

# A Comparison of Exponential Smoothing and ARIMA Modelling of the Infertility Rate among Women in Zaria Metropolis

Oyenike Mary Olanrewaju, Mukhtar Umar Shitu\*

*Computer Science Department, Federal University Dutsin-Ma, Katsina, Nigeria*

*\*Corresponding Author*

**Abstract.** Infertility has recently become a socially destabilizing condition worldwide and it is considered as a serious problem in sub-Saharan Africa. This problem has caused conflict and marital disharmony among couples due to stigma and other social discriminations. Although, infertility affects both men and women, the social and psychological trauma affects more women than men. The clinical pattern of infertility varies in different parts of the world, and Nigeria has recorded more cases in Africa. Diagnosis and treatment of infertility is often viewed as a low priority in national development policies and health strategies of the public hospitals. The lack of funding, effective allocation of available health resources and standard infrastructures by the government as well as by international sponsors to the public health sector remain a challenge in most of our societies today. This research compares the optimization of two time-series approaches namely, an autoregressive integrated moving average (ARIMA) and Holt-Winters exponential smoothing in predicting infertility. The research evaluates the forecasting performance of the models using the post forecast summary statistics which includes; mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). The result indicates that exponential smoothing out-performed the ARIMA model with a lower MAE (4.16), RSME (5.27) and MAPE (16.15). Although both models performed well, the selected model gave the optimal accurate prediction compared to the other model. The validity of the model was further checked by comparing the fitted values with the actual values.

**Key words:** Infertility, Autoregressive Integrated Moving Average (ARIMA), Holt-Winters Exponential Smoothing, Trends, mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE).

## I. INTRODUCTION

Infertility is a reproductive system disease that affects both male and female of childbearing age. It is defined as a medical condition characterized by the inability to conceive after 12 months of unprotected sexual intercourse [1]. Globally, it is estimated that 48.5 million couples experience infertility every year [2]. In Africa, because of the level of literacy, infertility is often linked with different superstition held by some people [3]. Infertility and other related complications like miscarriages can affect the person's health and quality of life generally. Couple who want to start a family early and are unable to conceive will definitely

experience psychological and interpersonal distress. This has a great impact on their families and communities. Infertility is one of the primary reasons for divorce among couples. Women in our societies today, especially where child bearing defines a woman's identity and motherhood, they are consistently blamed for the couples' infertility problem and they suffer psychologically and socially [4]. However, infertility as a matter of fact is not merely a health problem as most people assumed. It is a medical condition that may lead to childlessness, and that becomes a great problem to the couples leading to discrimination, inequality and most times divorce. The cause can be difficult to diagnose because there are other medical conditions that contribute to infertility.

Researches done in our environment made emphasis on clinical pattern, prevalence and identify the commonest causes of infertility among childless women in our societies. Majority of the articles cited by authors, are from researches carried outside African environment. There is scarcity of data on infertility in the North Western Nigeria as stated [5]. The aim of the study is to use two-time series models to forecast the future trends and prevalence rate of occurrence of infertility using available past clinical data. Time series models works efficiently and produces a good forecast if fitted to time series data. They have been applied in various domains such as; agriculture, banking, engineering and geology to make analysis and predictions. In medical domain, it has been used to predict mortality rates, disease rates, and epidemic outbreak. Time series models take great advantage of their capability to describe the relationship between variables to investigate the predictive model and give out results.

Most activities in different domains are subject to enormous changes that may affect the best establishment in the world. There is need for an accurate and practical reading in to the future, and this will help in predicting future trend [6]. Forecasts are becoming very crucial since they are the sign of survival and part way to a successful domain. A forecast is a science of estimating and making assumption about the future values of some studied variables [7]. These variables could be the prices of goods, temperature, disease rate, consumption rate etc.

A time series is nothing but observations according to the chronological order of time [8]. Time series forecasting models use mathematical techniques that are based on historical data to forecast. It is founded on the hypothesis that the future is an expansion of the past; that is why we can definitely use historical data to forecast future demand [9].

Many studies about forecasting by time series analysis have been done in several domains. They encircle forecasting inflation, consumption of livestock, product sales and global solar radiation etc.

According to a study conducted in a business domain [10], an approach was used to predict inflation in Ghana. Monthly inflation figures were collected covering a particular period of time to build the model. Inflation was found to integrated order of (6, 1, 6) and root square mean error (RMSE) of 0.115453 indicating the efficient predictability of the model.

Smart [11] explores the feasibility for application of Box-Jenkins approach to time series autoregressive integrated moving average (ARIMA) in modelling and forecasting maternal mortality ratios (MMR). Having been observed that the values fell within the confidence interval, it was concluded that the ARIMA (1, 0, 2) model was adequate for forecasting quarterly maternal mortality ratios at the hospital.

In [12], a time series model SARIMA was implemented to forecast the daily and monthly solar radiation. The goodness of fit was tested against the standardized residuals. The performance of the model was compared with Monte Carlo simulations using root mean square errors (RMSE) and coefficient of determination for evaluation. The suggested model in the study was given as ARIMA (1, 1, 2) having the best confidence level.

Another study aimed at implementing ARIMA model to forecast consumption of livestock product in Tanzania [13]. Models implemented were tested to find a model with good fit. He discovered that the results was in favour of ARIMA (3, 1, 0) having the lowest normal Bayesian Information Criteria (BIC).

Generally, there are many approaches for forecasting to get future values, trend and seasonality. In fact, a lot of research studies are available to justify a precise selection of ARIMA model. This study is also an attempt to forecast the future infertility rate by fitting ARIMA technique and exponential smoothening on past data from the hospital.

## II. MATERIALS AND METHODS

### A. Data Collection

Data of infertility cases of female patients (of child bearing age 18-40) currently attending or that have attended the gynaecology of department of Ahmadu Bello Teaching Hospital (ABUTH) were obtained. Information based on gynaecological cases, age and other clinical data relevant to infertility was obtained. Table I represents the gynaecological cases based on the classification of infertility and table II

represents the Monthly data of infertility cases. A total of 991 cases were recorded from the year 2017, 2018 and 2019.

Table I : Gynaecological Cases Based On Classification Of Infertility

Year	Cervical Factor	Endometriosis	Ovulatory Disorder	Tubal Factor	Uterine Factor	Unexplained Factor	Total Frequency
2017	50	16	72	79	66	30	313
2018	39	24	68	86	72	19	308
2019	59	27	85	97	70	32	370
	148	67	225	262	208	81	991

Table II : Monthly Data Of Infertility Cases From 2017-2019

Month	2017	2018	2019
January	33	29	38
February	26	29	34
March	25	27	29
April	32	28	33
May	26	26	24
June	21	28	32
July	31	30	30
August	32	23	28
September	20	21	29
October	29	21	37
November	18	24	26
December	20	22	30
Total	313	308	370

### B. Data Analysis

Data of infertility cases was analysed using two different time series models, namely, autoregressive moving average, exponential smoothing. Implementation for this research was performed using XLSTAT 2021.

### C. Unit Root Test

Augmented Dickey fuller test or unit root test is used to determine whether the data series is stationary or non-stationary. The null hypothesis in this test states that the series has unit root and alternative hypothesis states that the series has no unit roots. The test statistic of the dickey fuller test is determined and the null hypothesis is rejected, if its value is found to be less than the critical value concluding that the data series has a unit root.

### D. Auto Regressive Integrated Moving Average Model

The acronym ARIMA stands for Auto-Regressive Integrated Moving Average. It has a combination of two- time series models which are Autoregressive (AR) and Moving Average (MA). The models investigate using the available past data of a time series to forecast future trends and values.) The ARIMA model is applied in the case where the series is

non-stationary, an initial differencing step to correspond to the “integrated” part of the model. The model is generally written as ARIMA (p,d,q).

- p is the number of autoregressive terms or the of lag observations
- d is known as the degree of differencing
- q is the number of forecast errors in the model and also to as size of the moving average window.

The forecasting equation is constructed below. Let y denote the d<sup>th</sup> difference of Y, meaning that:

$$\text{If } d=0: y_t = Y_t \tag{1.2}$$

The second difference of Y (equation 1.3) is not the difference from two periods. Rather, it is the first-difference of the first difference. In general, the forecasting equation in term of y is:

In the equation above, the moving average parameters ( $\theta$ 's) are defined so that their signs are negative following the steps introduced by Box Jenkins. Parameters are often denoted by AR(1), AR(2), MA(1), MA(2) etc.

In order to identify the appropriate ARIMA model for Y, you have to begin by determining the differencing (d) to stationarize the series and remove the gross of seasonality. The stationarized series however, may still have autocorrelated errors, suggesting that some number of AR terms ( $p \geq 1$ ) or some number terms ( $q \geq 1$ ) are also needed in the forecasting.

*E. Exponential Smoothing*

Exponential smoothing, proposed in the late 1950s, is one of the time series methods of prediction. It was called exponential weighted moving average. In simple moving average (MA), the exponential smoothing makes use of exponentially decreasing weights over time by weighing observations equally. As the recent observations are added to the model, it increases the association of weight. The method is also a straightforward time series prediction method, which has the characteristics of simple calculation [14].

Here,

- $s_t$  = smoothed statistic, it is the simple weighted average of current observation  $x_t$
- $s_{t-1}$  = previous smoothed statistic
- $\alpha$  = smoothing factor of data;  $0 < \alpha < 1$
- t = time period

If the value of the smoothing factor is larger, then the level of smoothing will reduce. Value of  $\alpha$  close to 1 has less of a smoothing effect and give greater weight to recent changes in the data, while the value of  $\alpha$  closer to zero has a greater smoothing effect and are less responsive to recent changes.

Many alternatives formulas exist, which formula to use is usually a moot point, as most exponential smoothing is performed using software.

III. RESULTS AND DISCUSSION

A. Pattern of Infertility Rate

The Box Jenkins approach has been followed to analyze the infertility data to build the forecast model. The infertility cases recorded for the year 2017, 2018 and 2019 have been depicted in Figure 1.

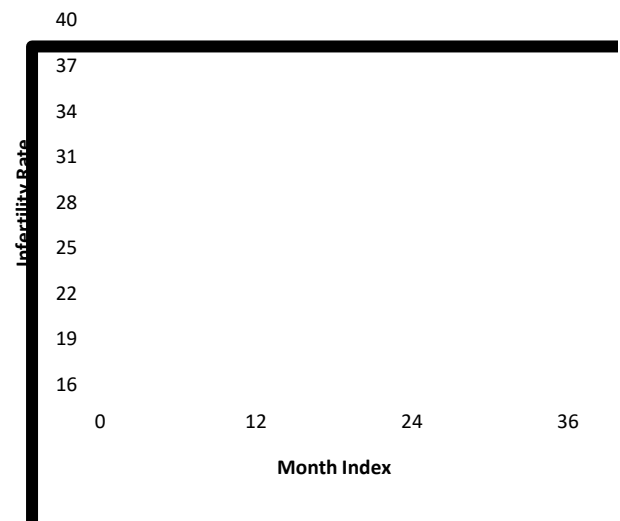


Figure. 1 Time Series Plot of Infertility Cases recorded for the year 2017, 2018 and 2019.

B. Identification of Model

The plots of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) residuals were almost within the confidence limits. The spikes observed in lag 1 and 6 of the ACF plots were out of the confidence limits of 95% (Figure 2). However, in the PACF plots spikes showed a gradual decay resembling a sinusoidal wave (no trend) as seen in Figure 3. The Ljung-Box statistic test did not reject the null hypothesis of independence in the residuals time series (P value = 0.93). The series was differenced to eliminate seasonality in the series and analysed to determine the seasonal and non-seasonal orders of auto-regressive and moving average components of the model as shown in Figures 4 and 5. This model was then adopted for 4 years ahead prediction using a step ahead approach.

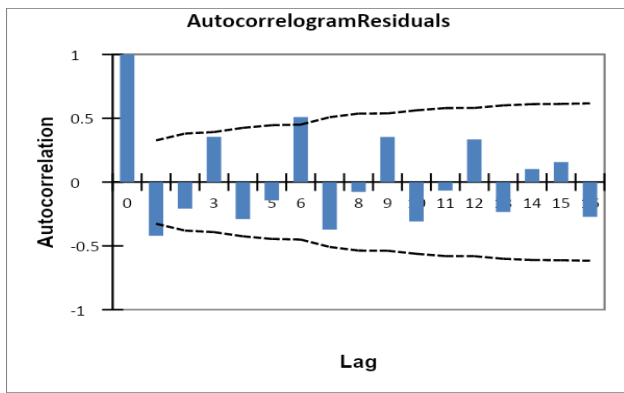


Figure. 2 Autocorrelation Function (ACF)

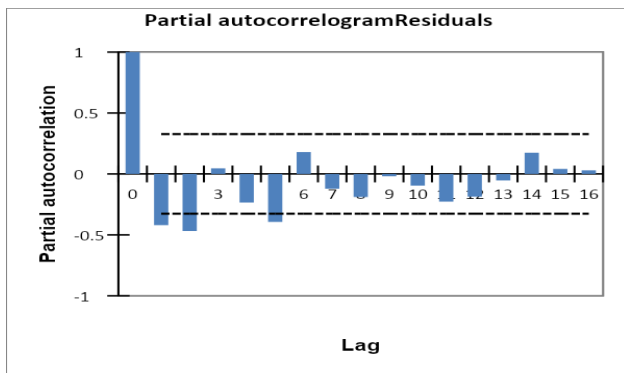


Figure. 3 Partial Autocorrelation Function (PACF)

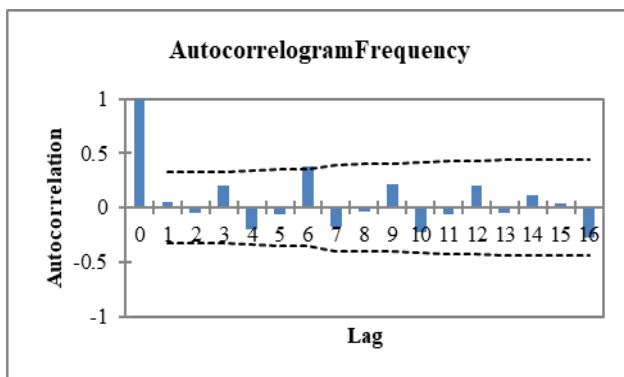


Figure. 4 ACF after Seasonal Differencing

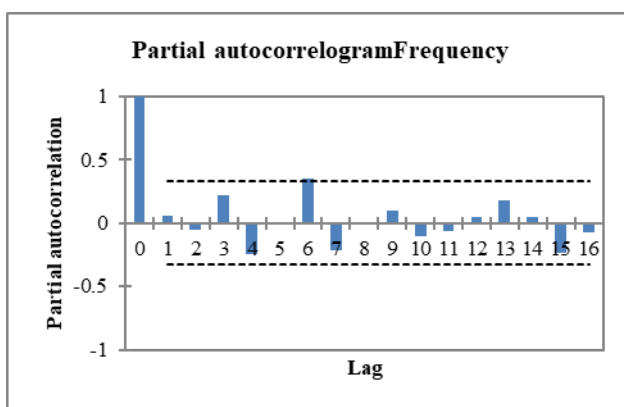


Figure. 5 PACF after Seasonal Differencing

C. Estimation of Model

Table III: Parameter Estimation Of The Models

Parameter		Value	Standard Error	Lower	Upper	Residual
ARIMA	AR (1)	0.941	0.144	0.445	0.602	67
SES	S1	0.123	0.56	0.398	0.535	5.5

D. Goodness Fit of Model

The goodness fit of model was checked using the multinomial goodness of fit test. In model diagnostics, it is suggested that the residuals of the model must be white noise with a constant mean zero and variance  $\sigma^2$ ; otherwise, a different model for the series must be searched (Box and Jenkins, 1976). The residuals of the models were plotted to check if the models were able to capture the necessary information from the data (Figure 6).

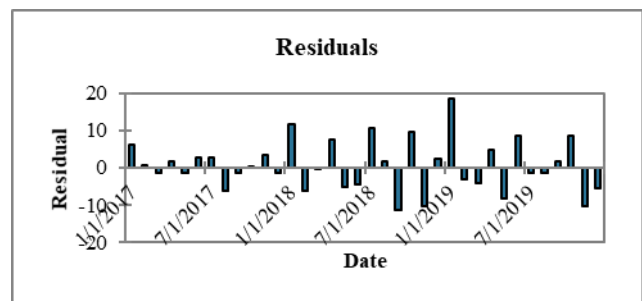


Figure. 6 Plots Residuals of Selected Model

We evaluate the forecasting performance of the models using the post forecast summary statistics which includes; mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). A comparison of performance of the models indicated that the exponential smoothing is better. The exponential model had a lower MAE, RMSE and MAPE values (Table IV). The number of cases predicted by exponential smoothing model has no negative numbers compared to that of ARIMA model. The model that has the smallest values of MAE, RMSE and MAPE was selected to be the most appropriate model for forecasting.

Table IV: Comparison Of The Two Time Series Model Prediction

Model	MAE	RSME	MAPE
ARIMA	4.21	5.65	16.99
Exponential Smoothing	4.16	5.27	16.15

Monthly cases of infertility from the year 2017 to 2019 showed an increase in the number of infertility. The total number of frequency drops in the year 2018 hence there was a rapid increase in the year 2019 compared to the number of cases in 2018. The prediction from ARIMA and exponential smoothing models are shown in figure 7 and 8 respectively.

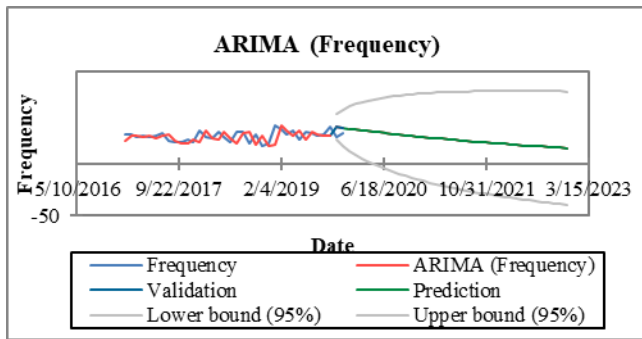


Figure. 7 Prediction from autoregressive integrated moving average of monthly infertility cases in Zaria Metropolis, Nigeria. Y-axis shows predicted monthly number of cases and x-axis shows year. Green curve in between the grey curved lines indicates the prediction.

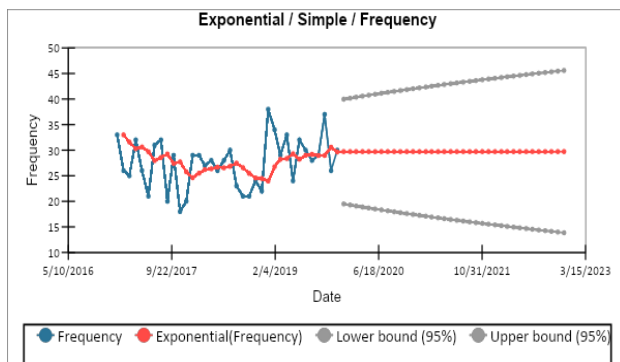


Fig. 8 Prediction from exponential smoothing model of monthly infertility cases in Zaria Metropolis, Nigeria. Y-axis shows predicted monthly number of cases and x-axis shows year. Red line indicates the prediction. While the grey lines represents the upper and lower bound respectively.

Forecast of 4 years ahead from the selected model is shown below in table V.

Table V: Forecast Values For 4 Years Infertility Rate (2020-2023)

Observations	Forecast	Confidence Limits	
		Lower	Upper
2020	324	270	369
2021	347	285	376
2022	331	278	367
2023	384	294	400

IV. CONCLUSION

In this study, we applied two different time series models and discovered that exponential smoothing model gives more accurate forecasting following the steps suggested by Box and

Jenkins methodology. The proposed model showed high performance in forecasting accuracy of (MAPE=16.15%). Prediction shows a precise future trend. Besides the prediction, infertility is medical condition having an increasing number of incidences in our society today. It is observed that the society pay more attention to stigmatising the individuals affected than finding a lasting solution to the problem. Prediction from our approach could provide and early warning signs to health practitioners and the government to put advanced strategies in the health challenge. Finally, adequate counselling and support should be adopted to encourage the individual.

REFERENCES

- [1]. "World Health Organization," 14 September 2021. [Online]. Available: <https://www.who.int/fertility-care>. [Accessed 11 March 2021].
- [2]. A. Agwal, A. Mulgund, and A. Hamada, "A unique view on male infertility around the globe," *Reprod Biol Endocrinol*, vol. 13, p. 37, 2015.
- [3]. E.L Ekpe, K.C. Osuji, and C.M. Ejikem, "Pattern of Hormonal Imbalance among Women of Child-bearing Age in a Tertiary Healthcare Centre in Southern Nigeria," *Research Journal of Obstetrics and Gynecology*, vol. 13(1), pp. 20-24, 2020.
- [4]. V. F. Balen, H. M. Bos, "The Social and Cultural Consequences of Being Childless in Poor-Resource Countries," *Facts, Views and Vision in Obstetrics and Gynaecology*, pp. 21-106, 1 January 2009.
- [5]. A. A. Panti and Y. T. Sununu, "The profile of infertility in a teaching hospital in North-West Nigeria," *Sahel Med J*, vol. 17, pp. 7-11, 2014.
- [6]. J. Fattah, I. Ezzine, Z. Aman, H. El Moussami, and A. Lachhab A, "Forecasting of demand using ARIMA model," *International Journal of Engineering Business Management*, vol. 10, pp. 1-9, 2018.
- [7]. R. S. Tsay, "Time series and forecasting: brief history research.," *J Am Stat Assoc*, vol. 95, pp. 638-643, 2000.
- [8]. B. C. Bozarth and Handfield R. B. *Introduction to operations and supply chain management*, North Carolina: 4th ed. Raleigh: North Carolina State University, 2016.
- [9]. J. D. Wisner, K. C. Tan and G. K. Leong, "Principles of supply chain management: a balanced approach," Thomson South-Western, Mason, 2011.
- [10]. E. A. Samuel & A. Ferdinand, "SARIMA approach to predicting inflation in Ghana," *Journal of Economics and International Finance*, vol. 3(5), no. ISSN 2006-9812 ©2011 Academic Journals, pp. 328-336, 2011.
- [11]. A. S. Smart, "Modeling and Forecasting Maternal Mortality; an Application of ARIMA Models," *International Journal of Applied Science and Technology*, Vol. 3 No. 1, vol. 3, p. 1, 2013.
- [12]. H. A. Mohammed, K.Y. Mohammad and K. Jeong, "Time Series ARIMA Model for South Prediction of Daily and Monthly Average Global Solar Radiation: The Case Study of Seoul, Korea," *Symmetr*, 2019.
- [13]. J. F. Mgaya, "Application of ARIMA models in forecasting livestock products consumption in Tanzania," *Cogent Food & Agriculture*, vol. 5(1), no. 1607430, 2019.
- [14]. P. Ji, D. Xiong, P. Wang and J. Chen, "A study on exponential smoothing model for load forecasting," *Proceedings of the 2012 Asia-Pacific Power and Engineering Conference, APPEEC 2012*.