International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 7, Issue 8, pp 512-518, August-2023
https://doi.org/10.47001/IRJIET/2023.708074

Making Use of ARIMA Model Predictions to Inform Maternal and Neonatal Healthcare Policies in the United States of America

[1]Dr. Smartson. P. NYONI, [2]Thabani NYONI

[1]ZICHIRe Project, University of Zimbabwe, Harare, Zimbabwe
[2]Independent Researcher & Health Economist, Harare, Zimbabwe

***Abstract* - Neonatal mortality remains a global health problem hence public health interventions must be designed to address this challenge. Efforts must be directed to improve the quality of healthcare services during antenatal, delivery and postnatal periods. This research uses annual time series data on neonatal mortality rate (NMR) for USA from 1968 to 2019 to predict future trends of NMR over the period 2020 to 2030. Unit root tests have shown that the series under consideration is an I (2) variable. The optimal model based on AIC is the ARIMA (0,2,1) model. The ARIMA model predictions indicate that neonatal mortality is expected to hover around 3 deaths per 1000 live births throughout the out of sample period. Hence, the US government should address various maternal and child health challenges existing in different parts of the country to keep neonatal mortality under control.**

Keywords: ARIMA, Forecasting, NMR.

I. INTRODUCTION

Surveillance mechanisms are critical in the management and control of disease incidences and other health related problems. Modelling techniques are widely applied as early surveillance tools in public health in order to facilitate planning, decision making and allocation of resources. Morbidity and mortality trends in several countries have been previously analyzed using various methods including modelling and forecasting approaches. The US is among the first world countries which have consistently reported low maternal and neonatal mortality rates as a result of high quality health service provision. Global estimates show that 2.6 million neonatal deaths occur every year (Basha *et al.* 2022). The majority of neonatal deaths are reported in low and middle income countries particularly in South Asia (39%) and Sub-Saharan Africa (38%) (Hug *et al*. 2017; Kolola *et al*.2016). The global under 5 mortality declined from 93 per 1000 live births in 1990 to 41 deaths per 1000 live births in 2016 whereas the global neonatal mortality rate declined from 37 deaths per 1000 live births in 1990 to 19 per 1000 live births in 2016 (WHO, 2018; Hug *et al.* 2017). The aim of this study is to model and project future trends of neonatal mortality rate (NMR) for the US using the popular Box-Jenkins ARIMA technique. This statistical and econometric approach is useful for modelling and forecasting linear time series data (Nyoni, 2018; Box & Jenkins, 1970). The findings of this study are expected to detect abnormal trends of neonatal mortality rate and stimulate a prompt response to keep neonatal deaths under control.

II. LITERATURE REVIEW

Neonatal mortality is a global issue which deserves to be researched on. There are many studies globally that have examined factors associated with neonatal mortality, however there are limited forecasting studies. Reis *et al*. (2021) evaluated the fetal and infant mortality rates due to congenital anomalies (CA) in Maranhão from 2001 to 2016 in Brazil.Data were obtained from the SINASC, and SIM databases. The study used simple linear regression, Poisson distribution, and ANOVA (Bonferroni's post hoc test) and analyzed the public data (2001–2016) of 1934858 births and determined the fetal, neonatal, perinatal, and post-neonatal mortality rates associated with CA by mesoregions. The results indicated mortality rates due to CA in Maranhão increased over the period 2001–2016 possibly as a result of improved maternal-infant health conditions eliminating other causes of death. Baroni *et al*. (2021) outlined an integrated dataset containing monthly data in a historical series from 1996 to 2017 with information on all births, neonatal deaths, and NMR (total, early and late components) enriched with information related to the municipality. It is adataset of historical data with information on the number of births, the number of neonatal deaths, the neonatal mortality rate (including early and late), and geographic information for each month (between January 1996 and December 2017) and Brazilian municipality. A retrospective review study was conducted by Falciglia *et al*. (2020) to investigate mortality in periviable neonates ≤23 weeks gestational age and calculate its impact on overall neonatal mortality rate over a 12-year period

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 8, pp 512-518, August-2023

https://doi.org/10.47001/IRJIET/2023.708074

(1998–2009). It was found that neonatal mortality rate from periviability was 96.2% and constituted half of the overall rate in the period (1998–2009). There was not significant reduction of periviable mortality between 2010 and 2015. Juarez *et al.* (2020) conducted a quality improvement study to increase the detection of neonatal complications by lay midwives in rural Guatemala, thereby increasing referrals to a higher level of care. A quality improvement team in Guatemala reviewed drivers of neonatal health services provided by lay midwives. Improvement interventions included training on neonatal warning signs, optimized mobile health technology to standardize assessments and financial incentives for providers. The primary quality outcome was the rate of neonatal referral to a higher level of care. It was found that structured improvement interventions, including mobile health decision support and financial incentives, significantly increased the detection of neonatal complications and referral of neonates to higher levels of care by lay midwives operating in rural home-based settings in Guatemala. Alexandra & Alkema (2018) applied the Bayesian Hierarchical model to estimate the global neonatal mortality using available data sources. The study findings revealed that the proportion of under 5 deaths which are neonatal are constant at 54%.

III. METHODOLOGY

The Autoregressive (AR) Model

A process X_t (annual NMR at time t) is an autoregressive process of order p, that is, AR (p) if it is a weighted sum of the past p values plus a random shock (Z_t) such that:

$$X_t = \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \emptyset_3 X_{t-3} + \cdots + \emptyset_p X_{t-p} + Z_t \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [1]$$

Using the backward shift operator, B, such that $BX_t = X_{t-1}$, the AR (p) model can be expressed as in equation [2] below:

$$Z_t = \emptyset(B) X_t \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [2]$$

where $\emptyset(B) = 1 - \emptyset_1 B - \emptyset_2 B^2 - \emptyset_3 B^3 - \cdots - \emptyset_p B^p$

The 1st order AR (p) process, AR (1) may be expressed as shown below:

$$X_t = \emptyset X_{t-1} + Z_t \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [3]$$

Given $\emptyset = 1$, then equation [3] becomes a random walk model. When $|\emptyset| > 1$, then the series is reffered to as explosive, and thus non-stationary. Generally, most time series are explosive. In the case where $|\emptyset| < 1$, the series is said to be stationary and therefore its ACF (autocorrelation function) decreases exponentially.

The Moving Average (MA) Model

A process is reffered to as a moving average process of order q, MA (q) if it is a weighted sum of the last random shocks, that is:

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [4]$$

Using the backward shift operator, B, equation [4] can be expressed as follows:

$$X_t = \theta(B) Z_t \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [5]$$

where $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$

Equation [4] can also be expressed as follows:

$$X_t - \sum_{j \leq 1} \pi_j X_{t-j} = Z_t \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [6]$$

for some constant π_j such that:

$$\sum_{j \leq 1} |\pi_j| < \infty$$

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 8, pp 512-518, August-2023

https://doi.org/10.47001/IRJIET/2023.708074

This implies that it is possible to invert the function taking the Z_t sequence to the X_t sequence and recover Z_t from present and past values of X_t by a convergent sum.

The Autoregressive Moving Average (ARMA) Model

While the above models are good, a more parsimonious model is the ARMA model. The AR, MA and ARMA models are applied on stationary time series only. The ARMA model is just a mixture of AR (p) and MA (q) terms, hence the name ARMA (p, q). This can be expressed as follows:

$$\emptyset(B)X_t = \theta(B)Z_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [7]$$

Thus:

$$X_t\left(1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p\right) = Z_t\left(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q\right) \ldots\ldots\ldots [8]$$

where $\emptyset(B)$ and $\theta(B)$ are polynomials in B of finite order p, q respectively.

The Autoregressive Integrated Moving Average (ARIMA) Model

The AR, MA and ARMA processes are usually not applied empirically because in most cases many time series data are not stationary; hence the need for differencing until stationarity is achieved.

$$\left.\begin{array}{c} \textit{The first difference is given by:} \\ X_t - X_{t-1} = X_t - BX_t \\ \textit{The second difference is given by:} \\ X_t(1-B) - X_{t-1}(1-B) = X_t(1-B) - BX_t(1-B) = X_t(1-B)(1-B) = X_t(1-B)^2 \\ \textit{The third difference is given by:} \\ X_t(1-B)^2 - X_{t-1}(1-B)^2 = X_t(1-B)^2 - BX_t(1-B)^2 = X_t(1-B)^2(1-B) = X_t(1-B)^3 \\ \textit{The } d^{th} \textit{ difference is given by:} \\ X_t(1-B)^d \end{array}\right\} \ldots [9]$$

Given the basic algebraic manipulations above, it can be inferred that when the actual data series is differenced "d" times before fitting an ARMA (p, q) process, then the model for the actual undifferenced series is called an ARIMA (p, d, q) model. Thus equation [7] is now generalized as follows:

$$\emptyset(B)(1-B)^d X_t = \theta(B)Z_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [10]$$

Therefore, in the case of modeling and forecasting NMR, equation [10] can be written as follows:

$$\emptyset(B)(1-B)^d X_t = \theta(B)Z_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [11]$$

The Box – Jenkins Approach

The first step towards model selection is to difference the series in order to achieve stationarity. Once this process is over, the researcher will then examine the correlogram in order to decide on the appropriate orders of the AR and MA components. It is important to highlight the fact that this procedure (of choosing the AR and MA components) is biased towards the use of personal judgement because there are no clear – cut rules on how to decide on the appropriate AR and MA components. Therefore, experience plays a pivotal role in this regard. The next step is the estimation of the tentative model, after which diagnostic testing shall follow. Diagnostic checking is usually done by generating the set of residuals and testing whether they satisfy the characteristics of a white noise process. If not, there would be need for model re – specification and repetition of the same process; this time from the second stage. The process may go on and on until an appropriate model is identified (Nyoni, 2018). The Box – Jenkins technique was proposed by Box & Jenkins (1970) and is widely used in many forecasting contexts, including medicine. In this paper, hinged on this technique; the researcher will use automatic ARIMA modeling for estimating equation [10].

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 8, pp 512-518, August-2023

https://doi.org/10.47001/IRJIET/2023.708074

Data Issues

This study is based on annual NMR in the US for the period 1968 to 2019. The out-of-sample forecast covers the period 2020 to 2030. All the data employed in this research paper was gathered from the World Bank online database.

Evaluation of ARIMA Models

Criteria Table

Table 2: Criteria Table

Model Selection Criteria Table
Dependent Variable: DLOG(Y, 2)
Date: 01/29/22 Time: 12:25
Sample: 1968 2019
Included observations: 50

Model	LogL	AIC*	BIC	HQ
(0,1)(0,0)	147.008157	-5.760326	-5.645605	-5.716640
(1,0)(0,0)	146.265783	-5.730631	-5.615910	-5.686945
(2,1)(0,0)	148.190965	-5.727639	-5.536436	-5.654828
(1,1)(0,0)	147.060793	-5.722432	-5.569470	-5.664183
(0,2)(0,0)	147.049522	-5.721981	-5.569019	-5.663732
(2,0)(0,0)	146.718395	-5.708736	-5.555774	-5.650487
(3,0)(0,0)	147.264743	-5.690590	-5.499387	-5.617779
(1,2)(0,0)	147.240307	-5.689612	-5.498410	-5.616801
(2,2)(0,0)	148.211403	-5.688456	-5.459013	-5.601083
(0,3)(0,0)	147.162751	-5.686510	-5.495308	-5.613699
(3,2)(0,0)	148.906612	-5.676264	-5.408581	-5.574329
(4,0)(0,0)	147.284616	-5.651385	-5.421942	-5.564012
(3,1)(0,0)	147.277203	-5.651088	-5.421645	-5.563715
(2,3)(0,0)	148.265568	-5.650623	-5.382940	-5.548687
(1,3)(0,0)	147.228137	-5.649125	-5.419683	-5.561752
(0,4)(0,0)	147.187251	-5.647490	-5.418047	-5.560117
(4,3)(0,0)	150.133697	-5.645348	-5.301184	-5.514288
(3,3)(0,0)	149.086909	-5.643476	-5.337553	-5.526979
(0,5)(0,0)	147.956411	-5.638256	-5.370573	-5.536321
(1,5)(0,0)	148.748176	-5.629927	-5.324003	-5.513430
(2,4)(0,0)	148.325368	-5.613015	-5.307091	-5.496517
(5,0)(0,0)	147.323457	-5.612938	-5.345255	-5.511003
(4,1)(0,0)	147.292085	-5.611683	-5.344000	-5.509748
(1,4)(0,0)	147.230561	-5.609222	-5.341539	-5.507287
(4,2)(0,0)	148.227989	-5.609120	-5.303196	-5.492622
(5,2)(0,0)	149.159199	-5.606368	-5.262204	-5.475308
(2,5)(0,0)	148.903763	-5.596151	-5.251986	-5.465091
(4,4)(0,0)	149.675221	-5.587009	-5.204604	-5.441387
(3,5)(0,0)	149.526740	-5.581070	-5.198665	-5.435448
(3,4)(0,0)	148.347312	-5.573892	-5.229728	-5.442833
(5,1)(0,0)	147.333097	-5.573324	-5.267400	-5.456826
(5,4)(0,0)	150.194532	-5.567781	-5.147136	-5.407597
(5,3)(0,0)	149.183583	-5.567343	-5.184939	-5.421721
(0,0)(0,0)	140.852872	-5.554115	-5.477634	-5.524991
(5,5)(0,0)	149.585195	-5.503408	-5.044522	-5.328662
(4,5)(0,0)	147.962946	-5.478518	-5.057873	-5.318334

International Research Journal of Innovations in Engineering and Technology (IRJIET)

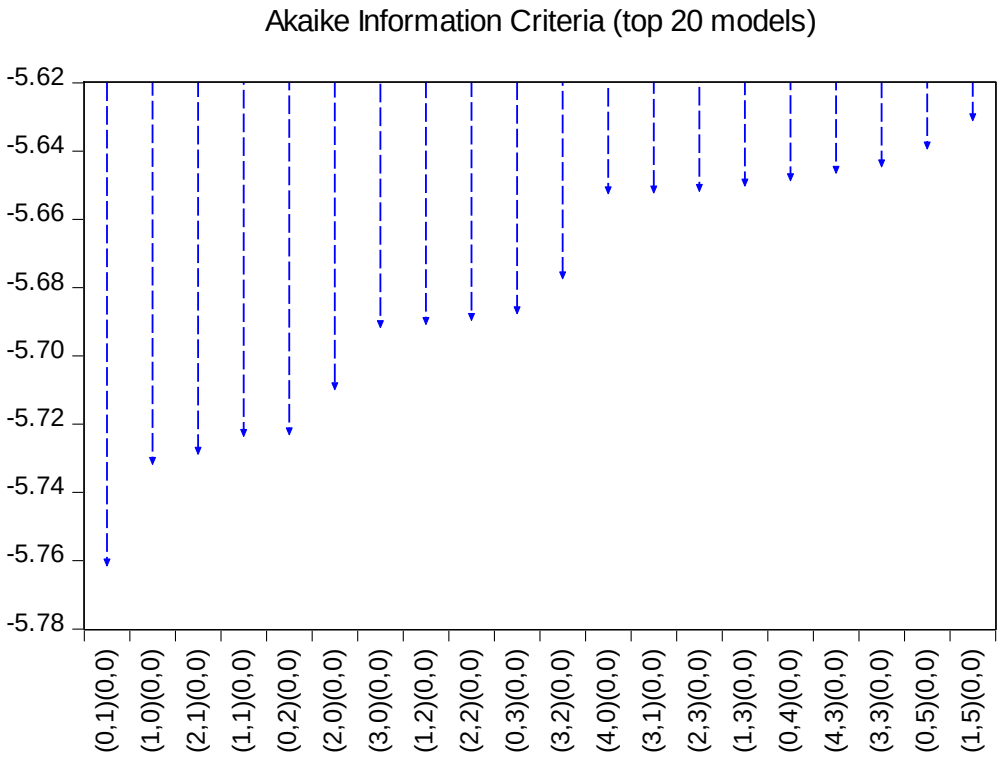
ISSN (online): 2581-3048

Volume 7, Issue 8, pp 512-518, August-2023

https://doi.org/10.47001/IRJIET/2023.708074

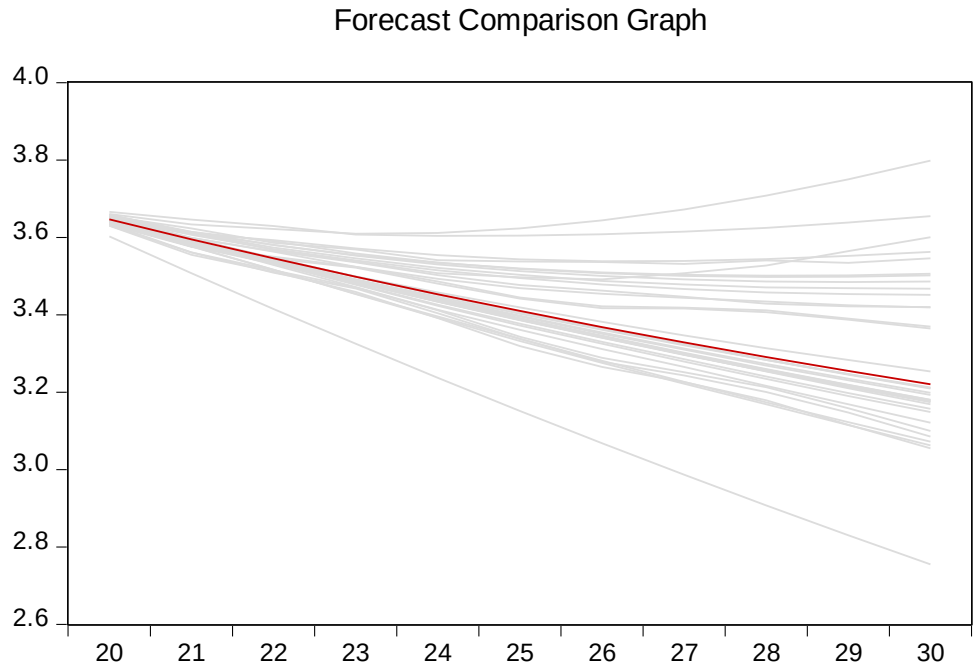
Criteria Graph

Figure 1: Criteria Graph



Forecast Comparison Graph

Figure 2: Forecast Comparison Graph



International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 7, Issue 8, pp 512-518, August-2023
https://doi.org/10.47001/IRJIET/2023.708074

Table 2 and Figure 1 indicate that the optimal model is the ARIMA (0,2,1) model. Figure 2 is a combined forecast comparison graph showing the out-of-sample forecasts of the top 25 models evaluated based on the AIC criterion. The red line shows the forecast line graph of the optimal model, the ARIMA (0,2,1) model.

IV. RESULTS

Summary of the Selected ARIMA () Model

Table 3: Summary of the Optimal Model

Automatic ARIMA Forecasting
Selected dependent variable: DLOG(Y, 2)
Date: 01/29/22 Time: 12:25
Sample: 1968 2019
Included observations: 50
Forecast length: 11

Number of estimated ARMA models: 36
Number of non-converged estimations: 0
Selected ARMA model: (0,1)(0,0)
AIC value: -5.76032626698

Main Results of the Selected ARIMA () Model

Table 4: Main Results of the Optimal Model

Dependent Variable: DLOG(Y,2)
Method: ARMA Maximum Likelihood (BFGS)
Date: 01/29/22 Time: 12:25
Sample: 1970 2019
Included observations: 50
Convergence achieved after 7 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000394	0.000875	0.450207	0.6546
MA(1)	-0.533160	0.144665	-3.685485	0.0006
SIGMASQ	0.000162	4.98E-05	3.263940	0.0021

R-squared	0.223455	Mean dependent var	-2.38E-05
Adjusted R-squared	0.190410	S.D. dependent var	0.014612
S.E. of regression	0.013148	Akaike info criterion	-5.760326
Sum squared resid	0.008125	Schwarz criterion	-5.645605
Log likelihood	147.0082	Hannan-Quinn criter.	-5.716640
F-statistic	6.762235	Durbin-Watson stat	1.966411
Prob(F-statistic)	0.002624		

Inverted MA Roots	.53

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 7, Issue 8, pp 512-518, August-2023
https://doi.org/10.47001/IRJIET/2023.708074

ARIMA () Model Forecast

Tabulated Out of Sample Forecasts

Table 5: Tabulated Out of Sample Forecasts

2020	3.646370671522282
2021	3.594934257562434
2022	3.54561958574185
2023	3.49835896397148
2024	3.453088026763214
2025	3.409745590128899
2026	3.368273514611752
2027	3.328616576023195
2028	3.290722343484712
2029	3.254541064399291
2030	3.220025556000517

Table 5 clearly indicates that neonatal mortality is expected to hover around 3 deaths per 1000 live births throughout the out of sample period.

V. POLICY IMPLICATION & CONCLUSION

Tracking progress towards achieving sustainable development goal targets by the end of 2030 should be a priority for every country. Even first world countries who have managed to achieve SDG-3 targets earlier on should utilize surveillance tools to detect abnormal trends of neonatal mortality to facilitate timeous implementation of neonatal strategies to keep neonatal deaths under control. This study proposed the popular Box-Jenkins ARIMA model to predict future trends of NMR for the US and the model projections indicated that neonatal mortality is expected to hover around 3 deaths per 1000 live births throughout the out of sample period. Hence, the US government should address various maternal and child health challenges existing in different parts of the country to keep neonatal mortality under control.

REFERENCES

[1] Box, D. E., and Jenkins, G. M. (1970). Time Series Analysis, Forecasting and Control, Holden Day, London.
[2] Nyoni, T. (2018). Box-Jenkins ARIMA Approach to Predicting net FDI Inflows in Zimbabwe, *University Library of Munich*, MPRA Paper No. 87737.
[3] Hug L., Sharrow D., and You D (2017). Levels & trends in child mortality: report 2017. Estimates developed by the UN Inter-agency Group for Child Mortality Estimation.
[4] Kolola T., Ekubay M., Tesfa E., and Morka W (2016). Determinants of Neonatal Mortality in North Shoa Zone, Amhara Regional State, Ethiopia. PloS one, 11, 10, e01644
[5] World Health Organization (2018). World health statistics 2018: monitoring health for the SDGs, sustainable development goals.

Citation of this Article:

Dr. Smartson. P. NYONI, Thabani NYONI, "Making Use of ARIMA Model Predictions to Inform Maternal and Neonatal Healthcare Policies in the United States of America" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 7, Issue 8, pp 512-518, August 2023. Article DOI https://doi.org/10.47001/IRJIET/2023.708074

****

