Comparative Performance of the ARIMA, ARIMAX and SES Model for Estimating Reported Cases of Diabetes Mellitus in Anambra State, Nigeria

C. V. Obi and C. N. Okoli

***Abstract* — This study examined the performance of the ARIMA, ARIMAX and the Single Exponential Smoothing (SES) model for the estimation of diabetes cases in Anambra State with the following specific objectives: to fit the model to the data, to determine the best fit model for estimating diabetes mellitus cases and forecast for expected cases for period of five years. The secondary data used for the study is sourced from records of Anambra state Ministry of Health. The Akaike information criterion is adopted for assessing the performance of the models. The R-software is employed for the analysis of data. The results obtained showed that the data satisfied normality and stationarity requirements. The finding of the study showed that ARIMA model has least value of AIC of 1177.92, following the ARIMAX model with value of AIC=1542.25 and SEM recorded highest value of 1595.67. The findings further revealed that the ARIMA has the least values across the measures of accuracy. More so, five years predictions of the cases of diabetes mellitus were obtained using the models under study. From the results of the findings, ARIMA model proved to be best alternative for estimating reported cases of diabetes mellitus in Anambra state. Based on the findings, we recommend there is need for medical practitioners /health planners to create awareness and inform patients about the possible related risk factors of death through early diagnosis and intervention.**

***Index Terms* — Diabetes mellitus, Arima, Arimax, Exponential smoothing, Akaike Information Criterion.**

I. Introduction

Diabetes has been identified as one the important non-communicable diseases (NCDs) that is rapidly attracting the attention of the international medical community, culminating in a United Nations political declaration on NCDs in September 2011 with follow‑up meeting on Political Declaration of the High‑level meeting of the General Assembly on the Prevention and Control of NCDs in May 2013 [1]. Diabetes mellitus (DM) refers to a metabolic disorder of chronic hyperglycemia characterized by disturbances in the metabolism of carbohydrates, proteins, and fats resulting from an absolute or relative insulin deficiency with dysfunction of the organic systems. This disease has shown a considerable increase in prevalence with a demographic transition in its epidemiology in recent years. Populations previously unaffected or minimally affected by DM are now reporting increasing prevalence figures, which poses a real challenge to the financing of health by governments and non-governmental organizations.

The prevalence of diabetes is increasing globally, and sub-Saharan Africa is no exception. Faced with various health problems, health authorities in sub-Saharan Africa and international donors need solid data on the epidemiology and impact of diabetes in order to plan and prioritize their health programs. Due to the high cost of treatment, ethno medical and alternative healing systems constitute primary and complementary health care for most Nigerians as in other African populations. Poor adherence to prescribed interventions also has a negative impact on the outcome of the disease. Based on the literature review so far no study has used the ARIMA, ARIMAX and the exponential smoothing model in estimating reported cases of diabetes mellitus hence the study in Anambra State

II. Review of Related Literature

Oguejiofor et al. [2] aimed at evaluating the impact of diabetes mellitus in Nigeria, challenges arising from the disease to individuals, families, societies, and communities and the way forward in managing this evolving national and global threat. They obtained the required data from the internet using Google chrome search engine and databases of PubMed, Medline, e-Medicine, and Medscape. Also, they used prevalence studies, hospital statistics, registry reports, WHO Reports, IDF declarations, and UN resolutions on diabetes. The result of their review showed that the prevalence of DM is rising to alarming levels. In Africa, Nigeria inclusive, prevalence has risen from < 1 % in the 1950s to the 1980s to current values of ≥ 4.5 %. The highest global prevalence is in the Middle East/North Africa region (11 %), while it is 10.7% and 6.7% in North America and Europe respectively. The prevalence of undiagnosed DM is even higher – 80 % in Africa compared to 35 % in Europe and North America. Populations of African origin have a much higher incidence of microvascular complications compared to macrovascular complications. DM is one of the commonest reasons for admission in Tertiary Hospitals in Nigeria with hyperglycemic emergencies and diabetic foot ulceration (DFU) being the commonest indications for admission. DFU is notoriously responsible for prolonged hospital stay, morbidity, and mortality. Findings from their review suggest the need that most Governments of African countries including Nigeria do not recognize the catastrophic potential of the diabetes epidemic and need to

Published on January 12, 2021.
C. V. Obi, Department of Statistics, Chukwuemeka Odumegwu Ojukwu University, Anambra State, Nigeria.
(e-mail: valley101marc@gmail.com)
C. N. Okoli, Department of Statistics, Chukwuemeka Odumegwu Ojukwu University, Anambra State, Nigeria.

reverse the current trend where DM occupies very low priority in their National Health care agenda before time runs out.

Oputa and Chinenye [3], in their study noted that Nigeria has the largest population in Africa (around 170 million); and of that number, the adult population (aged 20 to 79) is approximately 79 million. A third of all diabetes cases are in rural communities, while the rest are in urban centers. About two million cases of diabetes in Nigeria go undiagnosed. Diabetes-related deaths in Nigeria in 2013 were estimated at 105,091 cases. Nigeria has the highest diabetes burden in Africa, followed by South Africa with 2.6 million cases, Ethiopia 1.9 million, and Tanzania 1.7 million. The global (worldwide) prevalence of diabetes in 2013 is estimated at 382 million cases, including a large population (175 million) undiagnosed, and the global prevalence at around 8.3%. More than half of all deaths from diabetes occur in people under the age of 60. Studies in Nigeria show that up to 73% of diabetic patients are not self-monitoring. Therefore, due to the increased prevalence of hypertension and peripheral neuropathy in Nigeria, which was reported to be over 50%, while the prevalence of retinopathy was 35%, cataracts 25%, cardiovascular disease 5%, foot ulcers 16%, and kidney disease 3%.

Dahiru et al. [4] identified diabetes as one of the major causes of morbidity and mortality both in developing and developed countries. The objective of their study was to determine the prevalence and sex differential of diabetes. The findings of the study showed that the prevalence of diabetes ranged from 0.8% to 11% involving both urban and rural populations, with varying sampling schemes. It was found that a generally low prevalence of diabetes exists in Nigeria. The study suggests the need to undertake a nationally‑representative survey to assess the burden of diabetes in general population.

Sabir et al. [5] examined the prevalence of DM and its correlation in the suburban population of Northwest Nigeria. The study considered two hundred and eighty participants who were recruited using a multistage sampling technique. Data on anthropometric variables, fasting plasma glucose (FPG), and blood pressure measured using standard guidelines were obtained for the analysis. The result of the study found a mean age of 42.3±10.7 years. The overall prevalence of DM was 4.3% (males 4.5% and females 4.0%). The mean FPG was found to be higher in the females than males though the difference was found to be statistically insignificant. Obesity and increasing age were identified as the major risk factors for DM among the suburban population. Findings of the study indicate that DM is common in suburban areas of Northwest Nigeria and increased awareness of the epidemic potential of this public health problem even in suburban areas is suggested.

Adeloye et al. [6] aimed at estimating the prevalence of diabetes in Nigeria hospitalization and associated mortality. The study searched MEDLINE, EMBASE, Global Health, Africa Journals Online (AJOL), and Google Scholar for population and hospital-based studies on type 2 diabetes mellitus (T2DM) in Nigeria. The study equally conducted a random-effects meta-analysis on extracted crude estimates, and applied a meta-regression epidemiological model, using the United Nations demographics for Nigeria in 1990 and 2015 to determine estimates of diabetes in Nigeria for the two years. The result of the study showed that the age-adjusted prevalence rates of T2DM in Nigeria among persons aged 20–79 years increased from 2.0% in 1990 to 5.7% in 2015, accounting for over 874 000 and 4.7 million cases, respectively. The pooled prevalence rate of impaired glucose tolerance was 10.0%, while impaired fasting glucose was 5.8%. The hospital admission rate for T2DM was 222.6 per 100 000 population with hyperglycaemic emergencies, diabetic foot, and cardiovascular diseases being the most common complications. The overall mortality rate was 30.2 per 100 000 population, with a case fatality rate of 22.0%. The findings of the study suggest an increasing burden of T2DM in Nigeria with many persons currently undiagnosed and few known cases on treatment.

Aynalem and Zeleke [7] examined the prevalence of diabetes mellitus and its risk factors among individuals aged 15 years and above in Ethiopia. The Study employed the World Health Organization (WHO) stepwise approach for non-communicable disease surveillance to collect the required data. Glucose meter was used to check the fasting venous blood glucose level. The statistical tools used for the study were the descriptive and logistic regression analyses. The result of the study revealed that the prevalence of DM was 6.5%. Of which, the proportion of previously undiagnosed diabetes mellitus was 88.5%. The prevalence of pre-diabetes was also found to be 15.9%. The waist circumference (WC), body mass index, smoking habit, hypertension, and total cholesterol level were found to significantly associated with diabetes mellitus. The findings of the study indicate the presence of a higher prevalence of diabetes mellitus than the IDFA-projected estimate of DM for Ethiopia.

Uloko et al. [8] examined the prevalence of and risk factors for DM in Nigeria by performing a systematic review and meta-analysis. The findings of the study showed that the prevalence of DM was 5.77%. The prevalence of DM in the six geopolitical zones of Nigeria was 3.0% in the north-west, 5.9% in the northeast, 3.8% in the north-central zone, 5.5% in the south-west, 4.6% in the south-east, and 9.8% in the south-south zone. The study identified the risk factors for the prevalence of DM to include a family history of DM, urban-dwelling, unhealthy dietary habits, cigarette smoking, older age, physical inactivity, and obesity. The findings of the study indicate that there has been an increase in the prevalence of DM in Nigeria and all regions of the country have been affected, with the highest prevalence seen in the south-south geopolitical zone.

Onyenekwe et al. [9] evaluated the characteristics and practices of patients attending the outpatient diabetes clinic. The study also examined to what extent they achieved management goals and what practices by the patients impacted negatively on treatment outcome. The outcome of the study revealed that there were 193 subjects, 78 males and 115 females aged 35-82 years. Type 2 Diabetes cases were diagnosed in 93.4% of the patients while Hypertension was coexistent in 74%. Present or past foot ulcer was recorded in 11.9% while about 37% of the subjects had an exercise program. Also, 35% did the daily foot exam, and 45% had ophthalmology consult. Skipping medication was widespread (64%), mainly due to the self-titration of medications (44%) and cost (23%). Fear of hypoglycemia (83%) and hypotension (79%) was prevalent. Their body mass index was 18.2-41.2 kg/m^2. Subjects were prescribed a total of 2-14 medications. Metformin was the most commonly used glucose-lowering medication (88.6%), followed by sulfonylureas (64%) and insulin (27.5%). The findings of the study indicated that patients lacked the competence to manage their diabetes from day to day. Hence, the need for Diabetes Self- Management Education and Support (DSMES), and early use of insulin. Final Stage

When you submit your final version, after your paper has been accepted, prepare it in two-column format, including figures and tables.

III. Material and Methods

A. Method of Data Collection

The data for this study was sourced through the records of Anambra State Ministry of Health. The data comprises of reported cases of Diabetes mellitus and their corresponding glucose level of the patients from 2008 to 2019.

B. Methods of Data Analysis

This section presents the various tools to be used in analyzing the data.

C. Anderson Darling test for Normality

The Anderson-Darling (AD) test is a statistical test which is used to test whether a sample of data came from a population with specific distribution. It is a modification of the Kolmogorov-Smirnov (KS) test, but it gives more weight to the tails of the distribution than the KS test.

Anderson-Darling test makes used of the specific distribution in calculating the critical values. This has the advantage of allowing a more sensitive test and the disadvantage that the critical values must be calculated for each distribution.

The Anderson-Darling test for normality is defines as:

H_o: The data follows the normal distribution

H_a: The data do not follow the normal distribution

The test statistic is defined as:

$$A^2 = - N - S \tag{1}$$

where,

$$S = \sum_{i=1}^{N} \frac{((2i-1))}{N} [\ln F(Y_i) + \ln(1 - F(Y_{N+1-i}))] \tag{2}$$

F is the cumulative distribution function of the normal distribution while Y_i are the ordered data set of interest.

D. Unit Root Test

To fully explore the data generating process, we first examined the time series properties of model variables using the Augmented Dickey- Fuller test.

The ADF test regression equations with constant are:

$$\Delta Y_T = \alpha_0 + \alpha_1 Y_{T-1} + \sum_{j=1}^{k} a_j \, \Delta Y_{T-1} + \varepsilon_T \tag{3}$$

where Δ is the first difference operator ε_T is random error term that is independently identical dependent k = no of lagged differences Y = the variable. The unit root test is then carried out under the null hypothesis $\alpha = 0$ against the alternative hypothesis of $\alpha < 0$. Once a value for the test statistics $ADF_\tau = \frac{\hat{\alpha}}{SE(\alpha)}$ is computed, we shall compare it with the relevant critical value for the Dickey-Fuller Test. If the test statistic is greater (in absolute value) than the critical value at 5% or 1% level of significance, then the null hypothesis of $\alpha = 0$ is rejected and no unit root is present.

E. Auto-Regressive Integrated Moving Averages (ARIMA) Model

ARIMA models are also known as Box-Jenkins models which require historical chronological data of the underlying variables. The time series approach involves three stages, namely the process of identifying the model, estimating the parameters, and verifying the model. At the model identification stage, the data series is determined if the series is stationary before the development of the Box-Jenkins model or the ARIMA model. A stationary series in the Box-Jenkins model will have a constant mean, a constant variance, and a constant autocorrelation. For a non-stationary series, a differentiation on the non-stationary series one or more times will be made to achieve stationary.

The order (the p and q) of the autoregressive and moving average terms can be identified when the series has been test for stationarity. The tools employed for determining the order of the model are the autocorrelation function (ACF) and the partial autocorrelation function (PACF).

The procedure used to identify the terms of the autoregressive or moving average models is given below:

a. An ACF with steadily decline and a PACF which cut off suddenly after p lags indicates an autoregressive process at lag p, AR(p).

b. An ACF that cuts off suddenly after q lags and a PACF with steadily decline indicates a moving average process at lag q, MA(q).

Also, when the ACF and the PACF both exhibiting large spikes that gradually die out indicates that there are both autoregressive and moving averages processes.

In the present study, the Auto ARIMA approach was used to determine the appropriate ARIMA model for the study. The auto ARIMA function uses the generated AIC and BIC to determine the best combination of parameters for the model. The AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) values are estimators for comparing the models. The lower these values, the better the model.

Suppose we consider the following ARIMA model written as

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + \varepsilon + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \cdots + \beta_q \varepsilon_{t-q} \tag{4}$$

Equation (1) can also be expressed as:

$$\alpha(L) Y_t = \beta(L) \varepsilon_t \tag{5}$$

where,

L is the Log operator, $LY_t = Y_{t-1}$ and $L^2 Y_t = Y_{t-2}$

$$\alpha(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_p L^p \tag{6}$$

$$\beta(L) = 1 - \beta_1 L - \beta_2 L^2 - \cdots - \beta_q L^q \tag{7}$$

The degree of homogeneity d is determined in the identification process while Y_t is the response variable representing the stationary series. However, the model parameters are estimated using the least squares method. The response variable for the present study Y_t is the series of reported cases of diabetes mellitus from 2008-2019 in Anambra State.

F. Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) Model

An ARIMAX model is a kind of multiple regression model with one or more autoregressive (AR) terms and/or one or more moving average (MA) terms. In such situation, the autoregressive terms for a dependent variable are merely lagged values of that dependent variable that have a statistically significant relationship with its most recent value. Moving average terms is nothing more than residuals (i.e., lagged errors) resulting from previously made estimates. The general ARIMAX models are as follows:

Autoregressive model with exogenous variables (ARX):

$$Y_t = \phi(L)Y_t + \beta x_t + \varepsilon_t \tag{8}$$

Moving average model with exogenous variables (MAX):

$$Y_t = \beta x_t + \phi(L)\varepsilon_t \tag{9}$$

Autoregressive Moving Average with exogenous variables model (ARMAX):

$$\phi(L)Y_t = \beta x_t + \theta(L)\varepsilon_t \tag{10}$$

where, x_t is the exogenous variables, β is the coefficients of the exogenous variables, $\phi(L)Y_t$ is an AR model ($\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p}$), $\theta(L)\varepsilon_t$ is the MA model ($\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$).

The approach used in ARIMAX model involves two phases where the first phase deals with a linear regression model and the second phase deals with integration of AR and MA terms into a multiple regression model. In the first phase, the linear regression is used to identify the exogenous variables that are significant. Meanwhile, the second phase deals with iterative searching process to search the order of ARIMA part of the model. When there is large number of exogenous variables to be screened, the stepwise regression process is used in selecting and introducing a new exogenous variable into the ARIMAX model. This is followed by the iteration process of finding newly AR or MA or both terms to re-establish the random pattern of residuals in the ARIMAX models. The response variable for the present study Y_t is the series of reported cases of diabetes mellitus while the exogenous variable is the corresponding glucose level of the patients from 2008-2019 in Anambra State.

G. Single Exponential Smoothing (SES)

Exponential Smoothing is a technique of time series in which data over time is smoothed exponentially either by assigning exponentially increasing or decreasing weights with the data points. The Single Exponential Smoothing is a weighted average technique of time series analysis that uses historical values with weights associated with them and predicts future value [10]. It is used when the pattern shown by historical data is approximately horizontal which means the data is fluctuating around a constant mean. It assigns exponentially decreasing weights to the historical values as the observations get older. To obtain an exponentially smoothed time series the value of weight associated with historical values ranging between 0 and 1, called exponential smoothing constant and is calculated as follows:

$$F_t = \alpha Y_t + (1\text{-}\alpha)F_{t-1} \quad (\text{for } t \geq 2) \tag{11}$$

where, F_t is the exponentially smoothed forecasted value, Y_t is the response variable of the diabetes mellitus series at time t, F_{t-1} is the exponentially smoothed forecasted value at t-1, while α is the exponentially smoothed constant ranging $0 \leq \alpha \leq 1$.

H. Tools for assessing the performance of the Models

To evaluate the performance of the models in this study, the Akaike information criterion will be used for the selection of the best model.

The Akaike information criterion (AIC) is an estimator of the out-of-sample prediction error and therefore help in providing the relative quality of the statistical models for a given set of data. AIC estimates the relative amount of information lost by a given model. It should be noted that the less information a model loses, the higher the quality of that model. In estimating the amount of information lost by a model, AIC deals with both the risk of over-fitting and the risk of under-fitting in the model.

In this study, the AIC criterion used to compare the parametric models, defined as:

$$\text{AIC} = -2\text{LL} + 2\text{p} \tag{12}$$

where,

LL is the log-likelihood; p is the number of parameters in the model. Smaller value of AIC suggests a better model for data adaptation [11].

IV. Data Analysis and Results

This section presents the data analysis and results.

TABLE 1: DESCRIPTIVE ANALYSIS OF THE VARIABLES

	Number of Reported Diabetes cases (NRDC)	Average Glucose Level (AGL)
Mean	471.61	115.13
Median	467.00	113.79
Maximum	797.00	169.22
Minimum	275.00	60.72
Std. Dev.	105.51	21.60
Skewness	0.55	0.01
Kurtosis	0.17	-0.40
Observations	144	144

Source: Authors Analysis.

The result obtained in table 1 showed that the of number of reported diabetes cases has mean value of 471.61, median value of 467, standard deviation of 275, Skewness of 0.55 and Kurtosis value of 0.17. Similarly, average glucose level was found to have mean of 115.13, median value of 113.79, standard deviation of 60.72, Skewness of 0.01 and Kurtosis value of -0.40.

TABLE 2: SUMMARY OF TEST FOR NORMALITY OF THE VARIABLES

Variables	Anderson-Darling Test Statistic	P-value	Decision
NRDC	0.440	0.288	Normally distributed
AGL	0.189	0.900	Normally distributed

Source: Authors Analysis.

The test of normality of the variables was conducted using the Anderson-Darling test for normality. The null hypothesis that the variable is normally distributed was evaluated at 5% significance level. The null hypothesis was accepted in each if and only if the p-value of the test statistic is greater than the significant level, otherwise reject the null hypotheses. Table 2 indicated that all the variables were normally distributed.

TABLE 3: RESULT OF AUGMENTED DICKEY-FULLER UNIT ROOT TEST FOR THE VARIABLES

Variables	Level		1st Difference		Order of
	No Trend	With	No Trend	With	
NRDC	-0.266326	-2.306299	-2.355788	-5.595482	I(1)
AGL	-12.54185	-12.63128	-	-	I(0)
		Critical values			
1%	-3.480818	-4.029595	-3.480818	-4.029595	
5%	-2.883579	-3.444487	-2.883579	-3.444487	

Source: Authors Analysis.

The result of the unit root test on the variables using the Augmented Dickey-Fuller test statistic obtained in Table 3 found that the NRDC series has no unit root and stationary over time at first difference since it's test statistic value has a more negative value (-5.5955) than the critical values (-3.4444) assuming a 95% confidence level. While the AGL series has no unit root and stationary over time at zero difference since it's test statistic value has a more negative value (-12.6313) than the critical values (-3.4444) assuming a 95% confidence level. This result implies that the NRDC series is integrated of order 1 I(1) while the AGL series is integrated of order zero I(0). Hence, both series are stationary overtime and can be used to make forecast for future behaviour of the process.

TABLE 4: MODEL COEFFICIENTS FOR THE ARIMA (1, 1, 0) MODEL

	ar1	Sigma^2 estimate	Log likelihood	AIC
Coefficients	0.0306	382.7	-583.96	1177.92
s.e.	0.0023	-	-	-

The Table 4 showed that the ARIMA (1, 1, 0) model has a log likelihood value of -583.96 and AIC value = 1177.92.

TABLE 5: MEASURES OF ACCURACY FOR ARIMA (1, 1, 0) MODEL

RMSE	MAE	MAPE
18.3723	9.80049	2.14573

Result obtained in Table 5 showed that the ARIMA (1, 1, 0) model has a root mean square error (RMSE) of 18.3823, mean absolute error (MAE) of 9.80049 and mean absolute percentage error (MAPE) of 2.14573.

TABLE 6: MODEL COEFFICIENTS FOR THE ARIMAX (1, 1, 0) MODEL

	ar1	AGL	Sigma^2 estimate	Log likelihood	AIC
Coefficients	- 0.0398	0.0744	2711	-768.12	1542.25
s.e.	0.0843	0.1430	-	-	-

The result obtained in Table 6 showed that the ARIMAX (1, 1, 0) model has a log-likelihood value =-768.12 and AIC value of 1542.25.

TABLE 7: MEASURES OF ACCURACY FOR ARIMAX (1, 1, 0) MODEL

RMSE	MAE	MAPE
51.88718	28.78721	7.101198

Result obtained in Table 7 showed that the ARIMAX (1, 1, 0) model has a root mean square error (RMSE) of 51.88718, mean absolute error (MAE) of 28.78721 and mean absolute percentage error (MAPE) of 7.101198.

TABLE 8: SUMMARY RESULT OF SMOOTHING PARAMETERS

Alpha	Beta	Gamma	Sigma	AIC
0.9999	0.0001	0.0001	20.0146	1595.673

The result obtained in Table 8 showed that the exponential smoothing model has a gamma value of 0.0001 and AIC value of 1595.673.

TABLE 9: MEASURES OF ACCURACY FOR EXPONENTIAL SMOOTHING MODEL

RMSE	MAE	MAPE
18.86993	9.804107	2.200063

Result obtained in Table 9 showed that the exponential smoothing model has a root mean square error (RMSE) of 18.86993, mean absolute error (MAE) of 9.804107and mean absolute percentage error (MAPE) of 2.200063.

TABLE 10: SUMMARY RESULT OF THE ACCURACY MEASURES FOR THE ARIMA, ARIMAX AND THE EXPONENTIAL SMOOTHING MODEL

Measures	ARIMA Values	ARIMA Performance	ARIMAX Values	ARIMAX Performance	Exponential Smoothing Values	Exponential Smoothing Performance
AIC	1177.92	1	1542.25	2	1595.673	3
RMSE	18.37	1	51.8871	3	18.86993	2
MAE	9.800	1	28.78721	3	9.804107	2
MAPE	2.145	1	7.10119	3	2.200063	2

The result obtained in Table 10 showed that the ARIMA model performed best for estimating the number of reported diabetes cases in Anambra state since it has the least values across the measures of accuracy.

TABLE 11: FIVE YEARS FORECAST OF NUMBER OF REPORTED DIABETES CASES IN ANAMBRA USING THE ARIMA, ARIMAX AND THE EXPONENTIAL SMOOTHING MODEL

Year Month	ARIMA	ARIMAX	Exponential Smoothing
2020 JAN	563.946	807.52	628.691
FEB	579.679	809.11	652.351
MAR	597.676	810.447	670.213
APR	610.805	812.232	687.825
MAY	627.809	813.942	708.622
JUN	645.486	816.767	728.854
JUL	659.333	818.031	747.501
AUG	677.909	818.775	766.131
SEP	691.608	821.08	784.039
OCT	711.498	822.121	800.015
NOV	725.789	822.418	813.769
DEC	736.6	822.79	825.918
2021 JAN	596.351	809.481	657.61
FEB	611.316	810.745	681.269
MAR	629.405	812.827	699.131
APR	645.591	813.198	716.744
MAY	666.283	814.611	737.54
JUN	692.188	815.949	757.772
JUL	704.187	818.477	776.419
AUG	720.595	819.964	795.049
SEP	738.63	820.931	812.957
OCT	749.967	821.749	828.934
NOV	763.065	821.972	842.687
DEC	769.799	822.567	854.837
2022 JAN	563.946	807.52	628.691
FEB	579.679	809.11	652.351
MAR	597.676	810.447	670.213
APR	610.805	812.232	687.825
MAY	627.809	813.942	708.622
JUN	645.486	816.767	728.854
JUL	659.333	818.031	747.501
AUG	677.909	818.775	766.131
SEP	691.608	821.08	784.039
OCT	711.498	822.121	800.015
NOV	725.789	822.418	813.769
DEC	736.6	822.79	825.918
2023 JAN	600.311	813.075	659.86
FEB	615.276	815.157	683.519
MAR	633.365	815.528	701.381
APR	649.551	816.941	718.994
MAY	670.243	818.279	739.79
JUN	696.148	820.807	760.022
JUL	708.147	822.294	778.669
AUG	724.555	823.261	797.299
SEP	742.59	824.079	815.207
OCT	753.927	824.302	831.184
NOV	767.025	824.897	844.937
DEC	773.759	824.897	857.087
2024 JAN	602.651	815.185	661.08
FEB	617.616	817.267	684.739
MAR	635.705	817.638	702.601
APR	651.891	819.051	720.214
MAY	672.583	820.389	741.01
JUN	698.488	822.917	761.242
JUL	710.487	824.404	779.889
AUG	726.895	825.371	798.519
SEP	744.93	826.189	816.427
OCT	756.267	826.412	832.404
NOV	769.365	827.007	846.157
DEC	776.099	827.007	858.307

The result obtained from the forecast presented in Table 11 showed that by December 2024 the estimated number of reported diabetes case will be 776 cases, 827 cases, and 858 cases using the ARIMA, ARIMAX and exponential smoothing approaches respectively.

V. CONCLUSION

This study assessed the performances of the ARIMA, ARIMAX and the single exponential smoothing model for estimation of diabetes cases in Anambra State. The data obtained for the study met the requirements of normality and stationarity of the models.

The ARIMA model proved to be the best alternative model among the ARIMAX and the single exponential smoothing model for estimating the number of diabetes cases reported in Anambra, having recorded the least value of AIC 1177.92 and the lowest precision measures values of RMSE=18.3723, MAE=9.80049 and MAPE= 2.14573. Five year predictions of cases of Diabetes Mellitus were found using ARIMA, ARIMAX and SES model.

Based on the findings, we recommend there is need for medical practitioners /health planners to create awareness and inform patients about the possible related risk factors of death through early diagnosis and intervention.

REFERENCES

[1] United Nations, "Political declaration of the High-level Meeting of the General Assembly on the Prevention and Control of Non-communicable Diseases" New York: United Nations; 2011. p. 12-13.

[2] O. Oguejiofor, C. Odenigbo, and C. Onwukwe, "Diabetes in Nigeria: Impact, Challenges, Future Directions. *Endocrinol Metab Synd*, 2014, 3: 130. DOI:10.4172/2161-1017.1000130.

[3] R. N. Oputa and S. Chinenye, "Diabetes in Nigeria – a translational medicine approach" *African Journal of Diabetes Medicine*, 2015, 23(1): 1-10.

[4] T. Dahiru, A. A. Aliyu, and A. U. Shehu, "A review of population-based studies on diabetes mellitus in Nigeria" *Sub-Saharan Afr J Med*, 2016, 3:59-64.

[5] A. A. Sabir, S. Balarabe, A. A. Sani, S. A. Isezuo, K. S. Bello, A. O. Jimoh and S. O. Iwuala, "Prevalence of diabetes mellitus and its risk factors among the suburban population of Northwest Nigeria" *Sahel Med J*, 2017, 20:168-172.

[6] D. Adeloye, J. O. Ige, A. V. Aderemi, N. Adeleye, E. O. Amoo, A. Auta, and G. Oni, "Estimating the prevalence, hospitalisation and mortality from type 2 diabetes mellitus in Nigeria: a systematic review and meta-analysis" *BMJ Open*, 2017, 7:e015424. DOI:10.1136/ bmjopen-2016-015424.

[7] S. B. Aynalem and A. J. Zeleke" Prevalence of Diabetes Mellitus and Its Risk Factors among Individuals Aged 15 Years and Above in Mizan-Aman Town, Southwest Ethiopia, 2016: A Cross Sectional Study" *International Journal of Endocrinology*, Article ID 9317987: 2018, 1-7. https://doi.org/10.1155/2018/9317987.

[8] A. E. Uloko, B. M. Musa, M. A. Ramalan, I. D. Gezawa, H. Fabian, F. H. Puepet, A. T. Uloko, M. M. Borodo, and K. B. Sada, "Prevalence and Risk Factors for Diabetes Mellitus in Nigeria: A Systematic Review and Meta-Analysis" *Diabetes Ther*, 2018, 9:1307–1316 https://doi.org/10.1007/s13300-018-0441-1.

[9] B. M. Onyenekwe, E. E. Young, C. B. Nwatu, C. I. Okafor, C. V. Ugwueze, and K. O. Onuekwusi, "Diabetes Mellitus in adult Nigerians: patients' characteristics, laboratory profile, practices and management outcome" *International Journal of Research in Medical Sciences*, 2019, 7(10): 3844-3853. DOI: http://dx.doi.org/10.18203/2320-6012.ijrms20194320.

[10] K. Singh, S. Shastri, A. S. Bhadwal, P. Kour, M. Kumari, A. Sharma, and V. Mansotra" Implementation of Exponential Smoothing for Forecasting Time Series Data. International Journal of Scientific Research in Computer Science Applications and Management Studies, 2019, 1(1): 1-5.

[11] K. P. Burnham and D. R. Anderson "Model Selection and Multi model Inference: A Practical Information- Theoretical Approach. 2nd ed" New York: Springer, 2002.