J Biostat Epidemiol. 2017; 3(1): 13-9.

Original Article

Modeling births at a tertiary health-care facility in Ghana: Box-Jenkins time series approach

Raymond Essuman[1], Ezekiel N. N. Nortey[2], George Aryee[1*], Eunice Osei-Asibey[3], Ebenezer Owusu Darkwa[1], Robert Djagbletey[1]

[1] Department of Anaesthesia, School of Medicine and Dentistry, University of Ghana, Accra, Ghana
[2] Department of Statistics, School of Physical and Mathematical Sciences, University of Ghana, Accra, Ghana
[3] Department of Statistics, School of Graduate Studies, Regent University College of Science and Technology, Accra, Ghana

ARTICLE INFO

Received 01.06.2016
Revised 08.08.2016
Accepted 23.11.2016
Published 04.03.2017

Key words:
Forecasting;
Seasons;
Birth;
Models

ABSTRACT

Background & Aim: Changes in the trend of births among women have been studied worldwide with indications of peaks and troughs over a specified period. Periodic variations in the number of births among women are unknown at the Korle-Bu Teaching Hospital (KBTH). This study sought to model and predicts monthly number of births at the Department of Obstetrics and Gynaecology (O&G), KBTH.

Methods & Materials: Box-Jenkins time series model approach was applied to an 11-year data from the Department of (O&G), KBTH on the number of births from January, 2004 to December, 2014. Box-Jenkins approach was put forward as autoregressive integrated moving average (ARIMA) model. Several possible models were formulated, and the best model, which has the smallest Akaike information criterion corrected (AICc) was selected. The best model was then used for future predictions on the expected monthly number of births for the year 2015. Analysis was performed in R statistical software (version 3.0.3).

Results: Seasonal ARIMA $(2,1,1) \times (1,0,1)_{12}$ was selected as the best model because it had the smallest AICc. Furthermore, the forecasted values showed that the expected number of births were lowest in January (750 births) and highest in May (970 births) for the year 2015.

Conclusion: Seasonal ARIMA $(2,1,1) \times (1,0,1)_{12}$ was identified as the model that best describes monthly expected births and its use to forecast the expected number of births at the KBTH in Ghana will facilitate formulation of health policies and planning for safe maternal delivery and prudent use of hospital obstetric services and facilities.

Introduction

Trends and variations in births among women over a period of time have been reported worldwide especially in Europe, America, and Asia. Bohun-Chudynivy et al. (1) reported a 4.3 million births peak in 1960 in the United States of America but gradually declined to 3.1 million in 1970. However, the number of births peaked to over 4 million in the year 1989. Again, a study by Cohen (2) showed that the number of births declined from January to April. However, it kept increasing from May to mid-September and onward, depicted many sharp increases and decreases, although generally a smooth trend was observed. In Portugal, Caleiro (3) showed that months of May and September recorded an increase in birth compared to months of December and February.

* Corresponding Author: George Aryee, Postal Address: Department of Anaesthesia, Korle-bu Teaching Hospital, University of Ghana School of Medicine and Dentistry, College of Health Sciences, P.O. Box 4236, Accra, Ghana.
Email: garyee43@gmail.com

Please cite this article in press as: Essuman R, Nortey ENN, Aryee G, Osei-Asibey E, Darkwa EO, Djagbletey R. Modeling births at a tertiary health-care facility in Ghana: Box-Jenkins time series approach. *J Biostat Epidemiol. 2017; 3(1): 13-9*

At the Korle-Bu Teaching Hospital (KBTH), observations reveal that the month of April is the peak period for births. During such periods the facility's resources are overwhelmed, and patients may not be optimally managed resulting in increased maternal and perinatal mortality. Additional resources usually mobilized at the height of peaks in most instances become available after peak periods. This has financial implications as funds have been committed into these resources which may be under-utilized after the peak periods. Furthermore, monthly and yearly variation in birth at KBTH is unknown. Thus, the main objective of this study was to identify a time series model to describe birth data at the Department of Obstetrics and Gynaecology (O&G), KBTH. Specific objectives were to identify a model that best describes birth data and forecast the expected number of birth for each month for the year 2015.

Methods

Study design and site

A time series analysis of data consisting of 132 months of recorded number of births from 2004 to 2014 at the Department of O&G, KBTH was used. The Department of O&G is one of the clinical departments at the hospital that caters for pregnant women and their unborn babies and any other pregnancy-related health complications. The department comprises five sub-specialties which are gynecological oncology, gynecological urology, maternal-fetal medicine, reproductive medicine, and community gynecology. The department has 240 obstetric beds and 114 gynecologic beds.

Data analysis

Data obtained were captured in Microsoft Excel 2010 and analysis was performed in R statistical software version 3.0.3. Box-Jenkins (4) approach put forward as autoregressive integrated moving average (ARIMA) model was used for modeling.

Autoregressive (AR) model: The AR model of order *p*, AR (*p)* takes the form:

$$Y_t=Ø_1Y_{t-1}+Ø_2Y_{t-2}+\ldots\ldots+Ø_pY_{t-p}+w_t$$

Y_t denotes the current value of the series, Y_{t-1}, ……, Y_{t-p} denotes the prior values of the same series while w_t is the white noise and $Ø_1$,……, $Ø_p$ are the weights or coefficients.

Moving average (MA) model: The MA model also of order *q*, MA (q) takes the form:

$$Y_t=w_t+\theta_1w_{t-1}+\theta_2w_{t-2}+\ldots\ldots+\theta_qw_{t-q}$$

Y_t denotes the current value of the series, w_t, ……, w_{t-q} are the white noise or shock and $\theta_1\ldots\theta_q$ are the weights or regression coefficients.

ARMA model: ARMA is a blend of both the AR with order p and MA with order q in the form:

$$Y_t=Ø_1Y_{t-1}+Ø_2Y_{t-2}+\ldots\ldots+Ø_pY_{t-p}+w_t+\theta_1w_{t-1}+\theta_2w_{t-2}+\ldots\ldots+\theta_qw_{t-q}$$

The ARIMA model: The ARIMA model is expressed as $\phi\ (B)\ (1-B)^d\ Y_t=\theta\ (B)\ \omega_t$ where

$\phi\ (B)$ is the operator for the AR term and is given as $\phi\ (B) = 1-\phi_1B-\phi_2\ B^2-\ \ldots\ -\phi_PB^P$ and

θ (B) is the operator for the MA term and is given as $\theta\ (B) = 1+\theta_1B+\theta_2B^2+\ \ldots\ +\theta_qB^q$.

Where *p* and *q* represent the number of lags for the AR and MA terms, respectively, and *d* is the order for the integration term.

The Box-Jenkins approach involved model identification, parameter estimation, model diagnostics, and forecasting. A time series of the data was first plotted. The data were log transformed and plotted again. Non-stationarity was confirmed using augmented Dickey-Fuller (ADF) test on the transformed data. The transformed data were differenced once to attain stationarity accounting for the order d in the model. An autocorrelation function (ACF) plot was then used to determine the order (p) of AR and partial autocorrelation function (PACF) to determine the order (q) of MA. The model obtained was checked for seasonality and compared to other seasonal and non-seasonal models. The model with the smallest Akaike information criterion corrected (AICc) was selected as the best model. Diagnostic tests were done on the chosen models by performing a residual analysis to determine the adequacy of the models. The best model was used to forecast the expected number of birth for the next year (2015). Finally, predictive error (P.E) of the specified model was calculated to determine the robustness of the chosen model.

Results

Model identification and parameter estimation

A preliminary analysis was performed to ascertain the stationary or otherwise of the data as illustrated in figure 1.

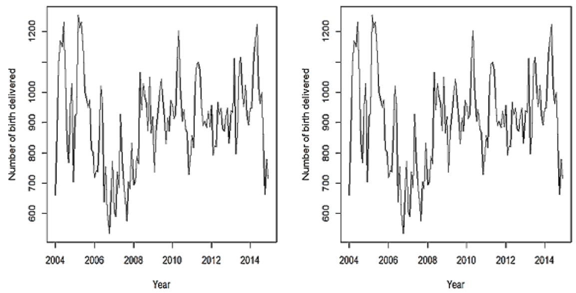

Figure 1. Time series graph of number of birth (left) and log-transformed of number of birth (right)

The log-transformed plot of the number of births series exhibits a trend, therefore non-stationary. The ADF test confirmed that the series was not stationary (ADF= −3.38; P = 0.061).

A first difference of the transformed data was then performed to achieve stationarity as illustrated in figure 2. The graph of the first difference appears stationary, and this was again confirmed by performing the ADF test (ADF = −6.82; P = 0.010).

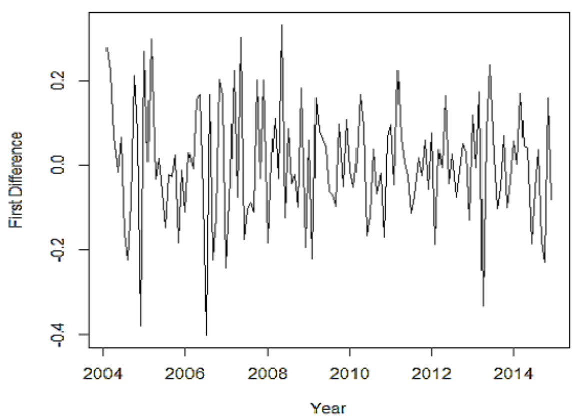

Figure 2. Time series graph of first difference

Figure 3 shows an ACF spike on lag 1 which exceeds the significance bounds (Figure 3). Likewise, from the PACF, there is also a spike on lag 1 which exceeds the significance bounds (Figure 3). Hence, the model ARIMA (1,1,1) was selected.

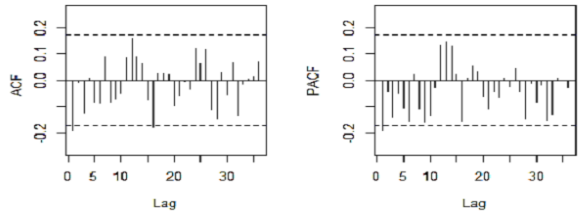

Figure 3. Sample autocorrelation function and partial autocorrelation function graph of first difference series

The estimated coefficient of the AR model 1 (AR1) was found to be 0.700, and that of the MA model 1 (MA1) was −1.000 with standard errors 0.066 and 0.024, respectively. The estimated value of the constant term was 7e-04 with a standard error of 9e-04. The AICc of the model was −153.38.

Model diagnostic of ARIMA (1,1,1)

Plot of the residuals: The plot was fairly random even though few of the observations are likely to be outliers as shown in figure 4. Further, tests of normality and autocorrelation of the residuals suggested that these outliers have no influence on the series.

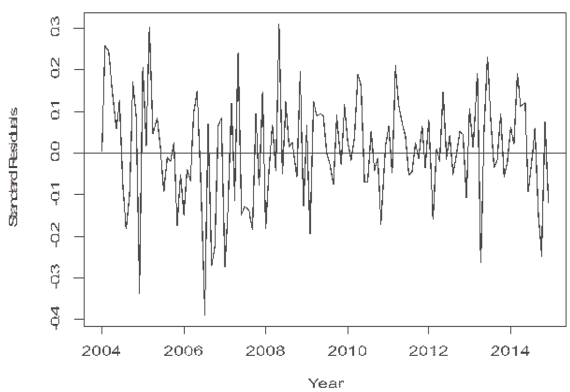

Figure 4. Graph of the standardized residuals of autoregressive integrated moving average (1,1,1)

Test of normality of the residuals: The normal Quantile-Quantile graph for the residuals from the model ARIMA (1,1,1) displayed some form of normality (Figure 5) and was confirmed using Shapiro-Wilk normality test (W = 0.99, P = 0.483).

Test of autocorrelation of the residuals: The plot of the residuals was fairly random and independent based on the sample ACF as shown in figure 6. The P values of the Ljung-Box test were all above the significant boundary (chi-squared = 21.106, P = 0.391).

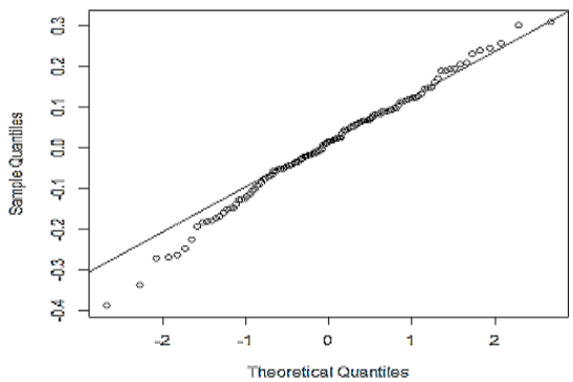

Figure 5. Normal Quantile-Quantile graph autoregressive integrated moving average (1,1,1) model

Test of other ARIMA Models: Several non-seasonal seasonal ARIMA models were estimated to determine whether any is likely to be superior to the ARIMA (1,1,1) obtained.

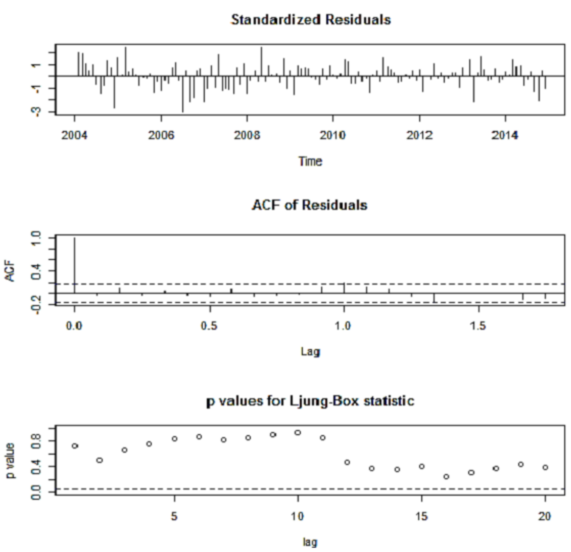

Figure 6. Diagnostic display of autoregressive integrated moving average (1,1,1) model

Non-seasonal ARIMA model estimation: ARIMA (1,1,1) had a minimum AICc (−153.38) as compared to all the other estimated non-seasonal ARIMA models; hence none of the non-seasonal ARIMA models estimated was found better than the ARIMA (1,1,1) model.

Seasonal ARIMA model estimation: SARIMA (2,1,1) × $(1,0,1)_{12}$ model had the least AICc and was thus selected as shown in table 1.

Model diagnostic of SARIMA (2,1,1) × $(1,0,1)_{12}$

Table 1. SARIMA models

Model	AIC	AICc	BIC
SARIMA $(3,1,1)(1,1,1)_{12}$	−130.00	−129.76	−111.30
SARIMA $(3,1,1)(1,0,1)_{12}$	−162.71	−161.80	−142.59
SARIMA $(3,1,1)(0,0,1)_{12}$	−157.71	−156.49	−139.91
SARIMA $(3,1,1)(1,0,0)_{12}$	−158.43	−157.75	−141.18
SARIMA $(3,0,1)(1,1,0)_{12}$	−119.16	−118.41	−102.43
SARIMA $(3,0,1)(0,1,0)_{12}$	−103.09	−102.56	−89.15
SARIMA $(3,0,1)(0,1,1)_{12}$	−140.15	−139.41	−123.42
SARIMA $(3,0,2)(1,1,1)_{12}$	−136.86	−135.56	−114.56
SARIMA $(3,0,2)(0,1,1)_{12}$	−138.17	−137.17	−118.66
SARIMA $(3,0,2)(0,1,0)_{12}$	−101.50	−100.76	−84.77
SARIMA $(3,0,2)(1,1,0)_{12}$	−118.39	−117.39	−98.88
SARIMA $(1,1,1)(1,0,0)_{12}$	−160.57	−160.25	−149.07
SARIMA $(1,1,1)(1,0,1)_{12}$	−163.98	−163.50	−149.60
SARIMA $(1,1,1)(0,0,1)_{12}$	−159.55	−159.23	−148.05
SARIMA $(1,1,2)(0,0,1)_{12}$	−158.8	−158.32	−144.42
SARIMA $(1,1,2)(1,0,0)_{12}$	−160.26	−159.73	−145.88
SARIMA $(2,1,1)(1,0,1)_{12}$	−164.21	−163.53	−146.96
SARIMA $(2,1,1)(1,0,0)_{12}$	−160.38	−159.90	−146.00
SARIMA $(2,1,1)(0,0,1)_{12}$	−158.95	−158.47	−144.58
SARIMA $(2,1,0)(1,0,1)_{12}$	−157.06	−156.58	−142.69
SARIMA $(2,1,0)(1,0,0)_{12}$	−151.10	−150.79	−139.60
SARIMA $(2,1,0)(0,0,1)_{12}$	−149.07	−148.75	−137.57

SARIMA: Seasonal autoregressive integrated moving average, AICc: Akaike information criterion corrected, AIC: Akaike information criterion, BIC: Bayesian information criterion

Residual analysis of SARIMA (2,1,1) × $(1,0,1)_{12}$

Plot of residuals: Figure 7 showed the graph of the standardized residuals (Figure 7) with fairly random observations through some of the observations are likely to be outliers.

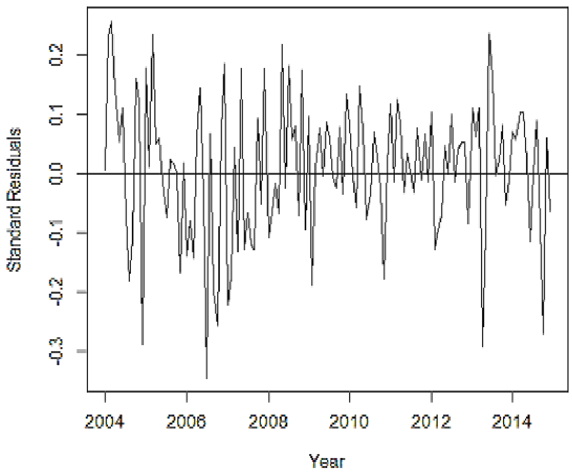

Figure 7. Graph standardized residuals of seasonal autoregressive integrated moving average (2,1,1) × $(1,0,1)_{12}$

Test of normality of the residuals of SARIMA (2,1,1) × $(1,0,1)_{12}$: The graph of the residuals displayed some form of normality as shown in figure 8 and this was further confirmed by Shapiro-Wilk normality test (W = 0.984, P = 0.120).

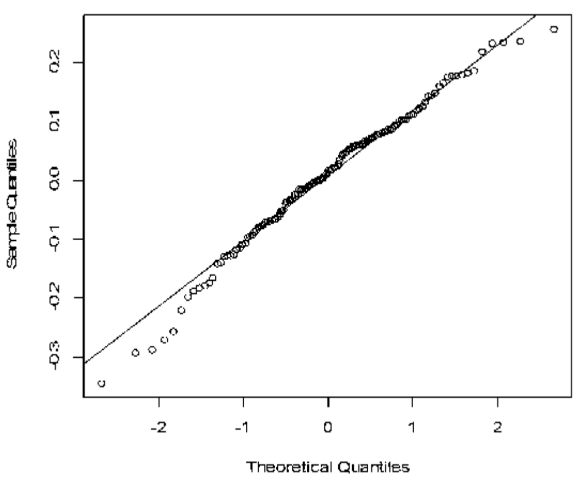

Figure 8. Normal Quantile-Quantile graph of seasonal autoregressive integrated moving average (2,1,1) × $(1,0,1)_{12}$ model

Test of autocorrelation of the residuals: It was shown in figure 9 that the plot of the residuals looks fairly random. The residuals appear independent based on the sample ACF. In addition, the P values of the Ljung-Box test were all above the significant boundary (chi-squared = 14.45; P value = 0.806).

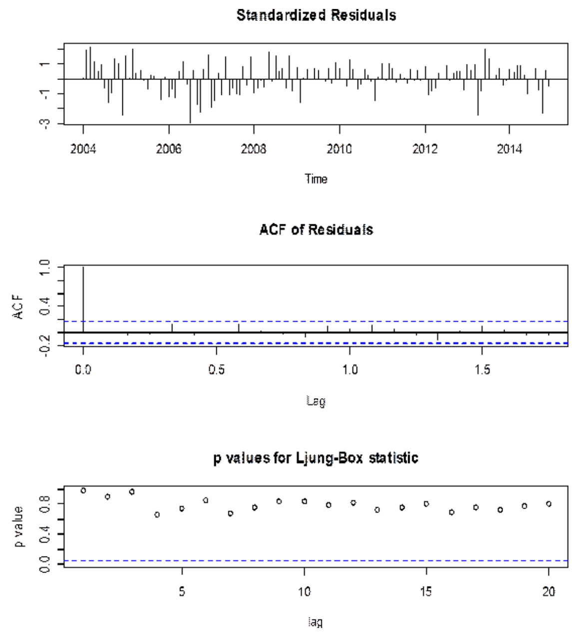

Figure 9. Test of autocorrelation function residuals of seasonal autoregressive integrated moving average (2,1,1) × $(1,0,1)_{12}$ model

Forecasting

Table 2 summarized the expected number of births per month for the O&G Department at the KBTH for the next 12 months in 2015 with 95% confidence interval (CI). The results from the forecast showed that the number of birth increased from the month of January through to the month of May. It tends to decline from the month of May through to the month of September and showed a slight increase from the month of September to November then declining in December. The forecasted values of the periods for 2015 are indicated in the shaded regions of figure 10.

Table 2. Expected number of births per month for the year 2015

Year 2015	Forecast	Lower limit at 95%	Upper limit at 95%
January	756	598	956
February	760	577	1000
March	864	640	1165
April	881	643	1205
May	967	701	1335
June	944	681	1310
July	890	639	1238
August	868	622	1210
September	822	589	1148
October	831	595	1161
November	842	603	1177
December	823	589	1151

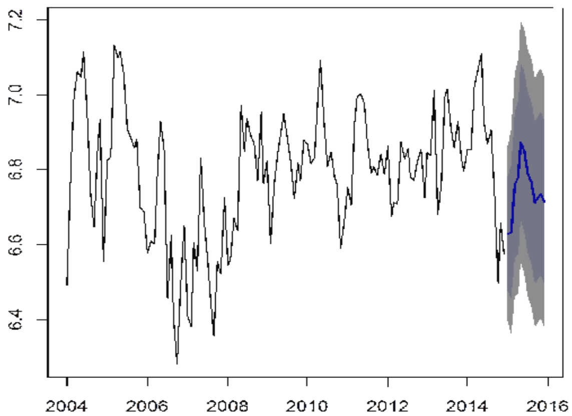

Figure 10. Forecasted values for seasonal autoregressive integrated moving average (2,1,1) × $(1,0,1)_{12}$ model

Robustness of the model (SARIMA [2,1,1] × $[1,0,1]_{12}$)

The P.E for SARIMA (2,1,1) × $(1,0,1)_{12}$ was 0.1 as shown in table 3.

Table 3. Predictive error of the seasonal ARIMA (2,1,1) × $(1,0,1)_{12}$

Year 2014	Actual value (A.V)	Predictive value (P.V)	$\text{P.E} = \frac{\text{P.V}-\text{AV}}{\text{AV.}} \times 100$
January	946	925	−2.2
February	948	888	−6.3
March	1124	950	−15.5
April	1175	887	−24.5
May	1224	934	−23.7
June	1018	969	−4.7
July	962	959	−0.3
August	999	919	−8.0
September	834	900	7.9
October	663	921	38.9
November	778	902	15.9
December	717	879	22.6
			P.E=0.1

ARIMA: Autoregressive integrated moving average

Discussion

The log-transformed plot of the number of births (right plot) as shown in figure 1 over the period studied exhibits an upward trend from 2007 to 2014 indicating non-stationary series (ADF = −3.38; P = 0.061). However, stationarity was attained (ADF = −6.8167; P = 0.010) on a first difference (Figure 2) of the transformed data.

ARIMA (1,1,1) was selected because it showed a spike on lag 1 for both ACF and PACF. Although few of the observations in the model were noted to be possible outliers (Figure 4), further tests of normality and autocorrelation of the residuals of the model suggested that these outliers had no influence on the series (W = 0.99, P = 0.483) (Figure 5).

The ARIMA (1,1,1) model was found to be a good model and does not exhibit lack of fit as all the P values of the Ljung-Box test were above the significant boundary (chi-squared = 21.106, P = 0.391) (Figure 6).

Other seasonal and non-seasonal models were formulated and compared to ARIMA (1,1,1). ARIMA (1,1,1) had the minimum AICc (−153.38) as compared to the other non-seasonal ARIMA models (Table 4). Hence, ARIMA (1,1,1) model was found superior to the other models estimated (Table 4).

Table 4. Non-seasonal ARIMA models

Model	AIC	AICc	BIC
ARIMA (1,1,0)	−143.28	−143.28	−134.65
ARIMA (0,1,1)	−144.06	−143.87	−135.44
ARIMA (2,1,0)	−141.06	−141.31	−130.13
ARIMA (2,1,1)	−152.12	−151.64	−137.75
ARIMA (2,1,2)	−150.83	−150.15	−133.58
ARIMA (3,1,0)	−142.68	−142.20	−128.31
ARIMA (3,1,1)	−150.87	−150.20	−133.62
ARIMA (3,1,2)	−152.05	−151.14	−131.92
ARIMA (1,1,2)	−152.05	−151.57	−137.67

ARIMA: Autoregressive integrated moving average, AICc: Akaike information criterion corrected, AIC: Akaike information criterion, BIC: Bayesian information criterion

Most SARIMA models (Table 1) have minimum AICc compared to ARIMA model (1,1,1). SARIMA (2,1,1) × $(1,0,1)_{12}$ had the least AICc (AICc = −163.53) and was thus chosen. It had estimated coefficients for the AR1 and AR2, 0.611 and 0.135, respectively, with standard errors (AR1 = 0.090; AR2 = 0.091).

The estimated coefficient for the MA1 was −0.999 with a standard error of 0.018. The seasonal autoregressive component of the model (SAR1) yielded estimated coefficient of 0.994 with a standard error of 0.053. Likewise, the estimated coefficient and standard error for seasonal moving average were −0.944 and 0.253, respectively. The general equation for the SARIMA model is:

$$(1-\psi_p B)(1-\Phi_p B^S)(1-B)(1-B^S)y_t=(1+\theta_q B)(1+\Theta_Q B^S)w_t$$

Where:

$(1-\psi_p B)$ and $(1-\Phi_p B^S)$ are the respective non-seasonal and SAR models component.

$(1-B)(1-B^S)$ is the difference factor for both non-seasonal and seasonal.

$(1+\theta_q B)$ and $(1+\Theta_Q B^S)$ are the MA models for non-seasonal and seasonal component, respectively, with w_t as the white noise.

Hence, the equation of the seasonal ARIMA $(2,1,1) \times (1,0,1)_{12}$ was found to be:

$$(1-0.6109B-0.1354B^2)(1-0.9941B^{12})(1-B)Y_t=(1+(-0.999B))+(1+(-0.9441B^{12}))w_t$$

A test of the SARIMA model obtained showed that the standardized residuals (Figure 7) were fairly random and further tests suggested that the residuals were normally distributed (W = 0.984, P = 0.120) (Figure 8). The residuals appeared independent based on the sample ACF. In addition, the *P* values of the Ljung-Box test were all above the significant boundary, which shows a good model (chi-squared = 14.45; P = 0.806) (Figure 9) not exhibiting lack of fit.

A P.E used in determining the robustness or applicability of the model yielded a value of 0.1 (Table 3). The minimal value yielded by the PE showed the robustness or usability of the model.

Based on the model we obtained, the monthly expected number of births for 2015 was forecasted (Figure 10). The highest expected number of births was predicted to occur in May (births = 967; CI = 701-1335) while the least expected births were predicted in January (births = 756; CI = 598-956).

Conclusion

Seasonal ARIMA $(2,1,1) \times (1,0,1)_{12}$ model was identified as the best model that describes monthly expected births. The model equation was found to be:

$$(1-0.6109B-0.1354B^2)(1-0.9914B^{12})(1-B)Y_t=(1+(-0.999B))(1+(-0.9441B^{12}))w_t$$

The monthly forecasted number of births for the year 2015 showed a peak in May and trough in January. The use of this model to forecast the expected number of births at the KBTH in Ghana will facilitate the formulation of health policies and planning for safe maternal delivery and prudent use of hospital obstetric services and facilities.

Conflict of Interests

Authors have no conflict of interests.

Acknowledgments

The authors wish to express their utmost thanks to the records unit and the entire Department of O&G, KBTH, for their permission to access the data.

References

1. Bohun-Chudynivy B, Curtis G, Knight A, McClave D, Osborne DL, Savada AM, et al. Domestic trends to the year 2015: Forecasts for the united states. Washington, DC: Federal Research Division Library of Congress; 1991.
2. Cohen A. Seasonal daily effect on the number of births in Israel. J R Stat Soc Ser C 1983; 32(3): 228-35.
3. Caleiro A. Forecasting the number of births in Portugal. Proceedings of the Conference of European Statisticians Statistical Office of The European Union (EUROSTAT); 2010 Apr. 28-30; Lisbon, Portugal.
4. Box GE, Jenkins GM. Time series analysis: Forecasting and control. Upper Saddle River, NJ: Prentice Hall; 1976.