Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.5, No.13, 2015
www.iiste.org

Forecasting Monthly Maternal Mortality in the Bawku Municipality, Ghana Using SARIMA

Gideon Mensah Engmann[1,3*] Emmanuel Thompson[2] Clement Abugri[1]

1. Department of Statistics, Faculty of Mathematical Sciences, University for Development Studies, P. O. Box 24, Navrongo, Ghana

2. Department of Mathematics, Southeast Missouri State University, USA

3. Department of Mathematics, Shanghai Jiao Tong University, Shanghai, China

Abstract

Maternal mortality is defined as the death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management. Maternal mortality which accounts for 14% of all female deaths is still the second largest cause of female deaths in Ghana. Maternal mortality is very high in the northern regions of Ghana hence the need to model and forecast maternal mortality data to give insight to public health workers so as to combat any expected high maternal mortality. This study therefore models and also forecast future maternal mortality trends in the Bawku Municipality of the Upper East Region of Ghana using Box-Jenkins Approach. Analyses were based on monthly data available at the Biostatistics department of the municipal hospital. Results show that SARIMA $(3, 0, 0) \times (1, 1, 2)_{12}$ adequately models the maternal mortality data. The forecasted values also revealed that, maternal mortality cases increased during the months of May to July and from September to December, which is an insight for public health workers.

Keywords: Bawku-Ghana, Maternal Mortality, SARIMA, Box-Jenkins Approach

1. Introduction

Maternal mortality is defined as the death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management (WHO–ICD 10).

Of all the health statistics compiled by World Health Organization, the largest discrepancy between developed and developing countries occurred in maternal mortality (Ujah et al., 2005). While 25 percent of females of reproductive age lived in developed countries, they contributed only 1 percent to maternal deaths worldwide (WHO, 2005). A total of 99 percent of all maternal deaths occur in developing countries. More than half of these deaths occur in sub-Saharan Africa and one third in South Asia (Ujah et al., 2005). An estimate of 529,000 women die each year worldwide from pregnancy-related complications, of which 90% occur in developing countries, the worst affected being Africa, including Ghana (Sarpong, 2013).

Maternal mortality which accounts for 14% of all female deaths is still the second largest cause of female deaths in Ghana (Asamoah, 2011).

Maternal hemorrhage, obstructed labour, postpartum sepsis, eclampsia, unsafe abortion and anemia are among the leading causes of death among pregnant women in developing countries (WHO, 2005). Contributory factors include lack of access to good quality maternal and neonatal health services and strong adherence to negative cultural beliefs and practices (AbouZahr and Wardlaw, 2001; WHO, 2005).

Accurate predictions for maternal mortality will aid health related policy makers to see ahead of time the possible future requirements to design strategies and effective policies to combat any expected high maternal mortality in the Bawku municipality.

A wealth of research has been made into maternal mortality in different countries using time series analysis. Sarpong (2013) modeled and forecasted maternal mortality in Kumasi, Ghana using ARIMA. Liu et al. (2011) conducted a study to explore the feasibility of applying time series autoregressive integrated moving average (ARIMA) model to predict maternal mortality ratio (MMR) in China so as to provide the theoretical basis for reducing the MMR. Additionally, Elard Koch (2009) used ARIMA models to analyze MMR and abortion mortality ratio (AMR) from 1960 to 2007 in Chile.

In this study, seasonal ARIMA (SARIMA) model was used to develop a model for forecasting maternal mortality in the Bawku Municipality of Ghana.

2. Materials and Methodology

Time series is a time dependent sequence Y_t, where t belongs to the set of integers and denotes the time steps. If a time series can be expressed as a known function, $Y_t = f(t)$, then it is said to be a deterministic time series. If it is however expressed as $Y_t = X(t)$, where X is a random variable then $\{Y_t\}$ is a stochastic time series.

2.1. Data and Source

Data was obtained from the Biostatistics department of Bawku Municipal Hospital in the Upper East (UE) region of Ghana. The data covered monthly reported cases of maternal mortality for the period January, 2000 to December, 2014. It is important to mention that, UE region is one of the poorest regions in Ghana. Also, Bawku, the district capital of the Bawku municipality, is known to be one of the major municipalities in the UE region.

2.2. Stationarity and Non Stationarity

Stationarity refers to the statistical equilibrium or stability in the data set. A time series is strictly stationary if the joint distribution of $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ is the same as the joint distribution of $X_{t_{1+T}}, X_{t_{2+T}}, \dots, X_{t_{n+T}}$ for all $t_{1+T}, \dots, t_{n+T}$. Thus shifting the time position by T periods has no effect on the joint distribution. However it is difficult to investigate strict stationarity empirically, therefore a weaker version of stationarity is assumed. The time series is said to be weakly stationary if it's mean, variance and covariance do not change with time or are time invariant. If a time series is not stationary, then it is said to be non-stationary. A simple non-stationary time series is given by:

$$Y_t = \mu_t + e_t \qquad (1)$$

2.2.1. Unit root test

Unit root test was derived in 1976 by Dickey and Fuller to test for the presence of unit root versus a stationary process. The test is based on the assumption that a time series data y_t follows a random walk:

$$Y_t = \rho y_{t-1} + e_t \qquad (2)$$

Hypothesis

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

H_1: $\rho < 1$ (series has no unit root or stationary)

where ρ is the characteristic root of an AR polynomial and e_t is purely a random process with mean zero and variance σ^2. If the test statistic value of the Augmented Dickey-Fuller (ADF) test is less than the critical value, we reject the null hypothesis that the data has a unit root. Similarly, the unit root test as proposed by Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test the hypothesis below:

H_0: Series is level or trend stationary

H_1: Series is level or trend non-stationary

If the test statistic value of the KPSS is less than the critical value, we do not reject the null hypothesis that the data is level or trend stationary.

2.3. SARIMA model

Seasonality in a time series is a regular pattern of changes that repeats over specific time periods. The SARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model. The model is written as: ARIMA (p, d, q) × $(P, D, Q)_s$ where p, d, q, are the non-seasonal orders AR, differencing and MA respectively. P, D, Q, are the seasonal AR order, seasonal differencing order and seasonal MA order respectively. S=time span of repeating seasonal pattern. This model can be written more formally as:

$$\varphi(B)\Phi(B)(1-B)^d(1-B^S)^D y_t = \theta(B)\Theta(B)\varepsilon_t \qquad (3)$$

The non-seasonal components are:

$$\text{AR: } \varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p \qquad (4)$$

$$\text{MA: } \theta(B) = 1 + \theta_1 B + \theta_2 B^2 \ \dots + \theta_q B^q \qquad (5)$$

The seasonal components are:

Seasonal AR: $\Phi(B) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \cdots - \Phi_P B^{PS}$ (6)

Seasonal MA: $\Theta(B) = 1 + \Theta_1 B^S + \Theta_2 B^{2S} \ldots + \Theta_Q B^{QS}$ (7)

2.4. The Box - Jenkins Methodology
Box and Jenkins (1976) proposed a three step iterative approach to modeling as follows:

- Model identification
- Model parameter estimation
- Model diagnostics (goodness of fit testing).

2.4.1. Model Identification
In the identification stage of model building, we determine the possible models based on the data pattern. The values of p, d, q, P, D and Q are determined. But before we can begin to search for the best model for the data, the first condition is to check whether the series is stationary or not.

2.4.2. Parameter Estimation
The second step is the estimation of the model parameters for the tentative models that have been selected. Here, the model with the minimum values of Akaike Information Criterion (AIC), modified Akaike Information Criterion (AICc), and Normalized Bayesian Information Criterion (BIC) were considered as the best model. In general, the AIC, AICc and BIC are computed as follows:

$$AIC = 2k - 2log\ L \quad (8)$$

or

$$AIC = 2k + nlog\left(\frac{RSS}{n}\right) \quad (9)$$

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} \quad (10)$$

$$BIC = -2Log\ L + klog\ n \quad (11)$$

where k: is the number of model parameters
L: is the maximum value of the likelihood function for the estimated model
RSS: is the residual sum of squares of the estimated model
n: is the number of observations
$\sigma^2{}_t$: is the error variance

2.4.3. Model Diagnostics (Goodness of fit)
Ideally, a model should extract all systematic information from the data. The diagnostic check is used to determine the adequacy of the chosen model usually based on the residuals of the model. One assumption of the ARIMA model is that, the residuals of the model should be white noise. A series $\{\varepsilon_t\}$ is said to be white noise if $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (i.i.d) random variable with finite mean and variance. In addition if $\{\varepsilon_t\}$ is normally distributed with mean zero and variance δ^2, then the series is called Gaussian White Noise. For a white noise series, all the ACFs are zero. In practice, if the residuals of the model is white noise, then the ACF of the residuals are approximately zero. If the assumptions are not fulfilled then different model for the series must be search for. An overall check of model adequacy is provided by the Ljung-Box Q statistic. The test statistic is given by;

$$Q_m = n(n+2)\sum_{k=1}^{m}\frac{r_k^2}{n-k} \quad (12)$$

r_k^2 = the residuals autocorrelation at lag k.

n= the number of residuals.

m= the number of time lags included in the test.

When the p-value associated with the Q is large the model is considered adequate, else the whole estimation process has to start again in order to get the most adequate model.

ARCH LM-test for conformity of the presence of, or otherwise ARCH effect was also performed.

The hypothesis is given by:

H_0 = There is no conditional heteroscedasticity in the residuals of the model

H_1 = There is conditional heteroscedasticity in the residuals of the model

The test statistic is given by

$$LM = TR^2 \quad (13)$$

Where T = number of observations

R^2 = coefficient of determination computed from the auxiliary regression

Normality test of the residuals were carried out by Jarque Bera test and Shapiro Wilk test. The null hypothesis is that the residuals follow a normal distribution against the alternative hypothesis that the residuals do not follow a normal distribution.

All statistical tests were controlled at 5% significance level (α).

3. Results and Discussion

3.1. Pattern of maternal mortality recorded from January, 2000 to December, 2014

Figure 1 shows that the series has a slightly noticeable decreasing trend hence appears to be non-stationary since the series is not constant in size overtime. From Figure 2, we observed existence of seasonal variations in the series which is constant over time, the random effect is also constant over time and the pattern (trend) of the series which seems not stationary thus not constant over time.

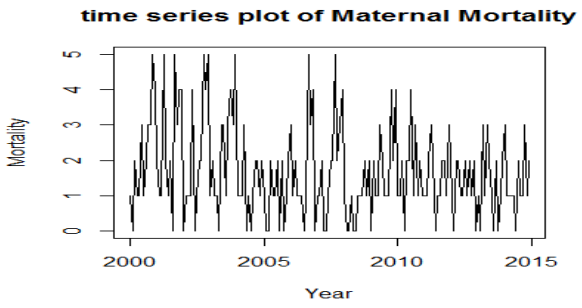

Figure 1. Time series plot of maternal deaths

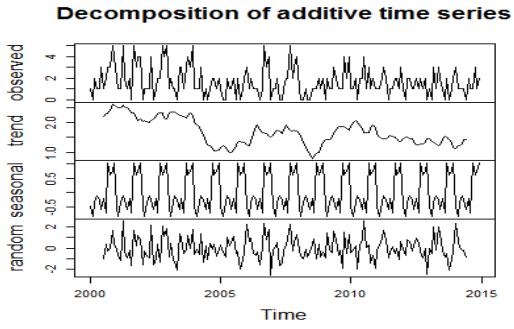

Figure 2. Decomposed time series plot

3.2. Model Identification

Figure 3 shows the ACF and PACF of the series with 95% confidence limits. From the correlogram, many spikes in both the ACF and PACF were outside the confidence interval. Also, the ACF shows a seasonal movement with significant spikes at the seasonal lags. KPSS test is used for verifying whether or not the series is stationary, while ADF test is used for verifying whether or not there is unit root. From Table 1, the KPSS test with a p-value of 0.01 indicates that the series is not stationary; this is confirmed by the ADF test with a p-value of 0.1.

Table 1. Unit Root and Stationarity test of the series

Test type	Test Statistic	Lag order	Alpha value	P-value
KPSS	0.4885	3	0.05	0.01
ADF	-5.8827	5	0.05	0.1

Table 2 shows the stationarity and unit root test for the first seasonal differenced series. Both the KPSS test and ADF test with p-values of 0.1 and 0.01 respectively showing the non-existence of unit root hence the first seasonal differenced series is stationary.

Table 2. Unit Root and Stationarity test for first differenced series

Test type	Test Statistic	Lag order	Alpha value	P-value
KPSS	0.0117	3	0.05	0.1
ADF	-8.3774	5	0.05	0.01

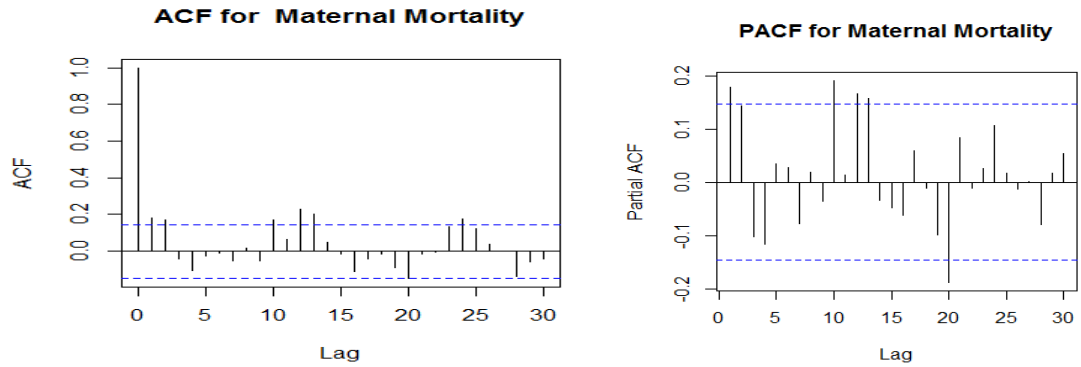

Figure 3: Correlogram of the series

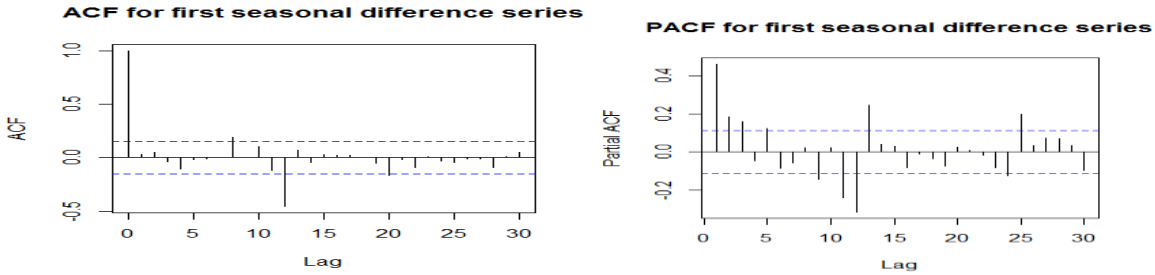

Figure 4. Correlogram of the first seasonal differenced series

The significant spikes in the PACF at lags 12 and 24 and also at lag 12 of the ACF of the seasonal differenced series is suggestive of a seasonal AR (2) and a seasonal MA (1) terms. In the non-seasonal lags, the PACF truncates after lag 3 suggesting a possible AR (3) term. The pattern in the ACF is not indicative of any simple model. From these plots, SARIMA (3, 0, 0) (P, 1, Q) has been identified. Consequently, other models while varying *P* and *Q* were fitted as shown in Table 3.

Table 3. Tentative SARIMA Models for the series

Models	AIC	AICc	BIC
SARIMA(3, 0, 0) × $(1, 1, 1)_{12}$	559.71	560.08	575.33
SARIMA(3, 0, 0)×$(1, 1, 2)_{12}$	557.17*	557.69*	575.21*
SARIMA(3, 0, 0) × $(1, 1, 3)_{12}$	558.96	559.06	580.83
SARIMA(3, 0, 0) × $(2, 1, 1)_{12}$	561.22	5561.75	579.97
SARIMA(3, 0, 0) × $(2, 1, 2)_{12}$	558.92	559.62	580.79
SARIMA(3, 0, 0) × $(2, 1, 3)_{12}$	564.31	565.21	589.30
SARIMA(3, 0, 0) × $(3, 1, 1)_{12}$	562.45	563.15	584.32

*:Best based on the selected criterion

Among those possible models, comparing their AIC, AICc and BIC as shown in Table 3, SARIMA (3, 0, 0) $(1, 1, 2)_{12}$ has the least values indicating that it is the appropriate model that fit the data well.

3.3. Model Estimation and Evaluation

Parameters of the model were estimated using the method of maximum likelihood. From Table 4, we conclude that the parameters are significant at 5% significant level.

Table 4. Parameter estimates of SARIMA (3, 0, 0) × $(1, 1, 2)_{12}$

Type	Coefficients	Standard error	T-value	P-value
AR(3)	0.1704	0.8693	2.2217	0.0010
SAR(1)	0.8693	0.0964	9.0204	0.0000
SMA(2)	0.9996	0.3111	3.2132	0.0000
		σ^2 =1.182		

3.4. Model Diagnostics (Goodness of fit)

Diagnostic checks are performed on the residuals to see if they are randomly and normally distributed. From Table 5 and Figure 5, the Ljung-Box test statistic had p-values greater than the alpha (α) value of 0.05 showing that there is no serial correlation in the residuals of the model since the p-values are all greater than the alpha (α) value, hence the model is adequate.

Table 5. Ljung-Box test statistic for model Diagnostics

Lags	Test statistic	P-value
12	15.4149	0.2195
24	31.3413	0.1442
36	48.1773	0.0844
48	62.1794	0.0802

The ARCH-LM test in Table 6 indicates that there is clearly no evidence of heteroscedasticity in the model residuals, since the p-values are all greater than the alpha (α) value. This makes the model adequate for predictions.

Table 6. ARCH-LM test of residuals of SARIMA (3, 0, 0) × $(1, 1, 2)_{12}$ of the series

Lag	Chi-square	Df	P-value
12	8.9209	12	0.7097
24	19.4524	24	0.7275
36	29.0896	36	0.7862
48	35.1220	48	0.9169

The Jarque Bera test shows that the residuals are normal since its p-values are greater than the 5% significant level. This is confirmed by the Shapiro Wilks test with a p-value of 0.2676 as shown in Table 7.

Table 7. Normality Test for Residual of Maternal Deaths

Normality test	Test statistic	P-value
Jarque Bera	0.9385	0.6255
Shapiro-wilk	0.9904	0.2676

The fitted model is therefore given by

$$(1 - 0.07B - 0.17B^2 - 0.22B^3)(1 - 0.86B^{12})(1 - B^{12})y_t = (1 - 1.93B^{12}\ 0.99B^{24})\varepsilon_t \qquad (14)$$

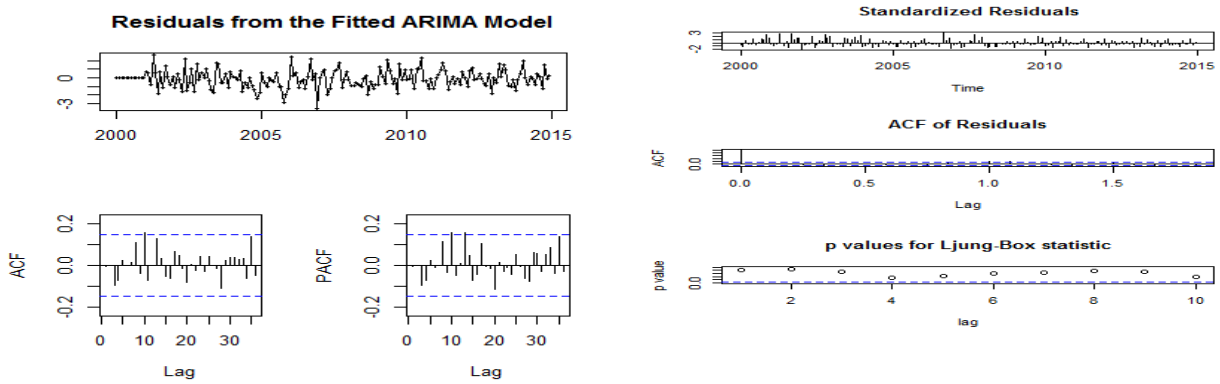

Figure 5. Diagnostic plots of SARIMA (3, 0, 0) × $(1, 1, 2)_{12}$ of maternal mortality

3.5. Forecasting

Finally we made a two year forecast with our model as shown in Figure 6. The forecasted maternal mortality cases generally showed a decreasing trend but increased during the months of May to July and from September to December for the forecasted period, which is an insight for public health workers.

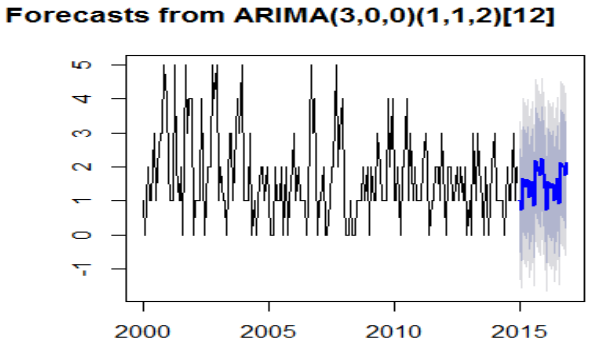

Figure 6. Time series plot of actual and forecast maternal deaths

4. Conclusion

Maternal mortality is one of the most sensitive indicators of health disparity between richer and poorer nations. The Health of a pregnant woman is very significant to the growth of our economy thus calling for major attention. In this study, we model maternal mortality cases from January, 2000 to December, 2014 at the Bawku Municipal Hospital, UE region of Ghana using SARIMA. The appropriate model identified was SARIMA (3, 0, 0) × $(1, 1, 2)_{12}$. The forecasted maternal mortality cases generally showed a decreasing trend however, increased during the months of May to July and from September to December, which is an insight for public health workers. Future research could integrate maternal mortality data from other regions of the country for an in-depth analysis.

Acknowledgement

Acknowledgements go to all and sundry who in diverse ways contributed to making this research a success especially Miss Kubura Issaka.

References

Abouzahr, C., and Wardlaw, T. (2001). Maternal mortality at the end of a decade: Signs of Progress?. *Bulletin of the World Health Organization.* 79(6), pp: 561-588.

Asamoah, B. O., Moussa, K. M., Stafström, M. and Musinguzi, G. (2001). Distribution of causes of maternal mortality among different socio-demographic groups in Ghana: A descriptive study", *BMC Public Health*, vol/issue: 11(159), doi: 10.1186/1471-2458-11-159.

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

Canova, F. and Hansen, B.E. (1995), Are Seasonal Patterns Constant Over Time? A Test for Seasonal Stability. *Journal of Business and Economic Statistics*. 13, pp: 237-252.

DFID (2004). "Reducing maternal deaths: Evidence and action"A strategy for DFID Published by the Department for International Development, September 2004 International Classification of Diseases (ICD), 10th Edition. Ghana Millennium Development Goals Report. (accessed: April 2015).

Dickey, D.A. and Fuller, W.A. (1979). Distribution of the Estimators for Autoregressive Time Series with Unit Root. *Journal of the American Statistical Association*. 74, pp: 427-431.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. & Shin, Y. (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. *Journal of Econometrics.* 54, pp: 159–178.

Liu, J., Gao Y., Qu, B., He, Q. (2011). Exploring the feasibility for application of time series ARIMA model to predict the maternal mortality ratio (MMR) in china. *The Journal of China Medical University*. 6, Pp: 22-49.

Pankratz, A. (1983). Forecasting with Univariate Box-Jenkins Models: Concepts and Cases. John Wily & Sons. Inc. USA

Sarpong, S. (2013). Modeling and Forecasting Maternal Mortality; an Application of ARIMA Models. *International Journal of Applied Science and Technology. 3(1),* pp: 19-28.

Schwarz, G. E. (1978). Estimating the Dimension of a Model. *Annals of Statistics.* 6 (2), pp: 461–464.

Ujah, I. A., Uguru, V. E., Aisien, A. O., Sagay, A. S., and Otubu J. A. M. (2005). Factors Contributing to Maternal Mortality in North-Central Nigeria: A Seventeen-year Review *African Journal Reproduction Health* Vol.9, No3, pp 27-40.

WHO (2005). Maternal Mortality in 2005 Estimates developed by WHO, UNICEF, UNFPA, and The World Bank. http://www.who.int/whosis/mme (accessed; March, 2015).