Statistics and Applications {ISSN 2454-7395 (online)}
Volume 19, No. 2, 2021(New Series), pp 101 –114

Forecasting Infant Mortality Rate of India Using ARIMA Model: A Comparison of Bayesian and Classical Approaches

Manika Agarwal, Praveen Kumar Tripathi and Sarla Pareek
Department of Mathematics and Statistics
Banasthali Vidyapith, Rajasthan-304 022, India

Received: 09 July 2020; Revised: 13 August 2020; Accepted: 09 October 2020

Abstract

This paper attempts to analyze the general autoregressive integrated moving average model under the classical and Bayesian paradigms. The paper aims to forecast the infant mortality rate of India under the two setups. A real data set is first examined for the presence of stationarity and is achieved by performing logarithmic scale transformation and then differencing it twice. After achieving stationarity, the most appropriate model is selected among the various competing models by using Akaike's information criterion and Bayesian information criterion. The selected model is analysed and the results in classical framework are obtained on the basis of maximum likelihood estimators. A complete Bayesian analysis is performed by using vague priors for the parameters and posterior inferences are drawn using Markov chain Monte Carlo simulation technique. The retrospective as well as prospective predictions are obtained, under the two paradigms, for infant mortality rate data and the results are, in general, found to be satisfactory.

Key words: Autoregressive integrated moving average model; Infant mortality rate; Stationarity; Akaike's information criterion; Bayesian information criterion; Markov chain Monte Carlo simulation.

AMS Subject Classifications: 37M10, 62F15, 65C05

1. Introduction

Forecasting demographic characteristics like fertility, morbidity, mortality, etc., is an important facet for the socio-economic planners as it facilitates them to analyze and regulate policies for the betterment of the human population. To forecast such characteristics require an appropriate model building so that a reliable result can be obtained. In this paper, we attempted to predict the infant mortality rate (IMR) of India in classical as well as in Bayesian paradigms. Truly speaking, IMR represents the number of deaths of children under one year of age per thousand live births. Being a vital demographic characteristic, IMR affects the population structure of a country and the projection of human population as well. Cruciality of the IMR data enforce us to model and forecast such a salient characteristic with utmost care.

Corresponding Author: Praveen Kumar Tripathi
Email: pkt2088@gmail.com

In the past few decades, a deep review of literature shows a remarkable contribution of the researchers to model and forecast mortality (see, for example, Keyfitz (1982), Pollard (1987), McNown and Rogers (1989), Lee and Carter (1992), etc.). More specifically, McNown and Rogers (1989) used a kind of parametrization of time series model to forecast mortality. Later on, McNown and Rogers (1992) employed the use of time series methods to forecast cause specific mortality. In their pioneering work Lee and Carter (1992) have proposed a probabilistic approach to model the age-specific mortality and made a long term forecast using time series methodology. Interestingly, Carter (1996) examined the stability of the 'Lee-Carter method' for structural change in a time series and made a comparison with the 'Box-Jenkins methodology' of autoregressive integrated moving average (ARIMA) process. Tuljapurkar and Boe (1998) have critically examined the change of mortality pattern and its forecasting. One should refer to Booth (2006) for a deep review of the methodologies to model and forecast the demographic components.

The use of Bayesian methodology is no more exception in time series analysis, specifically with demographic characteristics. Some of the recent works include Pedroza (2006), Reichmuth and Sarferaz (2008), Alkema and Ann (2011), Tripathi *et al.* (2018) among others. Pedroza (2006) applied a Bayesian approach in Lee-Carter model to forecast the mortality rates. Reichmuth and Sarferaz (2008) have reanalyzed the Lee-Carter model in Bayesian paradigm using the latent variable approach. Alkema and Ann (2011) used a hierarchical time series model, in Bayesian paradigm, to estimate the under-five mortality rate. Recently, Tripathi *et al.* (2018) used ARIMA model to predict the total fertility rate (TFR) of India using classical and the Bayesian approaches.

In this paper, we have applied the methodology discussed by Tripathi *et al.* (2018) for the time series based on IMR of India. In his classical work on IMR time series, Bishai (1995) has explained the issues of non-stationarity and co-integration of IMR data with the other socio-economic variables. In another study on IMR, Kurniasih *et al.* (2018) has discussed about the different methods of forecasting and their relative comparison. To forecast IMR time series data is always crucial for the view point of demographic planning and, hence, for the strategic development of the nations like India. With the same very spirit, we attempted to forecast IMR of India using ARIMA model.

Let $\{y_t\}$; $t = 1, 2, ..., T$, be a sequence of time series observations and $\{\epsilon_t\}$ is a sequence of independently and identically distributed (*iid*) error terms following normal distribution each with mean zero and a constant variance, say, σ^2, then the general form of the autoregressive moving average (ARMA) model of order (p,q) is given by:

$$y_t = \theta_0 + \sum_{i=1}^{p} \theta_i y_{t-i} + \sum_{j=1}^{q} \phi_j \epsilon_{t-j} + \epsilon_t, \tag{1}$$

where θ_0 represents the intercept term and θ_i's, ϕ_j's are the autoregressive (AR) and moving average (MA) coefficients respectively.

One can further introduce a generalization of ARMA models by taking the difference of a suitable order, say, d, of the original series y_t. This new generalization is known as the integrated form of ARMA model and is given by:

$$w_t = \theta_0 + \sum_{i=1}^{p} \theta_i w_{t-i} + \sum_{j=1}^{q} \phi_j \epsilon_{t-j} + \epsilon_t, \tag{2}$$

where w_t represents the d^{th} difference of y_t. Particularly, in ARIMA model, the order of differencing decides the level of stationarity of the time series. For more details one may refer to Box *et al.* (2015).

Although, there are a number of methods available in time series literature for the model assessment, we have adopted the techniques of autocorrelation function (ACF) and partial autocorrelation function (PACF) plots in our case. This technique is proposed by Box and Jenkins (1970) in their pioneering work on ARIMA model. Truly speaking, ACF can be defined as the correlation between the two observations in a time series. It measures the linear relationship between an observation at time t and the observation at some k (say) distance apart. Slightly different from ACF, the PACF measures the degree of association between the current and a previous observation, at a distance k (say), of a time series only after removing the effects of other intermediate observations in between.

This paper proceeds as follows. The next section elaborates the data structure those based on IMR of India and the model identification on the basis of ACF and PACF. The two model selection criteria, that is, Akaike's information criterion (AIC) and Bayesian information criterion (BIC) are also being discussed for choosing an appropriate model. The section finally ends with the numerical illustration of selected ARIMA model including both retrospective as well as prospective predictions. Section 3 explains the necessary priors setup and algorithm to draw the inferences under Bayesian paradigm. A detailed implementation of Markov chain Monte Carlo (MCMC) procedure, using Gibbs sampler and Metropolis algorithm, is also being discussed in a separate subsection. This section ends with the retrospective and prospective predictions for IMR data. The last section provides with a brief summary of the work done that concludes the whole paper.

2. Data Structure and Model Selection Criterion

We considered a real data set on IMR of India over the period of 48 years from 1971 to 2018. The data set, in the form of time series, has been collected from the SRS bulletin, Registrar General of India (see, `https://data.gov.in/resources/time-series-data-crude-death-rate-cdr-and-infant-mortality-rate-imr-srs-1971-2016`) and is framed in Table 1. To see the movement of time series, we have plotted the original data set in Figure 1. It is quite evident from the plot that the observed data on IMR shows a non-stationary pattern as it shows a consistent decline over the years. One has to ensure that the time series must achieve the stationarity to get a reliable result and for the further analyses.

To achieve the stationarity, we exercised the logarithmic scale transformation of the original series and then differenced the transformed series twice. The resultant time series, then, plotted for the different years as shown in Figure 2. The plot (Figure 2) confirms the stationarity of the time series as the mean level is constant over the years. To further strengthened our conclusion we shall provide some numerical evidences, for the differenced data set, those based on Augmented Dickey-Fuller (ADF) test and Kwaitkwoski-Phillips-Schmidt-Shin (KPSS) test. In case of ADF test the p-value is found to be 0.01 (less than 0.05) that rejects the null hypothesis. Truly speaking, the ADF test assumes the null hypothesis that a unit root is present in the AR process. KPSS test, on the other hand, assumes the null hypothesis that the process is stationary. The p-value, in KPSS test, is found to be 0.1 (greater than 0.05) that accepts the null hypothesis. On the basis of p-values, in the two tests, one may conclude that the data is stationary in nature and the values of the test statistics are not so significant which are calculated as –5.07 for ADF test and 0.06 for KPSS test, at 5% level of significance. It is important to mention here that an AR(p) model is said to be stationary if there is no unit root present in the process, that is, if all the roots of the characteristic polynomial lie outside the unit circle (see, for example, Tripathi *et al.* (2017)).

Table 1: IMR of India from 1971 to 2018 (from left to right)

129	139	134	126	140	129	130	127	120
114	110	105	105	104	97	96	95	94
91	80	80	79	74	74	74	72	71
72	70	68	66	63	60	58	58	57
55	53	50	47	44	42	40	39	37
34	33	32						

Source: SRS bulletin, Registrar General of India

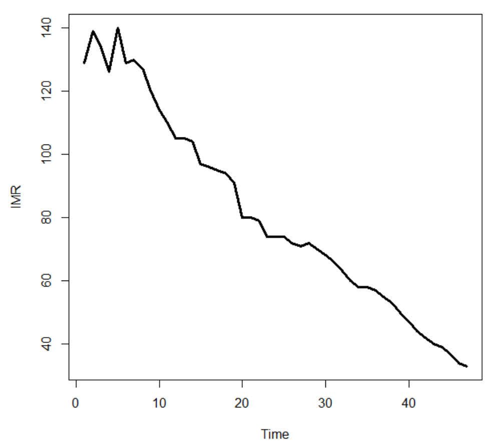

Figure 1: Time series plot for IMR data of India from 1971 to 2018

Once stationarity is achieved, we shall look forward for an appropriate model to get a reliable forecast. For this purpose, we start with the popular 'Box-Jenkins methodology' for the identification of order of ARIMA model, that is, p, d and q (see, for example, Box *et al.* (2015)). To estimate p and q we have plotted the values of ACF and PACF, respectively, against the lag values (see Figure 3). Following Box *et al.* (2015), we can observe from the Figure 3 that AR(2) and MA(1) are the most suitable choices for the components of ARIMA model. Since we have applied the second difference of the data set, therefore, the order of differencing d is fixed as 2. Consequently, a conclusive ARIMA model can be easily assessed as ARIMA(2,2,1). For more details of the procedure one may refer to Tripathi *et al.* (2018).

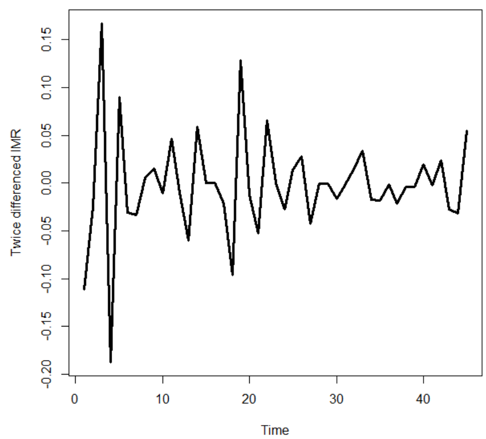

Figure 2: Time series plot for twice differenced transformed IMR data of India

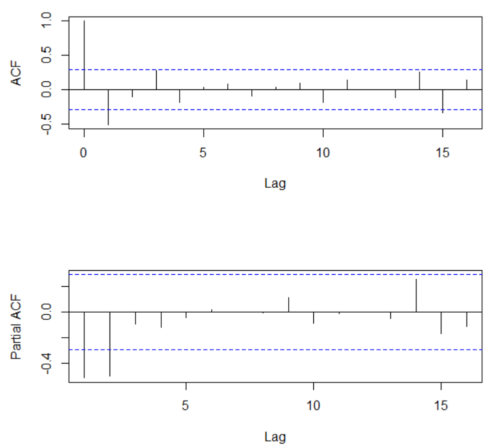

Figure 3: ACF and PACF plots for twice differenced IMR data

In order to avoid any fallacious conclusion, just on the basis of a tentative model assessed graphically, we shall consider some other nearby candidate models and a suitable model will be selected on the basis of some numerical findings. For this purpose, we used the two well known model selection criteria in statistics, namely; AIC and BIC (see, for example, Akaike (1974) and Schwarz (1978)). The two model selection criteria can be defined as below;

$$\text{AIC} = -2\,\log\hat{L} + 2k, \tag{3}$$

and

$$\text{BIC} = -2\,\log\hat{L} + k\,\log(T - p), \tag{4}$$

where $\hat{L}$ is the maximized likelihood function and k represents the number of parameters in the concerned model. These two criteria possess a common characteristic that they more disagree with the model that contains large number of parameters and, hence, increase the complexity of a model. The model corresponding to the least value of AIC (BIC) is considered to be a good model and can be chosen for the purpose of further analysis. We will not go into the details of these criteria due to space restriction, but the interested candidate may refer to Ghosh *et al.* (2007) for more information. Moreover, to proceed further we have to formulate the likelihood function of the candidate models which can be accomplished by the general form of the likelihood of the ARIMA model.

Likelihood function of ARIMA model (2) can be approximately written by using the conditional density of the differenced observations, $\underline{w} : w_1, w_2, ..., w_{T-d}$ (see, for example, Tripathi *et al.* (2018)), which is given by

$$f(w_t|w_{t-1}, w_{t-2}, ..., w_{t-p}; \theta_0, \Theta, \Phi, \sigma^2) \propto \left(\frac{1}{\sigma^2}\right) \exp\left(-\frac{1}{2\sigma^2}(w_t - \theta_0 - \sum_{i=1}^{p} \theta_i w_{t-i} - \sum_{j=1}^{q} \phi_j \epsilon_{t-j})^2\right) \tag{5}$$

Now, we can write the approximate likelihood function of model (2), using (5), as;

$$L(\theta_0, \Theta, \Phi, \sigma^2|\underline{w}) \propto \left(\frac{1}{\sigma^2}\right)^{(T-d-p)/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p+1}^{T-d}(w_t - \theta_0 - \sum_{i=1}^{p} \theta_i w_{t-i} - \sum_{j=1}^{q} \phi_j \epsilon_{t-j})^2\right), \tag{6}$$

where $\Theta = (\theta_1, \theta_2, ..., \theta_p)$ and $\Phi = (\phi_1, \phi_2, ..., \phi_q)$.

2.1. Model selection and prediction

As discussed in Section 2, we considered some nearby competing ARIMA models to validate our procedure. The competing models are; ARIMA(0, 2, 1), ARIMA(0, 2, 2), ARIMA(1, 2, 0), ARIMA(2, 2, 0), ARIMA(1, 2, 1), ARIMA(1, 2, 2) and ARIMA(2, 2, 2). By doing some small mathematical corrections in the expression (6), one can easily obtain the approximate maximum likelihood (ML) estimators for the parameters involved in the concerned model. All the competing models along with the ML estimates of their parameters and the values of log-likelihoods are reported in the Table 2. Since, the likelihood functions were not easily tractable, therefore, we have used numerical based approximation to obtain the ML estimates of the parameters.

Although the interpretation of results obtained in Table 2 is quite obvious still we shall highlight a few of them for the flow of analysis. The impact of intercept term on IMR data is not so considerable in all the considered models. Also, one may observe that the stationarized form of the data set shows less variability for error terms. Therefore, the fluctuations in the random component can be assumed to be constant over a period of time. We shall next proceed with the model selection for the observed stationarized form of the data.

The values of AIC and BIC, for each competing model, can be easily calculated by using the formulae (3) and (4) and the same are reported in Table 3. It is quite evident that both the criteria give their least value corresponding to ARIMA(2, 2, 0) that can be considered as the most appropriate model among others. Undoubtedly, a graphical assessment is always striking and reliable source of information still, we can not ignore the possibilities of numerical evidences which are more appealing as they consider any kind of loss due to fitting as well as complexity both. Certainly, we shall consider ARIMA(2, 2, 0) model for the further analyses and prediction of IMR of India.

Table 2: Classical estimates of the competing ARIMA models

Model	Parameter	ML estimate	log $\hat{L}$	Model	Parameter	ML estimate	log $\hat{L}$
	θ_0	0.0005			θ_0	−0.0010	
ARIMA(0, 2, 1)	ϕ_1	−0.9900	−89.4400	ARIMA(1, 2, 1)	θ_1	−0.2200	−96.1500
	σ^2	0.0012			ϕ_1	−1.2200	
					σ^2	0.0008	
	θ_0	−0.0001			θ_0	−0.0005	
ARIMA(0, 2, 2)	ϕ_1	−1.0100	−86.3700	ARIMA(2, 2, 1)	θ_1	−0.5100	−90.3700
	ϕ_2	0.1400			θ_2	−0.5900	
	σ^2	0.0013			ϕ_1	−1.1800	
					σ^2	0.0009	
	θ_0	−0.0010			θ_0	0.0001	
ARIMA(1, 2, 0)	θ_1	−0.5100	−75.5600	ARIMA(1, 2, 2)	θ_1	−1.0900	−88.2300
	σ^2	0.0020			ϕ_1	0.2100	
					ϕ_2	−1.1100	
					σ^2	0.0011	
	θ_0	−0.0012			θ_0	0.0019	
ARIMA(2, 2, 0)	θ_1	−0.9100	−71.8000	ARIMA(2, 2, 2)	θ_1	−1.1500	−79.4000
	θ_2	−0.5800			θ_2	−0.5300	
	σ^2	0.0011			ϕ_1	−0.3400	
					ϕ_2	−1.0000	
					σ^2	0.0015	

Table 3: Values of AIC and BIC for the competing ARIMA models

Model	AIC	BIC
ARIMA(0, 2, 1)	184.89	190.37
ARIMA(0, 2, 2)	180.75	188.07
ARIMA(1, 2, 0)	157.13	162.55
ARIMA(2, 2, 0)	**151.60**	**158.73**
ARIMA(1, 2, 1)	200.31	207.54
ARIMA(2, 2, 1)	190.74	199.66
ARIMA(1, 2, 2)	186.46	195.49
ARIMA(2, 2, 2)	170.80	181.51

Before we extend our study to the Bayesian analysis, let us perform the classical prediction of IMR based on the selected ARIMA(2, 2, 0) model. For this purpose we took only

43 observations out of 48 observations (see Table 1) and left rest of the values as the test sample. We have obtained the necessary classical estimates for the parameters of ARIMA(2, 2, 0) model as discussed in Section 2 and predicted for the next (44^{th}) observation using the ML estimates of the parameters. This predicted observation then forms a sample of size 44 and again the next (45^{th}) observation is predicted by obtaining the corresponding ML estimates using these 44 observations in a similar way. Proceeding in this way, we can predict for all the corresponding values in the test sample.

Theoretically, to predict the future values of the original time series of size T, we have predicted the very next value for the differenced time series (of size $T-2$), that is, w_{T-1}^{th} observation. Next, we shall obtain the future value corresponding to the scaled transformed time series, that is, $\log(\hat{y}_{T+1})$ which can be calculated using the recurrence relationship given below;

$$\log(\hat{y}_{T+1}) = \hat{w}_{T-1} + 2\log(y_T) - \log(y_{T-1}), \tag{7}$$

where $\hat{w}_{T-1}$ is the estimated predictive value corresponding to w_{T-1}^{th} observation obtained by using ARIMA(2, 2, 0) process that can comfortably be obtained from (2) for the error term having the distribution $N(0, \hat{\sigma}^2)$. Hence, the estimated predictive value for the original series of IMR data set can be obtained by performing the inverse logarithmic transformation on $\log(\hat{y}_{T+1})$. To get the estimated predictive intervals, for the corresponding future values of the original series, let us represent the ML estimates of the parameters as $\hat{\sigma}^2, \hat{\theta}_1$ and $\hat{\theta}_2$. We have calculated the predictive intervals for the differenced series $\{w_t\}$ by means of the formula;

$$\hat{w}_{T-1} \pm z_{1-\alpha/2}\sqrt{Var(\hat{w}_{T-1})}, \tag{8}$$

where $z_{1-\alpha/2}$ is the standard normal percentile and $Var(\hat{w}_{T-1}) = \dfrac{\hat{\sigma}^2}{(1-\rho_1\hat{\theta}_1 - \rho_2\hat{\theta}_2)}$ with $\rho_1 = \frac{\hat{\theta}_1}{(1-\hat{\theta}_2)}$ and $\rho_2 = \hat{\theta}_2 + \frac{\hat{\theta}_1^2}{(1-\hat{\theta}_2)}$. It is to be mentioned here that the confidence interval for the scaled transformed series can be obtained by using the expression (7), just by replacing $\hat{w}_{T-1}$ with $L\hat{w}_{T-1}$ and $U\hat{w}_{T-1}$ to get lower and upper limits, respectively. Here $L\hat{w}_{T-1}$ and $U\hat{w}_{T-1}$ are the lower and upper limits of predicted intervals respectively for the differenced data. Finally, the confidence interval for the original series is obtained by using the similar inverse logarithmic transformation. Truly speaking, the expression (8) gives a $100(1-\alpha)\%$ confidence interval. To get 95% confidence interval, one may use the critical value at 0.05 level of significance from the standard normal table.

The 95% predictive intervals for the estimated predictive values, from 2014 to 2018, are given retrospectively in Table 4. We have also calculated the width of predictive intervals (ω) to observe the consistency of the prediction. It can be easily visualized that the estimated predictive values of IMR data are not too far from the true values, also the true values are well within the range of corresponding predictive intervals. The retrospective prediction is quite satisfactory and, hence, we can predict the future values prospectively. For the prospective prediction we have applied the same strategy on the whole data set and have predicted for the next five years. The future predictions, for IMR of India, are reported in the Table 5.

Table 4: Retrospective predictions of IMR from 2014 to 2018 using the ML estimates

Year	True value	Estimated predictive value	95% Estimated predictive interval		ω
2014	39.00	38.08	34.33	42.24	7.91
2015	37.00	38.08	34.33	42.24	7.91
2016	34.00	38.14	32.39	42.31	9.92
2017	33.00	36.72	32.11	40.73	8.62
2018	32.00	37.71	31.00	41.83	10.83

Table 5: Future predictions of IMR for the next 5 years using ML estimates

Year	Estimated predictive value	95% Estimated predictive interval		ω
2019	30.31	27.30	33.65	6.35
2020	30.32	27.30	33.66	6.36
2021	30.36	27.35	33.71	6.36
2022	29.25	26.34	32.47	6.13
2023	30.03	27.05	33.34	6.29

It is to be noted that the future values of IMR go down, with a good consistency, which is a good sign for a developing nation like India. Before we set up a concrete opinion about these predicted values, let us extend this study to the advanced level and perform the Bayesian analysis in the next section.

3. Bayesian Inference

The conditional likelihood function of the selected ARIMA(2,2,0) model for the differenced data is given by

$$f(\underline{w}|\theta_0, \theta_1, \theta_2, \sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{(T-4)/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p+1}^{T-2}(w_t - \theta_0 - \theta_1 w_{t-1} - \theta_2 w_{t-2})^2\right), \quad (9)$$

where $w_t = \Delta^2 log(y_t)$. To perform the Bayesian analysis, a suitable choice of prior distribution is essential. Choosing the prior distributions is a vital aspect in Bayesian paradigm and one can use the information, if any, while selecting a prior distribution. Such priors are called informative priors. In case we have no information, the non-informative priors come into existence. For the present study, we shall consider the non-informative (vague) priors for σ^2, θ_0, θ_1 and θ_2 as we do not have any concrete information (see, for example, Tripathi *et al.* (2017)). The following prior distributions have been considered for the completion of Bayesian analysis;

$$g_1(\sigma^2) \propto \frac{1}{\sigma^2}; \qquad \sigma^2 \geq 0, \quad (10)$$

$$g_2(\theta_0) \propto U[-N_1, N_1]; \qquad N_1 > 0, \quad (11)$$

$$g_3(\theta_1) \propto U[-N_2, N_2]; \qquad N_2 > 0, \tag{12}$$

and

$$g_4(\theta_2) \propto U[-N_3, N_3]; \qquad N_3 > 0, \tag{13}$$

where N_1, N_2 and N_3 are the hyperparameters. One may choose the hyperparameters (large enough) in such a way that the priors remain vague over the range of parameters. We have considered the same set of values with opposite signs for the uniform range though, one can choose the different values which permit the vagueness of prior distributions. Moreover, the prior distribution for the σ^2, in (10), is a type of prior suggested by Jeffrey and is widely used by the researchers (see, for example, Marriott *et al.* (1996), Kleibergen and Hoek (2000) and Tripathi *et al.* (2018) among others).

Next, we shall obtain the joint posterior distribution by updating the prior distributions (from (10) to (13)) with the help of likelihood function (9) and it can be written up to proportionality as;

$$p(\theta_0, \theta_1, \theta_2, \sigma^2|\underline{w}) \propto \left(\frac{1}{\sigma^2}\right)^{(T-2)/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p+1}^{T-2}(w_t - \theta_0 - \theta_1 w_{t-1} - \theta_2 w_{t-2})^2\right) I_{[-N_1,N_1]}(\theta_0)$$
$$I_{[-N_2,N_2]}(\theta_1) I_{[-N_3,N_3]}(\theta_2), \tag{14}$$

where $I(.)$ denotes the usual indicator function that can take only two values; either zero or one. Truly speaking, if a value of the parameter falls in the interval, it will take the value one and zero otherwise.

It has been seen that the form of joint posterior (14) is analytically intractable, one has to go for the sample based approaches to get the required posterior samples from this. We, however, considered an MCMC approach and apply the Gibbs sampler with intermediate Metropolis steps. It has been seen that after a large number of iterations the sequence of parametric values converges in distribution to a random sample taken from the actual posterior distribution. For more details of the procedure one may refer to Gelfand and Smith (1991) and Upadhyay *et al.* (2001) among others. Once the posterior samples of desired size are obtained, the unobserved future value ($\hat{w}_{T-1}$) can be simulated, for each of the posterior samples, from the parent sampling distribution $p(w_{T-1}|\theta_0, \theta_1, \theta_2, \sigma^2, \underline{w})$. It can be easily verified that the predicted observation $\hat{w}_{T-1}$ follows an univariate normal distribution (see, for example, Tripathi *et al.* (2018)).

3.1. Full conditional distributions and MCMC implementation

To proceed for the MCMC implementation, let us calculate the full conditional distribution of each parameter, from the joint posterior (14), up to proportionality as below;

$$p(\theta_0|\theta_1, \theta_2, \sigma^2, \underline{w}) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p+1}^{T-2}(w_t - \theta_0 - \theta_1 w_{t-1} - \theta_2 w_{t-2})^2\right), \tag{15}$$

$$p(\theta_1|\theta_0, \theta_2, \sigma^2, \underline{w}) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p+1}^{T-2}(w_t - \theta_0 - \theta_1 w_{t-1} - \theta_2 w_{t-2})^2\right), \tag{16}$$

$$p(\theta_2|\theta_0, \theta_1, \sigma^2, \underline{w}) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p+1}^{T-2}(w_t - \theta_0 - \theta_1 w_{t-1} - \theta_2 w_{t-2})^2\right), \qquad (17)$$

$$p(\sigma^2|\theta_0, \theta_1, \theta_2, \underline{w}) \propto \left(\frac{1}{\sigma^2}\right)^{(T-2)/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p+1}^{T-2}(w_t - \theta_0 - \theta_1 w_{t-1} - \theta_2 w_{t-2})^2\right). \qquad (18)$$

Among all, the full conditional of σ^2 can be transformed into a gamma distribution by means of the transformation $\tau = 1/\sigma^2$. A simple mathematics shows that τ follows a gamma distribution with shape parameter $(T-4)/2$ and scale parameter $\frac{1}{2}\sum_{t=p+1}^{T-2}(w_t - \theta_0 - \theta_1 w_{t-1} - \theta_2 w_{t-2})^2$. It is also to be noted that the full conditionals (15), (16) and (17) are not easily available in close form and direct simulation is not possible. We, therefore, adopted the Metropolis algorithm to simulate from these full conditionals. To employ the Metropolis algorithm, a univariate normal density is proposed in each case with mean value corresponding to the ML estimate of the respected parameter and standard deviation is taken to be c times the Hessian based approximation at the value of ML estimate. The constant c behaves like a tuning parameter whose value often suggested to lie between 0.5 and 1.0 (see, for example, Tripathi *et al.* (2018)). Thus, we created a single long run of the simulated values from the posterior distribution (14) via the simple implementation of the Gibbs sampler. Posterior estimates were, then, obtained by choosing the posterior samples at a regular gap, after avoiding the initial transient behavior, so that the correlation between them is close to zero. For further reading of the algorithm one may refer to Smith and Roberts (1993), Upadhyay *et al.* (2001) among others.

3.2. Numerical illustration for Bayesian analysis

To illustrate the Bayesian methodology, discussed above, we took the same set of data reported in Table 1. As discussed, we have calculated the ML estimates of the selected ARIMA(2, 2, 0) model to initialize the Markov chain. The values of hyperparameters, N_1, N_2 and N_3, were set to be 100 in each case to maintain the vagueness of the prior distributions. An iterative procedure indicated us to choose the value of tunning parameter c as 0.7. It is important to mention here that these values are chosen to get a good acceptance probability in Metropolis algorithm. For successful implementation of the Gibbs sampler, we have considered a single long run of the chain up to 50K iterations. After avoiding the initial transient behavior of the chain at about 10K iterations, we took a sample of size 1K by maintaining a gap of 40 so that the serial correlation is negligibly small. We have provided the posterior summary, for the differenced data, on the basis of these 1K posterior samples in Table 6.

The results obtained in Table 6 are self explanatory and it reveals the fact that the estimated marginal posterior densities, for all the parameters, exhibit a normal trend, that is, almost symmetrical in nature. Also, as obvious, the posterior modes are close enough to the corresponding ML estimates which might be because of the vague consideration of priors. Although, we are not giving the densities plots for the estimated parameters due to space restriction still most of the inferences can be easily guessed from the table. Moreover, the length of highest posterior density (HPD) intervals tell the accuracy of the posterior estimates.

Table 6: Posterior summary for the parameters of ARIMA(2, 2, 0) model

Parameter	MLE	Posterior Mean	Posterior Median	Posterior Mode	0.95 HPD interval	
θ_0	−0.0012	−0.0034	−0.0035	−0.0035	−0.0143	0.0102
θ_1	−0.9110	−0.7719	−0.7710	−0.7622	−1.0235	−0.5528
θ_2	−0.5814	−0.5163	−0.5185	−0.5344	−0.7667	−0.2674
σ^2	0.0011	0.0016	0.0016	0.0016	0.0009	0.0023

Like the classical analysis, let us now work on the retrospective prediction of IMR in Bayesian framework. Again, we have considered only first 43 observations, out of 48 observations, as the informative data set (see Table 1) and rests are left to see the forecasting performance. We have applied the same strategy, as in classical prediction, to predict in Bayesian context. It is to be noted that the whole Bayesian analysis is performed repeatedly in each step of prediction until the last value is predicted. Moreover, to predict the next future value, we simulated 1K predictive samples based on 1K posterior samples and predictive summaries are drawn, for the next five values, in Table 7. It is important to know that the estimated predictive values are corresponding to the modal values of the predictive samples.

Table 7: Bayesian retrospective predictions of IMR from 2014 to 2018

Year	True value	Estimated Bayes predictive value	Estimated HPrD interval		ω
2014	39.00	38.83	35.01	41.08	6.07
2015	37.00	37.12	33.08	39.65	6.57
2016	34.00	34.46	30.61	36.80	6.19
2017	33.00	35.39	31.84	37.33	5.49
2018	32.00	38.18	31.32	39.74	8.42

It is nice to interpret that the predicted values are pretty close to the true values. Also, the 95% highest predictive density (HPrD) intervals are covering the corresponding true values nicely. Referring to Table 4, it can be inferred that the Bayes predictions, in general, appear to be more closer to the corresponding true values as compared to that on the basis of likelihood only. Also, the estimated predictive intervals in Bayesian paradigm appear to be more narrower than the classical paradigm (see the values of ω), that shows the accuracy of Bayesian analysis over the classical approach. Moreover, the widths of estimated predictive intervals ω in Table 7 look more consistent than those in Table 4.

Since the retrospective predictions (see Table 7) are found to be satisfactory, therefore, we did the prospective prediction of IMR of India using the same Bayesian methodology. For this prospective prediction, we considered the whole series (containing 48 observations) and apply the same strategy to forecast the next five observations. Table 8 provides the future values of IMR for the next five years. It can be seen that values of IMR (Table 8) will remain close, on an average, to 30.20 which, in fact, is a good sign and showing a decreasing trend of IMR values in upcoming years. Although, in this study, we did not consider any other demographic component which effects the IMR still, our findings are very hopeful and realistic for the developing countries like India.

Table 8: Bayesian prospective predictions of IMR for the next 5 years

Year	Estimated Bayes predictive value	Estimated HPrD interval		ω
2019	30.91	27.69	32.49	4.80
2020	29.84	25.90	30.37	4.47
2021	30.23	28.40	33.57	5.17
2022	30.26	28.22	33.88	5.66
2023	29.76	27.06	32.73	5.67

4. Conclusion

This paper has successfully modelled and analysed the ARIMA model under classical and the Bayesian paradigms. The analyses resulted in retrospective as well as prospective (for the next 5 years) predictions of IMR data of India. Stationarity of the data set has been examined carefully using ADF and KPSS tests. The likelihood based estimates have been used for the classical predictions whereas, for Bayesian predictions the corresponding modal values of the parameters have been used. It is found that the latter paradigm provided us with more accurate and reliable results as compared to the former. It is expected that such an analysis will be helpful for the policy makers and researchers to come across an appropriate planning.

Acknowledgements

The authors express their thankfulness to the Editor and the Reviewers for their valuable suggestions and comments that helped to improve the earlier version of the manuscript.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19(6)**, 716–723.

Alkema, L. and Ann, W. L. (2011). Estimating the under-five mortality rate using a Bayesian hierarchical time series model. *PloS One*, **6(9)**, e23954.

Bishai, D. (1995). Infant mortality time series are random walks with drift: Are they cointegrated with socioeconomic variables? *Econometrics and Health Economics*, **4(3)**, 157–167.

Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, **22(3)**, 547–581.

Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control, Revised ed.* Holden-Day.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley and Sons.

Carter, L. R. (1996). Forecasting US mortality: A comparison of Box-Jenkins ARIMA and structural time series models. *The Sociological Quarterly*, **37(1)**, 127–144.

Gelfand, A. E. and Smith, A. F. M. (1991). Gibbs sampling for marginal posterior expectations. *Communications in Statistics-Theory and Methods*, **20(5-6)**, 1747–1766.

Ghosh, J., Delampady, M., and Samanta, T. (2007). *An Introduction to Bayesian Analysis: Theory and Methods*. Springer Science and Business Media.

Keyfitz, N. (1982). Choice of function for mortality analysis: Effective forecasting depends on a minimum parameter representation. *Theoretical Population Biology*, **21(3)**, 329–352.

Kleibergen, F. R. and Hoek, H. (2000). Bayesian analysis of ARMA models. *Tinbergen Institute Discussion Paper* Amsterdam, the Netherlands.

Kurniasih, N., Ahmar, A. S., Hidayat, D. R., Agustin, H., and Rizal, E. (2018). Forecasting infant mortality rate for China: A comparison between α-sutte indicator, ARIMA and Holt-Winters. *Journal of Physics: Conference Series*, **1028(1)**, 012195.

Lee, R. D. and Carter, L. R. (1992). Modelling and forecasting US mortality. *Journal of the American Statistical Association*, **87(419)**, 659–671.

Marriott, J., Ravishanker, N., Gelfand, A., and Pai, J. (1996). Bayesian analysis of ARMA processes: Complete sampling-based inference under exact likelihoods. *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*, 243–256.

McNown, R. and Rogers, A. (1989). Forecasting mortality: A parametrized time series approach. *Demography*, **26(4)**, 645–660.

McNown, R. and Rogers, A. (1992). Forecasting cause-specific mortality using time series methods. *International Journal of Forecasting*, **8(3)**, 413–432.

Pedroza, C. (2006). A Bayesian forecasting model: Predicting US male mortality. *Biostatistics*, **7(4)**, 530–550.

Pollard, J. H. (1987). Projection of age-specific mortality rates. *Population Bulletin of the United Nations*, **No. 21/22**, 55–69.

Reichmuth, W. H. and Sarferaz, S. (2008). Modelling and forecasting age-specific mortality: A Bayesian approach. *SFB 649 Discussion Paper*, **(No. 2008, 052a)**.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6(2)**, 461–464.

Smith, A. F. M. and Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society*, **B 55(1)**, 3–23.

Tripathi, P. K., Mishra, R. K., and Upadhyay, S. K. (2018). Bayes and classical prediction of total fertility rate of India using autoregressive integrated moving average model. *Journal of Statistics Applications and Probability*, **7(2)**, 233–244.

Tripathi, P. K., Ranjan, R., Pant, R., and Upadhyay, S. K. (2017). An approximate Bayes analysis of ARMA model for Indian GDP growth rate data. *Journal of Statistics and Management Systems*, **20(3)**, 399–419.

Tuljapurkar, S. and Boe, C. (1998). Mortality change and forecasting. *North American Actuarial Journal*, **2(4)**, 13–47.

Upadhyay, S. K., Vasishta, N., and Smith, A. F. M. (2001). Bayes inference in life testing and reliability via Markov chain Monte Carlo simulation. *Sankhya: The Indian Journal of Statistics*, **A 63(1)**, 15–40.