International Journal of Statistics and Applied Mathematics 2023; SP-8(5): 1009-1018

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; SP-8(5): 1009-1018

https://www.mathsjournal.com
Received: 02-08-2023
Accepted: 03-09-2023

Bheemanna
Research Scholar, Department of Statistics, Karnataka University's Karnataka Arts College, Dharwad, Karnataka, India

MN Megeri
Professor, Department of Statistics, Karnataka University's Karnataka Arts College, Dharwad, Karnataka, India

Huchesh H Budihal
Assistant Professor (Temp), Department of Agricultural Statistics, University of Agricultural Sciences, Dharwad, Karnataka, India

Corresponding Author:
MN Megeri
Professor, Department of Statistics, Karnataka University's Karnataka Arts College, Dharwad, Karnataka, India

Forecast of demographic variables using the ARIMA Model in India

Bheemanna, MN Megeri and Huchesh H Budihal

DOI: https://doi.org/10.22271/maths.2023.v8.i5Sn.1322

Abstract
Life expectancy at birth reflects the overall mortality level of a population. Three important demographic indicators—Life Expectancy at Birth, Death Rate, and Infant Mortality Rate (IMR)—were examined for 1971 to 2020 and projected in this study for the years 2021 to 2030. The projections that went along with the forecasts were created using statistical models. The Auto-regressive Integrated Moving Averages (ARIMA) is discussed in this article for the selected demographic variables. We also used the AIC and BIC to find the best-fitting ARIMA model for the data and provide the life expectancy at birth, Death rate and IMR forecasts for future years. The ARIMA (0, 2, 1), (3, 1, 0), and (3, 1, 0) models were also found to be the best-fitting models for India's Life expectancy at birth, Death rate and IMR respectively. The life expectancy at birth is best fits compared to other variables based on the MAPE values.

Keywords: Life expectancy, Death rate, IMR, ARIMA, AIC, BIC, MAPE

Introduction

In 2019, the average life expectancy at birth was 72.8 years, an increase of about 9 years since 1990. Current predictions indicate that additional advancements in survival will lead to an average life expectancy of around 77.2 years worldwide in 2050. According to World Population Prospects (2022), the number of fatalities worldwide is predicted to rise during the following decades, from 67 million in 2022 to 92 million in 2050, as a result of the world population's rapid growth and aging.

Box and Jenkins (2015) [4] created the Autoregressive Integrated Moving Average (ARIMA) model for univariate forecasting. In this approach, a time series is defined in terms of its present and past lagged values of a white noise error term (the moving average component), as well as its previous lagged values (the autoregressive component). According to Peter Pflaumer (1992) [12], the Box-Jenkins technique is equivalent to a straightforward trend model for creating long-term population estimates in the United States, and it hasn't underperformed when compared to more intricate demographic models. There have been various time-series analyses of the German unemployment rate, according to Michael Funke (1992) [9]. In order to fit and predict the German unemployment rate, the multiple-impact ARIMA model outperforms the univariate ARIMA model. According to du Preez and Witt (2003) [6], ARIMA models perform better in forecasting than multivariate models. In the case of Pakistan, Zakria and Muhammad (2009) [14] used Box-Jenkins ARIMA models to study population dynamics using a data set spanning 1951 to 2007, coming to the conclusion that the ARIMA (1, 2, 0) model was the most accurate.

Ayele and Zewdie (2017) [1] used yearly data from 1961 to 2009 using Box-Jenkins ARIMA models to explore the size and distribution of the human population in Ethiopia. They came to the conclusion that the ARIMA (2, 1, 2) model was the most effective model for population prediction and forecasting in Ethiopia. In Nyoni. T (2019) [11], we model and estimate the total population for the next three decades using the Box-Jenkins ARIMA approach, utilizing yearly time series data on the total population in India from 1960 to 2017. According to Najla Salah Madlul *et al.*'s (2020) [10] statistical evaluations of the accuracy of prediction models, the findings of the model ARIMA (1.0.1) are the most practical for predicting the development of wheat output in Iraq until 2028.

The best forecasting model is the NNAR (4, 4, 4, 4, 4, 4), (4, 4), (4, 4), (11, 6), (10, 6, 10, 6, 10, 6), (5, 6, 10, 6), (6, 4) models by Bheemanna and Megeri M N (2023) [2]. This Neural Network Auto-regression (NNAR) model is used to predict demographic and economic factors for the next ten years. The study reveals that except for GDP all the selected variables were AM model fitted well and comparison shows rural population is best fitted as compared to the entire demographic and economic variable based on the MAPE by using Fuzzy time series model by Megeri M N and Bheemanna (2023) [2].

In this study, an attempt is made to study the model and forecast of important demographic variables using the ARIMA model. The first section contains the introduction, the second contains the methods and materials, the third contains a summary and discussion, and the final contains the conclusions.

Methods and Materials

The Sample Registration System (SRS) publication was used to obtain information on the life expectancy at birth, the death rate, and the infant mortality rate (IMR). In order to explore the forecasting of this variable, the autoregressive models were used to survey the sample and predict the variable's life expectancy at birth, death rate, and IMR from 1971 to 2020. Time series data are analyzed using the statistical models of autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), and autoregressive integrated moving average (ARIMA), which combine the two.

The Box-Jenkins approach has become more well-liked. This approach is based on the output of models like moving average (MA), integrated (I), and autoregressive (AR). Databases that are not stationary must be differentiated until they are. After that, we checked that our data was stationary using the ADF and PP tests before running Box-Jenkins on it. Models are found using AIC and BIC. After estimating the model's parameters using Maximum Likelihood Estimation, diagnostic tests should be carried out using R programming and the Box-Ljung test. Forecasts are produced if the model is acceptable; if not, it is important to look into other models. MAPE stands for Mean Absolute Percentage Error.

Theoretical Background

Time series analysis can produce a reasonably accurate short-term forecast given a big enough amount of data, as Granger and Newbold (1986) [7] showed. The ARIMA model is versatile and popular in univariate time series analysis. The ARIMA model combines three processes: the moving average (MA), the autoregressive (AR), and the differencing process.

Autoregressive (AR) model

The Autoregressive model of order p, AR (p), can be expressed as:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon_t$$

Where β_0 is constant and $\beta_1, \beta_2, \ldots, \beta_p$ are the coefficients

In this model, all previous values can have additive effects on this level X_t, so ε_t a random error process consisting of independently and identically distributed (iid) random variables with E (ε_t) = 0 and V (ε_t) = σ^2 i.e. $\varepsilon_t \sim N(0, \sigma^2)$.

Moving Average (MA) model: The Moving Average model of order q, MA (q), can be expressed as:

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

Where $\theta_1, \theta_2, \ldots, \theta_q$ are coefficients

The independent variables in this model are the past errors as ε_t a random error process.

Autoregressive Moving Average (ARMA) model

A time series $\{Y_t\}$ is said to follow an Autoregressive Moving Average model of order p and q, can be expressed as

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

Autoregressive Integrated Moving Average (ARIMA) model

By allowing the data series to differ from one another, the ARIMA model can be extended to non-stationary series. The ARIMA (p, d, q) is the name of the generic non-seasonal model. In this model, p is the order of Autoregressive, d is the differencing, and q is the order Moving Average order.

$$\left(1 - \sum_{i=1}^{p'-d} \beta_i L^i\right)(1-L)^d Y_t = \beta_0 + \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)\varepsilon_t$$

Where p $= p' - d$

In addition to the fact that the time series literature provides a range of methods for model evaluation, we decided to employ autocorrelation function (ACF) and partial autocorrelation function (PACF) plots in our example. This approach was created by Box and Jenkins (1976) in their pioneering study on the ARIMA model. The correlation between two successive observations in a time series can be used to define ACF. It determines the linear relationship between two observations—one made at time t and the other at, say, a distance of k—made at different points in space. The PACF varies from the ACF in that it ignores the impact of other intermediate data in between and solely assesses the correlation between the current and previous observations of a time series at a distance of k (say).

Akaike Information Criterion (AIC) and Bayesian Information Criteria (BIC)

The Akaike Information Criterion (AIC) and Bayesian Information Criteria (BIC), which are the two most widely used model selection criteria, are each defined as follows:

$$AIC = 2m - 2\ln(L)$$

$$BIC = \ln(n)m - 2\ln(L)$$

Where L is the model's maximum likelihood function value, m is the number of parameters estimated by the model, and n is the number of observations (sample size).

Box-Ljung test

A diagnostic technique for evaluating the lack of fit in a time series model is the Box-Ljung test (1978) [5]. A time series' residuals are subjected to the test after an ARMA (p, q) model has been fitted to the data. The analysis determines if the residuals exhibit autocorrelation. If the autocorrelations are very modest, then the model does not show a substantial lack of fit, we find.

In general, the Box-Ljung test is defined as follows:

H_0: The model does not exhibit a lack of fit.

H_1: The model exhibits a lack of fit.

Test Statistic: Given a time series Y of length n, the test statistic is defined as:

$$Q = n(n+1)\sum_{k=1}^{m}\frac{\hat{\rho}^2}{n-k}$$

Where $\hat{\rho}^2$ the series estimated auto correlation at lag k, and m denotes the number of lags being tested for desired level of significance α.

Critical Region: The Box-Ljung test rejects the null hypothesis (indicating that the model has significant lack of fit)

If $Q > \chi^2_{1-\alpha,h}$

Where $\chi^2_{1-\alpha,h}$ is the chi-square distribution table value with h degrees of freedom and significance level α. Since the test is run on residuals, the degrees of freedom must take into account the estimated model parameters for h to equal m-p-q, where p and q represent the number of parameters from the ARMA (p, q) model that were used to fit the data.

The Mean Absolute Percentage Error (MAPE)

In the present study, the Mean Absolute Percentage Error (MAPE) between the predicted value and the actual value is also assessed. For calculating mean absolute percentage error (MAPE), use the equation

$$MAPE = \frac{1}{n}\left(\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|\right)*100$$

Where A_t denotes the actual value, F_t is the predicted value, and n denotes the number of fitted values.

Results and Discussion

The variables considered for analysis in this study include Life expectancy at birth, Death rate, and IMR variables from 1971 to 2020. The below Fig. 1, demonstrates the life expectancy at birth, which has increased by 20 years in the last 50 years, initially from 48.8 years to the current 70 years. Additionally, the death rate and IMR have both declined during the past 50 years.

Stationary of the series: One can refer to a statistical series as stationary if its mean, variance, and covariance do not fluctuate over time or are not time-dependent.

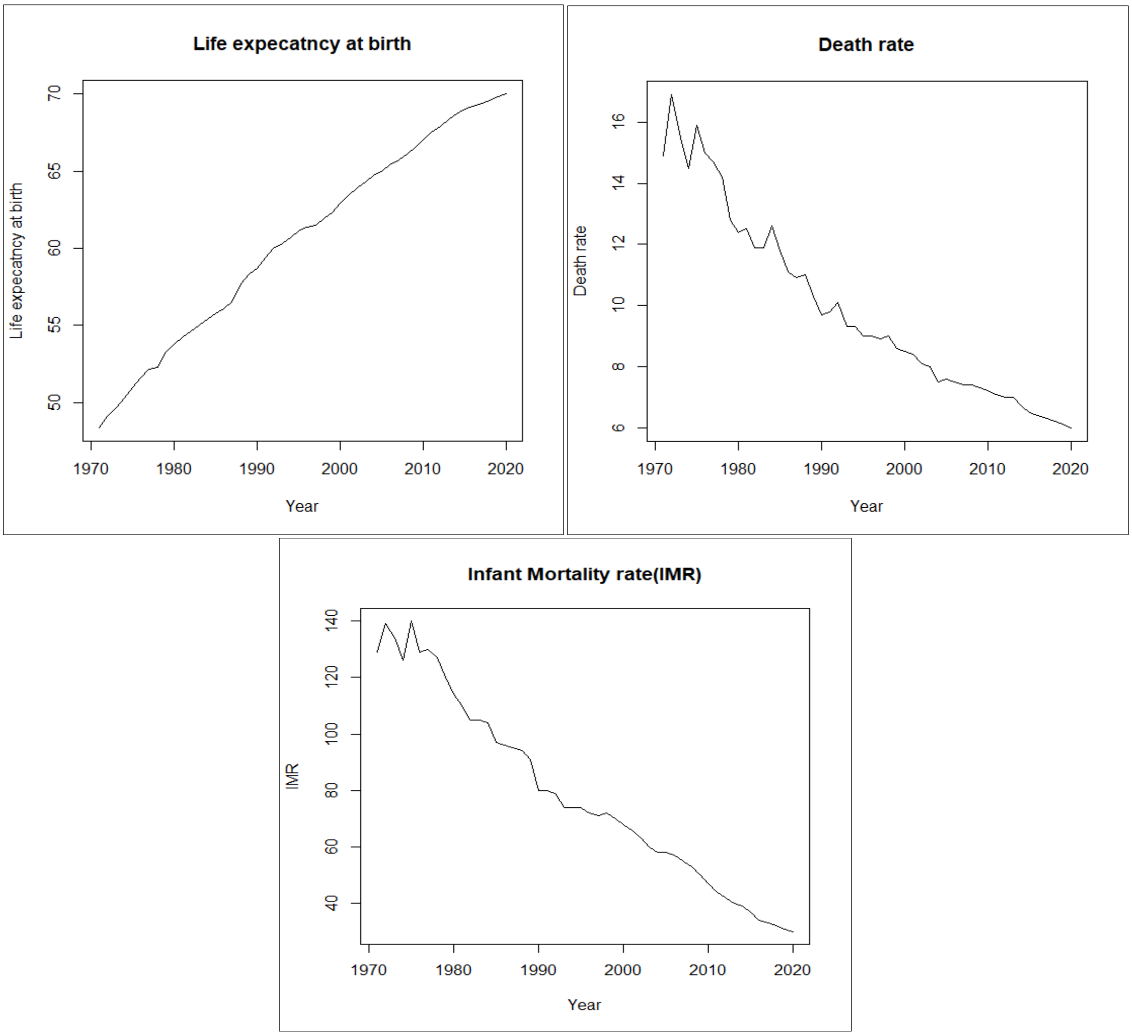


Fig 1: The observed data for life expectancy at birth, Death rate and IMR variables from 1971 to 2020.

When applying ARIMA modeling, the potential for stationary is visually assessed in Figure 1. The time series' upward trends in movement suggest that some of its features may be non-stationary. Particularly, the series displays distinctive average values and fluctuations in various sample sub-periods.

The above-mentioned intuitive conclusion is put to the test using a correlogram of the series that shows the autocorrelation function (ACF) for 16 lags for these three variables (Fig. 2). The series has non-stationarity because the correlogram begins with a very high correlation coefficient of 0.99, the ACF plot gradually declines, and the series has a well stated autocorrelation even for several lags.

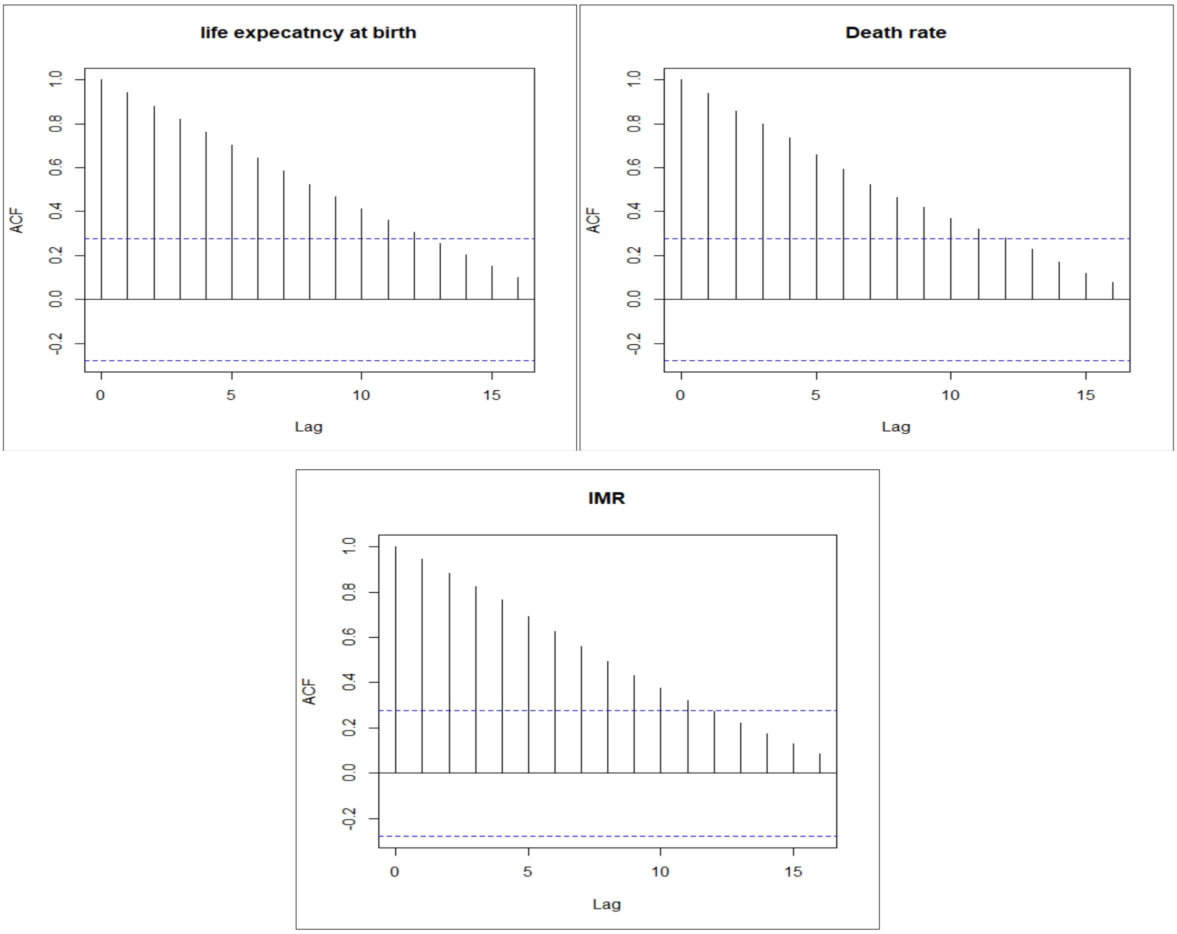


Fig 2: Autocorrelation Function for Life expectancy at birth, Death rate and IMR variables (Actual values)

The study also examines the sample autocorrelation coefficients' statistical significance to determine whether or not they accurately depict the population's actual ACF plot. According to statistical theory, when dealing with a random process, the ACF can be generally characterized by the normal distribution, which has a zero mean and a variance of (1/n), where n is the sample size (Gujarati, D.N., 1995) [8]. The standard error of the ACF can then be calculated as 1/50 = 0.1414. According to the tabular values for the normal distribution, the 95% confidence interval for the ACF is equal to 1.96*0.1414 = 0.277.

If the computed ACF is inside the confidence interval, it cannot be ruled out that the population's true ACF is zero (H0: k = 0). For these three varaibles of the ACF, the first twelve lags are statistically significant (i.e., different from zero), the coefficients from lags 13 to 16 lie within the confidence interval, and then these coefficients are once more different from zero, as shown in Fig. 2. The high number of statistically significant coefficients serves as more evidence of the series' non-stationary character.

Table 1: Augmented Dickey Fuller (ADF)and Phillips-Perron(PP) test for stationarity (Actual values)

Variables	ADF	P values	PP	P values
Life expecatncy at Birth	-0.72901	0.9619	-2.1263	0.9625
Death rate	-1.5786	0.7433	-9.4597	0.5472
IMR	-1.8833	0.6211	-15.119	0.2002

The results of the ADF test for each variable are shown in the first column of Table 1, which shows that the null hypothesis that a unit root exists cannot be rejected. This test indicates that the series is non-stationary. The results are additionally tested using the Phillips Perron (PP) test because the ADF test is known to have a p value higher than the significance level of 0.05.

In all of the variables of the PP test, the p values obtained is greater than the significance level of 0.05, which suggests that the null hypothesis of a unit root cannot be rejected, as shown in the third column of Table 1. As a result, this test also indicates that the series is non-stationary. As previously stated, the Box–Jenkins technique cannot be used if the time series is non-stationary. This means that the series must be transformed in order to become stationary, which is executed by differencing the originaseries (see Table 1).

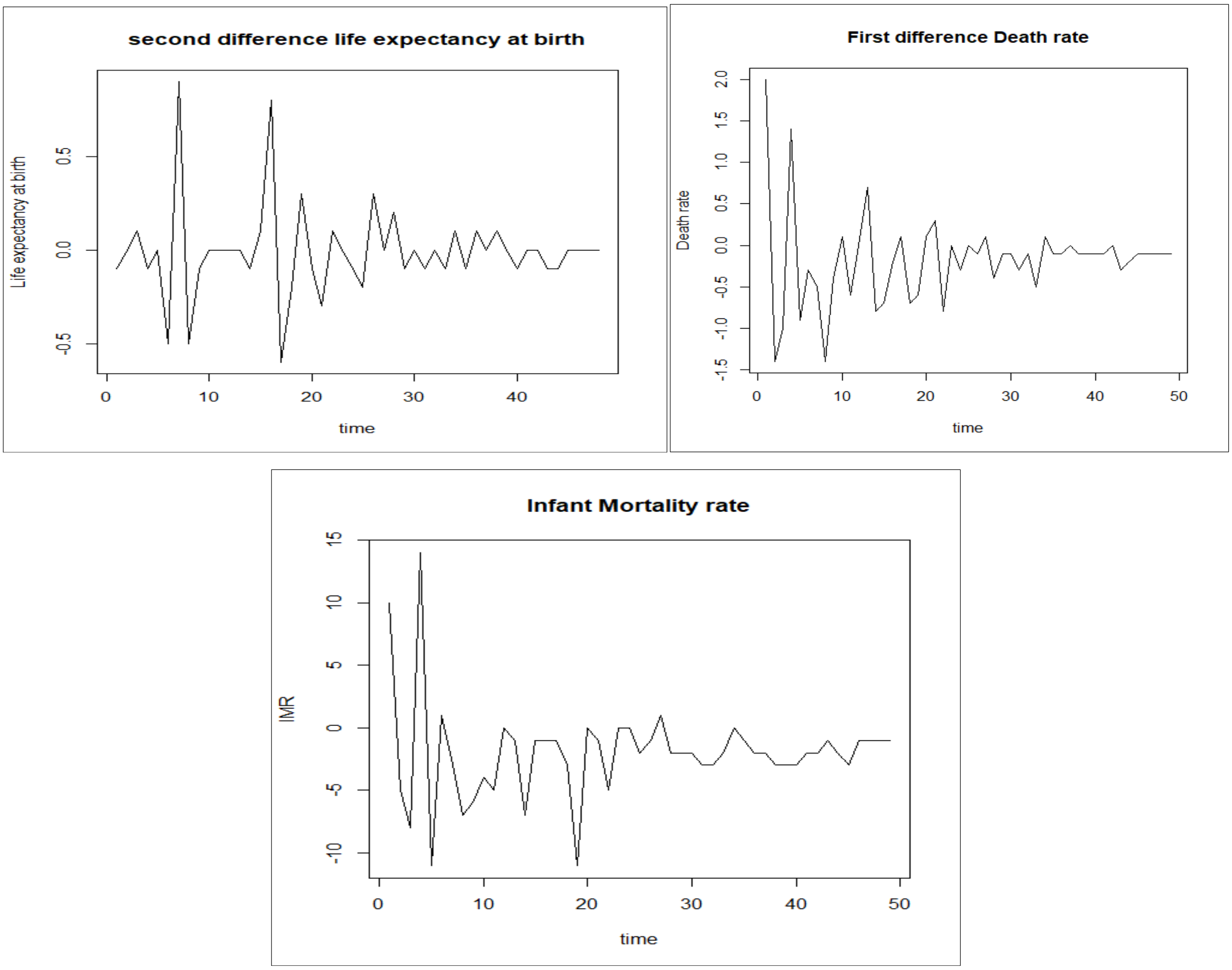
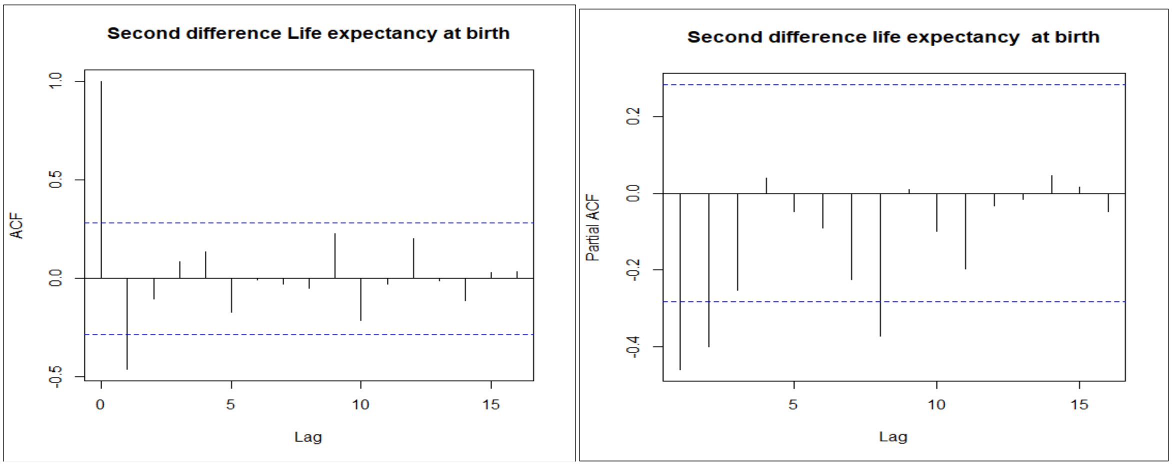


Fig 3: The Second difference data for Life expectancy at birth and first difference data for Death rate and IMR variables.

Fig. 3 illustrates that all the variables of the time series X_t are stationary by a second difference for life expectancy at birth, first difference for Death rate and IMR and do not have a trend, their movements are wave-like.

When the series is second and first differenced, one cannot observe a regular movement of the autocorrelation coefficients, which begin with low values, decreasing quickly to zero, and then, up to lag 16, moving in a wave style. The preceding findings illustrate that stationary can be achieved by second-diffencing for life expectancy at birth and first diffencing for Death and IMR, the original time series. Next, we examine the appearance of the ACF and PACF in the differenced series and compare the two plots in Fig.4 to determine the value of the other two ARIMA model parameters, p and q.



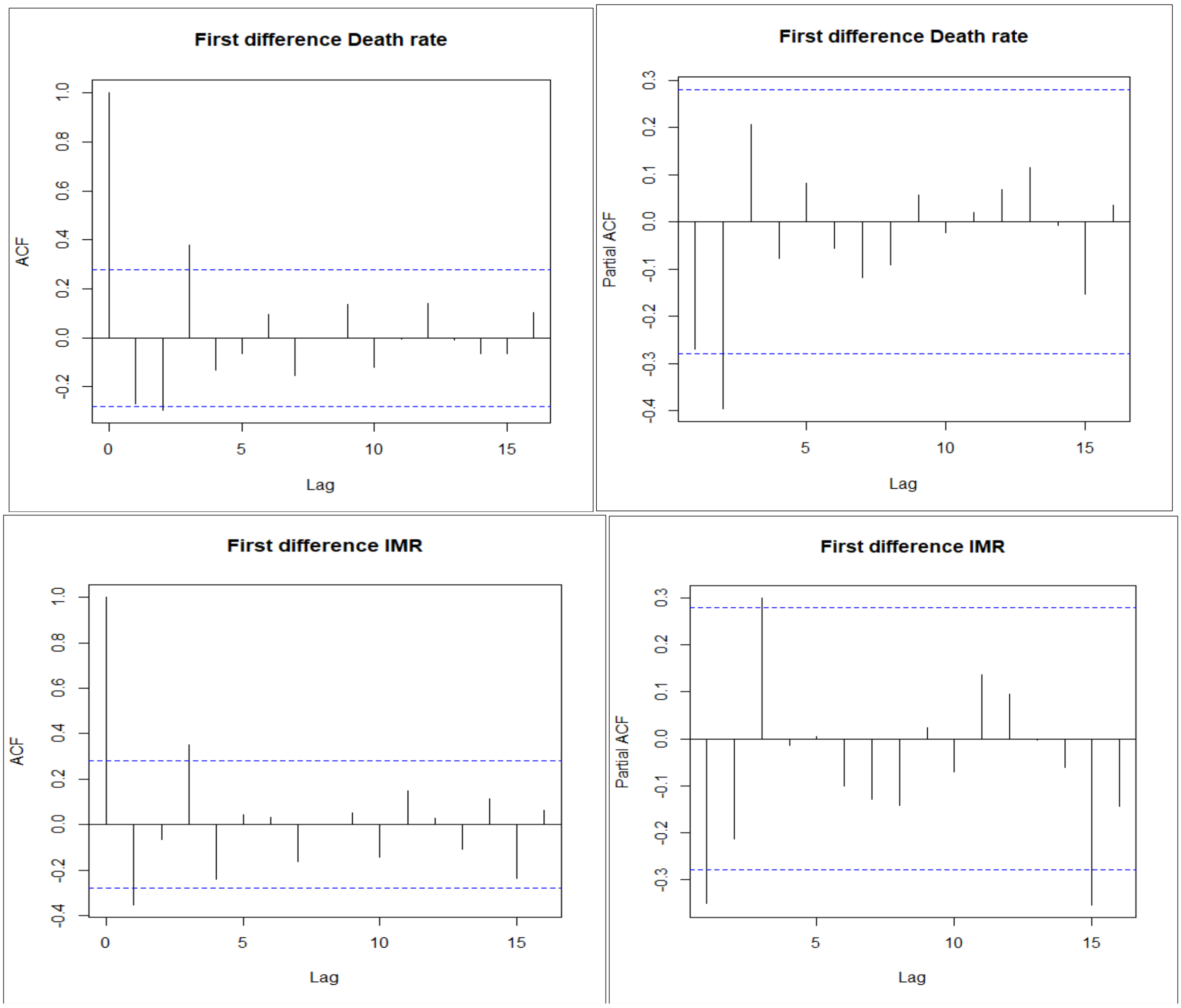


Fig 4: Autocorrelation Function and Partial Autocorrelation Function for Life expectancy at birth (second difference), Death rate and IMR (first difference)

The ADF test and the PP test are used to compare the results. Table 2 shows that the p values less than the significance level of 0.05, implying that the second-differenced series is stationary. d = 2 for life expectancy at birth, the first differenced series is stationary for Death rate and IMR in the ARIMA model. To determine the same the Box–Jenkins methodology can be used.

Table 2: Augmented Dickey Fuller(ADF) and Phillips-Perron (PP) test for stationary (second difference for Life expectancy at birth and fourth difference for Death rate and IMR).

Variables	ADF	P values	PP	P values
Life expecatncy at Birth	-4.2414	0.01	-57.812	0.01
Death rate	-5.0302	0.01	-56.249	0.01
IMR	-4.9184	0.01	-62.077	0.01

Model selection: For model selection the study uses the AIC and BIC, the minimum values of these test gives the model which is best fitted to all selected variables.

The AIC and BIC are then used to find the best ARIMA parameters (p, d, q). The study computed the AIC and BIC using various combinations of (p, d, q). Then, from all the models analyzed, choose the best-fitting ARIMA model with the lowest AIC and BIC.

The ARIMA (0, 2, 1) is the best-fitted model for the life expectancy at birth and the ARIMA (3, 1, 0) is the best-fitted model for Death rate and IMR according to the AIC and BIC (see Table 3). Table 3 shows the ARIMA (0, 2, 1), (3, 1, 0) and (3, 1, 0) parameters estimates. Further, models estimated these coefficients and are presented in Table 4 below.

Table 3: Evaluation of various ARIMA models (without a constant)

Variables	Model	AIC	BIC
Life expectancy at Birth	(0, 2, 1)	-16.7777	-13.04
	(0, 2, 2)	-14.95	-9.33603
	(0, 2, 3)	-12.96	-5.47592
Death rate	(2, 1, 0)	89.59356	95.26902
	(3, 1, 0)	74.64524	82.21252
	(4, 1, 0)	76.59962	86.05872
IMR	(2, 1, 0)	288.6968	294.3722
	(3, 1, 0)	269.6528	277.22
	(4, 1, 0)	271.5975	281.0566

Estimation of Coefficient for selected Model: The estimation of the coefficients from Maximum Likelihood Estimator for these three variables.

The Maximum Likelihood Estimator estimated the modeling results of an ARIMA (0, 2, 1), (3, 1, 0) and (3, 1, 0) process, which are provided in Table 4. The estimated coefficient of MA (1), MA (2), MA (3), MA (4) and MA (5) components for these three are statistically significant at a 1% and 5% level of significance.

Table 4: The parameters estimated by Maximum likelihood Estimator for these three variables

Variables	Model		coefficients	Standard Error	Z value	Pr(>|z|)
Life expectancy at Birth	(0, 2, 1)	MA(1)	-0.8293	0.10046	-8.2548	2.2e-16 ***
Death rate	(3, 1, 0)	AR(1)	-0.097531	0.129535	-0.7529	0.4515
		AR(2)	-0.168646	0.135292	-1.2465	0.2126
		AR(3)	0.635398	0.128268	4.9537	7.282e-07***
IMR	(3, 1, 0)	AR(1)	-0.23689	0.11866	-1.9964	0.04589 *
		AR(2)	0.22579	0.1148	1.9668	0.04920 *
		AR(3)	0.64757	0.11878	5.4518	4.987e-08 ***

Note: - *** 1% and * 5% level significance

Diagnostic check

Diagnostic checking would be required once the models have been calculated to ensure that they are reasonably acceptable and statistically significant for predicting.

The Box-Ljung test is a statistical analysis used to examine if serial correlation or independently spread distribution characterizes the residuals from a time series model. The Box-Ljung test results for the three variables are explained as follows: The life expectancy at birth Box-Ljung test statistic is 17.799 with 15 degrees of freedom.

Table 5: The Box-Ljung test for residuals are Independent

Variables	Statistic test	DF	p value	Chi-square table value
Life expecatncy at Birth	17.799	15	0.2734	19.67514
Death rate	13.585	13	0.4037	22.36203
IMR	19.014	13	0.1227	22.36203

The p-value for this is 0.2734. We cannot decide to reject the null hypothesis since the p-value is higher than the significance level of 0.05. This implies that the life expectancy residuals are probably independent. The death rate has a Box-Ljung test statistic of 13.585 with 13 degrees of freedom. The corresponding p-value is 0.4037. We are unable to reject the null hypothesis since the p-value exceeds the 0.05 criterion of significance. This means that the death rate residuals are likely to be independent. With 13 degrees of freedom, the Infant Mortality Rate (IMR) Box-Ljung test statistic is 19.014. The corresponding p-value is 0.1227. We fail to reject the null hypothesis since the p-value exceeds the 0.05 criterion of significance. As a result, it seems possible that the IMR residuals are independent(see Table 5).

Box-Ljung test statistic for all the variables and the corresponding p values, which is much larger than 0.05 and the Box-Ljung test is less than the Chi-square critical values, Then, we fail to reject the null hypothesis of the test and conclude that our residuals are independent, which indicates that our fitted model is adequate.

Table 6: Life expectancy at birth forecast for the next 30 years using ARIMA model.

	Life expectancy at Birth		
Year	**Forecasted**	**Lower**	**Upper**
2021	70.27	69.89	70.65
2022	70.27	69.9	71.13
2023	70.83	70.05	71.6
2024	71.1	70.14	72.06
2025	71.38	70.23	72.53
2026	71.66	70.31	73
2027	71.93	70.39	73.48
2028	72.21	70.45	73.97
2029	72.49	70.52	74.46
2030	72.76	70.57	74.95

Table 6 shows that India's life expectancy at birth was 70 years in 2020, it is an increase to 71.38 years in 2025 with 1.97 per cent increase from the last five years, and predicted to 72.76 years by 2030. 3.94 per cent increase from the last decade. Over the next 10 years, life expectancy in India has 3.94% increases; the average life expectancy is 72.6 years. The projected increase in life expectancy at birth from 70.27 years in 2021 to 72.76 years in 2030 usually indicates an upward trend. This predicts that over the following decade, health outcomes and living conditions will generally improve. The projected life expectancy steadily rises each year, which points to ongoing advancements in healthcare and medical conditions. For instance, life expectancy has increased by about 2.5 years in the last ten years. Given that more people would be living longer and needing healthcare and retirement benefits, the rising life expectancy suggests a potential strain on healthcare systems and pension schemes. Longer life expectancy creates a higher return on human capital, thereby encouraging more investment in education and health, thus stimulating economic growth and also lowering the mortality rate.

Figure 5 shows that the life expectancy at birth will be slowly increasing due to the death rate and IMR variable slowly decreasing because of the better health care ad hygiene, healthier lifestyles, diet, and improved medical care. We have access to antibiotics and vaccines, clean water, plentiful and more nutritious food, and we know that exercise and smart life choices improve our quantity and quality of life.

Table 7 shows that in 2020, the death rate in India was 6.0 per 1,000 people; it is predicted to fall by 5 per 1,000 people in the last half decade, a 2.83% decrease, and 5 per 1,000 people by last decade, a 3.66% decrease from 2020. A broad decrease trend from 5.96 in 2021 to 5.78 in 2030 may be seen in the predicted death rate. This shows that a general improvement in living standards, healthcare, and public health will reduce mortality. The lower and upper bounds offer a range of potential death rates, reflecting the forecasts' inherent uncertainty.

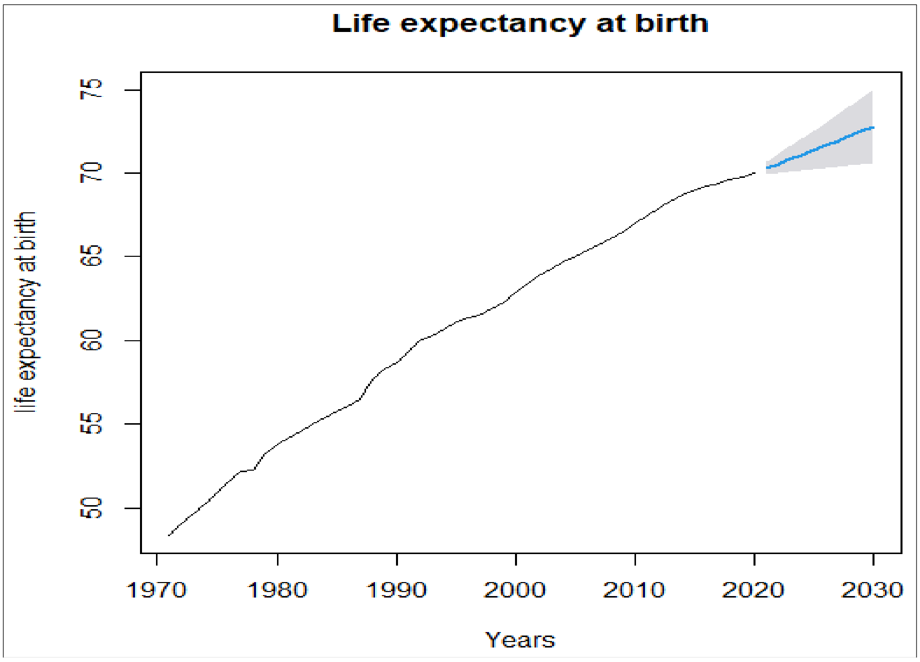


Fig 5: Life expectancy at birth forecast for the next 30 years using ARIMA model

Table 7: Death rate forecast for the next 30 years using ARIMA model.

	Death rate		
Year	**Forecasted value**	**Lower**	**Upper**
2021	5.96	5.04	6.88
2022	5.92	4.68	7.15
2023	5.86	4.45	7.27
2024	5.85	3.94	7.77
2025	5.83	3.57	8.1
2026	5.8	3.32	8.28
2027	5.8	2.94	8.67
2028	5.8	2.62	8.97
2029	5.78	2.39	9.17
2030	5.78	2.08	9.48

As we get closer to 2030, this range gets smaller over time, indicating greater confidence in the projections. The width of the range (difference between upper and lower bounds), similar to the life expectancy analysis, diminishes over time, showing less uncertainty or unpredictability in estimating the death rate. The decreasing death rate suggests improvements in public health, illness prevention, and healthcare practices.

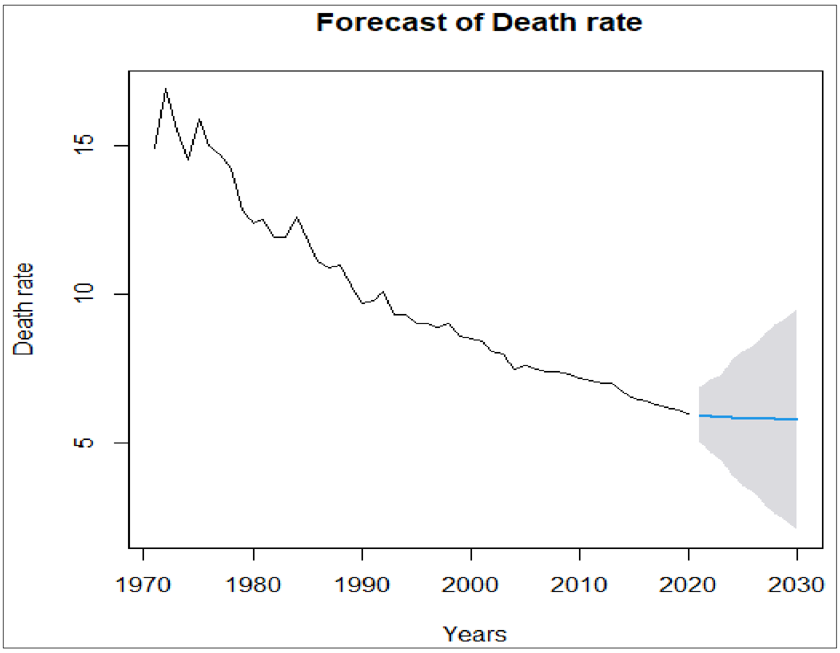


Fig 6: Death rate forecast for the next 30 years using ARIMA model

It is anticipated that mortality reduction efforts will continue to be effective. For the next ten years, death rate will be slowing decreasing that is approximately 4 per cent due to the melioration medical care, female education, and economic growth. When the death rate will be decreases then life expectancy at birth will be increasing, which means there is a correlation between the two variables and there is a negative correlation between the life expectancy at birth and death rate. Figure 7 show that death rate will be slowly deceasing for 10 years due to the improved healthcare and advances in medicine**:** There are several advances in healthcare and medicine that have increased life expectancy. One of the most important is the development of vaccines. Before the development of vaccines, diseases killed millions of people each year. Vaccines have eradicated these diseases in many parts of the world, drastically reducing mortality rates. Late marriage has helped in reducing the death rate of women and children. People are now getting more and more nutritive and balanced diet than before. Government is also paying more attention to it. This has helped in reducing death rate.

Table 8: Infant Mortality rate forecast for the next 30 years using ARIMA model.

	Infant Mortality Rate (IMR)		
Year	**Forecasted**	**Lower**	**Upper**
2021	29.36	22.26	36.09
2022	28.64	20.17	37.1
2023	28.02	17.01	39.02
2024	27.59	12.32	42.86
2025	27.08	9.29	44.87
2026	26.7	5.51	47.89
2027	26.4	1.58	51.22
2028	26.06	-1.58	53.71
2029	25.83	-5.25	56.92
2030	25.61	-8.68	59.9

In India in 2020, the infant mortality rate was 30 deaths per 1,000 children; it is projected to drop by 27 deaths per 1,000 children in the last five years, a 16.6 per cent decrease from 2020, and 25 deaths per 1,000 children by 2030, indicating 33.3% decrease in the previous decade. For the next ten years, IMR will be a decline because of the improvement in IMR has coincided with an improvement in public health spending, female education, and economic growth. A general declining trend from 29.36 in 2021 to 25.61 in 2030 could be seen in the predicted IMR. This shows that public health programs aimed at lowering infant mortality and improving access to healthcare services have improved overall. The lower and upper bounds provide the IMR a range of potential values, reflecting the forecasts' inherent uncertainty. As we get closer to 2030, this range gets smaller over time, indicating greater confidence in the projections. In line with earlier findings, the range's width (the difference between upper and lower bounds) gets less over time, indicating less fluctuation or uncertainty in IMR prediction. The lowering IMR suggests improvements in maternal care, newborn healthcare, immunization campaigns, and public health initiatives focusing on baby health.

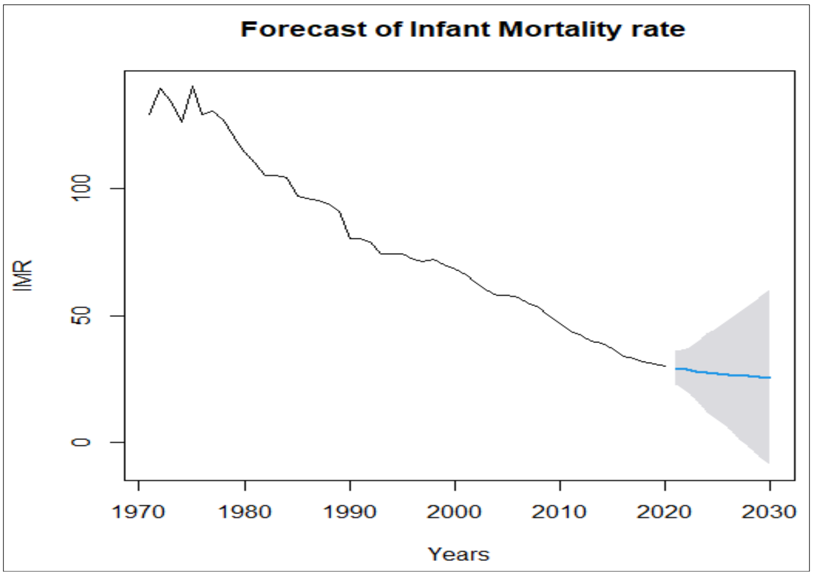


Fig 7: IMR forecast for the next 30 years using ARIMA model

Figure 7 show that the infant mortality rate will be slowly declining for the next ten year. When the IMR has decline then life expectancy at birth has increasing that indicating there is correlation between two variables and here negative correlation between the IMR and life expectancy at birth, which means the one variable is increases then other variable is decreases, that indicting the opposite direction. Literacy among woman is progressing rapidly. Educated women bring up their children with utmost care. This rapidly brings down the infant mortality rate. Working women enjoy better economic status and as such they are healthier.

Forecast Accuracy

The Mean Absolute Percentage Error (MAPE) is used for testing accuracy of the forecasting result.

Table 9 show that the value of MAPE is lower, which is that variable is best fit for ARIAM model. The ARIMA model is a better fit for three the variables. The Life expectancy at birth is best fits model than death rate and IMR.

Table 9: Forecast of the Accuracy

Variables	MAPE
Life expectancy at Birth	0.2159
Death rate	3.059
IMR	2.9278

Among three variables, Life expectancy at birth is better than other variables based on the MAPE values. Except Life expectancy at birth, Infant mortality rate is better than death rate based on the MAPE value.

Conclusion

In this paper, the ARIMA model is applied to analyze and forecast of three important demographic variables such as Life expectancy at birth, Death rate and IMR in India, we are able to gain important knowledge about current trends until the year 2030. Here, we conclude that the ARIMA (0, 2, 1), (3, 1, 0), and (3, 1, 0) are the best-fitting model for the Life expectancy at birth, Death rate and IMR respectively. The ARIMA models are predicting an increase in life expectancy at birth and predict a decrease in Death rate and IMR for the next 30 years. This model appears to represent the data accurately. The value of MAPE of ARIMA model is lower, which shows that the ARIMA model is a better fit. For life expectancy at birth, which MAPE value is lower, this variable is best fit than other variables. The next 10 years life expectancy at birth will be increasing because of the corresponding the Death rate and IMR values are slowly decline, which indicates the health care facility improving and female education developed. For policymakers, healthcare workers, and researchers to plan and make educated decisions on public health programs, retirement planning, healthcare infrastructure, and social support systems for the population, they need the forecasts and their accompanying ranges. Understanding trends in life expectancy might assist in predicting healthcare requirements and societal changes brought on by an aging population.

References

1. Ayele AW, Zewdie MA. Modeling and forecasting Ethiopian human population size and its pattern, International Journal of Social Sciences, Arts and Humanities. 2017;4(3):71-82.
2. Bheemanna, Megeri MN. Prediction of India's demographic and economic variables using the neural network auto-regression model, International Journal of Statistics and Applied Mathematics. 2023;Sp-8(4):574-582.
3. Megeri MN, Bheemanna. Forecasting of demographic and economic variables in India using the Fuzzy time series model, International Journal of Research in engineering Science. 2023;11(2):44-55.
4. Box GEP, Jenkins GM. Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day, Cambridge; c2015.
5. Box G, Ljung G. On a Measure of Lack of Fit in Time Series Models, Biometrika. 1978;65(2):297-303.
6. Du Preez J, Witt SF. Univariate and multivariate time series forecasting: An application to tourism demand, International Journal of Forecasting. 2003;19(3):435-451.
7. Granger CW, Newbold P. Forecasting Economic Time Series, Academic Press; c1986.
8. Gujarati DN. Basic econometrics (2nd ed.), New Delhi: McGraw-Hill; c1995.
9. Funke M. Time-series forecasting of the German unemployment rate, Journal of Forecasting. 1992;11(2):111-125.
10. Madlul NS, Al-Najjar EY, Baker YT, Irhaim FI. Using using the ARIMA models to predict wheat crop production in Iraq, Int. J Agricult. Stat. Sci. 2020;16(1):121-127.
11. Nyoni T. Predicting total population in India: A Box-Jenkins ARIMA approach, University of Munich Library – Munich Personal RePEc Archive (MPRA), Paper No. 92436; c2019.
12. Pflaumer P. Forecasting US population totals with the Box-Jenkins approach. International Journal of Forecasting. 1992;8(3):329-338.
13. World Population Prospects. United Nations New York; c2022. https://population.un.org
14. Zakria M, Muhammad F. Forecasting the population of Pakistan using ARIMA models, Pakistan Journal of Agricultural Sciences. 2009;46(3):214-223.