European Journal of Statistics and Probability

Vol.10, No.2, pp., 29-38, 2022

Print ISSN: 2055-0154(Print),

Online ISSN 2055-0162(Online)

On Forecasting Infant Mortality Rate by Sex using ARIMA Model: A Case of Nigeria

[1*]Christogonus Ifeanyichukwu Ugoh, [2] Udochukwu Victor Echebiri, [3]Ikuomola Emiloju Abiodun, [4]Momoh Besiru, [5]Emwinloghosa Kenneth Guobadia

[1]Department of Statistics, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka, Nigeria
[2,3]Department of Statistics, Faculty of Physical Sciences, University of Benin, Benin, Nigeria
[4]Department of Statistics, Federal Polytechnic, Auchi, Nigeria
[5]Department of Administration, Federal Medical Centre, Asaba, Delta State, Nigeria

Citation: Christogonus Ifeanyichukwu Ugoh, Udochukwu Victor Echebiri, Ikuomola Emiloju Abiodun, Momoh Besiru, Emwinloghosa Kenneth Guobadia(2022) On Forecasting Infant Mortality Rate by Sex using ARIMA Model: A Case of Nigeria, *European Journal of Statistics and Probability*, Vol.10, No.2, pp., 29-38

ABSTRACT: *This paper examines the application of ARIMA model on forecasting Infant Mortality Rate (IMR) in Nigeria. It undertakes a comparison of Male and Female. The data used were obtained from the website of the World Bank. The data consist of annual Infant Mortality Rate (per 1000 live births) on Male and Female from 1980 to 2019. Akaike's Information Criterion (AIC) was used to select the best model and Time Series Plot, Residual Plot and the Histogram for Residuals were used to check the forecast adequacy of the selected models. The results of this study showed that the Infant Mortality Rate (IMR) on Male and Female attain stationarity after the second differencing. ARIMA (2,2,0) with AIC of -9.94 and ARIMA (1,2,0) with AIC of -13.10 were selected for forecasting Infant Mortality Rate for Male and Female respectively. The results further showed that the selected ARIMA models are adequate for forecasting male and female Infant Mortality Rate, and that by 2030, Male infant mortality rate will decline to 58.54 per 1000 live births while Female infant mortality rate will decline to 44.50 per 1000 live births.*

KEYWORDS: ARIMA, Forecasting, Infant Mortality Rate, Male, Female, Nigeria

INTRODUCTION

In the health care sector, forecasting of demographic characteristics such as mortality, morbidity, fertility etc., is very important to socio-economic planners, for it helps them to evaluate, plan, and to regulate policies that will better the lives of people in the society. However, making better decisions based on forecasting on health issues, it will require an adequate model so that a liable result can be obtained.

Infant mortality is the death of an infant before his or her first birthday [1]. Infant Mortality Rate (IMR) represents the number of deaths of children under one year of age per thousand live birth [2,3]. It is one of the indicators commonly used in determining public health. According to [4], high infant mortality rates are generally indicative of unmet human health needs in sanitation, medical care, nutrition, and education.

Infant Mortality Rate (IMR) is one of the health problems facing Nigeria. Nevertheless, Nigeria has been showing a little decrease in the IMR in Male and Female alike over the period review. According to [5], Infant Mortality Rate is an important tool for the evaluation of a population's health and health care system and a barometer for the measurement of a country's wellbeing and the state of health and health facilities. This paper attempts to obtain best forecast model that will be used to forecast IMR in Nigeria from ARIMA dynamics, and undertaking a comparison of male and female.

[6] analyzed the Infant Mortality Rate in Nigeria using annual data from 1964 to 2018, using ARIMA modeling. They selected ARIMA (1,1,1) as the best model and their forecast revealed that by 2030, there will be a reduction of 30% in the IMR in Nigeria. [7] forecasted the India Infant Mortality Rate using ARIMA (2,1,1), the result showed that by 2025, there will be a reduction of IMR to 15 per 1000 live births.

MATERIALS AND METHODS

Autoregressive Integrated Moving Average (ARIMA) Process

If the data exhibits no apparent deviations from stationarity (a stochastic process X_t is said to be stationary if its properties are unaffected by a change of time origin) and has a rapidly decreasing autocorrelation function, then it is appropriate to seek a suitable ARMA (Autoregressive Moving Average) process to represent the mean-correlated data. However, if there is an apparent deviation from stationarity, then it will be appropriate to difference or transform the data to generate a new series which leads to ARIMA (Autoregressive Integrated Moving Average) process, where AR is the Autoregressive Process and MA is Moving Average Process.
An $ARMA(p,q)$ process can be expressed as

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q}; \quad where\ \{W_t\} \sim WN(0, \sigma^2) \qquad (1)$$

Differencing of Stochastic Process $\{X_t\}$: The d difference of the stochastic process X_t is denoted by $\nabla^d X_t$ or X_t' where d is the number of differencing. It be written as

$$\nabla^d X_t = X_t' = (1-B)^d X_t = \sum_{k=0}^{d} \binom{d}{k} (-1)^k X_{t-k}\,; \quad where \quad \binom{d}{k} = \frac{d!}{k!\,(d-k)!} \qquad (2)$$

For example, the first differencing is obtained as

$$\nabla X_t = \binom{1}{0}(-1)^0 X_{t-0} + \binom{1}{1}(-1)^1 X_{t-1} = X_t - X_{t-1} \quad (3)$$

and the second differencing is obtained as

$$\nabla^2 X_t = \binom{2}{0}(-1)^0 X_{t-0} + \binom{2}{1}(-1)^1 X_{t-1} + \binom{2}{2}(-1)^2 X_{t-2} = X_t - 2X_{t-1} + X_{t-2} \quad (4)$$

An $ARIMA\ (p, d, q)$ process with a drift δ can be expressed as

$$X_t' = \delta + \phi_1 X_{t-1}' + \cdots \phi_p X_{t-p}' + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q} + W_t \quad (5)$$

An $ARIMA\ (p, d, q)$ process without a drift δ can be expressed as

$$X_t' = \phi_1 X_{t-1}' + \cdots \phi_p X_{t-p}' + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q} + W_t \quad (6)$$

Autoregressive (AR) Process: An $AR(p)$ process is a stationary process that satisfies

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = W_t \quad \text{where } \{W_t\} \sim WN(0, \sigma^2) \quad (7)$$

Moving Average (MA) Process: An $MA(q)$ process is a stationary process that satisfies

$$X_t = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q} \quad \text{where } \{W_t\} \sim WN(0, \sigma^2) \quad (8)$$

ARIMA Model Fitting

ARIMA model is developed using the Box-Jenkins methodology, which is described by four steps: Model Identification, Parameter Estimation, Diagnostic Checking, and Forecasting.

Step 1: Model Identification: In order to identify the specified model, the following steps are considered:

- Time series plot to check stationarity. If the plot shows stationarity, ARMA model is applied, if not, then
- Unit root test for confirmation of differencing using the Augmented Dickey-Fuller Test
- Computation of sample ACF and PACF
- ACF and PACF plots to confirm the specified model
- If the ACF plot cuts off rapidly, then MA(q) model is suggested, but if it has an (or a mixture of) exponential decay(s) pattern, then AR(p) model is suggested.
- If two or more models are obtained, the Akaike Information Criterion (AIC) will be used to obtain the best model among them (which is the smallest AIC)

$$\text{AIC} = log(\hat{\sigma}^2) + 2(p + q)T^{-1} \quad (9)$$

where T is the total sample size, p is the AR model order, q is the MA model order

Step 2: Parameter Estimation: This study adopts the least square method to obtain the specified model

Step 3: Diagnostic Checking: The adequacy and acceptance of the specified model are actualized through the following:

- Computation of Modified Box-Pierce (Ljung-Box)

$$Q_{BP} = T\sum_{k=1}^{K} r_k^2 \sim \chi^2_{\alpha,K} \tag{10}$$

- Time series plot of the residuals
- Residuals plot, and
- Histogram of the residuals

Step 4: Forecasting: When the specified ARIMA model is obtained, it is used to obtain forecasts for future observations. The forecast t-period-ahead is obtained using:

$$X_{T+t} = \delta + \sum_{i=1}^{p+d} \phi_i X_{T+t-i} + W_{T+t} - \sum_{i=1}^{q} \theta_i W_{T+t-i} \tag{11}$$

RESULT/FINDINGS

Figure 1 shows the time plot for the distributions of Male and Female Infant Mortality Rate for the period of 1990-2019.

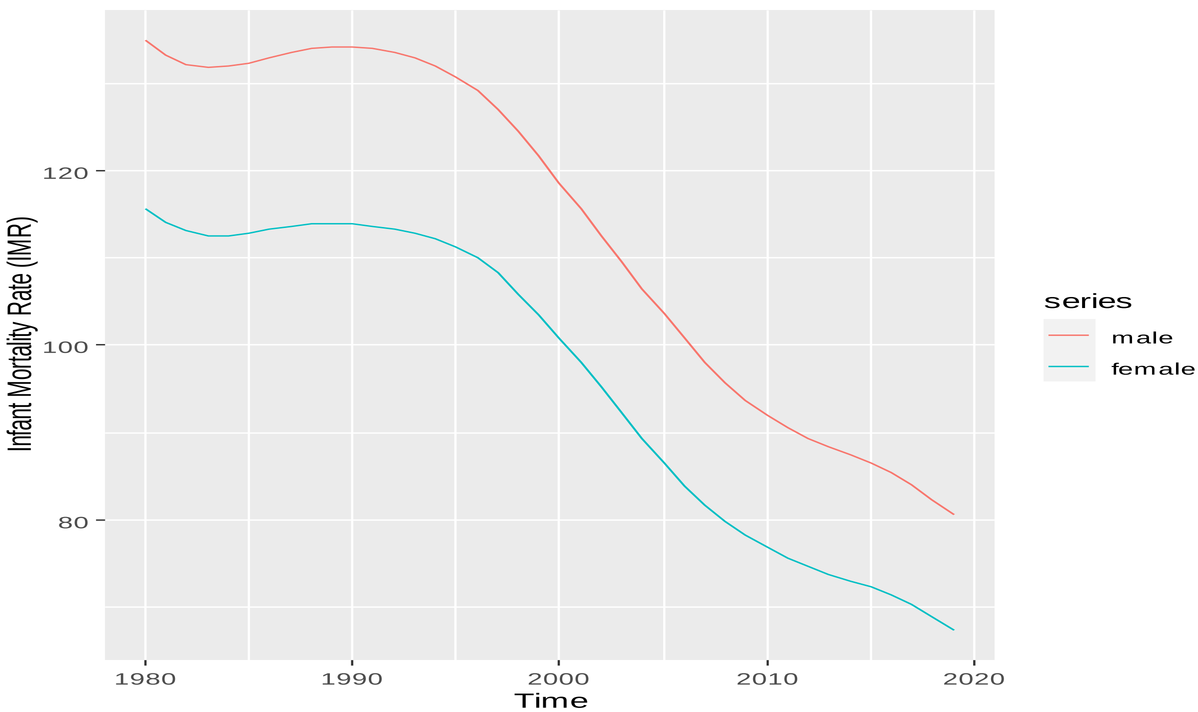

Figure 1. Time Plot Showing the Male and Female Infant Mortality Rate

The time plot in Figure 1 shows that the distribution of both Male Infant Mortality Rate and Female Infant Mortality Rate data series show no evidence of constant variances. However, this is a case of non-stationarity. The data requires to be differenced in order to attain stationarity.

Table 1. Descriptive Statistics for Male IMR and Female IMR

	Male IMR	Female IMR
1st Quartile	93.28	77.85
Median	120.20	102.20
Mean	113.81	96.26
3rd Quartile	132.53	112.95

In Table 1, the average Infant Mortality Rate for male is 113.81, at 75% quantile it is 132.53, and at 25% quantile it is 93.28. While the average Infant Mortality Rate for female is 96.26, at 75% quantile it is 112.95, and at 25% quantile it is 77.85.

Table 2. Augmented Dickey-Fuller (ADF) Test

	DF	p-value
Male	-3.4276	0.06775
Female	-3.2300	0.09687

In Table 2, the Male IMR data series gives a Dickey-Fuller (DF) of -3.4276 and a p-value of 0.06775 (greater than 0.05), implying there is need to difference the data. While the female IMR data series gives a Dickey-Fuller (DF) 0f -3.2300 and a p-value is 0.09687 (greater than 0.05), showing evidence of need for differencing.

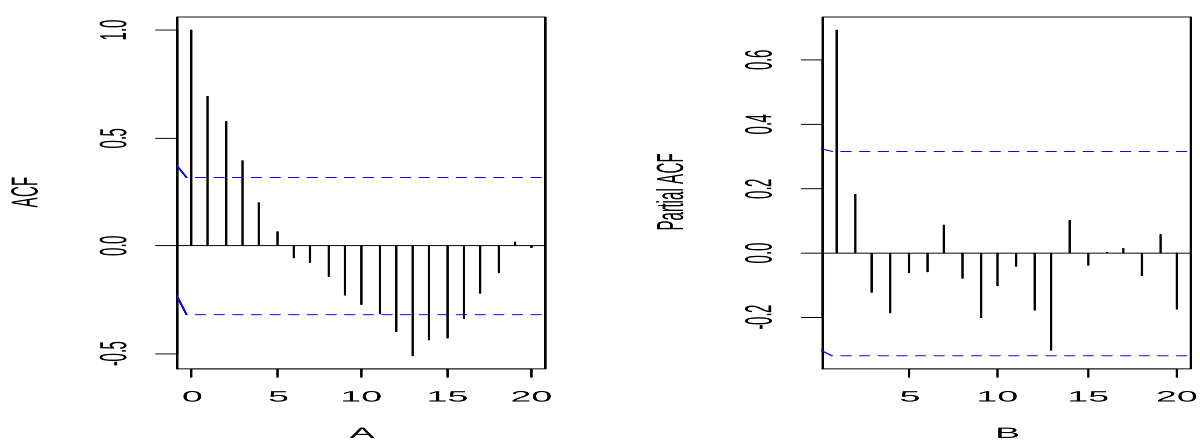

Figure 2. (A) 2nd Difference ACF Plot (B) 2nd Difference PACF Plot for Male IMR

Figure 2A shows a mixture of exponential decay pattern, implying that the required model is an ARIMA(p,2,0). In Figure 2B, lag 1 is the only significant lag. The best model is obtained by comparing ARIMA(1,2,0) and ARIMA(2,2,0) using AIC as shown in Table 3.

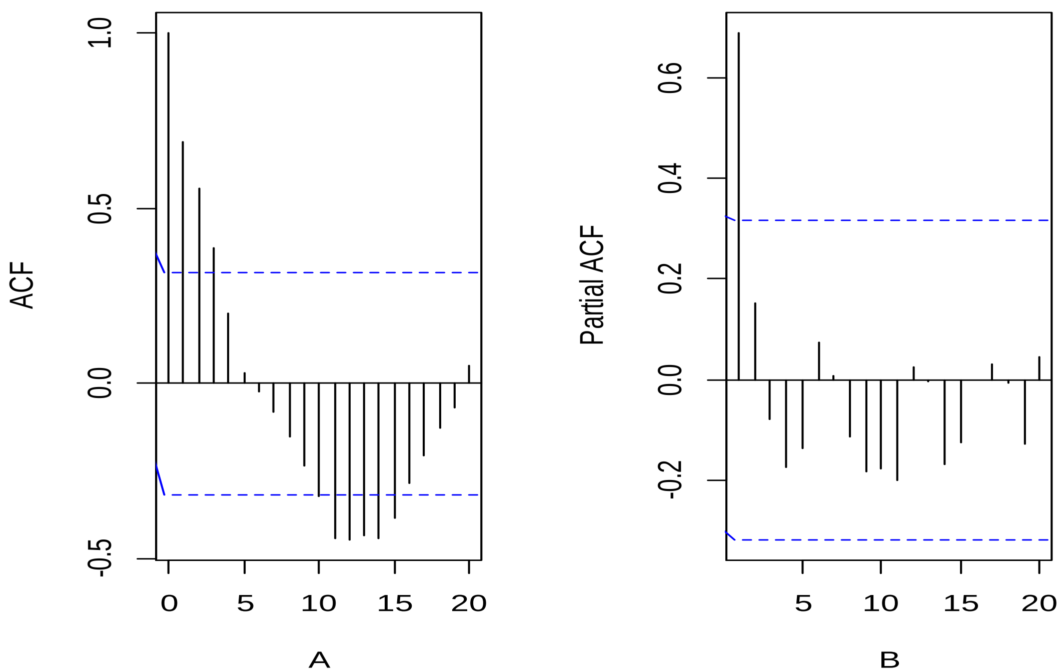

Figure 3. (A) 2nd Difference ACF Plot (B) 2nd Difference PACF Plot for Female IMR

Figure 3A shows a mixture of exponential decay pattern, implying that the required model is an ARIMA(p,2,0). While in Figure 3B, lag 1 is the only significant lag. The best model is obtained by comparing ARIMA(1,2,0) and ARIMA(2,2,0) using AIC as shown in Table 3.

Table 3. Model Comparison using AIC

Male IMR		Female IMR	
Model	AIC	Model	AIC
ARIMA(1,2,0)	-9.24	ARIMA(1,2,0)	-13.10
ARIMA(2,2,0)	-9.94	ARIMA(2,2,0)	-13.03

In Table 3, Under Male IMR, ARIMA(2,2,0) has the least AIC of -9.94, thus making it the best model to forecast Male IMR. An Under Female IMR, ARIMA(1,2,0) has the least AIC of -13.10, making it the best model to forecast Female IMR.

Table 4. Estimated Coefficients of ARIMA(2,2,0) Model and ARIMA(1,2,0) Model

ARIMA(2,2,0)	AR1	AR2
AR	0.5984	0.2847
Standard Error	0.1595	0.1680

ARIMA(1,2,0)	AR1
AR	0.7675
Standard Error	0.1103

ARIMA(2,2,0) model is expressed as $X_t = 0.5984X_{t-1} + 0.2847X_{t-2} + W_t$ (12)
ARIMA(1,2,0) model is expressed as $X_t = 0.7675X_{t-1} + W_t$ (13)

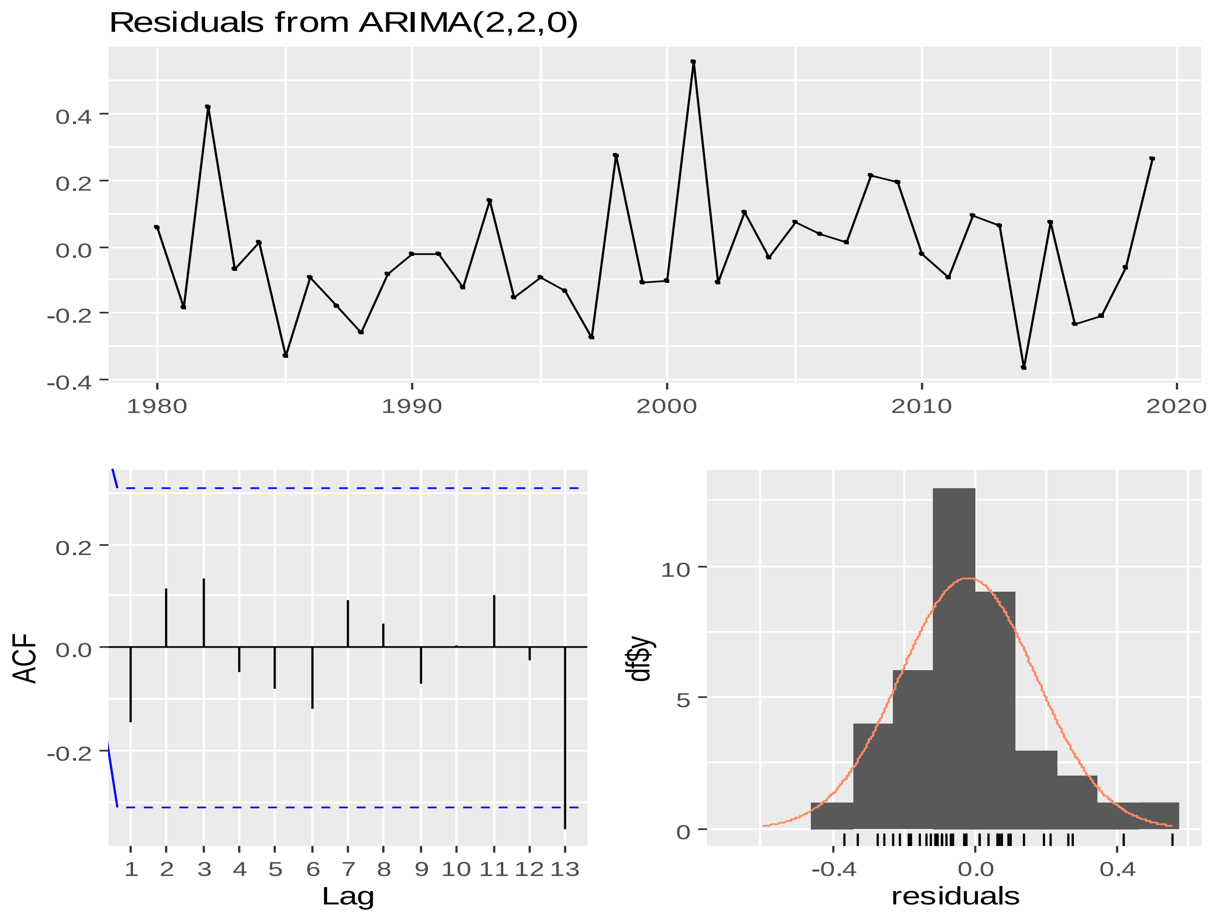

Figure 4. Plot Showing the Diagnostic Check for Forecast Errors of ARIMA(2,2,0) Model

In Figure 4, the time plot of the in-sample forecast errors (residuals) shows that the variance of the forecast errors is constant over time. The histogram of the time series shows that the forecast errors are normally distributed and the mean is close to zero. Therefore, it is plausible that the forecast errors are normally distributed with mean zero and constant variance. And in the ACF plot of the in-sample forecast errors, lag 13 exceeds the significance bounds, then confirming whether there is evidence of non-autocorrelations in the in-sample forecast error, we refer to Table 5. However, in Table 5, the Ljung-Box test statistic is 3.9405 and p-value is 0.6847, which is greater than 0.05, therefore, the successive forecast errors do not seem to be correlated. Since the forecast errors are normally distributed with mean zero and constant variance, and the successive forecast errors are not correlated, then the ARIMA(2,2,0) does seem to provide an adequate predictive model for the Male Infant Mortality Rate (IMR).

Table 5. Ljung-Box Test for forecast errors of ARIMA(2,2,0) Model

Ljung-Box test
data: Residuals from ARIMA(2,2,0)
Q* = 3.9405, df = 6, p-value = 0.6847
Model df: 2. Total lags used: 8

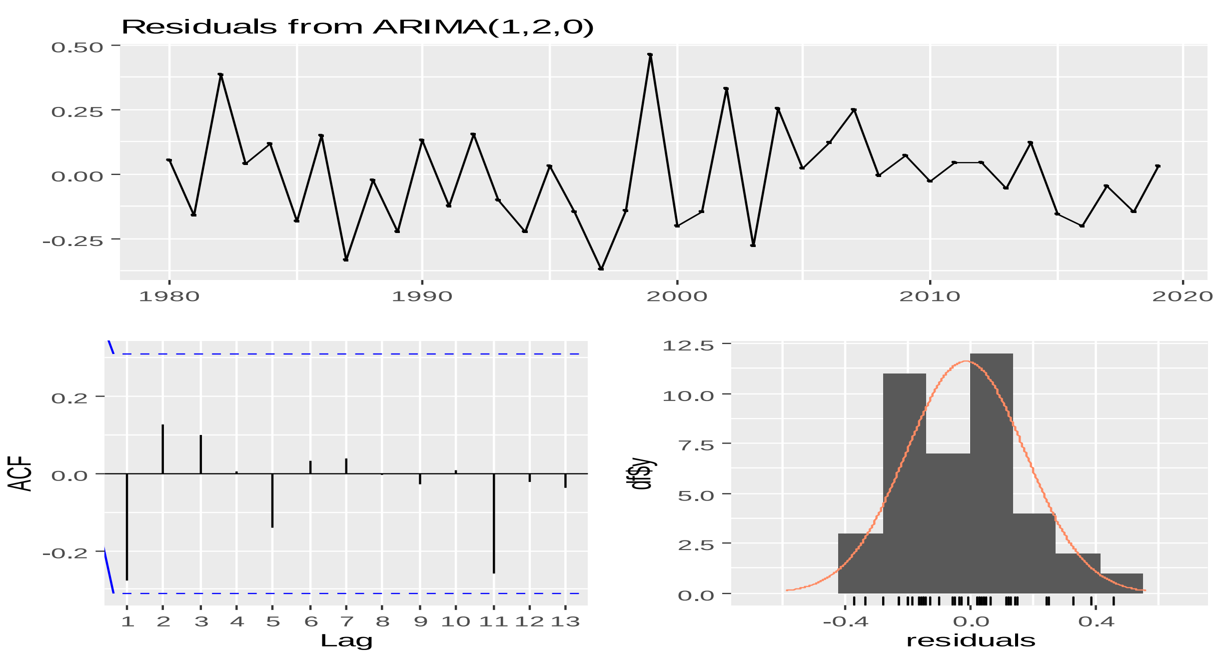

Figure 5. Plot Showing the Diagnostic Check for Forecast Errors of ARIMA(1,2,0) Model

In Figure 5, the time plot of the in-sample forecast errors (residuals) shows that the variance of the forecast errors is constant over time. The histogram of the time series shows that the forecast errors are normally distributed and the mean is close to zero. Therefore, it is plausible that the forecast errors are normally distributed with mean zero and constant variance. And in the ACF plot of the in-sample forecast errors, there is no significant lag, therefore, there is evidence of non-autocorrelations in the in-sample forecast error. Since the forecast errors are normally distributed with mean zero and constant variance, and the successive forecast errors are not correlated, then the ARIMA(1,2,0) does seem to provide an adequate predictive model for the Female Infant Mortality Rate (IMR).

Table 5. Out-of-Sample Forecast for the Period 2020-2030 using ARIMA(2,2,0) and ARIMA(1,2,0)

Year	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
Male	78.81	76.98	75.09	73.15	71.16	69.14	67.08	64.99	62.87	60.72	58.54
Female	65.55	63.68	61.71	59.68	57.60	55.47	53. 32	51.14	48.94	46.72	44.50
Difference	13.26	13. 30	13. 38	13.47	13.56	13.67	13.76	13.85	13.93	14.00	14.04

From Table 5, the forecast value of Male IMR by 2030 is 58.54, showing 42.07% decrease in IMR, while the forecast value for Female IMR, by 2030 is 44.50 showing 39.80% decrease.

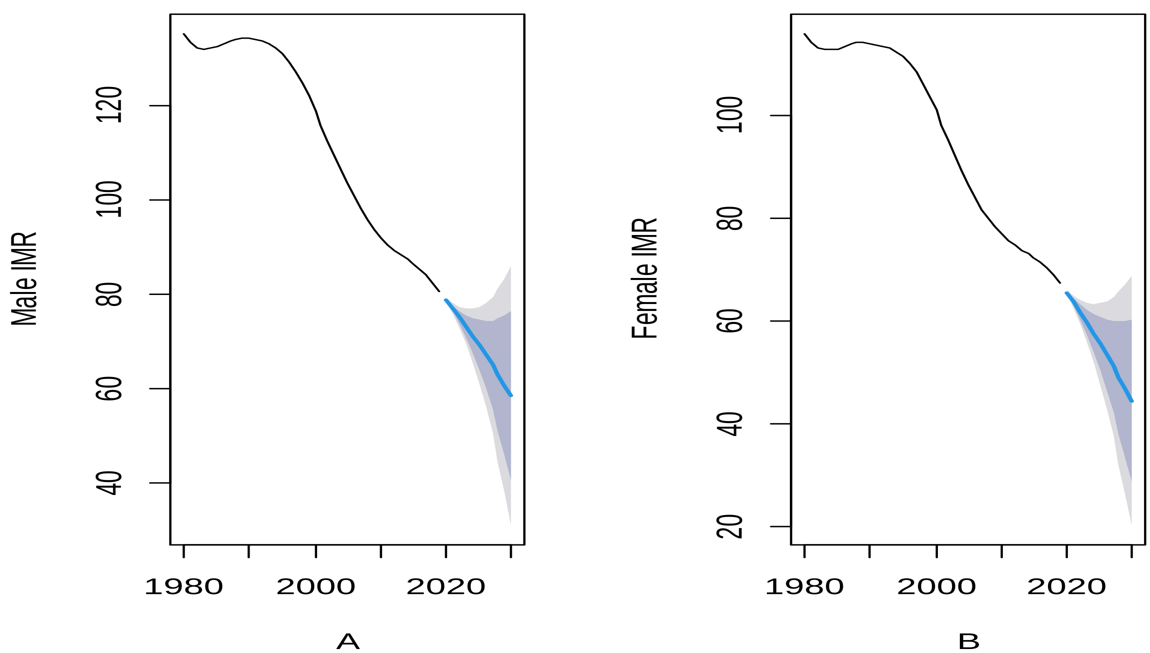

Figure 6. In-Sample Data and Out-of-Sample Forecast for (A) Male IMR (B) Female IMR

Vol.10, No.2, pp., 29-38, 2022

Print ISSN: 2055-0154(Print),

Online ISSN 2055-0162(Online)

CONCLUSION

The paper aimed at obtaining the best ARIMA models that will be used to forecasting both male and female infant mortality rates (IMR) in Nigeria using annual data. Infant mortality rate is one of the health challenges Nigeria is facing. It was observed that the data on male and female infant mortality rate showed a slight decrease. The analysis of the data showed that ARIMA (2,2,0) is better for male infant mortality rate forecasting and ARIMA (1,2,0) is better for female IMR forecasting. The forecast with the selected ARIMA models revealed that there will be a decline in male infant mortality rate to 58.54 per 1000 live births and a decline in female infant mortality rate to 44.5 per 1000 live birth by year 2030.

REFERENCES

[1] Mathews TJ, MacDorman MF (January 2013). "Infant Mortality Statistics from 2009 period linked birth/infant death data set". National Vital Statistics Reports. 61 (8): 1-27

[2] Infant Mortality Rate. Geneva, Switzerland: World Health Organization. 2015. Retrieved October 22, 2020

[3] CDC. "Reproductive Health". Centers for Disease Control and Prevention. Retrieved September 8, 2021

[4] Treiber, Linda, A. "Infant Mortality Rate". Encyclopedia Britannica, 16 Aug. 2017, https:www.britannica.com/science/infant-mortality-rate. Accessed May 20, 2022

[5] Madise, N., Magadi, M. and Diamond, I. "Factors associated with unfavourable birth outcomes in Kenya", Journal of Biosocial Sciences, 2001, 33(20), 199-225

[6] Ewere Friday and Donalben O. Eke, "Time series analysis and forecast of infant mortality rate in Nigeria: An ARIMA modeling approach", Canadian Journal of Pure and Applied Sciences, 2020, 14(2),5049-5059

[7] Mishra A.K., Chandar Sahanaa and Mani Manikandan, "Forecasting Indian Infant Mortality Rate: Application of autoregressive integrated moving average model", Journal of Family and Community Medicines, 2019, 26(2), 123-126