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 (TID) 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 TID 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

1. To investigate the months during which newly diagnosed DM cases mostly occur at South Medical Hospital.
2. To forecast DM cases for the out-of sample period.
3. 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; Chirombwe 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 warranties 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 Citimted 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 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 generalised Box-Jenkins SARIMA model is as specified in equation [1] below:

\[ \Phi_p(B) \Phi_p(B^2) X_t = \Theta_q(B) \Theta_q(B^2) \varepsilon_t \]  

Where \( B \) is the backshift operator, \( \Phi_p, \Phi_q, \Theta_q \) and \( \Theta_q \) are polynomials of order \( p, P, q \) and \( Q \) respectively. \( \varepsilon_t \) is a white noise process and \( X_t = \nabla \Delta \Delta^d Y_t \) 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 1 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>
<thead>
<tr>
<th>Variable</th>
<th>Constant</th>
<th>Constant + Trend</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_t$</td>
<td>-10.06067***</td>
<td>-11.32536***</td>
<td>-1.356611</td>
</tr>
<tr>
<td>$D(X_t)$</td>
<td>-7.260087***</td>
<td>-7.203000***</td>
<td>-7.285994***</td>
</tr>
</tbody>
</table>

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:

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>z</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_0$</td>
<td>-0.961883</td>
<td>0.0861965</td>
<td>-11.16</td>
<td>0.0000***</td>
</tr>
<tr>
<td>$\theta_4$</td>
<td>-0.672588</td>
<td>0.163843</td>
<td>-4.105</td>
<td>0.000404***</td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Year: Month</th>
<th>Prediction</th>
<th>Standard Error</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019:01</td>
<td>0.908762</td>
<td>9.47673</td>
<td>(-17.6653, 19.4828)</td>
</tr>
<tr>
<td>2019:02</td>
<td>-0.106975</td>
<td>9.48361</td>
<td>(-18.6945, 18.4806)</td>
</tr>
<tr>
<td>2019:03</td>
<td>3.29998</td>
<td>9.49049</td>
<td>(-15.3010, 21.9010)</td>
</tr>
<tr>
<td>2019:04</td>
<td>-1.65079</td>
<td>9.49736</td>
<td>(-20.2653, 16.9637)</td>
</tr>
<tr>
<td>2019:05</td>
<td>-0.879043</td>
<td>9.50423</td>
<td>(-19.5070, 17.7489)</td>
</tr>
<tr>
<td>2019:06</td>
<td>13.6069</td>
<td>9.51109</td>
<td>(-5.03453, 32.2483)</td>
</tr>
<tr>
<td>2019:08</td>
<td>1.18019</td>
<td>9.52480</td>
<td>(-17.4881, 19.8485)</td>
</tr>
<tr>
<td>2019:09</td>
<td>-1.55090</td>
<td>9.53165</td>
<td>(-20.2326, 17.1308)</td>
</tr>
<tr>
<td>2019:10</td>
<td>3.32792</td>
<td>9.53849</td>
<td>(-15.3672, 22.0230)</td>
</tr>
<tr>
<td>2020:01</td>
<td>-0.201819</td>
<td>10.1609</td>
<td>(-20.1167, 19.7131)</td>
</tr>
<tr>
<td>2020:02</td>
<td>-1.21756</td>
<td>10.1722</td>
<td>(-21.1546, 18.7195)</td>
</tr>
<tr>
<td>2020:03</td>
<td>2.18940</td>
<td>10.1835</td>
<td>(-17.7698, 22.1486)</td>
</tr>
<tr>
<td>2020:05</td>
<td>-1.98962</td>
<td>10.2060</td>
<td>(-21.9931, 18.0138)</td>
</tr>
</tbody>
</table>
Graphical Presentation of the Predicted Monthly DM Cases at Citimed

Figure 5: Graphical presentation of out-of-sample forecasts

<table>
<thead>
<tr>
<th>Year</th>
<th>Prediction</th>
<th>Actual</th>
<th>Prediction Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020:06</td>
<td>12.4963</td>
<td>10.2173</td>
<td>(-7.52921, 32.5218)</td>
</tr>
<tr>
<td>2020:07</td>
<td>-3.31098</td>
<td>10.2285</td>
<td>(-23.3585, 16.7365)</td>
</tr>
<tr>
<td>2020:08</td>
<td>0.0696119</td>
<td>10.2398</td>
<td>(-19.9999, 20.1392)</td>
</tr>
<tr>
<td>2020:09</td>
<td>-2.66148</td>
<td>10.2510</td>
<td>(-22.7530, 17.4301)</td>
</tr>
<tr>
<td>2020:10</td>
<td>2.21734</td>
<td>10.2622</td>
<td>(-17.8962, 22.3308)</td>
</tr>
<tr>
<td>2020:11</td>
<td>-2.16669</td>
<td>10.2734</td>
<td>(-22.3021, 17.9688)</td>
</tr>
<tr>
<td>2020:12</td>
<td>1.08726</td>
<td>10.2846</td>
<td>(-19.0701, 21.2446)</td>
</tr>
</tbody>
</table>

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 \((0,1,1)(01,1,1)_{12}\) 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:
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.

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.

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


