ARIMA FORECASTING OF THE PREVALENCE OF ANEMIA IN CHILDREN IN SIERRA LEONE

Dr. Smartson. P. NYONI
ZICHIRe Project, University of Zimbabwe, Harare, Zimbabwe

Mr. Thabani NYONI
Department of Economics, University of Zimbabwe, Harare, Zimbabwe

ABSTRACT

Using annual time series data on the prevalence of anemia in children under 5 years of age in Sierra Leone from 1990 – 2016, the study makes predictions for the period 2017 – 2025. The study applies the Box-Jenkins ARIMA methodology. The diagnostic ADF tests show that the AS series under consideration is an I (0) variable. Based on the AIC, the study presents the AR (4) model, also known as the ARIMA (4, 0, 0) model as the optimal model. The diagnostic tests further show us that the presented model is stable and its residuals are not serially correlated. The results of the study indicate that the prevalence of anemia in children in Sierra Leone will rise over the out-of-sample period. By 2025, the country will be having a prevalence of anemia in children of approximately 78.6%.

1.0 INTRODUCTION

The prevalence of anemia is a severe public health problem for children in Sierra Leone (Wirth *et al*., 2017). Anemia, defined as low hemoglobin concentration, is estimated to globally affect 43% of children under five years of age, 29% of non-pregnant women (aged 15-49 years), and 38% of pregnant women (Stevens *et al*., 2013). Due to poor nutritional status (Wirth *et al*., 2017), necessitated by rampant poverty (UNDP, 2015), anemia prevalence is very high among Sierra Leonean children under five years of age (Stevens *et al*., 2013). Motivated by the scale of the problem of anemia in Sierra Leone, this study will forecast the prevalence of anemia in children under the age of 5 over the period 2017 – 2025. The paper is envisioned to strengthen anemia programming in the country in order to address this severe public health problem.

2.0 LITERATURE REVIEW

Wirth *et al*. (2017) analyzed the prevalence of anemia in children in Sierra Leone based on a cross-sectional survey. The study found out that anemia prevalence was very high in the country and that it was largely driven by iron deficiency. In a related West African study carried out in The Gambia, Petry *et al*. (2019) assessed the prevalence of anemia, as well as its risk factors; based on multivariate analysis. The study established that among children aged 6 – 59 months, the prevalence of anemia was 50.4%. No research has attempted modeling and forecasting the prevalence of anemia in children under the age of five in the country. Given the existing dearth of such studies in literature, the contribution of this study is largely hinged on fill-up this information gap. This piece of work will be the first of its kind in Sierra Leone and is expected to go a long way in accelerating the reduction of anemia prevalence in the country.

3.0 METHODODOLOGY

3.1 The Box – Jenkins (1970) Methodology

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, 2018c). This approach will be used to analyze, AS, the series under consideration.

3.2 The Applied Box – Jenkins ARIMA Model Specification

If the sequence $\Delta^d AS_t$ satisfies an ARMA (p, q) process; then the sequence of AS_t also satisfies the ARIMA (p, d, q) process such that:

$$\Delta^d AS_t = \sum_{i=1}^{p} \beta_i \Delta^d L^i AS_t + \sum_{i=1}^{q} \alpha_i L^i \mu_t + \mu_t \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots [1]$$

where Δ is the difference operator, vector $\beta \in \mathbb{R}^p$ and $\alpha \in \mathbb{R}^q$.

3.3 Data Collection

This study is based on annual observations (that is, from 1990 – 2016) on the prevalence of anemia in children under the age of 5 in Sierra Leone [denoted as AS]. Prevalence of anemia in children under 5 years of age in Sierra Leone, means, the percentage of children under the age of 5 whose hemoglobin level is less than 110 grams per liter at sea level. Out-of-sample forecasts will cover the period 2016 – 2025. All the data was gathered from the World Bank online database.

3.4 Diagnostic Tests & Model Evaluation

3.4.1 The ADF Test in Levels

Table 1: with intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
AS	-2.967187	0.0891	-3.737853	@1%	Non-stationary
			-2.991878	@5%	Non-stationary
			-2.635542	@10%	Stationary

Table 1 shows that AS is stationary in levels.

3.4.2 Evaluation of ARIMA models (with a constant)

Table 2: Evaluation of ARIMA Models (with a constant)

Model	AIC	U	ME	RMSE	MAPE
ARIMA (1, 0, 1)	-7.493454	0.70113	-0.01844	0.50274	0.2738
ARIMA (2, 0, 2)	-43.94921	0.31517	0.086507	0.5071	0.18474
ARIMA (1, 0, 0)	8.551662	0.99878	-0.091132	0.52763	0.34121
ARIMA (2, 0, 0)	-35.29225	0.4089	0.071874	0.48965	0.20496
ARIMA (3, 0, 0)	-42.32236	0.32652	0.084662	0.49631	0.18467
ARIMA (3, 0, 2)	-42.42485	0.31157	0.086363	0.50435	0.18391
ARIMA (3, 0, 1)	-43.54583	0.31743	0.08823	0.50213	0.18447
ARIMA (4, 0, 0)	**-44.38625**	0.31175	0.0894	0.5058	0.18588
ARIMA (5, 0, 0)	-43.2626	0.30605	0.088258	0.506	0.1875

A model with a lower AIC value is better than the one with a higher AIC value (Nyoni, 2018b) Similarly, the U statistic can be used to find a better model in the sense that it must lie between 0 and 1, of which the closer it is to 0, the better the forecast method (Nyoni, 2018a). In this research paper, only the AIC is used to select the optimal model. Therefore, the ARIMA (4, 0, 0) model, which is also called the AR (4) process; is finally chosen.

3.5 Residual Tests

3.5.1 Correlogram of the Residuals of the ARIMA (2, 2, 0) Model

Figure 1: Correlogram of the Residuals

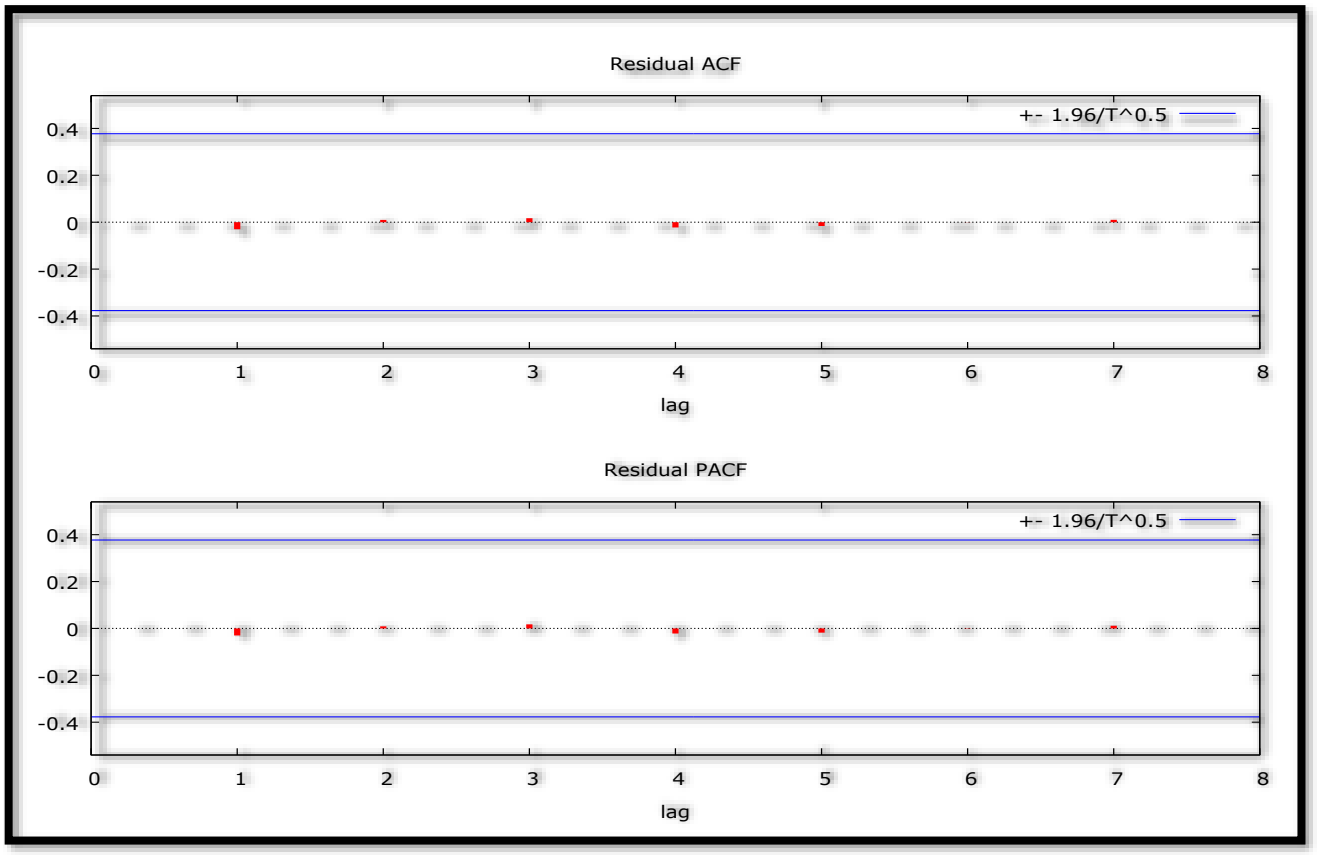

Figure 1 indicates that the estimated parsimonious model is adequate since ACF and PACF lags are quite short and within the bands.

4.0 FINDINGS OF THE STUDY

4.1 Results Presentation[1]

Table 3: Main Results

ARIMA (4, 0, 0) Model:

The chosen optimal model, the ARIMA (4, 0, 0) model can be expressed as follows:

$$AS_t = 79.5975 + 1.18798AS_{t-1} + 0.275264AS_{t-2} - 0.210863AS_{t-3} - 0.281584AS_{t-4} \ldots \ldots [2]$$

Variable	Coefficient	Standard Error	z	p-value
constant	79.5975	0.0715748	1112	0.0000***
β_1	1.18798	0.191724	6.196	0.0000***
β_2	0.275264	0.298841	0.9211	0.357
β_3	-0.210863	0.302095	-0.698	0.4852
β_4	-0.281584	0.192005	-1.467	0.1425

Table 3 shows the main results of the ARIMA (4, 0, 0) model.

Forecast Graph

Figure 2: Forecast Graph – In & Out-of-Sample Forecasts

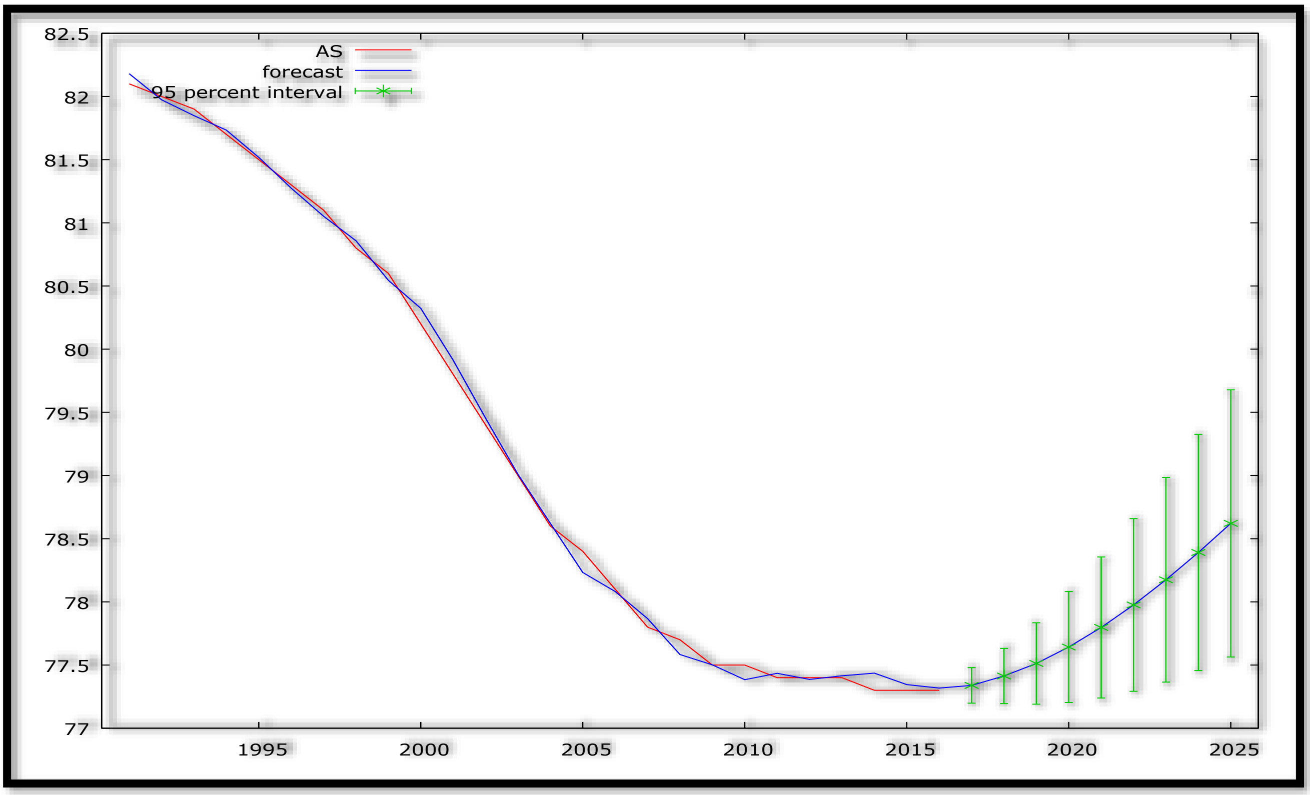

Figure 2 shows the in-and-out-of-sample forecasts of the AS series. The out-of-sample forecasts cover the period 2017 – 2025.

Predicted AS– Out-of-Sample Forecasts Only

Table 4: Predicted AS

Year	Predicted AS	Standard Error	95% Confidence Interval
2017	77.3389	0.0717527	(77.1983, 77.4796)
2018	77.4133	0.111420	(77.1950, 77.6317)
2019	77.5124	0.164497	(77.1900, 77.8348)
2020	77.6424	0.224038	(77.2033, 78.0816)
2021	77.7975	0.284793	(77.2393, 78.3557)

[1] *The *, ** and *** imply statistical significance at 10%, 5% and 1% levels of significance; respectively.*

2022	77.9757	0.348708	(77.2922, 78.6592)
2023	78.1747	0.413001	(77.3653, 78.9842)
2024	78.3909	0.476861	(77.4563, 79.3255)
2025	78.6213	0.539432	(77.5640, 79.6785)

Figure 3: Graphical Analysis of Out-of-Sample Forecasts

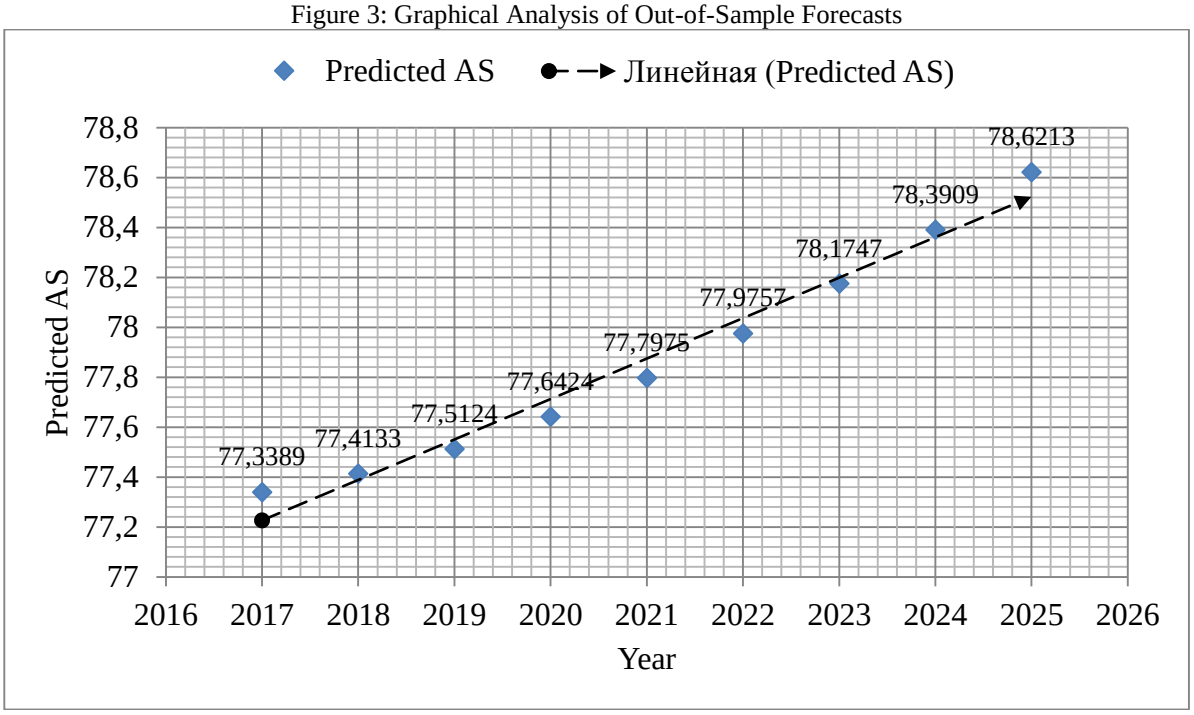

Table 4 and figure 3 show the out-of-sample forecasts only. The prevalence of anemia in children in Sierra Leone is forecasted to rise from the estimated 77.3% in 2017 to approximately 78.6% by 2025.

5.0 CONCLUSION

The study shows that the ARIMA (4, 0, 0) model is not only stable but also the most suitable model to forecast the prevalence of anemia in Sierra Leone over the period 2017 – 2925. The model predicts a sharp increase in the prevalence of anemia in the country, from 73.3% in 2017 to about 78.6% by 2025. This points to the notion that anemia in Sierra Leonean children is far from being eradicated in the country. The study recommends that the government of Sierra Leone should intensify nutritional supplementation and food fortification programmes, especially in rural areas where significant groups of households are economically disadvantaged. Also given that Sierra Leone is one of the poorest countries in the world, and that most anemia cases are linked to iron deficiency, nutritional supplementation programmes should be available to everyone in the country, for free. There is need also for continued early diagnosis and treatment of pediatric HIV and TB. Furthermore, health workers should have refresher courses on Integrated Management of Childhood Illnesses (IMCI).

REFERENCES

[1] Nyoni, T (2018b). Modeling and Forecasting Inflation in Kenya: Recent Insights from ARIMA and GARCH analysis, *Dimorian Review*, 5 (6): 16 – 40.

[2] Nyoni, T. (2018a). Modeling and Forecasting Naira/USD Exchange Rate in Nigeria: A Box-Jenkins ARIMA Approach, MPRA Paper No. 88622, *University Library of Munich*, Munich, Germany.

[3] Nyoni, T. (2018c). Box – Jenkins ARIMA Approach to Predicting net FDI inflows in Zimbabwe, MPRA Paper No. 87737, *University Library of Munich*, Munich, Germany.

[4] Petry, N. *et al*. (2019). Micronutrient Deficiencies, Nutritional Status and the Determinants of Anemia in Children 0 – 59 Months of Age and Non-pregnant Women of Reproductive Age in The Gambia, *Nutrients*, 11 (2275): 1 – 20.

[5] Stevens, G. A. *et al*. (2013). Global, Regional, and National Trends in Hemoglobin Concentration and Prevalence of Total and Severe Anemia in Children and Pregnant and Non-pregnant Women for 1995 – 2011: A Systematic Analysis of Population Representative Data, *Lancet Global Health*, pp: 1 – 9.

[6] UNDP (2015). Human Development Reports, *UNDP*, New York.

[7] Wirth, J. P., *et al*. (2017). Anemia, Micronutrient Deficiencies, and Malaria in Children and Women in Sierra Leone Prior to the Ebola Outbreak – Findings of a Cross-sectional Study, *PLoS ONE*, 11 (5): 1 – 22.