Detection of Anticipated Future Values of Annual Neonatal Mortality Rates for Mauritania Using the Box-Jenkins ARIMA Model

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

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

Abstract - **This study uses annual time series data on neonatal mortality rate (NMR) for Mauritania from 1960 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 (1) variable. The optimal model based on AIC is the ARIMA (5,1,4) model. The study findings indicate that neonatal mortality will gradually decline from approximately 31 in 2020 to around 27 deaths per 1000 live births by the end of 2030. Therefore, Health authorities are encouraged to formulate country specific neonatal policies that will help to address local causes of neonatal mortality.**

Keywords: ARIMA, Forecasting, NMR.

I. INTRODUCTION

Neonatal mortality continues to be a significant global health problem with 4 million neonates dying annually (Weddih *et al*. 2019). In 2013 approximately 2.8 million neonates died across the globe (UNICEF report 2013-2014; Liu *et al*. 2012). The majority of neonatal deaths were reported in low and middle income countries (Liu, 2015). In the case of Mauritania a country with an estimated population size of 4 million inhabitants, neonatal mortality rate (NMR) has remained stagnant at around 40 per 1000 live births during the period from 1990-2010 (Weddih *et al*. 2019). The main causes of neonatal mortality in this country are prematurity (37%), birth asphyxia (24%) and neonatal sepsis (Liu, 2015). The aim of this paper is to project future trends of neonatal mortality rate for Mauritania. The study employs the popular Box-Jenkins ARIMA model because it is appropriate for analyzing linear data (Nyoni, 2018; Box &Jenkins, 1970).The findings of this paper are expected to guide maternal & Child Health (MNCH) policy formation and implementation with a goal to achieve the set SDG-3 target 3.2 by the end of 2030.

II. LITERATURE REVIEW

Neonatal mortality is a public health issue of concern in Mauritania. Previous studies for the country and region have thoroughly investigated causes and risk factors of mortality in neonates. A retrospective and descriptive study was conducted by Lahmini and Bourrous (2020) to investigate epidemiology of the pediatric mortality in the pediatric emergency department (PED), to determine its rate and identify its most frequent causes. The study was conducted over five years (1st January 2012 and 31st December 2016) including all children aged from 0 to 15 years old who died at the PED in the Mohamed VI Hospital in Marrakech. It was revealed that the most frequent causes of neonatal mortality were neonatal infections and prematurity. Sougou & Diouf(2020) conducted a secondary analysis of the 2017 DHS for Senegal to analyze the factors associated with neonatal deaths in Senegal in 2017. The study results indicated that significant predictors of neonatal mortality were newborns with a low birth weight < 2500 g, newborns who are considered "very small" by their mother at birth and birth by caesarean section. Weddih *et al*.(2019)applied logistic regression to find the significant predictors of neonatal mortality at the Referral Hospital in Nouakchott, Mauritania. The findings of the study indicated that significant predictors of neonatal mortality were low birth weight, hypothermia, and birth outside the NRH. Doukkali *et al*.(2016) examined the characteristics of pregnant women in the hospital center in the urban commune of Missour as well as the morphological data of the newborn and identify the risk factors implicated in neonatal morbidity. The retrospective study of collected data from records of women who gave birth in the Maternity of Missour during 2012. The authors found out that the risk factors associated to neonatal morbidity are gestational age and weight birth (in 50% of the cases, 85.7% of deaths are premature infants less than 1500 g) and 11 (0.9%) congenital malformations.

III. METHODOLOGY

The Autoregressive (AR) Model

A process M_t (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:

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

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

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

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:

$$M_t = \emptyset M_{t-1} + Z_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [3]$$

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:

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

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

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

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

Equation [4] can also be expressed as follows:

$$M_t - \sum_{j \leq 1} \pi_j M_{t-j} = Z_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [6]$$

for some constant π_j such that:

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

This implies that it is possible to invert the function taking the Z_t sequence to the M_t sequence and recover Z_t from present and past values of M_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:

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

Thus:

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

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.

$$\left.\begin{array}{c} The\ first\ difference\ is\ given\ by: \\ M_t - M_{t-1} = M_t - BM_t \\ The\ second\ difference\ is\ given\ by: \\ M_t(1 - B) - M_{t-1}(1 - B) = M_t(1 - B) - BM_t(1 - B) = M_t(1 - B)(1 - B) = M_t(1 - B)^2 \\ The\ third\ difference\ is\ given\ by: \\ M_t(1 - B)^2 - M_{t-1}(1 - B)^2 = M_t(1 - B)^2 - BM_t(1 - B)^2 = M_t(1 - B)^2(1 - B) = M_t(1 - B)^3 \\ The\ d^{th}\ difference\ is\ given\ by: \\ M_t(1 - B)^d \end{array}\right\} \cdots [9]$$

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:

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

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

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

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 public health. In this paper, hinged on this technique; the researcher will use automatic ARIMA modeling for estimating equation [10].

Data Issues

This study is based on annual NMR in Mauritania for the period 1960 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: D(M)
Date: 01/23/22   Time: 17:54
Sample: 1960 2019
Included observations: 59

Model	LogL	AIC*	BIC	HQ
(5,4)(0,0)	42.002177	-1.050921	-0.663584	-0.899720
(2,3)(0,0)	37.213640	-1.024191	-0.777704	-0.927972
(4,2)(0,0)	38.103478	-1.020457	-0.738757	-0.910493
(3,2)(0,0)	36.941793	-1.014976	-0.768489	-0.918757
(3,4)(0,0)	38.854190	-1.012006	-0.695094	-0.888297
(3,0)(0,0)	34.841544	-1.011578	-0.835515	-0.942850
(1,2)(0,0)	34.755448	-1.008659	-0.832597	-0.939932
(4,0)(0,0)	35.710164	-1.007124	-0.795849	-0.924651
(3,1)(0,0)	35.560231	-1.002042	-0.790767	-0.919569
(5,3)(0,0)	39.528121	-1.000953	-0.648828	-0.863498
(5,2)(0,0)	38.374928	-0.995760	-0.678848	-0.872050
(2,2)(0,0)	35.207309	-0.990078	-0.778803	-0.907605
(3,3)(0,0)	37.111480	-0.986830	-0.705130	-0.876866
(2,5)(0,0)	37.941344	-0.981062	-0.664150	-0.857353
(4,3)(0,0)	37.897502	-0.979576	-0.662664	-0.855867
(1,3)(0,0)	34.888847	-0.979283	-0.768008	-0.896810
(5,5)(0,0)	40.879700	-0.978973	-0.556423	-0.814026
(4,1)(0,0)	35.759156	-0.974887	-0.728399	-0.878668
(5,0)(0,0)	35.756915	-0.974811	-0.728323	-0.878592
(1,4)(0,0)	35.419374	-0.963369	-0.716881	-0.867150
(2,4)(0,0)	36.411346	-0.963096	-0.681396	-0.853132
(1,5)(0,0)	35.851401	-0.944115	-0.662415	-0.834151
(5,1)(0,0)	35.761969	-0.941084	-0.659384	-0.831119
(2,1)(0,0)	32.681945	-0.938371	-0.762309	-0.869643
(4,5)(0,0)	38.537644	-0.933479	-0.546142	-0.782279
(2,0)(0,0)	30.842637	-0.909920	-0.769070	-0.854938
(3,5)(0,0)	36.733353	-0.906215	-0.554090	-0.768760
(0,4)(0,0)	31.031744	-0.848534	-0.637259	-0.766061
(1,1)(0,0)	28.722920	-0.838065	-0.697215	-0.783083
(1,0)(0,0)	26.415091	-0.793732	-0.688094	-0.752495
(4,4)(0,0)	32.987989	-0.779254	-0.427129	-0.641799
(0,3)(0,0)	26.168184	-0.717566	-0.541503	-0.648838
(0,2)(0,0)	18.559967	-0.493558	-0.352708	-0.438576
(0,5)(0,0)	10.803388	-0.128928	0.117559	-0.032710
(0,1)(0,0)	-2.358277	0.181637	0.287274	0.222873
(0,0)(0,0)	-27.326106	0.994105	1.064530	1.021596

Criteria Graph

Figure 1: Criteria Graph
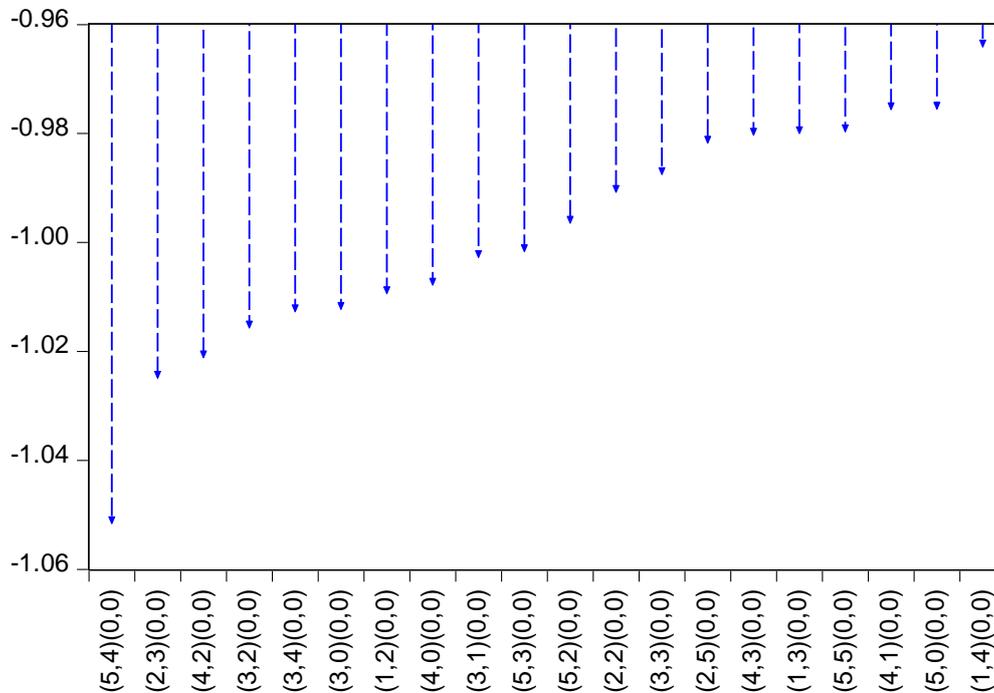
Forecast Comparison Graph
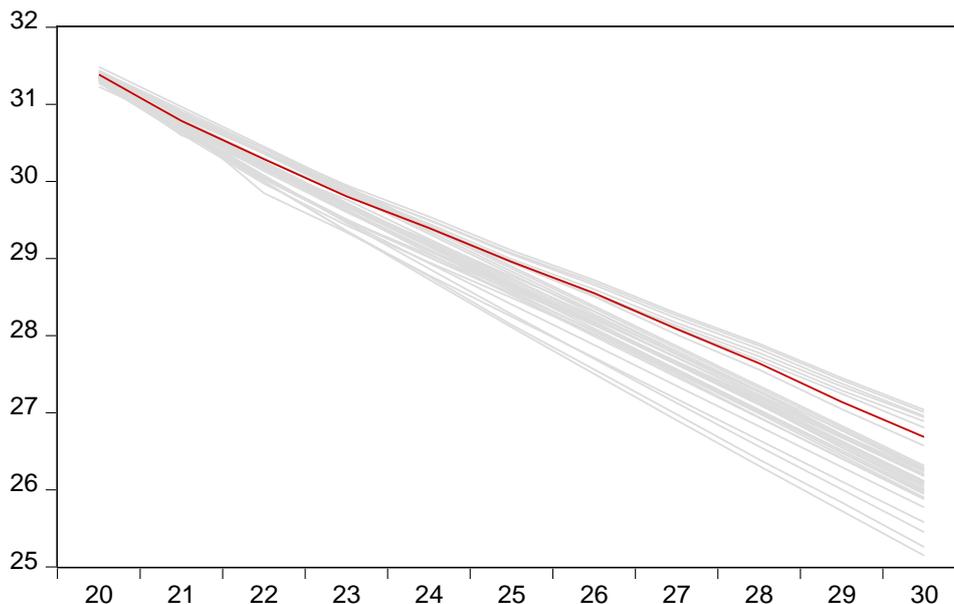
Figure 2: Forecast Comparison Graph

Table 2 and Figure 1 indicate that the optimal model is the ARIMA (5,1,4) 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 (5,1,4) model.

IV. RESULTS

Summary of the Selected ARIMA () Model

Table 3: Summary of the Optimal Model

Automatic ARIMA Forecasting
Selected dependent variable: D(M)
Date: 01/23/22   Time: 17:54
Sample: 1960 2019
Included observations: 59
Forecast length: 11

Number of estimated ARMA models: 36
Number of non-converged estimations: 0
Selected ARMA model: (5,4)(0,0)
AIC value: -1.05092124163

Main Results of the Selected ARIMA () Model

Table 4: Main Results of the Optimal Model

Dependent Variable: D(M)
Method: ARMA Maximum Likelihood (BFGS)
Date: 01/23/22   Time: 17:54
Sample: 1961 2019
Included observations: 59
Convergence achieved after 176 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.556700	0.024310	-22.89967	0.0000
AR(1)	2.506157	0.049719	50.40692	0.0000
AR(2)	-1.430499	0.071477	-20.01339	0.0000
AR(3)	-1.625319	0.042004	-38.69438	0.0000
AR(4)	2.417748	0.057148	42.30647	0.0000
AR(5)	-0.893589	0.026551	-33.65566	0.0000
MA(1)	-1.438181	0.098977	-14.53039	0.0000
MA(2)	-0.252464	0.000908	-277.9813	0.0000
MA(3)	1.437973	0.098985	14.52715	0.0000
MA(4)	-0.747311	0.002781	-268.7445	0.0000
SIGMASQ	0.012103	0.002765	4.376485	0.0001

R-squared	0.918140	Mean dependent var	-0.516949
Adjusted R-squared	0.901086	S.D. dependent var	0.387811
S.E. of regression	0.121969	Akaike info criterion	-1.050921
Sum squared resid	0.714068	Schwarz criterion	-0.663584

Log likelihood	42.00218	Hannan-Quinn criter.	-0.899721	
F-statistic	53.83681	Durbin-Watson stat	1.911397	
Prob(F-statistic)	0.000000			
Inverted AR Roots	.96-.19i	.96+.19i	.79-.55i	.79+.55i
	-1.00			
Inverted MA Roots	1.00	.72-.48i	.72+.48i	-1.00

ARIMA () Model Forecast

Tabulated Out of Sample Forecasts

Table 5: Tabulated Out of Sample Forecasts

2020	31.38447964171337
2021	30.78390102778052
2022	30.28873679911986
2023	29.80379214385389
2024	29.3960508925473
2025	28.95647756714054
2026	28.55159207991675
2027	28.08420650333497
2028	27.63982847336537
2029	27.14018849779012
2030	26.68303501041353

Table 5 clearly indicates that neonatal mortality gradually decline from approximately 31 in 2020 to around 27 deaths per 1000 live births by the end of 2030.

V. POLICY IMPLICATION & CONCLUSION

The challenge of neonatal deaths in Mauritania needs special attention as the country's NMR is unacceptable. These newborns die from preventable causes therefore the government should address maternal, newborn and health system related factors which influence neonatal mortality. This study employs the Box-Jenkins ARIMA approach to predict NMR for Mauritania and the findings indicate that neonatal mortality gradually decline from approximately 31 in 2020 to around 27 deaths per 1000 live births by the end of 2030. Therefore, Health authorities are encouraged to formulate country specific neonatal policies that will help to address local causes of neonatal mortality.

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] Weddih A., Ahmed MLCB., Sidatt M., Abdelghader N., Abdelghader F., Ahmed A.,

[4] Regad S., Makhalla K., Heukelbach J., and BarkatA (2019). Prevalence and factors associated with neonatal mortality among neonates hospitalized at the National Hospital Nouakchott, Mauritania, Pan African Medical Journal, 34, 152.

[5] UNICEF/WHO/The World Bank/UN Pop Div. Levels and Trends in Child Mortality. Report 2013 and 2014.

[6] Liu (2012). Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. Lancet, 379, 9832, 2151-61

[7] Liu (2015). Global, regional, and national causes of child mortality in 2000-13, with projections to inform post-2015 priorities: an updated systematic analysis. Lancet, 385, 9966, 430-40.

****