

Access this article online

Quick Response Code:



Website:

www.jfcmonline.com

DOI:

10.4103/jfcm.JFCM\_51\_18

# Forecasting Indian infant mortality rate: An application of autoregressive integrated moving average model

Amit K. Mishra, Chandar Sahanaa, Mani Manikandan

## Abstract:

**BACKGROUND:** The Infant Mortality Rate (IMR) reflects the socioeconomic development of a nation. The IMR was reduced by 28% between 2015 and 2016 (National Family Health Survey-4 [NFHS-4]) as compared to 2005–2006 (NFHS-3), from 57/1000 to 41/1000 live births. The target fixed by the Government of India for IMR in 2019 is 28/1000 live births (National Health Policy, 2017). One of the most common methods of forecasting this is the autoregressive integrated moving average (ARIMA) model. A forecast of IMR can help implementation of interventions to reduce the burden of infant mortality within the target range.

**MATERIALS AND METHODS:** The objective of the study was to give a detailed explanation of ARIMA model to forecast the IMR (2017–2025). Secondary data analysis and forecast were done for the available year and IMR data extracted from “open government data platform India” website.

**RESULTS:** The forecast of the sample period (1971–2016) showed accuracy by the selected ARIMA (2, 1, 1) model. The postsample forecast with ARIMA (2, 1, 1) showed a decreasing trend of IMR (2017–2025). The forecast IMR for 2025 is 15/1000 live births.

**CONCLUSION:** In the current study, long-time series IMR data were used to forecast the IMR for 9 years. The data showed that IMR would decline from 33/1000 live births in 2017 to 15/1000 live births in 2025. When the actual data for another year (2017) are available, the model can be checked for validity and a more accurate forecast can be performed.

## Keywords:

Forecast, infant mortality rate, live births, National Health Policy

## Introduction

Globally, infant mortality appears to have declined on average by about 23/1000 live births and mortality of older infants by about 25/1000 live births.<sup>[1]</sup> In 2000, when the UN adopted eight-millennium developmental goals (MDG), the infant mortality rate (IMR) in India was more than 25 times, with rural India being higher than rural areas in the developed countries.<sup>[2]</sup>

IMR reflects the socioeconomic development of a nation. India made significant moves to

reduce the IMR (41/1000 live births)<sup>[3]</sup> but had been unable to achieve the MDG in 2015.<sup>[4]</sup> The IMR fell by 28% in 2015–2016 (National Family Health Survey-4 [NFHS-4]) as compared to 2005–2006 (NFHS-3), from 57/1000 to 41/1000 live births.<sup>[5,6]</sup> As per the NFHS-4, the lowest IMR was reported from Kerala and Uttar Pradesh reported the highest.<sup>[7]</sup> The target fixed by the Government of India for IMR in 2019 is 28/1000 live births (National Health Policy, 2017).<sup>[8,6]</sup> One of the targets in sustainable developmental goals is to reduce the under-5 mortality to 25/1000 live births by 2030.<sup>[9]</sup> Infant mortality accounts for over 80% of under-5 mortality rate.<sup>[10]</sup>

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

**How to cite this article:** Mishra AK, Sahanaa C, Manikandan M. Forecasting Indian infant mortality rate: An application of autoregressive integrated moving average model. *J Fam Community Med* 2019;26:123-6.

Department of Community  
Medicine, Pondicherry  
Institute of Medical  
Sciences, Puducherry,  
India

### Address for correspondence:

Dr. Chandar Sahanaa,  
Department of Community  
Medicine, Pondicherry  
Institute of Medical  
Sciences, Kalapet,  
Puducherry - 605 014,  
India.  
E-mail: csahanaa@gmail.  
com

IMR fluctuates according to health status of the country, which is a dynamic process. In this regard, if the forecast IMR is available, intervention could be planned and implemented effectively at the right time. One of the most common methods of forecasting is the autoregressive integrated moving average (ARIMA) model, widely used in the field of health and agricultural sciences being simple and easy to use.<sup>[11]</sup> As infant mortality is one of the indicators of the health status of the country, an attempt was made, in the current study, to forecast the IMR of India up to 2025 with a detailed stepwise explanation of ARIMA model.

## Materials and Methods

The current study was based on the secondary data analysis of IMR of India between 1971 and 2016, collected from the “open government data platform India” website.<sup>[12]</sup> The data available to public are not individually identifiable (the data referred to in the present study were the annual mortality rates and not the individual data). Ethical approval was not obtained for this study as there was no direct involvement of any human subject. For the analysis (forecast), the ARIMA model introduced by Box and Jenkins (also called as Box–Jenkins model) in the 1960s for forecasting a time series variable was used.

The ARIMA method is an extrapolation method for forecasting data/variable, and like other methods of forecasting, it requires only long-time series data. To forecast any event, the best ARIMA model has to be selected/prepared as per the following three steps (i.e., model identification, parameter estimation, and model validation). Once the model is validated, the procedure could be carried out with the available data. In the current study, the data for IMR for the period of 1971–2016 (46 years) were used to fulfill the need for long-term series data set for analysis using the ARIMA model. The expression for the ARIMA ( $p, d, q$ ) model is  $Y_t = b_0 + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + e_t$ , where  $Y_t$  and  $e_t$  are actual and random error at period  $t$ , respectively, and the other variables involved are the order of autoregressive (AR) part ( $p$ ), the degree of differencing involved ( $d$ ), and the order of the moving average (MA) part ( $q$ ).

The first step (model identification) in forecasting any time series data using ARIMA model is to check how stationary the available data is because the ARIMA model is used when the data are stationary. A stationary data series is that whose values vary over time around a constant mean and variance. Hence, to apply the ARIMA model to forecast data that are nonstationary, it should be transformed into stationary data series. There are several ways to ascertain stationarity of data. The

most common method to check stationarity of data is by examining its graph/time plot. Figure 1 shows that the data are nonstationary as there is a downward trend and the values did not vary over time around a constant mean and variance. Another method of checking for stationarity of data is the autocorrelation function (ACF) and partial ACF (PACF) of the IMR time series data. Figure 2 shows that the values are not within the upper and lower confidence limits, which signify that the data set is nonstationary.

The transformation of data from nonstationary to stationary can be done by considering appropriate differencing of the data. The newly constructed variable after considering the difference of order ( $d$ ) is “ $X_t$ ” which can be examined for stationarity. For each difference of order ( $d$ ), the final hypothetical time series data should be checked for stationarity. The stationarity of the new time series hypothetical data can be checked by ACF and PACF. In the current study, the difference of order ( $d$ ) 1 was sufficient to achieve stationarity in mean. The stationarity of the hypothetical time series data after difference of order 1 was checked by ACF and PACF of IMR data.

After transforming the data to a stationary set, the next step (parameter estimation) is to identify the values of  $p$  and  $q$ , where  $p$  is the order of AR component and  $q$  is the order of MA component. When AR and MA are combined, it is known as ARMA. However, in time series data, two other components besides AR and MA are the “trend (long-term variation)” and “seasonality (short-term variation).” When the ARMA model is integrated with the trend and seasonality component, it is known as the ARIMA model. For each ARIMA model to forecast the data under consideration, a set of values, i.e., ( $p, d, q$ ) is required. The value of “ $d$ ” is already estimated as 1 for the current forecast. Using different values (0, 1, 2) for  $p$  and  $q$ , with permutations and combinations, many models can be suggested. The criteria for the best model of forecasting for the available data are the minimum Akaike information criterion (AIC) and Bayesian information criterion (BIC).

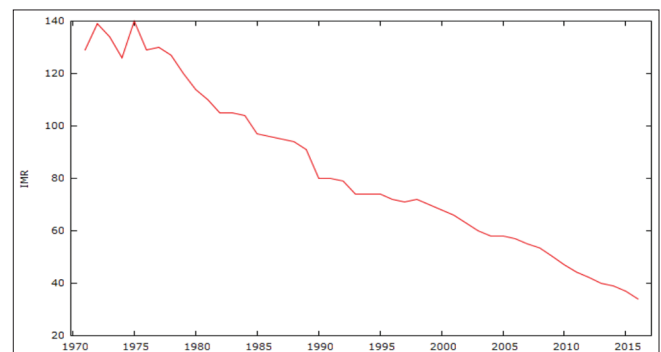


Figure 1: Year-wise infant mortality rate in India (1971–2016)

For the current study, to consider different values for  $p$  and  $q$ , the AIC and BIC values were estimated and the model (ARIMA [2, 1, 1]) with lowest values was selected for the forecast [Table 1].

The last step (model validation) in ARIMA model estimation is the validation of the selected model. There are many methods of validating the selected model, but the most common (goodness of fit) is by examining the ACF and PACF of the residuals of IMR. As shown in Figure 3, the ACF and PACF of residuals of IMR at different lag are within the confidence level. This proves that the selected ARIMA model is appropriate for the forecast of the available data or data under consideration.

### Results

Using the selected ARIMA (2, 1, 1) model, the IMR was forecasted for postsample period of 9 years (2017–2025). The results showed a declining trend of IMR in the predicted period [Table 2 and Figure 4]. The predicted infant mortality in India is 28/1000 live births in 2019 (19.6, 35.5) and 15/1000 live births by 2025 (1.2, 27.3).

### Discussion

According to the available literature, forecasting for a specific period is the fourth and final step<sup>[13,14]</sup> in the ARIMA model. There are two kinds of forecasts: sample period (the period for which data are available) and

postsample period (actual forecasting period) forecast.<sup>[11]</sup> The sample period forecast gives confidence to the model as the gap between actual data and forecast data could be discovered. This is an indirect indicator of the accuracy of the model. The postsample forecast generates a genuine

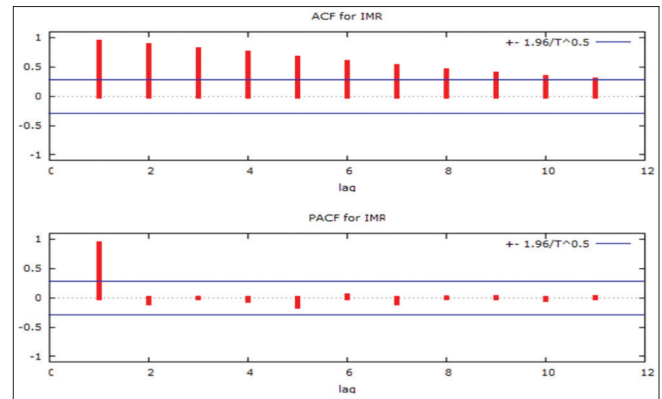


Figure 2: Autocorrelation function and partial autocorrelation function of infant mortality rate

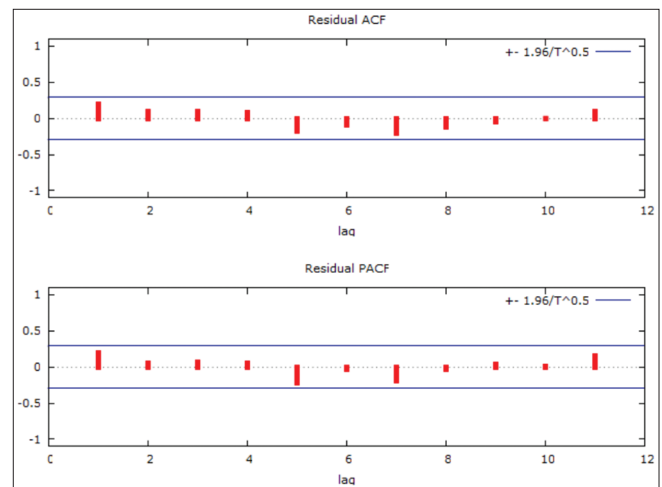


Figure 3: Validation of model for current forecasting

Table 1: Model identification for infant mortality rate forecasting

Model	AIC	BIC
ARIMA (0, 1, 0)	258.29	261.90
ARIMA (0, 1, 1)	252.18	257.60
ARIMA (1, 1, 0)	252.66	258.08
ARIMA (1, 1, 1)	253.4	260.65
ARIMA (2, 1, 1)	247.4	256.43
ARIMA (2, 1, 2)	247.18	258.02

AIC=Akaike information criterion, BIC=Bayesian information criterion, ARIMA=Autoregressive integrated moving average

Table 2: Postsample forecast (2017-2025) infant mortality rate using autoregressive integrated moving average (2,1,1)

Years	Forecast IMR	95% CI
2017	32.4253	25.8812-38.9695
2018	29.8150	22.5130-37.1171
2019	27.6017	19.6359-35.5675
2020	25.5605	15.9108-35.2102
2021	23.0332	13.0519-33.0146
2022	21.0216	9.94521-32.0980
2023	18.6641	6.95228-30.3758
2024	16.4129	4.15880-28.6670
2025	14.2532	1.19033-27.3160

CI=Confidence interval, IMR=Infant mortality rate

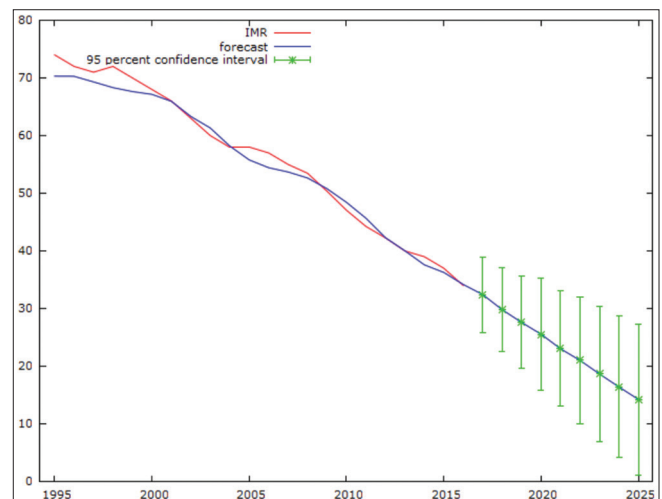


Figure 4: Forecast infant mortality rate in sample and postsample period

forecast for a specific period for use in planning and intervention. The selected ARIMA model is used in the current study (2, 1, 1), and the IMR has been forecasted for both sample and postsample periods [Table 2 and Figure 4].

Figure 4 shows the forecast of IMR. The gap between the actual data (red line) and forecast data (blue line) is visible but negligible and gives confidence to the selected model for further forecasts. The postsample forecast has been plotted with 95% confidence interval. As the period of forecast increases, the confidence range becomes broader. In the current study, the postsample forecast is for 9 years and the change in the confidence interval range from 25.8812–38.9695 to 1.19033–27.3160 becomes noticeable. There is a decreasing trend of IMR in the period of forecast (2017–2025) and the predicted IMR by 2025 is 15/1000 live births. The predicted data show that the nation will achieve the target of IMR 28/1000 live births in 2019 as fixed by the National Health Policy 2017, Government of India.<sup>[8]</sup>

It is obvious that many factors directly or indirectly impact on infant mortality, thereby reflecting in the rates. An analysis of these annual rates indirectly takes into account the factors impacting on them. One of the limitations of the current study tool is that direct consideration of factors is not possible in the use of the ARIMA model for analysis.

## Conclusion

The ARIMA model is the most commonly used method of forecast because of its simplicity and its usefulness for any time series data, the only requirement being the long-time series data. The forecast is not 100% perfect; however, if the current data are available and the best model has been selected, the accuracy of forecasting of any variable is improved. In the current study, long-time series IMR data had been used to forecast the IMR for 9 years. The predicted data show that IMR will decline from 33 in 2017 to 15/1000 live births by 2025. According to the data forecast, the target fixed by National Health Policy, 2017 (28/1000 live births) in 2019 is achievable. When the data for another year (2017) are available, the model can be checked for validity, and the forecast can be done more accurately. The current study reveals the application of a statistical tool for forecasting an event to help in the proper planning of intervention. Similarly, other health events could be forecasted so that proper intervention can be planned at the appropriate time.

## Financial support and sponsorship

Nil.

## Conflicts of interest

There are no conflicts of interest.

## References

1. Infant and Child Mortality: levels, Trends and Demographic Differentials. Available from: <https://www.popline.org/node/399179>. [Last accessed on 2018 Jan 01].
2. UNICEF. Levels and Trends in Child Mortality. UNICEF; 2011. Available from: [http://www.unicef.org/media/files/Child\\_Mortality\\_Report\\_2011\\_Final.pdf](http://www.unicef.org/media/files/Child_Mortality_Report_2011_Final.pdf). [Last accessed on 2018 Mar 30].
3. India Fact Sheet. National Family Health Survey-4 2015–16. Ministry of Health and Family Welfare, Government of India. Available from: <http://www.rchiips.org/NFHS/pdf/NFHS4/India.pdf>. [Last accessed on 2018 Jan 18].
4. SRS Bulletin; September 2014. Available from: [http://www.censusindia.gov.in/vital\\_statistics/SRS\\_Bulletins/SRS%20Bulletin%20September%202014.pdf](http://www.censusindia.gov.in/vital_statistics/SRS_Bulletins/SRS%20Bulletin%20September%202014.pdf). [Last accessed on 2018 Jan 15].
5. India-Report. National Family Health Survey-4 (2015–16). Ministry of Health and Family Welfare, Government of India. Available from: <http://www.rchiips.org/NFHS/NFHS-4Reports/India.pdf>. [Last accessed on 2018 Jan 18].
6. Voluntary National Review Report on the Implementation of Sustainable Development Goals. United Nations High Level Political Forum; 2017. Available from: [http://www.niti.gov.in/writereaddata/files/Final\\_VNR\\_report.pdf](http://www.niti.gov.in/writereaddata/files/Final_VNR_report.pdf). [Last accessed on 2018 Jan 18].
7. State/UT-wise Infant Mortality Rate during 2015–16. National Family Health Survey-4. Open Government Data Platform India. Available from: <https://www.community.data.gov.in/stateut-wise-infant-mortality-rate-during-2015-16/>. [Last accessed on 2018 Jan 15].
8. National Health Policy 2017. Ministry of Health and Family Welfare, Government of India. Available from: <http://www.164.100.158.44/index1.php?lang=1&level=1&sublinkid=6471&lid=4270>. [Last accessed on 2018 Jan 28].
9. Sustainable Development Goals: A Handbook. Available from: [http://www.4dj7dt2ychlw3310xlowzop2.wpengine.netdna-cdn.com/wp-content/uploads/2016/12/SDG\\_Booklet\\_English.pdf](http://www.4dj7dt2ychlw3310xlowzop2.wpengine.netdna-cdn.com/wp-content/uploads/2016/12/SDG_Booklet_English.pdf). [Last accessed on 2018 Jan 20].
10. Sustainable Development Goals: Agenda 2030 A Civil Society Report INDIA 2017. Wada Na Todo Abhiyan. Available from: <http://www.wadanatodo.net/wp-content/uploads/2017/07/Civil-society-Report-on-SDGs.pdf>. [Last accessed on 2018 Jan 20].
11. Mandal BN. Forecasting Sugarcane Production in India with ARIMA Model. Statistics 1-13. Available from: <http://www.interstat.statjournals.net/YEAR/2005/articles/0510002.pdf>. [Last accessed 2018 Jan 03].
12. Infant Mortality Rate of India. Open Government Data Platform India. Available from: <https://www.data.gov.in/catalog/infant-mortality-rate-india>. [Last accessed on 2018 Jan 07].
13. Upadhyay R. Step-by-Step Graphic Guide to Forecasting through ARIMA Modeling using R – Manufacturing Case Study. Available from: <http://www.ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arima-modeling-in-r-manufacturing-case-study-example/>. [Last accessed on 2018 Feb 02].
14. Kasyoki A. Simple steps for fitting ARIMA model to time series data for forecasting using R. Int J Sci Res 2015;4:318-21.