International Journal of Applied Science and Mathematics
Volume 7, Issue 5, ISSN (Online): 2394-2894

Time Series Analysis of Monthly Birth Rate in Ghana

Buah Ahoba Masha [1*], Zigli David Delali [2] and Annan Wobir Reuben [3]
[1, 2, 3] Department of Mathematical Sciences, University of Mines and Technology, Tarkwa, Ghana.
Corresponding author email id: afsab4ever@yahoo.com

Date of publication (dd/mm/yyyy): 17/10/2020

Abstract **– This project seeks to model and forecast the monthly Birth rate in Ghana using Box- Jenkins method. Monthly data ranging from January, 2014 through to December 2019 was collected from Ghana Statistical Service. SPSS [12] and E-views statistical software were used for the analysis of the data. ARIMA (1, 1, 1) model was found to be the most suitable model with the least normalised Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC) values amidst the model test in the work. The developed model was then used to forecast for the 12 months and the results showed that the monthly Birth rate will decrease in the subsequent years. Hence the forecasted results showed a decrease in the monthly Birth rate in Ghana for subsequent years. In conclusion, it was established that there was a negative relationship between the dependent variable (Birth rate) and the independent variable (time/months).**

Keywords **– Time Series, Birth Rate, Box-Jenkins Method.**

I. Introduction

One of the main challenges Ghana faced in 1950 was high fertility. Ghana introduced its first Population Policy in 1969 to address the problem of high fertility and escalating growth rates in the face of declining mortality [4]. After 25 years of little progress, the Population Policy was revised in 1994 to include a systematic integration of population in development planning with renewed emphasis on fertility deceleration to keep pace with resource generation. Since then, Ghana has made significant progress in reducing its fertility. An important aspect of the demographic and health surveys conducted in the country has been the collection of birth history information to enhance data availability for monitoring the progress in fertility decline.

Birth rate in simple terms is defined as the total number of live births per 1000 of population in a year. High birth rates could lead to a significant increase in population, which could be good for Ghana because labour force increases and vice versa [9].

This important factor has been neglected as a result of the poor documentation of birth rate in the district and the nation at large. It has become difficult in determining the trend and planning for the economy of the nation.

This work therefore seeks to examine the trend pattern of birth rates as well as fitting an ARIMA model to forecast the monthly birth rates in Ghana.

II. Preliminaries

2.1. *Time Series*

2.1.1. *Definition*

Models for time series data has many forms and represent different stochastic process. There are two most widely used models in literature, and they are Autoregressive (AR) and Moving Average (MA) models. Combining these two models, Autoregressive Moving Average (ARMA) model is achieved [13]. When the

Integrated component is added Autoregressive Integrated Moving Average (ARIMA) model is obtained and its different variations are based on the famous Box-Jenkins principle and so these are also broadly known as the Box- Jenkins models.

2.1.2. *Seasonal Variation*

Many time series data exhibit a seasonal variation. Seasonal variations describe any regular variation or almost identical patterns that a time series appears to follow with a period less than one year. They are the changes which repeats themselves within a fixed time [8]. Examples of such are the sale of umbrella, the sale of firecrackers and Christmas cards, Traffic on roads in the morning and evening hours etc. [2].

2.1.3. *Secular Trends*

The secular trend is the main component of a time series which describes the direction in which the plot of a time series appears to be going over a long interval of time as a result from long term effects of socio-economic and political factors. This is the type of tendency which continues to persist for a very long period. The trend can be straight or curvilinear with rise or decline in a time series over a long period. Examples of straight-line trends are increase of vehicular usage in Ghana, increase in the population of Ghana from 1960 to 2010. Example of curvilinear relationship is a life circle of a new product [6].

2.1.4. *Cyclic Variation*

These are long term oscillations or swings occurring in a time series. These cyclic variations may or may not be periodic and are mostly observed in economics data and periods of such oscillations are generally extended from five to twelve years or more. Examples are economic cycles representing intervals of propensity, recession, depression and depression [1].

2.1.5. *Random Variation*

These refer to sporadic motions (occurring at irregular points in time) in a time series which are unlikely to be repeated. Random variation comprises the irregular changes in a time series and cannot be explained by the trends, seasonal or cyclic movements. These variations, though accidental in nature, can cause a continual change in the trends, seasonal and cyclical oscillations during forthcoming period. Examples are floods, fire, elections, employee strikes, earthquakes, revolutions, epidemic etc [5].

2.1.6. *Time Series Tools*

Box-Jenkins Method: The Box-Jenkins method to time series forecasting is one of the most widely used methodology for analysing a time series data [7]. It can handle any series whether stationary or not as well as a series with or without seasonal elements. This time series forecasting is a step by step approach which apply Autoregressive Moving Average (ARMA) or Autoregressive Integrated Moving Average (ARIMA) to find the best fit of time series model to past values of a time series data [11].

2.1.7. *Differencing to Achieve Stationarity*

The time plot of the series is checked to see whether the series is stationary or not. A time plot which shows presence of either an upward trend or downward trend clearly is not stationary. The mean and variance of the ti- -me series should not be a function of time. Thus, must satisfy the assumption of homoscedasticity. The graph of the correlogram of the series can also be used to check for stationarity of a time series.

2.1.8. *Stationarity Test*

Augmented Dickey-Fuller (ADF) test was performed on the time series data to check whether the time series data is stationary or not.

The hypothesis for the test:

: Not Stationary.

: Stationary.

Level of significance (α) = 5%

Conclusion on the hypothesis testing is based on the p-value of the ADF test. In general, a p-value of less than 5% means that the null hypothesis is rejected and vice versa.

Model identification and parameter estimation. The ACF and PACF are informative in determining the order of an ARMA model. Once the series is stationary, the ACF and PACF are examined to decide on the appropriate orders of AR and MA components [3]. Identification of the AR and MA orders are the key aspect of the Box-Jenkins forecasting approach.

III. Main Result

3.1. *Examining the Stationarity of the Monthly Birth Rate*

The time series plot of the monthly birth ratio data is shown in Fig 1. The time horizon of the dataset comprises January, 2014 to December, 2019.

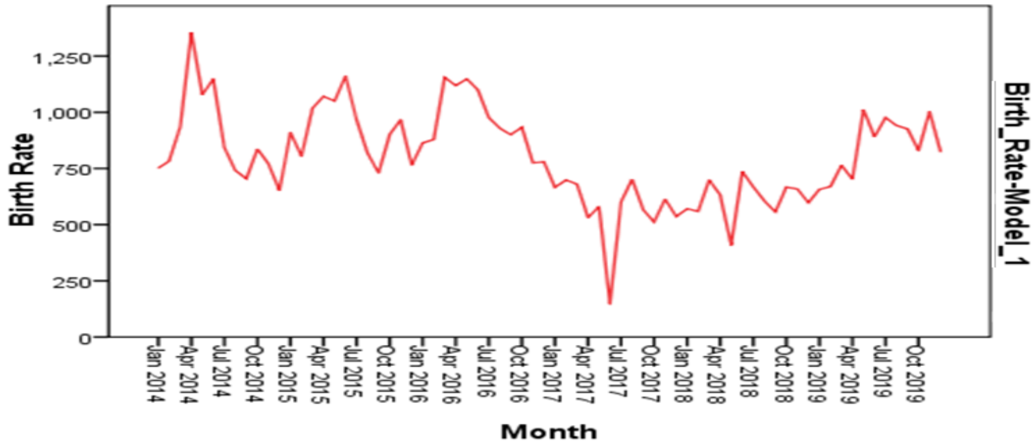

Fig. 1. Time Series Plot of Birth Rate in Ghana from 2014 to 2019.

From Fig. 1, the general form of the pattern of the monthly birth rate in Ghana changes over time. In other words, the general trend of birth rate does not look similar across the months, so it can be regarded as a non-stationary series.

3.2. *ANOVA Test of Stationarity*

Regression analysis is used to show if there is a relationship between the dependent variable of the series and the independent variable of the raw data [10].

- Null Hypothesis: Birth Rate is Stationary.
- Alternate Hypothesis: Birth Rate is not stationary.

Significance level = 5%.

Decision Rule: If the p value is greater than the level of significance, do not reject the null hypothesis i.e. the birth is stationary otherwise, it is not stationary.

Table 1. ANOVA Test of Stationarity.

Model		Sum of Squares	Df	Mean Square	F	Sig. (p -value)
1	Regression	419082.248	1	419082.248	11.006	0.001
	Residual	2665315.071	70	38075.930		
	Total	3084397.319	71			

From Table 1, the p-value is 0.001 which is less than the 5% significance level. Therefore, the null hypothesis is rejected indicating that the of birth rates series is not stationary.

3.3. *Augmented Dickey Fuller (ADF) Root Test*

The Augmented Dickey Fuller unit root test is carried out by setting the hypothesis as:

- Null Hypothesis: Birth Rate is Not Stationary (i.e. it has a unit root).
- Alternate Hypothesis: Birth Rate is Stationary (i.e. it has no unit root).

The chosen Level of Significance is 5%

Decision Criteria: if the absolute value of the t-statistic is greater than the absolute of the critical value (given for the level of significance), then the null hypothesis is rejected, i.e. the birth rate is stationary otherwise, it is not stationary.

Table 2. Augmented Dickey Fuller Unit Root Test.

Augmented Dickey-Fuller test statistic		-2.445520
Test critical values	1% level	-3.527045
	5% level	-2.903566
	10% level	-2.589227

From Table 2, the ADF t-statistic value for direct values is less than the critical value at 5% level of significance, so we reject the null hypothesis. Therefore, the monthly birth rate series is not stationary.

3.4. *Differencing the Series to Achieve Stationarity*

According to Box-Jenkins approach, time series data are assumed to be stationary. In order to achieve that, the correlogram of the time series is studied to check whether the time series is stationary or not. If not the method of differencing is applied on the data to remove the variations in the series so that the mean and the variance will be constant over time.

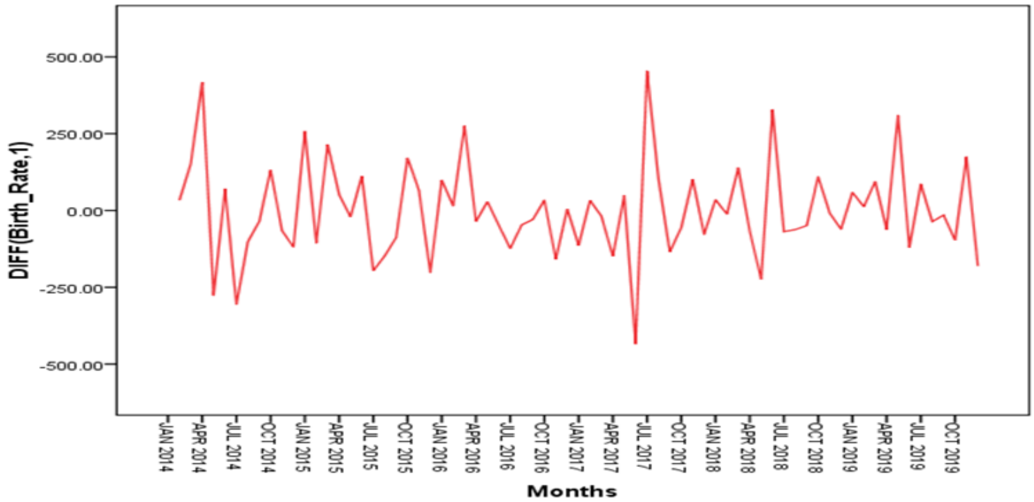

Fig. 1. A Correlogram of the first Differencing of the Time Series Plot.

Regression analysis is used to show if there is a relationship between the dependent variable of the first order differenced data and the independent variable of the first ordered differenced.

- Null Hypothesis, H_0 : Birth Rate is Stationary.
- Alternate Hypothesis, H_1 : Birth Rate is not Stationary.

Significance level = 5%.

Decision Rule:

If the p – value is greater than the level of significance, do not reject the null hypothesis i.e. the birth is stationary otherwise, it is not stationary.

Significance level = 5%.

Decision Rule:

If the p – value is greater than the level of significance, do not reject the null hypothesis i.e. the birth is stationary otherwise, it is not stationary.

Table 3. Stationarity Test of Transformed Birth Rates.

Model		**Sum of Squares**	**Df**	**Mean Square**	**F**	**Sig. (p -value)**
1	Regression	37.042	1	37.042	0.001	0.970
	Residual	1742211.944	69	25249.448		
	Total	1742248.986	70			

From Table 3, the p-value is 0.970 which is greater than the 5% significance level. Therefore, the null hypothesis is not rejected indicating that the first order difference of birth rates is stationary.

3.5. *Identification of Tentative Model*

Box Jenkins approach can be used to identify the possible Autoregressive (AR) and the Moving Average (MA) and their order. Since the series was differenced, it implies that Autoregressive Integrated Moving

Average (ARIMA) model, ARIMA (p, d, q) model would be appropriate for time series data. Since the data was differenced only one time, it can be suggested that ARIMA (p, 1, q) is going to be used for the data where p denotes the autoregressive (AR) that can be obtained from the ACF and q denotes the moving average (MA) which can be obtained from the PACF. After differencing procedure to get a stationary time series is achieved, correlation is examined to decide on appropriate orders of the AR and MA components and this is done by examining ACF and PACF plots of the differenced data.

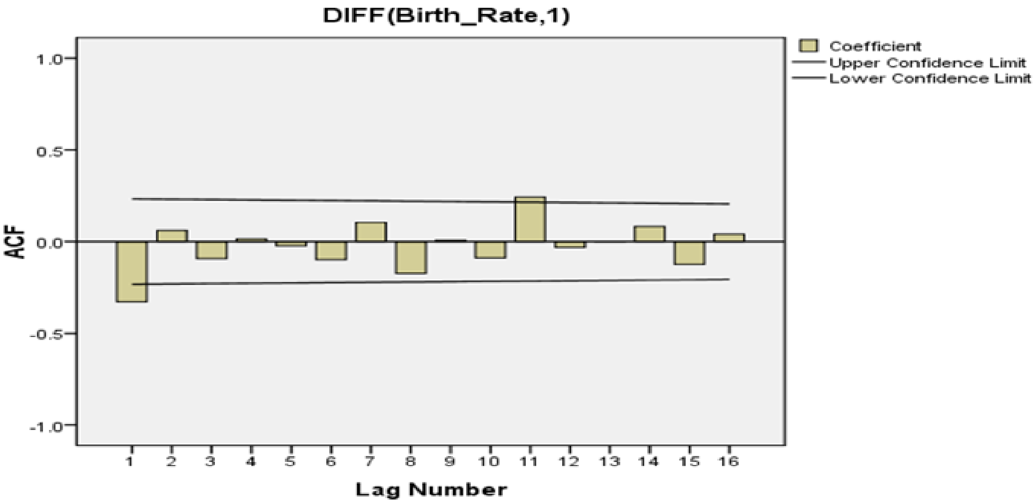

Fig. 2. A Correlogram of ACF with First Order Differencing.

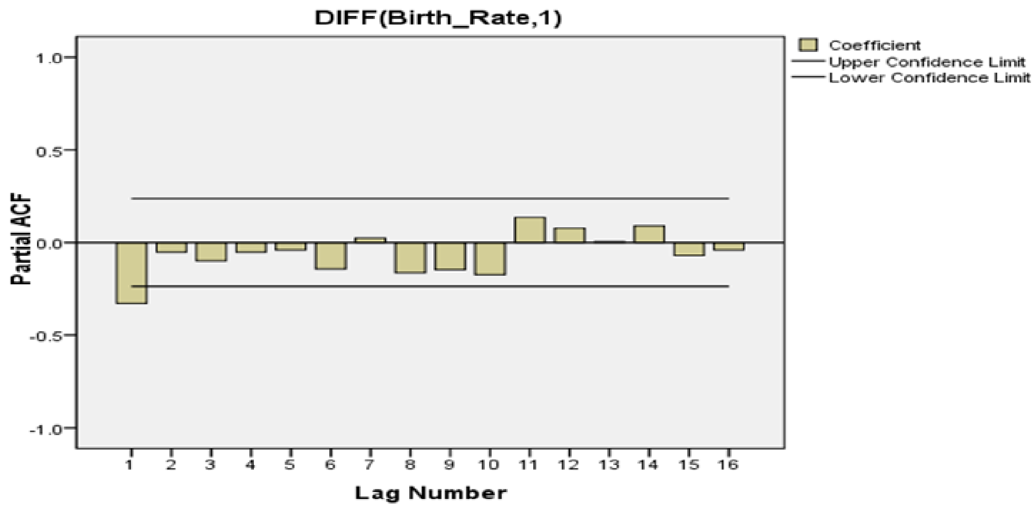

Fig. 3. A Correlogram of PACF with First Order Differencing.

From the ACF graph, the plot shows the autocorrelation function of first differencing of birth rates at various lags. It was found out that, the ACF has a negative significant spike at lag 1 and a positive significant spike at lag 11. From the above PACF plot, it has a negative significant spike at lag 1.

Table 3. ARIMA Model Selection Criteria.

ARIMA (p, d, q)	AIC	BIC
(1,1,1)	0.132987	0.259468
(2,1,0)	0.156197	0.282678
(1,1,2)	0.160614	0.318716

ARIMA (p, d, q)	AIC	BIC
(2,1,1)	0.160629	0.318731
(3,1,0)	0.166629	0.324731
(4,1,0)	0.184385	0.374108
(1,1,3)	0.187673	0.377395
(3,1,1)	0.187955	0.377677
(2,1,2)	0.188401	0.378123
(3,1,3)	0.196033	0.448996
(1,1,0)	0.199539	0.294400

The final model can be selected using a penalty function statistic such as Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC). The AIC and BIC are measures of the goodness of fit of an estimated statistical model. From Table 3, the criterion values suggested that ARIMA (1, 1, 1) is the best among the competing models since it has the least Bayesian Information Criterion (BIC). It is the model that best explains the data with a minimum of free parameters but also includes a penalty that is an increasing function of the number of estimated parameters.

3.6. *Ljung-Box Test of Serial Correlation for ARIMA (1, 1, 1)*

The hypothesis for this test is outlined as:

- Null Hypothesis $H_0: \beta_0 = 0$ (No Serial Correlation).
- Alternate Hypothesis $H_1: \beta_0 \neq 0$ (Serial Correlation).

Level of Significance, $\alpha = 5\%$.

Decision Rule: Reject the null hypothesis if the significant value is less than the significance level, $\alpha = 5\%$, otherwise do not reject the null hypothesis.

Table 4. Model Statistics for ARIMA (1, 1, 1).

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)		
		Stationary R-squared	Statistics	DF	Sig. (p-value)
DIFF (Birth_Rate, 1) - Model_1	1	0.00002	21.250	18	0.267

From Table 4, since $0.267 > 0.05$ (i.e. level of significance), we do not reject the null hypothesis, which implies that the Ljung-Box Q(18) statistics is not statistically significant at 5% level of significance. Hence, the ARMA (1, 1, 1) model is independently distributed which signifies that there are no serial correlations in the model and therefore it is the preferred model.

Table 5. ARIMA (1, 1, 1) Model Parameters.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Constant	-8.706	0.159924	41.64688	0.0000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.463	0.130147	6.707471	0.0000
MA(1)	0.814	0.180884	-2.799738	0.0067

From Table 5, the ARIMA (1, 1, 1) model with constant term would have the prediction equation defined by equation (1).

$$B_t = \mu + B_{t-1} + \alpha\left(B_{t-1} - B_{t-2}\right) - \phi\xi_{t-1} + \xi_t \tag{1}$$

Where, $B_t =$ Birth Rate in month t

$\mu =$ Constant Term

$B_{t-1} =$ Lag 1 of B_t

$B_{t-2} =$ Lag 2 of B_t

$\xi_t =$ Random Error Term

$\xi_{t-1} =$ Lag 1 of the Random Error Term

$\alpha =$ Auto Regressive (AR) Coefficient

$\phi =$ Moving Average (MA) Coefficient

The Table 5 above shows the parameter estimate of ARIMA (1,1,1) with their respective significance level. Putting the values of the constants into Equation (1) yields Equation (2).

$$B_t = 6.660352 + B_{t-1} + 0.872960\left(B_{t-1} - B_{t-2}\right) + 0.506429\xi_{t-1} + \xi_t \tag{2}$$

Equation (2) represents the trend of the monthly birth rate in Ghana and hence, the forecasting equation.

3.7. *Model Forecasting*

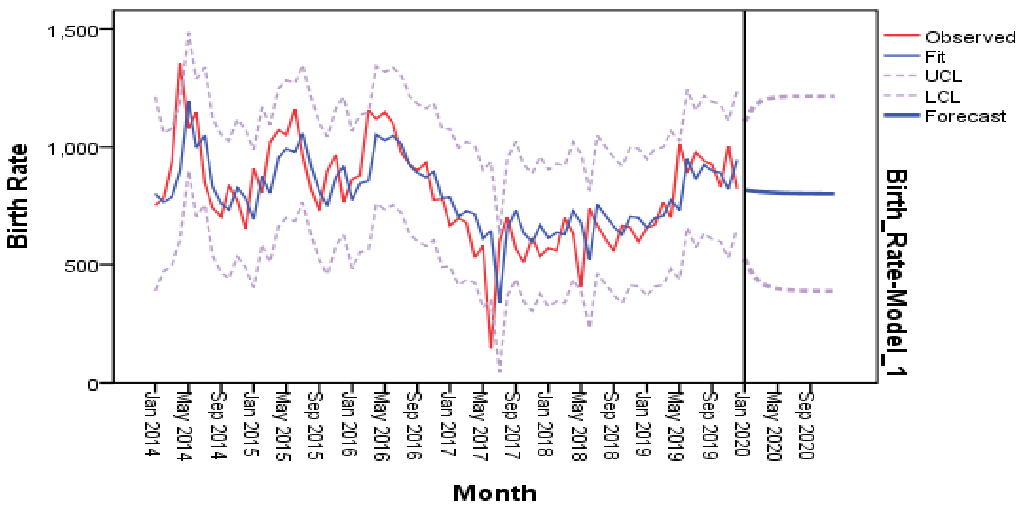

Fig. 5. Forecasted Values of the Birth Rate in Ghana for 2020.

Table 6. Forecasted Values of the Birth Rate in Ghana for 2020.

		95% Confidence Interval	
Month	**Forecasted**	**Lower Level**	**Upper Level**
Jan-20	816.4865	524.8075	1108.1660
Feb-20	811.8737	454.4564	1169.2910
Mar-20	808.6068	422.4091	1194.8040
Apr-20	806.2932	406.4399	1206.1460
May-20	804.6547	398.1251	1211.1840
Jun-20	803.4943	393.6571	1213.3310
Jul-20	802.6725	391.1864	1214.1590
Aug-20	802.0905	389.7799	1214.4010
Sep-20	801.6783	388.9548	1214.4020
Oct-20	801.3864	388.4559	1214.3170
Nov-20	801.1797	388.1454	1214.2140
Dec-20	801.0333	387.9470	1214.1200

IV. Conclusion

In this study, it was observed that the original series data for the monthly birth rate in Ghana was not stationary, so first differencing was done to attain stationarity. The ADF and KPSS test was carried out to ascertain the stationarity of the first differenced data. ARIMA (1, 1, 1) was found to have a good fit and hence appropriate for the trend in the monthly birth rate in Ghana from January, 2014 to December, 2019.

References

[1] Goldbeter, A., (1997) "Modelling Biochemical Oscillations and Cellular Rhythms." *Current Science*, Vol. 73, no. 11, pp. 933–939.
[2] Hamilton, J.D. (1990), "Analysis of Time Series Subject to Changes in Regime", *Journal of econometrics*, Vol. *45*, pp.39-70.
[3] Helga, Silaghi & Claudiu, Costea. (2008), "Wind Speed Prediction using Box-Jenkins Method", Journal of Computer Science and Control Systems. 1.
[4] Kwankye, S.O. (2013), "Growing old in Ghana: Health and Economic Implications", Postgraduate *Medical Journal of Ghana*, Vol. 10, pp. 88-97.
[5] Mackas, D.L., Greve, W., Edwards, M., Chiba, S., Tadokoro, K., Eloire, D., Mazzocchi, M.G., Batten, S., Richardson, A.J., Johnson, C. and Head, E. (2012), "Changing Zooplankton Seasonality in a Changing Ocean: Comparing Time Series of Zooplankton Phenology" Progress *in Oceanography*, Vol. 97, pp. 31-62.
[6] Miranda, L.C. and Lima, C.A. (2010), "On Trends and Rhythms in Scientific and Technological Knowledge Evolution", International *Journal of Technology Intelligence and Planning*, Vol. 23 pp.76-109.
[7] Nuzula, Isnaini. (2019). The Approach of box Jenkins Time Series Analysis for Predicting Stock Price on LQ45 Stock Index. Journal of profit. pp. 18-25
[8] Nwogu, E. & Iwueze, Iheanyi & Nlebedim, Valentine. (2016). Some Tests for Seasonality in Time Series Data. Journal of Modern Applied Statistical Methods. 15. 382-399. 10.22237/jmasm/1478002920.
[9] Sackey, H.A. (2005), Female *labor force participation in Ghana: The effects of education*. The African Research Economic Research Consortium, Research paper 150, Nairobi, Kenya.
[10] Schneider, A., Hommel, G. and Blettner, M. (2010), "Linear regression analysis: part 14 of a series on evaluation of scientific publications*", Deutsches Ärzteblatt International*, Vol. 44, pp.776.
[11] Singh, A. and Mishra, G.C. (2015), "Application of Box-Jenkins method and Artificial Neural Network procedure for time series forecasting of prices" Statistics *in Transition new series*, *1* Vol. 16, pp.83-96.
[12] www. statsghana.gov.gh, www.statistica.com , (Accessed Feb, 2020).
[13] Zhang Z., Moore, J. C., (2015), "Autoregressive Moving Average Models" Mathematical and Physical Fundamentals of Climate Change.

AUTHOR'S PROFILE

First Author
Buah Masha Ahoba is a lecturer in the Department of Mathematical Sciences, University of Mines and Technology, Ghana and currently pursuing her PhD in Mathematics. Her research interest lies in Asset Pricing, Derivatives, Portfolio Optimimsation and Risk Management.

Second Author
Zigli David Delali is a Lecturer in the Department of Mathematical Sciences, University of Mines and Technology, Ghana. He holds MPhil in Pure Mathematics and currently pursuing PhD in Mathematics. His research interest lies in Algebraic Topology and more specifically Fundamental Groups.

Third Author
Annan Wobir Reuben is a student in the Department of Mathematical Sciences, University of Mines and Technology, Ghana.