

4.7 Forecasting.

Forecasting is the procedure of estimating the unknown situations, which is usually used in discussion of time-series data. It is a planning tool which helps decision makers to predict the future uncertainty based on the behaviour of past and current observations. For intuitive notion, short-term forecasting should be more reliable than long-term forecasting.

Table 4.6: **ARIMA (3,0,1) Forecasting Results for In-Sample and Out- of-Sample.**

Year	**Months**	**Actual MMR**	**Predicted MMR**	95 % Intervals	
				Lower limit	Upper limit
2014	Sep	853	453.15	-443.73	1350.03
	Oct	493	527.75	-369.18	1404.67
	Nov	0	497.86	-400.90	1396.62
	Dec	622	494.41	-405.55	1394.37
2015	Jan		521.38	-380.00	1422.77
	Feb		485.56	-416.28	1387.39
	Mar		520.79	-381.21	1422.81
	Apr		493.86	-408.15	1395.87
	May		510.09	-391.94	1412.13
	Jun		504.68	-397.46	1406.83
	Jul		501.93	-400.30	1404.17
	Aug		509.72	-392.59	1412.03
	Sep		500.27	-402.06	1402.61
	Oct		508.94	-393.39	1411.28
	Nov		502.63	-399.71	1404.97
	Dec		506.09	-396.25	1408.43

Using the estimated model in equation (4.1), we forecast MMR for the last four months of the data set from the Aug 2014 to Dec 2014 and compared to its actual values to measure the forecast accuracy of the model. Since the actual value for the month of Oct (493) MMR is near to it predicted value (527.75) and since both actual and the predicted values of the forecast data fall between the 95% confidence interval, we rather concluded that the ARIMA

(3,0,1) model is adequate to be used to forecast monthly MMR at the JTH.

The results in table 4.5 and figure 4.5 summarize the forecasting values of monthly MMR over the period of Aug 2014 to Dec 2015 with their respective 95% confidence intervals. When converted to one year, the estimated MMR for forecast period was 505.495 per 100,000 live births, showing an increase of 5.7% from an annual figure of 476.57 per 100,000 live births for the study period.

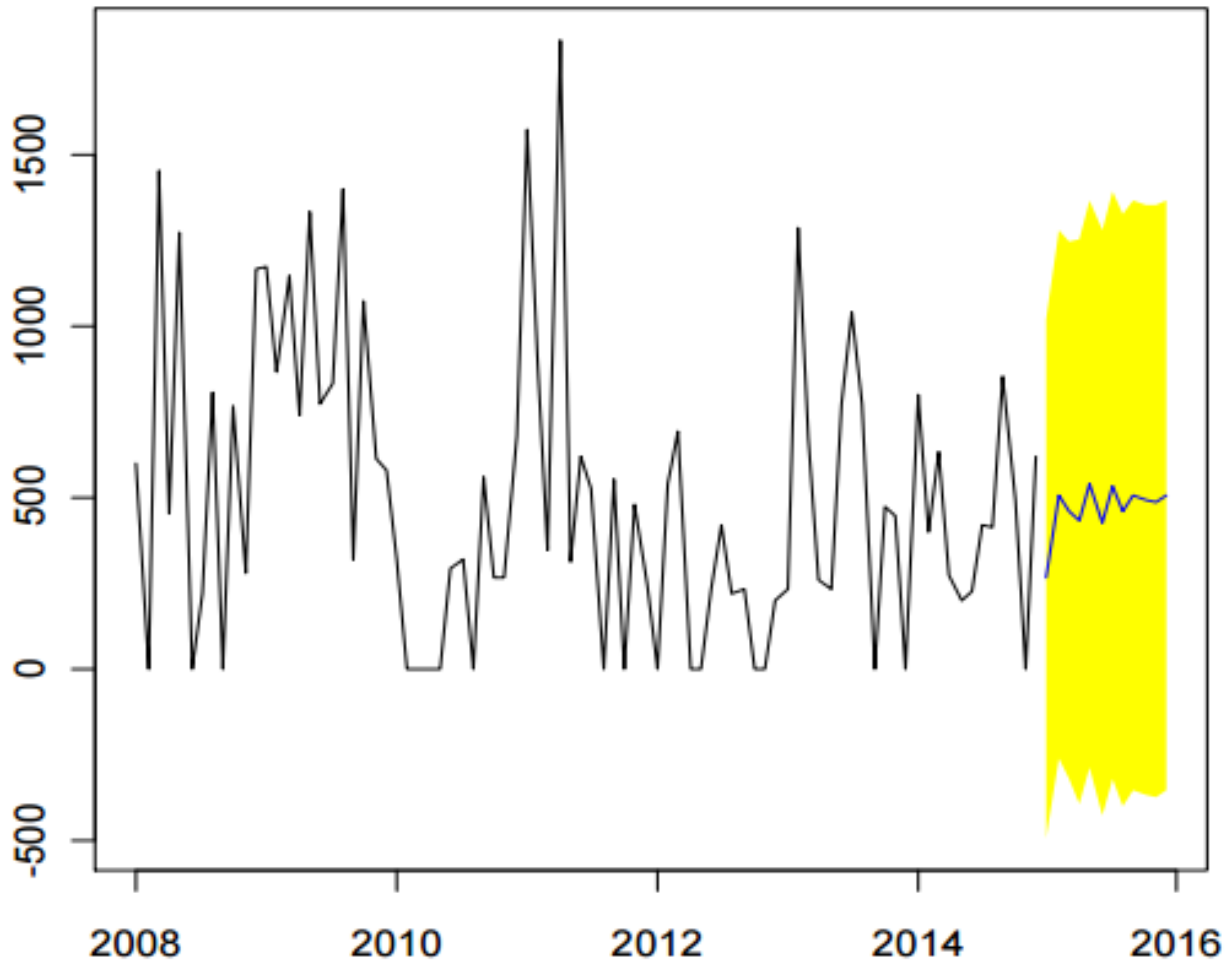

Figure 4.5: Forecasts for 12 Months from January 2015 to December 2015 (Highlighted part).

The results from table 4.6, show that the mean forecast error is -247.64 and the mean absolute deviation is 265.01 and the conclusion was that the forecast model tends to be over-forecast,with an average absolute error of 265 units.

Table 4.7: **Measures of Forecast Accuracy for (2014) Out-of-Sample Forecast.**

Months	A_t	F_t	**ME**	**MAE**	**MSE**
Sep	853	453.15	-399.85	399.85	159,880.02
Oct	493	527.75	34.75	34.75	1,207.56
Nov	0	497.86	-497.86	497.86	247,864.58
Dec	622	494.41	-127.59	127.59	16,279.21
Total			-990.55	1,060.05	425,231.37
			MAD = 265.01	MFE = -247.64	MSE=106.31

4.8 Discussion.

This study shows that there were 135 maternal deaths and about 29,711 live birth deliveries, which account for the total maternal mortality ratio of 454 per 100,000 live births, within the study period of January 2008 to December 2014. The average annual MMR recorded for the study period was 476.6 per 100,000 live births, which is lower than the average annual figure of the country. The MDG report (2014) reveals that MMR for South Sudan was 730 per 100,000 live births. Nevertheless, the figure from this study is still higher for the institution when compared to both continental and global estimates in the same report. The study further found that ARIMA (3, 0, 1) model is best for the prediction of monthly MMR at the JTH in South Sudan for the period of January 2015 and December 2015. The integrated order of zero (0) found in this study is similar to the findings of Sarpong, (2013). Who constructed an ARIMA (1, 0, 2) and ARIMA (2, 0, 1) and found that ARIMA (1,0,2) model was the best to predict MMR for the period of 20 quarters since ARIMA (2,0,1) model was insignificant at the Okomfo Anokye Teaching Hospital in Kumasi (Ghana).

Chapter 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusion.

ARIMA (3, 0, 1) is an appropriate and parsimonious model to forecast the MMR at Juba Teaching Hospital for the next twelve (12) months from January 2015 to December 2015 .

5.2 Recommendations.

Since, the ARIMA (3, 0, 1) model is adequate to be used to forecast monthly MMR at JTH for the period of January 2015 to December 2015. A similar approach like the one presented in this report, could be used to model, maternal mortality ratio at both institutional,Ngos and national level in South Sudan for future planning.

Bibliography

[1] Afzal, M., Rehman, H. U. R., and Butt, A. R. (2002). FORECASTING: A DILEMMA OF MODULES (A comparison of theory based and theory free approaches). Paskistan Economic and Social Review, **40**(1), 1-18.

[2] Ahmed, S., Hill, K., Ahmed, S., and Hill, K. (2011). Maternal mortality estimation at the subnational level: a model- based method with an application to Bangladesh. Bull World Health Organization, **89**, 12-21.

[3] Asia, S. (2013). Improve maternal health. The Millennium Development Goals Report 2013,United Nations 2013,UNDP,UNFPA,UNICEF,UN Women,WHO.

[4] Atkinson, T. (2005). A tkinson Reviews:Final Report-Measurement of Government out put productivity for the National Accounts.

[5] Baby, H. A. (2010). Institutional factors affecting maternal mortality summary: introduction: Bangladesh Coll Phys Surg, **28**(1), 5-9.

[6] Bhutta, Z. A., Cabral, S., Chan, C. W., and Keenan, W. J. (2012). Reducing maternal, newborn, and infant mortality globally: an integrated action agenda. International Journal of Gynecology and Obstetrics,**119**, S13-S17.

[7] Box, G. E. P., and Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control.Holden-Day,San Francisco., 467-509.

[8] Cham, M. (2003). Maternal Mortality in the Gambia: Contributing factors and what can be done to reduce them., 8-139.

[9] Chatfield, C. (2000). Time-series forecasting.Chapman and Hall/CRC,USA.

[10] Chatfield, C. (2004). The Analysis of Time Series: an Introduction.6th Ed, Chapman and Hall/CRC,USA.

[11] Conde-Agudelo, A., and Belizán, J. M. (2000). Maternal morbidity and mortality associated with interpregnancy interval: cross sectional study. BMJ, **321**, 1255-1259.

[12] David A. Dickey, W. A. F. (1979). Distribution of the eestimators for autoregressive time series with a unit root. Journal of the American Statistical Association, **74**(366), 427431.

[13] Dobre, I., and Alexandru, A. A. (2008). Quantitative methods inquires modelling unemployment rate using Box-Jenkins procedure plot series is it stationary? Yes diagnostics No. Journal of Applied Quantitative Methods, **3**, 156-166.

[14] Fawole.A.O, Shah, A., Fabanwo, A.O., Adewunmi, A.A., Eniayewun, A.B., Dara, K., El-Ladan, A.M., Umezulikw, A.C., Alu, F.E., Adebayo, A.A., Obaitan, F.O., Onala, O.E., Usman, Y., Sullayman, A.O., Kailani, S., and Said, M. (2012). Predictors of maternal mortality in institutional deliveries in Nigeria. African Health Sciences, **12**(1), 32-40.

[15] Gor, R. M. and Mohan. M. (2009).Industral Statistics and operational Management6:Forecasting techniques,http://nsdl.niscarir.res.in/jspui/hadle/123456789/829.

[16] Greene, W. H. (2012). Econometric Analysis. 7th Ed, Pearson Education, London.

[17] Hogan, M. C., Foreman, K. J., Naghavi, M., Ahn, S. Y., Wang, M., Makela, S. M., Murray, C. J. L. (2010). Maternal mortality for 181 countries, 1980-2008: a systematic analysis of progress towards Millennium Development Goal 5. Lancet, **375**(9726), 1609-6736.

[18] Høj, L., Hedegaard, K., and Sandstro, A. (2003). Maternal mortality: only 42 days? BJOG:an International Journal of Obstetrics and Gynaecology,**110**, 995-1000.

[19] Hyndman, R. J., and Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, **22**(4), 679-688.

[20] Kitui, J., Lewis, S., and Davey, G. (2013). Factors influencing place of delivery for women in Kenya: an analysis of the Kenya demographic and health survey , 2008 - 2009. BMC Pregnancy and Childbirth, **13**(1),13-40.

[21] Nielsen, H. B. (2005). Univeriate time series Analysis: ARIMA models, Econometrics 2-Fall ,1-41.

[22] Mojekwu, J., and Ibekwe, M. (2012). Maternal Mortality in Nigeria: Examination of Intervention Methods. International Journal of Humanities and Social Science,**2**(20), 135-149.

[23] Mujumdar, P. P., and Kumar, D. N. (1990). Stochastic models of streamflow: some case studies. Hydrological Sciences Journal, **35**(4), 395-410.

[24] Office for National Statistics (ONS). (2008). From Holt-Winters to ARIMA Modelling: Measuring the Impact on Forecasting Errors for Components of Quarterly Estimates of Public Service Output. United Kingdom Centre for the Measurement of Government Activity, 1-43.

[25] Osoro, A.A., Ngang'a, Z., Mutugi, M., and Wanzala, P. (2014). Maternal Mortality among Women Seeking Health Care Services in Kisii Level 5 Hospital. American Journal of Public Heath Research, **2**(5), 182-187.

[26] HILL,R.C., Griffiths, E. W., and Lim,C. G. (2011). Principles of Econometrics. 4th Ed,John Wiley and Sons Inc,USA.

[27] RMF.(2013).Real Medicine Foundation Annual Report:South Sudan, 21-27

[28] Robert H.S., and David, S. S .(2006). Time Series Analysis and Its Applications with R examples.2nd Ed, Springer Science+Business Media,LLC, USA.

[29] Romero, C. X., Koch, E., Thorp, J., Bravo, M., and Ahlers, I. (2012). Women s education level , maternal health facilities , abortion legislation and maternal deaths: a natural experiment in Chile from 1957 to 2007. PLoS ONE, **7**(5), 1-16.

[30] Sarpong, S. A. (2013). Modeling and forecasting maternal mortality; an application of ARIMA models. International Journal of Applied Science and Technology, **3**(1), 19-28.

[31] Savadogo, L. G. B., Zombra, A., Tamini, C., and Kinda, M. (2014). Maternal mortality risk factors in regional hospital of Burkina Faso. Open Journal of Epidemiology, **4**, 5762.

[32] Shrivastav, A. K., and Ekata. D. (2012). Applicability of Box Jenkins ARIMA Model in Crime Forecasting: A case study of counterfeiting in Gujarat State. International Journal of Advanced Research in Computer Engineering and Technology, **1**(4), 2278-1323.

[33] United Nation (UN). (2014). The Millennium Development Goals Report 2014. New York, 28-33.

[34] Wei, W. (1990). Time series analysis Univariate and Multivariate Methods .2nd Ed,Pearson Addison Wesley, USA.

[35] World Health Organization (WHO). (2014). Trends in Maternal Mortality: 1990 to 2013. Estimates by WHO, UNICEF, UNFPA, the World Bank and the United Nations Population Division. 1-68

Appendix 1

Table1. Calculated Monthly MMR for Period January 2008 to December 2014

Years	Maternal Deaths	Live Births	MMR per 100,000 live births
2008	1	166	602
2008	0	185	0
2008	3	206	1456
2008	1	221	452
2008	3	235	1277
2008	0	230	0
2008	1	455	220
2008	2	247	810
2008	0	293	0
2008	2	260	769
2008	1	360	278
2008	3	257	1167
2009	3	256	1172
2009	2	230	870
2009	3	261	1149
2009	2	270	741
2009	4	300	1333
2009	2	259	772
2009	2	240	833
2009	2	143	1399
2009	1	311	322
2009	4	373	1072
2009	2	327	612
2009	2	345	580
2010	1	314	318
2010	0	270	0
2010	0	297	0
2010	0	289	0
2010	0	310	0
2010	1	340	294
2010	1	315	317
2010	0	394	0
2010	2	356	562
2010	1	372	269
2010	1	374	267
2010	2	294	680
2011	4	254	1575
2011	2	219	913
2011	1	288	347
2011	6	327	1835
2011	1	321	312

Years	Maternal Deaths	Live Births	MMR per 100,000 live births
2011	2	324	617
2011	2	380	526
2011	0	372	0
2011	2	363	551
2011	0	418	0
2011	2	414	483
2011	1	408	245
2012	0	401	0
2012	2	369	542
2012	2	288	694
2012	0	403	0
2012	0	363	0
2012	1	435	230
2012	2	475	421
2012	1	459	218
2012	1	435	230
2012	0	522	0
2012	0	468	0
2012	1	502	199
2013	1	431	232
2013	4	311	1286
2013	3	445	674
2013	1	381	262
2013	1	424	236
2013	3	388	773
2013	4	383	1044
2013	3	388	773
2013	0	440	0
2013	2	423	473
2013	3	670	448
2013	0	288	0
2014	2	250	800
2014	1	251	398
2014	2	317	631
2014	1	369	271
2014	1	498	201
2014	1	444	225
2014	2	476	420
2014	2	483	414
2014	4	469	853
2014	3	609	493
2014	0	528	0
2014	3	482	622
Total	**135**	**29,711**	**454**