



PREDICTING DIABETES MELITUS CASES AT A PRIVATE HOSPITAL IN ZIMBABWE USING THE BOX-JENKINS “CATCH ALL” MODEL

Dr. Smartson. P. NYONI

ZICHIRE Project, University of Zimbabwe, Harare, Zimbabwe

Mr. Thabani NYONI

Department of Economics, University of Zimbabwe, Harare, Zimbabwe

ABSTRACT

This paper employs monthly time series data on confirmed Diabetes Mellitus (DM) cases at South Medical Hospital (also known as Southmed or Citimed Hospital) from January 2012 to December 2018; to forecast DM cases over the period January 2019 to December 2020. Unit root tests have indicated that the DM series is basically I (1). The paper applied the SARIMA (0,1,1)(01,1,1)₁₂ model. The model has been found to be stable and adequate in forecasting DM cases at Citimed Hospital over the out-of-sample period. The results of the study basically show a seasonal pattern in the trends of predicted DM cases; characterized by repeats in June each year in the out-of-sample period. In order to improve service provision and resource allocation at Citimed Hospital, the study offers a 3-fold policy prescription.

1. INTRODUCTION

Diabetes mellitus (DM) is a serious life-long disease in which levels of glucose in the blood and urine become too high because the body's ability to produce or respond to the hormone insulin released by the pancreas gets impaired and cannot move the glucose into the cells (Singye & Unhapipat, 2018). Diabetes disease is generally categorized into three categories, diabetes type I (insulin-dependent diabetes mellitus) and diabetes type II (non-insulin-dependent diabetes mellitus) as well as Gestational Diabetes Mellitus (GDM). These three different types of diabetes have different effects and can be caused by different factors. However, they all can cause serious health complications and it can increase the risk of blindness, blood pressure, heart disease, kidney disease and nerve damage (Temurtas *et al.* 2009) and need to be treated and managed properly (Singye & Unhapipat, 2018).

Type I diabetes (T1D) is the more severe variant of diabetes, where the pancreas has stopped producing insulin, a vital protein to regulate blood glucose levels. Hence, patients with T1D have to normally inject insulin using either syringes or insulin pumps (Khashei

et al. 2012). Type II diabetes is a condition when the body does not produce enough insulin or there is extremely less insulin to move the glucose to the cells; more over it is also a case when the insulin fails to work properly. Therefore, the sugar level in the blood is risen (Singye & Unhapipat, 2018). GDM is diabetes in pregnant women. The number of patients with diabetes has increased significantly during the last three decades (Meijner & Persson, 2017). In 1980, approximately 108 million adults in the world suffered from diabetes. Today that figure is estimated to be around 422 million (WHO, 2016). Hence, diabetes is now called the “silent killer” (Khashei *et al.* 2012). Diabetes type II is more common than diabetes type I (Khashei *et al.* 2012): in fact, of those with diagnosed diabetes, 90-95% have type 2 diabetes (Hjelm *et al.* 2003) and this is the case with Zimbabwe, where DM continues to pose a significant economic burden (Mutowo *et al.* 2016). This paper will focus on confirmed DM cases recorded at South Medical Hospital in Chitungwiza, Zimbabwe.



Objectives of the Study

- i. To investigate the months during which newly diagnosed DM cases mostly occur at South Medical Hospital.
- ii. To forecast DM cases for the out-of sample period.
- iii. To examine the influence of past DM cases at South Medical Hospital to the present time.

Relevance of the Study

The prevalence of DM is increasing (Magliano *et al.* 2009) and is becoming one of the major health issues of both developed (Tabish, 2007) and the developing world (Meijner & Persson, 2017). In Zimbabwe, DM is now a serious public health problem (Ministry of Health and Child Welfare, 2009 & 2016; Hjelm & Mufunda, 2010; Mufunda *et al.* 2012; Mutowo *et al.* 2014; Mandewo *et al.* 2014; Mutowo *et al.* 2015; Mutowo *et al.* 2016; Mufunda *et al.* 2018; Nyoni *et al.* 2018; Chirombe *et al.* 2018; Mukona *et al.* 2019; Nkomani *et al.* 2019); especially given the fact that there are also high numbers of cases of diabetes which remain undiagnosed (International Diabetes Federation, 2014). Actually, DM is the 5th among the ten most common diseases in Zimbabwe (Mudiayi *et al.* 1997). This could be attributed to urbanization and industrialization, leading to changes in life style from “traditional” and active life to a “modern” sedentary life with unhealthy dietary habits and obesity in combination with increased longevity (Hjelm *et al.* 2003). Hence, DM in Zimbabwe can be thought of as a collision between the modern life style and our ancient genes built for a life as a hunter-gatherer (Zimmet, 2000). In developing countries such as Zimbabwe, diabetes is predicted to overtake other communicable diseases as the major cause of death by 2020 (Murray & Lopez, 1997).

The prevalence of DM in Zimbabwe is approximately 9.7% and the country is ranked 4th amongst African countries with highest prevalence of diabetes (Mukona *et al.* 2016). This is too high a prevalence and warrants the need for modeling and forecasting DM cases for policy formulation. In fact, DM has become a chronic disease that affects between 2% and 4% of the global population and its avoidance and effective treatment are undoubtedly crucial public health and health economics issues in the 21st century (Khashei *et al.* 2012). Hence, the prediction of DM cases is increasingly recognized as a valuable tool to facilitate service provision and resource allocation (Soyiri & Reidpath, 2013). In order to improve service provision and resource allocation at Citimed Hospital,

this study examines the trends of DM cases using the Box-Jenkins “catch-all” model. The paper will go a long way in enhancing the management of DM at South Medical Hospital as well as in other similar health facilities dotted around the country.

2. LITERATURE REVIEW

In Iran, Khashei *et al.* (2012) analyzed diabetes type II using a soft intelligent binary classification model and basically found out that the hybrid model is generally better than linear or nonlinear, soft or hard and classic or intelligent classification models presented for diabetes classification. Villani *et al.* (2017) forecasted Prehospital Emergency Medical Services (EMS) demand for diabetic emergencies using SARIMA models with a data set covering the period 2009 to 2015. Their results indicate that the SARIMA (0,1,0)(0,1,2)₁₂ model provided the best fit, with a MAPE of 4.2% and predicted a monthly caseload of approximately 740 by the end of 2017. Singye & Unhapipat (2018), in Thailand, studied diabetes patients using time series analysis, with a data set covering the period January 2006 to December 2016. Their results showed that the ARIMA (0,1,1) model is the best model to describe and predict future trends of diabetes incidences.

In a Spanish study, Rodriguez-Rodriguez *et al.* (2019) examined the use of big data in predicting short-term blood glucose levels in type 1 diabetes mellitus through machine learning techniques and found out that very accurate short-term prediction can be achieved by only monitoring interstitial glucose data over a very short time period and using a low sampling frequency. In another Spanish study, Rodriguez-Rodriguez *et al.* (2019) investigated the possibility of predicting Glycaemia with constrained IoT devices in type 1 diabetes mellitus patients. Their results basically indicate it is possible to forecast, in a smartphone, a 15min horizon with RMSE of 11.65 mg/dL in just 16.5s, employing 10min sampling of the past 6 h of data and the RF algorithm. This study follows the intuition of Singye & Unhapipat (2018). However, we use the seasonal ARIMA instead of the generalized ARIMA model used by Singye & Unhapipat (2018); on the basis that the former performs better than the latter.

3. METHODOLOGY

Box – Jenkins ARIMA modeling has demonstrated successful prediction of a range of specific health disease events (Medina *et al.* 2007; Wah *et al.* 2014) and has been recognized for its simplicity and easy of administration (Earnest *et al.* 2012). The Box – Jenkins



type models belong to Box & Jenkins (1970) and in this study; it will be adopted for analyzing newly diagnosed monthly DM cases at South Medical Hospital. A generalized Box-Jenkins SARIMA model is as specified in equation [1] below:

$$\phi_p(B)\phi_p(B^s)X_t = \theta_q(B)\theta_q(B^s)\varepsilon_t \dots \dots \dots [1]$$

Where B is the backshift operator, ϕ_p, ϕ_p, θ_q and θ_q are polynomials of order p, P, q and Q respectively. ε_t

is a white noise process and $X_t = \nabla_d \Delta_s^D Y_t$ is the differenced X series.

Data Issues

This study is based on newly diagnosed monthly DM cases [X] (all age groups) at South Medical Hospital, from January 2012 to December 2018. The out-of-sample forecast covers the period January 2019 to December 2020. All the data employed in this paper was gathered from the DHIS2 system for Chitungwiza urban city.

**Diagnostic Tests and Model Evaluation
Stationarity Tests: Graphical Analysis**

Figure 1: Graphical Analysis

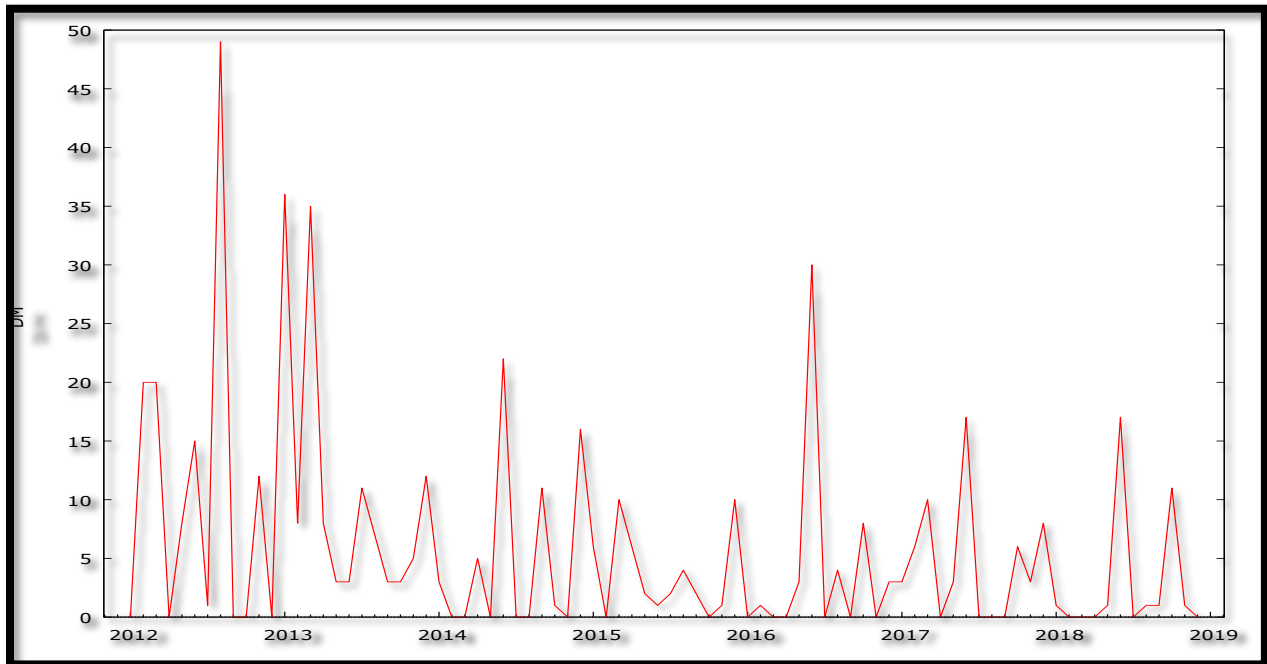


Figure I above shows that the plotted series does not follow any explicit or particular trend and therefore, it is reasonable to suspect non-stationarity. The striking feature of figure 1 is the possible existence of seasonality after every 12 months as suggested by the highest number of DM cases that tend to be experienced

in June of almost every year under study. For the 7 years studied, June 2014, June 2016, 2017 and 2018 have the highest number of DM cases, that is; 22, 30, 17 and 17, respectively. This gives us a clue to the possible need for seasonal differencing, hence the suitability of the Box-Jenkins “catch all” model.



Unit Root Tests

Table 1: Unit root tests

Augmented-Dickey-Fuller test			
Test Statistic			
Variable	Constant	Constant + Trend	None
X _t	-10.06067***	-11.32536***	-1.356611
D(X _t)	-7.260087***	-7.203000***	-7.285994***

*NB: ***, ** and * imply rejection of null hypothesis at 1%, 5% and 10% levels of significance, respectively.*

The null hypothesis of non-stationarity is rejected under all the three circumstances and we conclude that X is essentially an I (1) variable. This implies that we

can proceed to estimate the Box-Jenkins “catch all” model.

Analysis of the Residuals of the SARIMA (0, 1, 1)(0, 1, 1)₁₂ Model

Residual Correlogram of the SARIMA (0, 1, 1)(0, 1, 1)₁₂ Model

Figure 2: Residual Correlogram

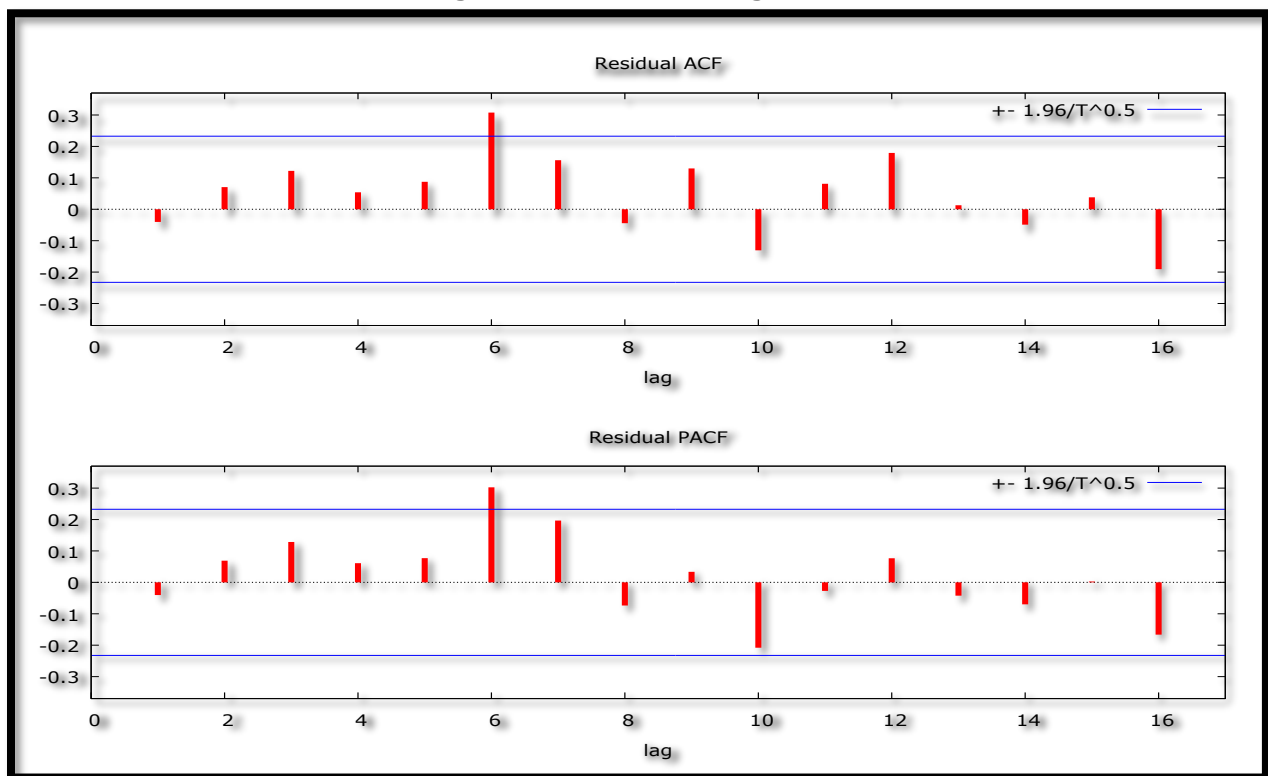


Figure 2 shows that the Box-Jenkins “catch all” model applied in this study is adequate.



4. FINDINGS OF THE STUDY

Descriptive Statistics

Figure 3: Descriptive statistics

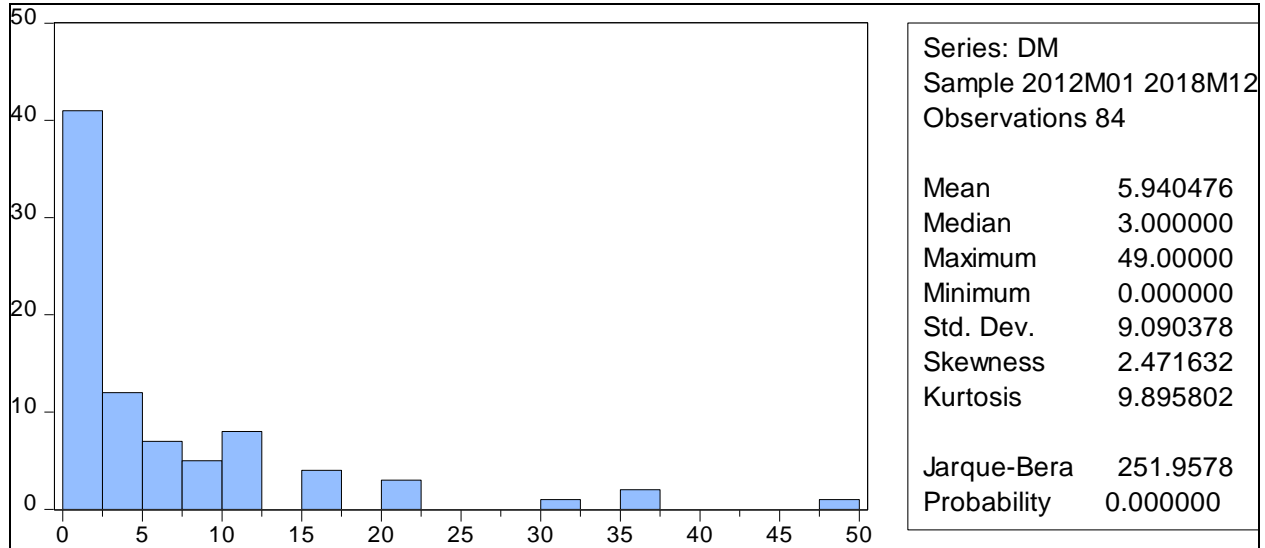


Figure 3 above shows that the maximum number of newly diagnosed DM cases at South Medical Hospital, over the study period, is 49 and this was realized in August 2012. The minimum is zero and this has been experienced in quite a number of months, that is January, April, September, October and December 2012; February, March, May, July, August and November 2014; February and October 2015; January,

March, April, September and November 2016; April, July, August and September 2017 as well as February, March, April and December 2018. The average number of DM cases over the period under study is approximately 6 cases per month. Worthy to note as well is that the series is not normally distributed but rather positively skewed with a skewness statistic of approximately 2.5.

Results Presentation

Table 2: Main Results of the SARIMA (0, 1, 1)(0, 1, 1)₁₂ Model

The SARIMA (0, 1, 1)(0, 1, 1)₁₂ model may be presented as follows:

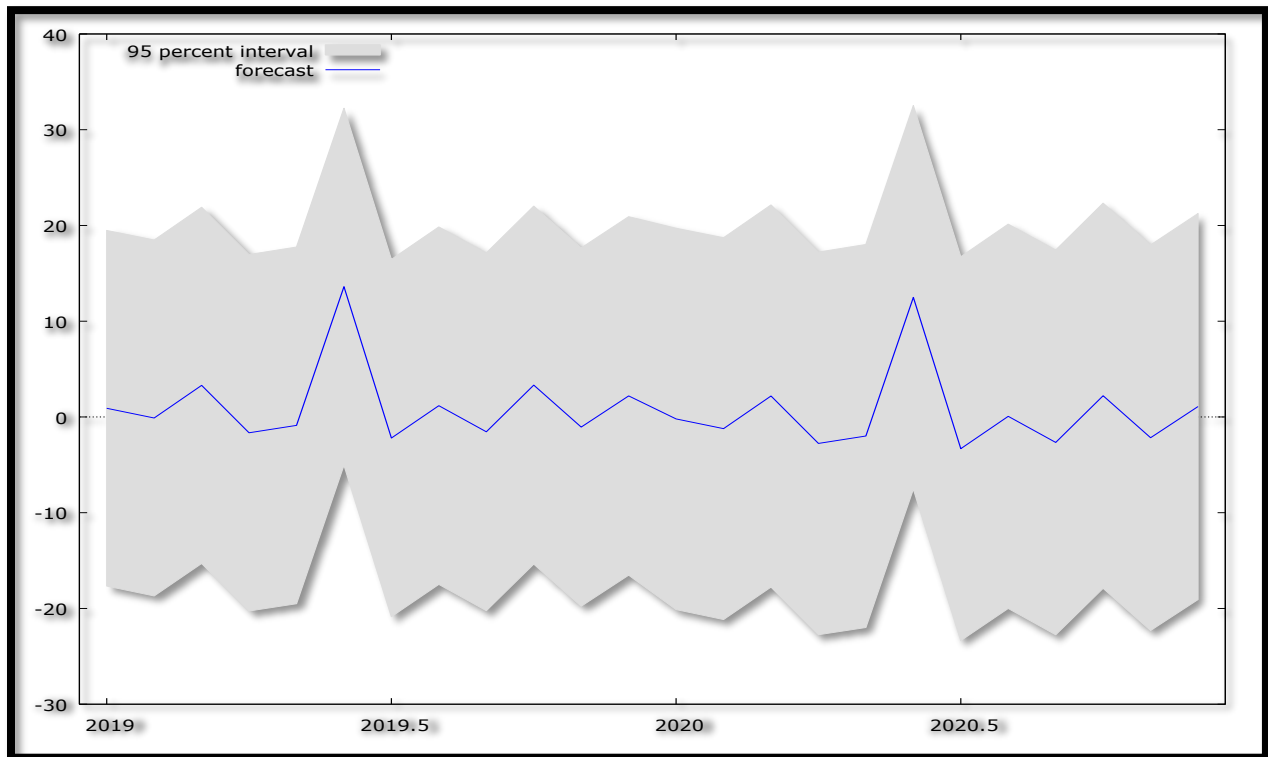
$$(1 - B)(1 - B^{12})X_t = (1 - 0.961883B)(1 - 0.672588B^{12})\varepsilon_t \dots \dots \dots [2]$$

Variable	Coefficient	Standard Error	z	p-value
θ_q	-0.961883	0.0861965	-11.16	0.0000***
θ_q	-0.672588	0.163843	-4.105	0.000404***

NB: ***, ** and * means significant at 1%, 5% and 10% level of significance, respectively.



Forecast Graph
Figure 4: Forecast Graph



Out of Sample Forecasts

Table 3: Out-of-sample forecasts (January 2019 – December 2020)

Year: Month	Prediction	Standard Error	95% Confidence Interval
2019:01	0.908762	9.47673	(-17.6653, 19.4828)
2019:02	-0.106975	9.48361	(-18.6945, 18.4806)
2019:03	3.29998	9.49049	(-15.3010, 21.9010)
2019:04	-1.65079	9.49736	(-20.2653, 16.9637)
2019:05	-0.879043	9.50423	(-19.5070, 17.7489)
2019:06	13.6069	9.51109	(-5.03453, 32.2483)
2019:07	-2.20040	9.51795	(-20.8552, 16.4544)
2019:08	1.18019	9.52480	(-17.4881, 19.8485)
2019:09	-1.55090	9.53165	(-20.2326, 17.1308)
2019:10	3.32792	9.53849	(-15.3672, 22.0230)
2019:11	-1.05611	9.54533	(-19.7646, 17.6524)
2019:12	2.19784	9.55216	(-16.5240, 20.9197)
2020:01	-0.201819	10.1609	(-20.1167, 19.7131)
2020:02	-1.21756	10.1722	(-21.1546, 18.7195)
2020:03	2.18940	10.1835	(-17.7698, 22.1486)
2020:04	-2.76137	10.1947	(-22.7427, 17.2200)
2020:05	-1.98962	10.2060	(-21.9931, 18.0138)



2020:06	12.4963	10.2173	(-7.52921, 32.5218)
2020:07	-3.31098	10.2285	(-23.3585, 16.7365)
2020:08	0.0696119	10.2398	(-19.9999, 20.1392)
2020:09	-2.66148	10.2510	(-22.7530, 17.4301)
2020:10	2.21734	10.2622	(-17.8962, 22.3308)
2020:11	-2.16669	10.2734	(-22.3021, 17.9688)
2020:12	1.08726	10.2846	(-19.0701, 21.2446)

Graphical Presentation of the Predicted Monthly DM Cases at Citimed
Figure 5: Graphical presentation of out-of-sample forecasts

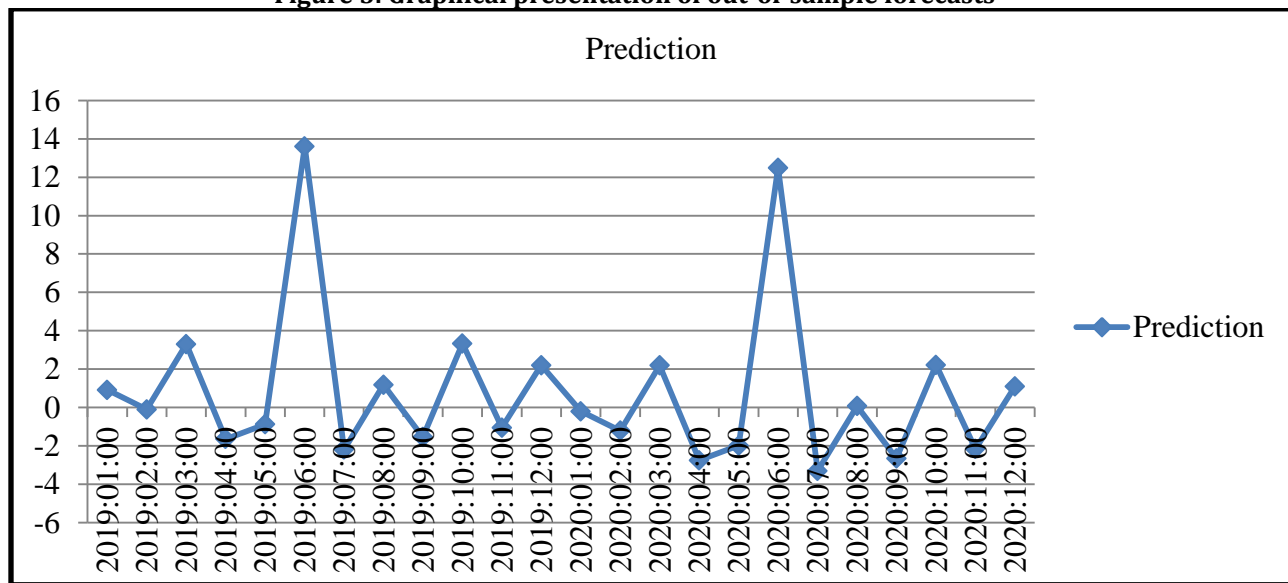


Table 2 shows the main results of the Box-Jenkins “catch-all” model. In that table, equation [2] is the mathematical expression of the model. All parameters of the model are statistically significant at 1% level of significance. The model predicts that most DM cases at South Medical Hospital will be received during the month of June, that is; June 2019 and June 2020. Hence, the seasonal pattern of DM cases at Citimed Hospital repeats in the month of June every year. Figure 4, table 3 and figure 5 display out-of-sample forecasts of the Box-Jenkins “catch-all” model. Generally, the forecasts show that DM cases at South Medical Hospital are anticipated to be less than 4 cases per month, with exception of the month of June, which is projected to be characterized by a higher number of DM cases of approximately 12 cases. South Medical Hospital usually carries out Health Expos in the months of April, May and June each year leading to relatively higher numbers of new DM cases around this period.

5. CONCLUSION & RECOMMENDATIONS

Private health institutions such as South Medical Hospital have a role to play in the prevention and management of DM in Zimbabwe. Hence, studies like these should be used for policy formulation and should steer a scholarly debate in the private health sector. This study was hinged on the Box-Jenkins modelling approach, particularly the SARIMA (0,1,1)(0,1,1)₁₂ model, which is also called the Box-Jenkins “catch-all” model; probably due to its wide applicability in empirical works. The results of this study are specifically meant for Citimed Hospital but may be generalized to other similar health institutions. Further research should look into constructing separate Box-Jenkins models for type 1, type 2 and gestational DM. Separate models for each type of DM could uncover new insights into the management of each specific type of DM. The paper recommends the following:



- i. There is need for the South Medical Hospital to organize adequate finance for the acquisition of the much needed pharmacotherapies for the management of DM.
- ii. Health workers at South Medical Hospital, just like in any other health institution in the country, ought to be properly trained in the detection and treatment of DM. This will help in reducing the number of undiagnosed DM cases.
- iii. There is need to, at least, briefly educate patients on DM to not only improve patients' self-care but also promote good care-seeking behaviour. In this regard, patients should be encouraged to stick to health life styles and diets.

REFERENCES

1. Chirombe, M., Ngara, B., Chibvongodze, R., Charuka, V., & Zhou, D. T. (2018). Glucose Control in Diabetic Patients Attending Parirenyatwa Group of Hospitals in Zimbabwe, *The Open Clinical Biochemistry Journal*, 8: 12 – 19.
2. Earnest, A., Tan, S. B., Smith, A., & Machin, D. (2012). Comparing Statistical Methods to Predict Dengue Fever Notifications, *Computational Mathematical Methods and Medicine*, 6: 1 – 10.
3. Hjelm, K., & Mufunda, E. (2010). Zimbabwean Diabetics' Beliefs About Health and Illness: An Interview Study, *BMC International Health & Human Rights*, 10 (7): 1 – 11.
4. Hjelm, K., Mufunda, E., Nambozi, G., & Kemp, J. (2003). Preparing Nurses to Face the Pandemic of Diabetes Mellitus: A Literature Review, *Journal of Advanced Nursing*, 41: 424 – 434.
5. International Diabetes Federation (2014). *IDF Diabetes Atlas, 6th Edition*, IDF, Brussels.
6. Khashei, M., Eftekhari, S., & Parvizian, J. (2012). Diagnosing Diabetes Type II Using a Soft Intelligent Binary Classification Model, *Review of Bioinformatics and Biometrics*, 1 (1): 9 – 23.
7. Magliano, D. J., Peeters, A., Vos, T., Sicree, R., Shaw, J., Sindall, C., Haby, M., Begg, S. J., & Zimmet, P. Z. (2009). Projecting the Burden of Diabetes in Australia – What is the Size of the Matter? *Australian National Journal of Public Health*, 33 (6): 540 – 543.
8. Mandewo, W., Dodge, E. E., Chideme-Munodawafa, A., & Mandewo, G. (2014). Non-Adherence to Treatment Among Diabetic Patients Attending Outpatients Clinic at Mutare Provincial Hospital, *Manicaland Province, Zimbabwe, International Journal of Scientific & Technology Research*, 3 (9): 66 – 86.
9. Medina, D. C., Findley, S. E., Guindo, B., & Doumbia, S. (2007). Forecasting Non-stationary Diarrhoea, Acute Respiratory Infection and Malaria Time Series in Niono, Mali, *PLoS ONE*, 2 (11): 1 – 8.
10. Meijner, C., & Persson, S. (2017). *Blood Glucose Prediction for Type 1 Diabetes Using Machine Learning – Long Short-term Memory Based Models for Blood Glucose Prediction*, Masters Thesis, Department of Computer Science and Engineering, Chalmers University of Technology, University of Gothenburg, Gothenburg.
11. Ministry of Health and Child Welfare (2009). *Equity and Equality in Health – A People's Right, National Health Strategy 2009 – 2013*, Ministry of Health and Child Welfare, Harare.
12. Ministry of Health and Child Welfare (2016). *Equity and Quality in Health: Leaving No One Behind, National Health Strategy for Zimbabwe 2016 – 2020*, Ministry of Health and Child Welfare, Harare.
13. Mudiayi, T. K., Onyanga-Omara, A., & Gelman, M. L. (1997). Trends of Morbidity in General Medicine at United Bulawayo Hospitals, *Bulawayo, Zimbabwe, Central Africa Journal of Medicine*, 43: 213 – 219.
14. Mufunda, E., Albin, B., & Hjelm, K. (2012). Differences in Health and Illness Beliefs in Zimbabwean Men and Women With Diabetes, *The Open Nursing Journal*, 6: 117 – 125.
15. Mufunda, E., Ernersson, A., & Hjelm, K. (2018). Limited Knowledge of Diabetes in Patients Attending an Outpatient Diabetic Clinic at a Referral Hospital in Zimbabwe: A Cross-Sectional Study, *Pan African Medical Journal*, pp: 1 – 13.
16. Mukona, D. M., Munjanja, S. P., Zvinavashe, M., Dodzo, L. G., & Chituku, S. (2019). The Subjective Impact of Diabetes in Pregnancy: A Quantitative Study, *EC Diabetes and Metabolic Research*, 3 (3): 76 – 82.
17. Mukona, D., Munjanja, S. P., Zvinavashe, M., & Stray-Pederson, B. (2016). Gestational Diabetes Mellitus (GDM): Challenges for Adherence to Anti-Diabetic Therapy and Perinatal Outcomes in Pregnant Women With GDM – A Proposal for Research, *IOSR Journal of Dental and Medical Sciences (IOSR-JDMS)*, 15 (3): 58 – 62.
18. Murray, C. J. L., & Lopez, A. D. (1997). Mortality By Cause for Eight Regions of the World: Global Burden of Disease Study, *The Lancet*, 349 (9061): 1269 – 1276.
19. Mutowo, M. P., Lorgelly, P.K., Laxy, M., Renzaho, A. M. N., Mangwiwo, J. C., & Owen, A. J. (2016). The Hospitalization Costs of Diabetes and Hypertension Complications in Zimbabwe:



- Estimations and Correlations, Journal of Diabetes Research*, pp: 1 – 10.
20. Mutowo, M. P., Mangwiro, J. C., Lorgelly, P. K., & Owen, A. (2015). *Hypertension in Zimbabwe: A Meta-Analysis to Quantify its Burden and Policy Implications*, *World Journal of Meta-Analysis*, 3 (1): 54 – 60.
21. Mutowo, M., Gowda, U., Mangwiro, J. C., Lorgelly, P., Owen, A., & Renzaho, A. (2014). *Prevalence of Diabetes in Zimbabwe: A Systematic Review With Meta-Analysis*, *International Journal of Public Health*, 60 (1): 1 – 11.
22. Nkomani, S., Ruskaniko, S., & Blaauw, R. (2019). *The Impact of Existing Diabetes Self-Management Education Interventions on Knowledge, Attitudes and Practices in Public Health Care Institutions in Harare, Zimbabwe*, *South African Journal of Clinical Nutrition*, pp: 1 – 7.
23. Nyoni, A. M., Chiwaridzo, M., Tadyanemhandu, C., January, J. & Dambi, J. M. (2018). *Profiling the Mental Health of Diabetic Patients: A Cross-Sectional Survey of Zimbabwean Patients*, *BMC Research Notes*, 11 (772): 1 – 7.
24. Rodriguez-Rodriguez, I., Chatzigiannakis, I., Rodriguez, J., Maranghi, M., Gentili, M., & Zamora-Izquierdo, M. (2019). *Utility of Big Data in Predicting Short-term Blood Glucose Levels in Type 1 Diabetes Mellitus Through Machine Learning Techniques*, *Sensors*, 19 (4482): 1 – 14.
25. Rodriguez-Rodriguez, I., Rodriguez, J., Chatzigiannakis, I., & Izquierdo, M. A. Z. (2019). *On the Possibility of Predicting Glycaemia “On the Fly” With Constrained IoT Devices in Type 1 Diabetes Mellitus Patients*, *Sensors*, 19 (4538): 1 – 19.
26. Singye, T., & Unhapipat, S. (2018). *Time Series Analysis of Diabetes Patients: A Case Study of Jigme Dorji Wangchuk National Referred Hospital in Bhutan*, *Journal of Physics – IOP Conference Series*, 1039: 1 – 12.
27. Soyiri, I. N., & Reidpath, D. D. (2013). *An Overview of Health Forecasting*, *Environmental Health and Preventive Medicine*, 18 (1): 1 – 9.
28. Tabish, S. A. (2007). *Is Diabetes Becoming the Biggest Epidemic of the Twenty-First Century?* *International Journal of Health Science*, 1 (2): V – VIII.
29. Temurtas, H., Yumusak, N., & Temurtas, F. (2009). *A Comparative Study on Diabetes Disease Diagnostics Using Neural Networks, Expert Systems With Applications*, 36: 8610 – 8615.
30. Villani, M., Earnest, A., Nanayakkara, N., Smith, K., Courten, B., & Zoungas, S. (2017). *Modeling to Forecast Prehospital EMS Demand for Diabetic Emergencies*, *BMC Health Services Research*, 17 (332): 1 – 9.
31. Wah, W., Da, S., Earnest, A., Lim, L. K. Y., Chee, C. B. E., Cook, A. R., Wang, Y. T., Win, K. M. K., Ong, M. E. H., & Hsu, L. Y. (2014). *Time Series Analysis of Demographic and Temporal Trends of Tuberculosis in Singapore*, *BMC Public Health*, 14 (1): 1 – 10.
32. WHO (2016). *Global Report on Diabetes – 2016*, WHO, Geneva.
33. Zimmet, P. (2000). *Globalization, Coca-Colonization and the Chronic Disease Epidemic: Can The Doomsday Scenario Be Averted?* *Journal of Internal Medicine*, 247: 301 – 310.