Modeling and Forecasting TB Incidence in Bolivia Using the Multilayer Perceptron Neural Network

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

[1]ZICHIRe Project, University of Zimbabwe, Harare, Zimbabwe
[2]Department of Economics, University of Zimbabwe, Harare, Zimbabwe

Abstract - **In this paper, the ANN approach was applied to analyze TB incidence in Bolivia. The employed annual data covers the period 2000-2018 and the out-of-sample period ranges over the period 2019-2023. The residuals and forecast evaluation criteria (Error, MSE and MAE) of the applied model indicate that the model is stable in forecasting TB incidence in Bolivia. The results of the study indicate that TB incidence in the country will continue on a downward trend although it will remain high around 106 cases/100 000/year over the period 2019-2023. The government is encouraged to intensify TB surveillance and control programs amongst other measures.**

Keywords: ANN, Forecasting, TB incidence.

I. INTRODUCTION

Tuberculosis (TB) is caused by an infectious agent called mycobacterium tuberculosis. The disease has been in existence for decades causing illnesses and deaths among the people worldwide (WHO, 2015; Dheda et al, 2016; Nyoni & Nyoni (2019a & b). Epidemic forecasting models are regarded as an important tool to predict the occurrence of infectious diseases and to formulate reasonable precautions (Wang, 2017). Various methods have been in applied to predict the occurrence of infectious diseases such as Autoregressive integrated moving average (ARIMA), exponential smoothing, the grey model, generalized regression models ,Markov and machine learning methods (Tomaszewski et al ,2015; Zhang et al, 2014; Yang et al, 2014; Li et al, 2012; Li et al, 2015). In this paper we will briefly discuss the two most widely used forecasting models namely the artificial neural networks and the Box-Jenkins ARIMA models. The widely used artificial neural networks (ANNs) are the multilayer perceptrons (MLPs) with a single hidden layer feed forward network (FNN). The models consist of 3 layers, that is the input, hidden and output layers connected by acyclic links called connection weights. The model is biologically inspired and thus mimics the function of the human brain (Wang et al, 2017; Kaushik &Sahi, 2018). ANN models can approximate a nonlinear mapping with any degree of complexity and without prior knowledge of problem solving, however they cannot capture both linear and nonlinear patterns of data equally well (Taskaya-Temizel et al, 2005). Therefore the hybrid model ARIMA-ANN have been shown to have better forecasting performance as compared to ARIMA or ANNS (Jain et al ,2007; Yu et al, 2014). The ARIMA model was proposed by Box and Jenkins in the 1970s.The ARIMA model (p,d,q)(P,D,Q)s is the seasonal ARIMA model. p and q represent the non-seasonal autoregressive (AR) and moving (MA) parts. P and Q are the corresponding seasonal AR and MA parts .d represents the nonseasonal differencing degree and D is the seasonal differencing degree, S is the period of seasonality (Nyoni & Nyoni 2019 a &b). Box and Jenkins proposed a 3 step iterative procedure of ARIMA model building which are model identification, parameter estimation and diagnostic checking. The optimal model is chosen and then used for prediction. The predictions are important in planning and decision making to control and prevent infectious diseases. In this paper we choose to apply the ANN (9,12,1) model with a more efficient hyperbolic tangent function to predict the annual incidence of TB in Bolivia. The results of the study will help in decision making and allocation of resources to the TB /HIV program in the country.

II. LITERATURE REVIEW

There are several studies which have been done to model and forecast TB incidence globally. Ribeiro et al (2019) applied 3 models; ARIMA, simple exponential smoothing (SES) and the Holt-Winters' Exponential smoothing (HWES) to model and forecast TB incidence cases in Brazil using data covering the period January 2001-June 2018.The study concluded that the HWES (0.2,0.1,0.1) was the best model. Wang et al (2020) compared the forecasting performance of the SARIMA and the Error trend seasonal (ETS) framework. The study used monthly TB incidence data from January 1997 to August 2019 in China .The study concluded that the ETS framework has the ability to conduct long term forecasting for TB incidence which may be beneficial for prevention and control of TB. In a related study, Wang et al (2020) compared the TB forecasting performance of the SARIMA and the SARIMA-NNNAR models .The study utilized monthly TB incidence data covering the period January 2004 to December

2016 for Qinghai Province, China. The study indicated that the hybrid model outperformed the SARIMA model. Nyoni & Nyoni (2019a) developed a SARIMA model to predict the monthly TB notifications at Zengeza clinic in Zimbabwe. The study utilized monthly TB notification data covering the period January 2013 to December 2018. The optimal model SARIMA $(2,0,2)$ $(1,0,1)_{12}$ predicted that TB notifications would decline over the out of sample period. In a similar study, Nyoni & Nyoni (2019b) constructed a SARIMA model to project TB notifications at Silobela District Hospital in Zimbabwe. The study utilized the data covering the period January 2014 to December 2018. The optimal model SARIMA $(0,1,1)$ $(1,0,1)_{12}$ projected that TB notifications would decline over the out of sample period.

III. METHOD

The Artificial Neural Network (ANN), which we intend to apply in this paper; is a data processing system consisting of a large number of simple and highly interconnected processing elements resembling a biological neural system. It has the capability of learning from an experimental or real data set to describe the nonlinear and interaction effects with great accuracy. ANN-based curve fitting technique is one of the extensively applied artificial intelligence methods that are used for forecasting and prediction purpose. It consists of basically three layers i.e., input layer, hidden layer, and output layer, the present work includes the number of years as input layer and the annual TB incidence in Bolivia as output data for the network. In this paper, our ANN is based on the hyperbolic tangent function.

Data Issues

This study is based on TB incidences (new cases per 100 000 population/year) [referred to as B series in this study]. The annual data covers the period 2000-2018 while the out-of-sample forecast covers the period 2019-2023. All the data employed in this research paper was gathered from the World Bank online database.

IV. FINDINGS OF THE STUDY

DESCRIPTIVE STATISTICS

Table 1: Descriptive statistics

Mean	Median	Minimum	Maximum
141.84	139.00	108.00	184.00
Std. Dev.	C.V.	Skewness	Ex. Kurtosis
24.040	0.16948	0.25756	-1.1779
5% Perc.	95% Perc.	IQ range	Missing obs.
Undefined	184.00	43.000	0

ANN MODEL SUMMARY FOR TB INCIDENCE (cases per 100 000 population/year) IN BOLIVIA

Table 2: ANN model summary

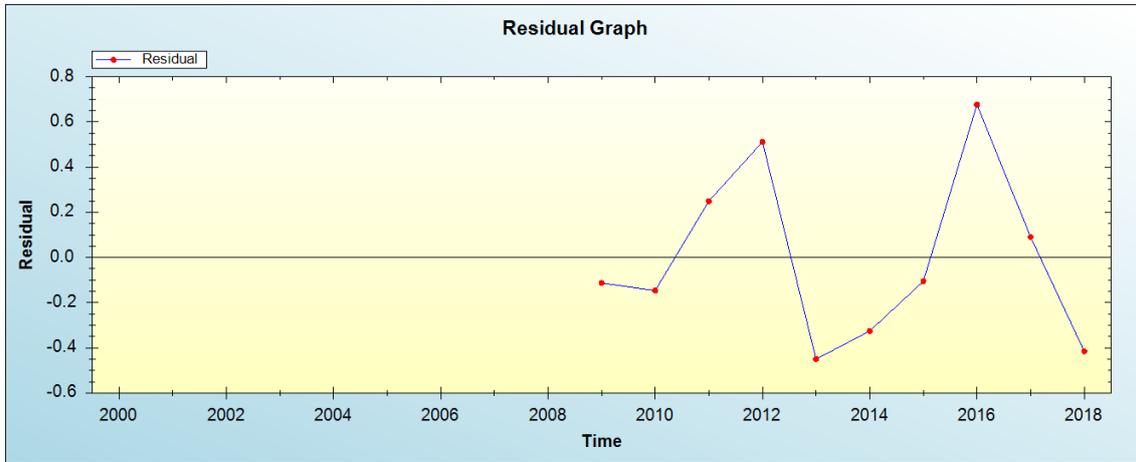| Variable	B
Observations	10 (After Adjusting Endpoints)
Neural Network Architecture:	
Input Layer Neurons	9
Hidden Layer Neurons	12
Output Layer Neurons	1
Activation Function	Hyperbolic Tangent Function
Back Propagation Learning:	
Learning Rate	0.005
Momentum	0.05
Criteria:	
Error	0.008594
MSE	0.131678
MAE	0.308414

Residual Analysis for the ANN model

Figure 1: Residual analysis
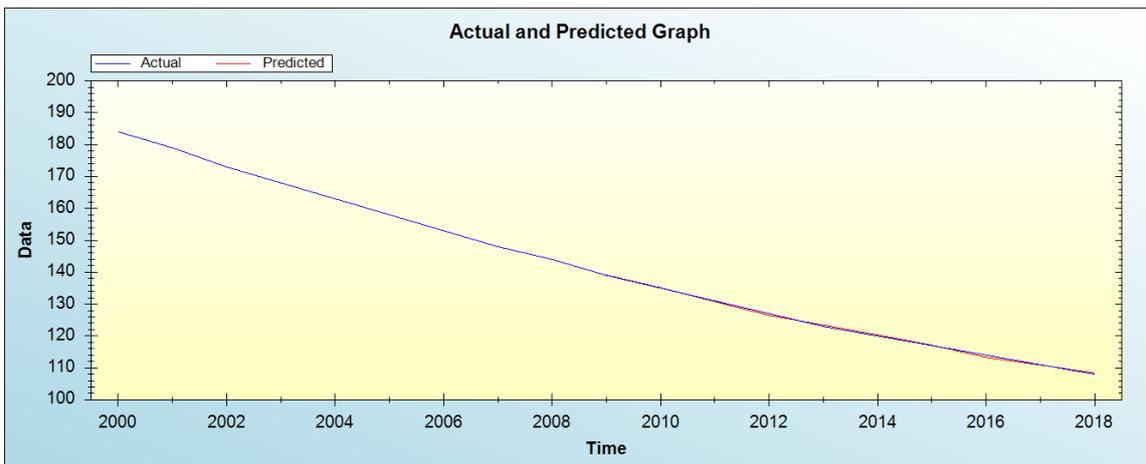
In-sample Forecast for B

Figure 2: In-sample forecast for the B series

Figure 2 shows the in-sample forecast for B series.

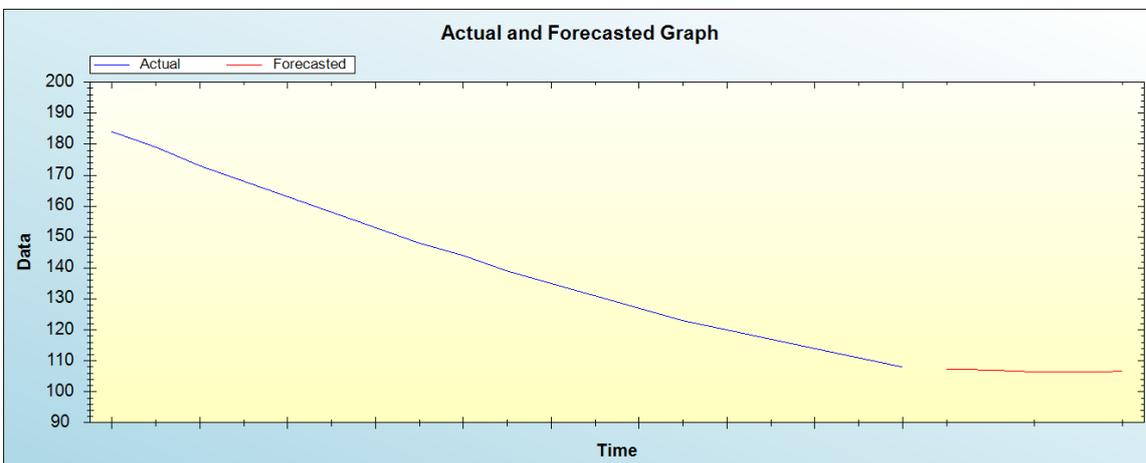*Out-of-Sample Forecast for B: Actual and Forecasted Graph*

Figure 3: Out-of-sample forecast for B: actual and forecasted graph

Out-of-Sample Forecast for B: Forecasts only

Table 3: Tabulated out-of-sample forecasts

Year	Forecasts
2019	107.3687
2020	107.0873
2021	106.4657
2022	106.3604
2023	106.6641

Table 1 shows that over the study period 2000-2018 the minimum and maximum TB incidence was 108 and 184 cases per 100 000 populations /year respectively and the average incidence was 141 cases per 100 000 populations /year. In addition, the TB incidence has been gradually decreasing. The residual graph and evaluation criteria indicate the applied ANN (9,12,1) model is adequate, stable and suitable for forecasting TB incidence in Bolivia. The in-sample forecast demonstrates that the model simulates the observed (actual)values very well. The out of sample forecasts in Figure.4 imply that the projected TB incidence over the period 2019-2023 will continue on a downward trend though high at level of approximately 106 cases per 100 000 population/year.

V. CONCLUSION & RECOMMENDATIONS

Bolivia has shown its commitment in the fight against TB over the period 2000-2018 as the country recorded a downward trend in the incidence of TB. The model predicted a continued decline in incidence although it will remain high around 106 cases per 100 000/year. The government is encouraged to strengthen TB/HIV collaboration, intensify TB surveillance and control programs and allocate more resources to TB/HIV programs in order to significantly reduce TB incidence in the country.

REFERENCES

[1] Box GEP, Jenkins G. Time Series Analysis, Forecasting and Control. San Francisco, CA: Holden-Day, 1970

[2] Dheda, K., Barry, C. E., & Maartens, G. (2016). Tuberculosis, Lancet, 387 (10024): 1211 – 1226.

[3] Jain A, et al. Hybrid neural network models for hydrologic time series forecasting. Applied Soft Computing 2007; 7: 585–592.

[4] K. W. WANG, C. DENG, J. P. LI, Y. Y. ZHANG, X. Y. LIAND M. C. WU*Hybrid methodology for tuberculosis incidence time-series forecasting based on ARIMA and a NAR neural networkEpidemiol. Infect. (2017), 145, 1118–1129.

[5] Kaushik AC &Sahi. S (2018). Artificial neural network-based model for orphan GPCRs.Neural.Comput.Appl. 29,985-992

[6] Li K, et al. Daily temperature change in relation to the risk of childhood bacillary dysentery among different age groups and sexes in a temperate city in China. Public Health 2015.

[7] Li Q, et al. Application of an autoregressive integrated moving average model for predicting the incidence of hemorrhagic fever with renal syndrome. American Journal of Tropical Medicine and Hygiene 2012; 87: 364–370.

[8] Nyoni S. P &Nyoni T (2019). Forecasting TB notifications at Silobela District Hospital, Zimbabwe.IJARIIE 5(6)2395-4396.

[9] Nyoni S. P &Nyoni T (2019). Forecasting TB notifications at Silobela District Hospital, Zimbabwe.IJARIIE 5(6)2395-4396.

[10] Nyoni S. P &Nyoni T (2019). Forecasting TB notifications at Zengeza clinic, Zimbabwe. Online at https://mpra.ub.uni-muenchen.de/97331/ MPRA Paper No. 97331, posted 02 Dec 2019 10:13 UTC

[11] Nyoni S. P &Nyoni T (2019). Forecasting TB notifications at Zengeza clinic, Zimbabwe. Online at https://mpra.ub.uni-muenchen.de/97331/ MPRA Paper No. 97331, posted 02 Dec 2019 10:13 UTC

[12] Tabaszewski M, et al. Using a set of GM (1,1) models to predict values of diagnostic symptoms. Mechanical Systems and Signal Processing 2015; 52: 416–425.

[13] Taskaya-Temizel T, et al. A comparative study of autoregressive neural network hybrids. Neural Networks 2005; 18: 781–789.

[14] World Health Organization. Global TB report 2015. Geneva: World Health Organization, 2015. (http://www.who.int/tb/publications/global report/en/)

[15]   Yang W, et al. Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics. PLoS Computational Biology 2014; 10.

[16]   Yu L, et al. Application of a new hybrid model with seasonal auto-regressive integrated moving average (ARIMA) and nonlinear auto-regressive neural network (NARNN) in forecasting incidence cases of HFMD in Shenzhen, China. PLoS ONE 2014; 9: e98241.

[17]   Zhang X, et al. Applications and comparisons of four time series models in epidemiological surveillance data. PLoS ONE 2014; 9: e88075.

Citation of this Article:

Dr. Smartson. P. NYONI, Thabani NYONI, "Modeling and Forecasting TB Incidence in Bolivia Using the Multilayer Perceptron Neural Network" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 5, Issue 3, pp 301-305, March 2021. Article DOI https://doi.org/10.47001/IRJIET/2021.503051

****