Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/070202

Bayes and Classical Prediction of Total Fertility Rate of India Using Autoregressive Integrated Moving Average Model

Praveen Kumar Tripathi[1], *Rahul Kumar Mishra*[1] * *and S. K. Upadhyay*[1,2]

[1] Department of Statistics, Banaras Hindu University, Varanasi-221005, India.
[2] DST Centre for Interdisciplinary Mathematical Sciences, Banaras Hindu University, Varanasi-221005, India.

Received: 9 Sep. 2017, Revised: 21 Mar. 2018, Accepted: 26 Mar. 2018
Published online: 1 Jul. 2018

Abstract: The paper illustrates a simple methodology to predict total fertility rate of India through an approximate Bayes analysis using a particular case of general autoregressive integrated moving average model. The corresponding results based on classical paradigm are also obtained especially using maximum likelihood estimators. The study first examines the data for its stationarity and the same is achieved by differencing the data twice. Once the stationarity is achieved, some specific cases of general autoregressive integrated moving average model are examined for the given time series data to find the most appropriate candidate. This is being done using Akaike's information criterion and Bayes information criterion. The selected specific case of the model is analyzed both in Bayesian and classical frameworks, the former using vague prior for the parameters. The posterior computation in Bayesian paradigm is done using Markov chain Monte Carlo simulation. The two paradigms ultimately focus on drawing relevant inferences including the short term predictions, both retrospectively and prospectively. The results are, in general, found to be satisfactory.

Keywords: Autoregressive integrated moving average model, Total fertility rate, Age-specific fertility rate, Stationarity, Akaike's information criterion, Bayes information criterion, Gibbs sampler, Metropolis algorithm.

1 Introduction

Forecasting the demographic characteristics of a human population such as the fertility, mortality and migration is an important aspect of any socio-economic planning. Normally when such forecast is needed, one may go either for a short term or a long term forecasting based on the requirement of the plan. In fact, short term forecasts are chosen in situations where one is interested in obtaining an estimate about the number of people requiring services like education, medical facilities or other basic amenities related to human lives whereas for planning purposes a long term forecast is often preferred. Many times it is also desired to find the uncertainty attached with the forecast so as to know the likelihood of events under consideration. Obviously, for all this, a proper modelling of underlying forces is required so as to develop a precise forecast framework. The objective in this paper, however, confines us to forecast the fertility of a population. Obviously, for the purpose, we need to use some summary index for the fertility. The total fertility rate (TFR), being a composite measure of fertility, is an important component for fertility projection. TFR represents the average number of children a woman would bear if she survived throughout the range of her reproductive span, experiencing at each age duration the age-specific fertility rate (ASFR) of that period.

Demographers have had a long history of interest in forecasting fertility. Saboia [17] and McDonald [14] used a time series method to forecast the total number of births that have a significant role in the population growth. Miller [15] employed a bivariate autoregressive model, which actually acted as a transfer function model, to forecast the total fertility and the mean age of childbearing. Extending a similar approach, Ortega and Poncela [16] used a dynamic factor model with common and country-specific factors to forecast TFR for a homogeneous cluster of countries. To overcome the problem of dimensionality encountered during the forecast of ASFR, Thompson *et al.* [20] and Keilman and Pham [8] employed a parametric model of ASFR and instead of modelling ASFR for each age of a reproductive span directly,

* Corresponding author e-mail: rahul.stats11@gmail.com

they used a multivariate time series to model the parameters of these parametric models. Bozik and Bell [5] employed the principle component approach to forecast the fertility using ASFR. They, however, used only first four principal components of the model, in order to reduce its dimensionality, and then employed the multivariate ARIMA method. Their work can be considered better than the parametric approach used by Thompson *et al.* [20] in the sense that the error in the fitted curve may be seen to be less than what was previously obtained by the authors. Hyndman and Ullah [7] used the functional data approach to develop a similar strategy.

Taking a cue from the illustrious paper of Lee and Carter [10] in which the authors developed a probabilistic approach to project mortality in a long term, Lee [11] modelled fertility over time using a single time-varying fertility index, that is, TFR. The author, however, observed that the long term fertility forecasts yielded in a large width of its prediction interval. To overcome this problem, which is observed as a result of structural change in the fertility, a restriction was imposed on the values of lower and upper bounds and that of the average value of TFR. Lee and Tuljapurkar [12] used the same model with a different value for the average level of TFR and they imposed no restriction on the limits of its bounds. Booth [3] is an interesting work that has presented a good review discussion of various major demographic forecasting methods developed till date and discussed approaches for the forecast of cohort fertility.

The problem inherent with most of the methods discussed so far is that they are developed to study the fertility of the countries which have already passed through their fertility transition and are currently having low to very low fertility pattern and, owing to which, one can assume that the fertility will almost remain stationary in a way. Besides, all these methods rely on classical paradigm to draw the desired inferences. Recently, Alkema *et al.* [2] undertook this problem and they distinctly modelled the TFR with respect to the pre-transition period, transition period and post-transition period using Bayes paradigm. Working on the same very spirit, we aim to model TFR for India via ARIMA model and obtain its short-term forecast by harnessing the attributes of Bayesian paradigm.

The general form of an ARIMA(p,d,q) model is given by

$$w_t = \theta_0 + \sum_{i=1}^{p} \phi_i w_{t-i} + \sum_{j=1}^{q} \psi_j \varepsilon_{t-j} + \varepsilon_t \tag{1}$$

where $w_t = \Delta^d y_t$ is the d^{th} differenced time series corresponding to the observed time series data y_t, $t = 1,2,\ldots,T$. θ_0 is the intercept, ϕ_i's and ψ_j's represent the AR and MA coefficients, respectively, and ε's are the error terms, distributed in accordance with independent and identically distributed normal variates with mean zero and a constant variance σ^2. The parameters p and q identify the order for the seasonal autoregressive (the number of lag observations in the model) and seasonal moving average (size of moving average window) terms. From Eq. 1, it is quite evident that an ARIMA model is nothing but the d^{th} differenced stationary autoregressive-moving average (ARMA) model in which d of the roots of characteristic polynomial of autoregressive (AR) process are all unity and the remainder lie outside the unit circle (see Box *et al.* [4]).

The plan of the paper is as follows. The next section provides a real data set on TFR of India and ascertains its stationarity by differencing the data twice. A preliminary assessment of the subclasses of ARIMA model is provided based on Box-Jenkins' criterion, especially using the values of autocorrelation function (ACF) and the partial autocorrelation function (PACF) for a tentative identification of MA and AR components. Section 3 provides approximate likelihood function corresponding to the proposed ARIMA model to obtain the corresponding maximum likelihood (ML) estimators. The section also comments briefly on Akaike information criterion (AIC) and Bayes information criterion (BIC) as the tools for model comparison. A separate subsection provides the corresponding classical results including the retrospective predictions after choosing an appropriate model based on the two information criteria. Section 4 provides Bayesian model formulation for the chosen ARIMA model using vague priors for the parameters. The section advocates the use of Gibbs sampler algorithm for getting samples from the corresponding posterior distribution though some of the full conditionals are generated using the Metropolis algorithm. A separate subsection provides the numerical illustration where the posterior corresponding to final selected model is explored completely and both retrospective and future predictions are provided for TFR of India. The paper ends with a brief conclusion given in the last section.

2 The Data and the Approximate Model Assessment

Let us consider a yearly data set on TFR of India for the years 1971 to 2011 reported by Sample Registration System, Registrar General of India. The data set is taken from the web site https://nrhm-mis.nic.in/PubFWStatistics%202013/Complete%20Book.pdf and is reproduced below in Table 1 for a ready reference. Since our objective includes future prediction of TFR, we first try to find out an approximate model for the prevailing fertility pattern of India. The time series plot for the observed data set, given in Table 1, is shown in Figure 1. Obviously, the plot shows a regular non-increasing, non-stationary pattern of the considered data set on TFR.

Table 1: TFR of India from 1971 to 2011

5.2	5.2	4.9	4.9	4.9	4.7	4.5	4.5	4.4	4.4	4.5	4.5	4.5	4.5	4.3	4.2
4.1	4.0	3.9	3.8	3.6	3.6	3.5	3.5	3.5	3.4	3.3	3.2	3.2	3.2	3.1	3.0
3.0	2.9	2.9	2.8	2.7	2.6	2.6	2.5	2.4							

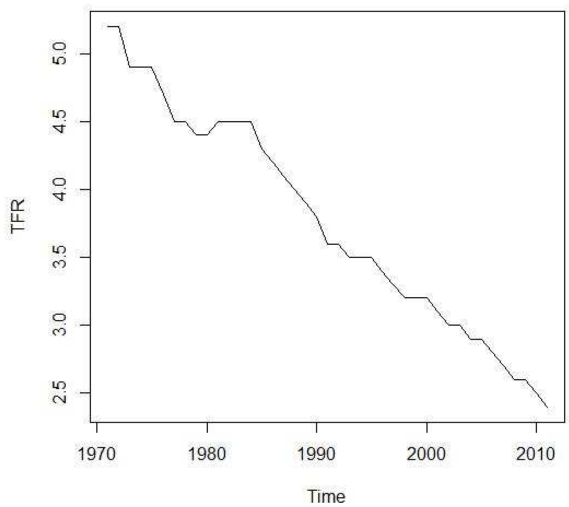

Figure 1: Time series plot showing the TFR of India from 1971 to 2011.

Since stationarity is an important condition in any time series analysis, the same may be achieved by a number of approaches, the simplest being the one based on differencing the data. Thus, in order to achieve the required stationary pattern in the data, we proceeded by differencing the data twice. The time series plot for the differenced so obtained is shown in Figure 2 which evidently shows stationarity behaviour of the differenced data. The model assessment can be done by identifying the parameters of ARIMA model at least approximately for the considered data set. Since twice differenced TFR data is showing stationarity behaviour, the parameter d can be considered as 2 although this assessment is based on graphical tool only. For assessing the parameters p and q, we rely on Box-Jenkins methodology and use, in particular, the ACF plot and the PACF plot. It is to be noted that the autocorrelation explains the way the observation in the time series are related to each other and is measured by the simple correlation between the current observation and some previous observation at specific lag, say p. Similarly, the partial autocorrelation is used to measure the degree of association between the current observation and some previous observation at lag p before the current observation, after removing the effects of intermediate observations such as those at lags $1,2,...,p-1$. The model is AR(MA) if the ACF(PACF) trails off after a lag and has a hard cut-off in the PACF(ACF) after a lag. This lag is taken as the value for $p(q)$. The model is a mixture of both AR and MA, viz. ARMA, if both ACF and PACF trail off. Truly speaking, the identification of AR and MA components based on these two plots may often be tentative and involve a kind of approximate personal judgement. It has been often suggested that identification of AR model can be best done on the basis of PACF plot. Say, for instance, note down the lag after which the PACF plot shuts off, that is, partial autocorrelation becomes zero after that point. Similarly, suggestion involves relying on ACF plot to identify MA model. The ACF plot will show non-zero autocorrelation only at lags involved in the model.

The ACF and PACF plots of twice differenced TFR data are given in Figure 3. It is evident from the figure that the ACF plot trails-off to zero after one lag, which implies that the given time series follow a MA(1) process. Similarly, the PACF plot trails-off to zero after two lags so a AR(2) process can be a suitable candidate for this double differenced time series data. Obviously, our overall conclusion suggests ARIMA(2,2,1) model for the considered data set. We, however, consider some other models such as 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) so that all nearby models can also be looked upon and any misleading conclusion based on tentative assessment of ACF and PACF plots can be ruled out. It is important to mention that some of these

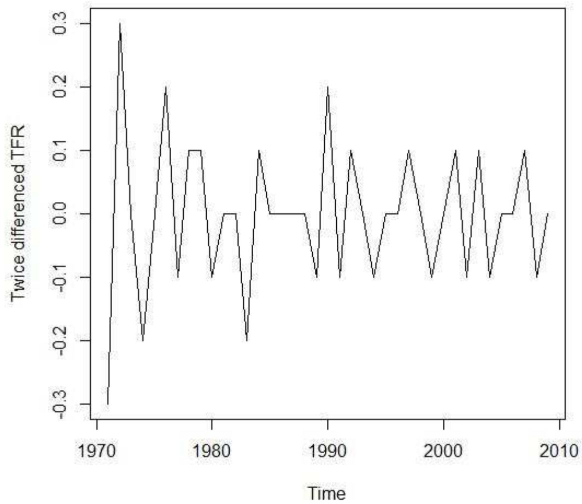

Figure 2: Time series plot based on twice differenced TFR data of India from 1971 to 2011.

models will represent under-fitted and over-fitted scenarios for the considered data set. Once all these models are entertained, a final conclusion for recommending a model may be drawn based on their comparison using a few standard tools. In this paper, we have considered AIC and BIC as the tools for comparing the various entertained models.

3 The Likelihood Function and the Model Selection Criteria

Let $\underline{w} : w_1, w_2, ..., w_{T-d}$ be the entertained observations from model (1). The conditional density of w_t, conditioned on $w_{t-1}, w_{t-2}, ..., w_{t-p}$, is given by

$$f(w_t|w_{t-1},w_{t-2},...,w_{t-p};\theta_0,\Phi,\Psi,\sigma^2) \propto \left(\frac{1}{\sigma^2}\right)\exp\left(-\frac{1}{2\sigma^2}(w_t-\theta_0-\sum_{i=1}^{p}\phi_i w_{t-i}-\sum_{j=1}^{q}\psi_j\varepsilon_{t-j})^2\right). \tag{2}$$

Using (2), the likelihood function corresponding to model (1) can be approximated by its conditional form as

$$L(\underline{w}|\theta_0,\Phi,\Psi,\sigma^2) \propto \prod_{t=p+1}^{T-d} f(w_t|w_{t-1},w_{t-2},...,w_{t-p};\theta_0,\Phi,\Psi), \tag{3}$$

which, on simplification, reduces to

$$L(\underline{w}|\theta_0,\Phi,\Psi,\sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{(T-d-p)/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{t=p+1}^{T-d}(w_t-\theta_0-\sum_{i=1}^{p}\phi_i w_{t-i}-\sum_{j=1}^{q}\psi_j\varepsilon_{t-j})^2\right), \tag{4}$$

where $\Phi = (\phi_1,...,\phi_p)$ and $\Psi = (\psi_1,\psi_2,...,\psi_q)$. It is important to note that if one has a sample of size T to estimate an ARMA(p,q) process by means of conditional maximum likelihood (ML) estimation from equation (4), one will use only $(T-d-p)$ observations of this sample. This is the approximation that has been entertained due to dependence structure and non-availability of data before w_1. Obviously, (4) can be used to obtain approximate ML estimators of the model parameters. Let us denote these by $\hat{\theta}_0$, $\hat{\Phi}$, $\hat{\Psi}$ and $\hat{\sigma^2}$, respectively, for the parameters θ_0, Φ, Ψ and σ^2 and let us denote the corresponding maximized likelihood function by $\hat{L}$.

Once the maximized likelihood function is obtained, the AIC (see Akaike [1]) and BIC (see Schwarz [18]) can be defined as

$$AIC = -2\log\hat{L} + 2k, \tag{5}$$

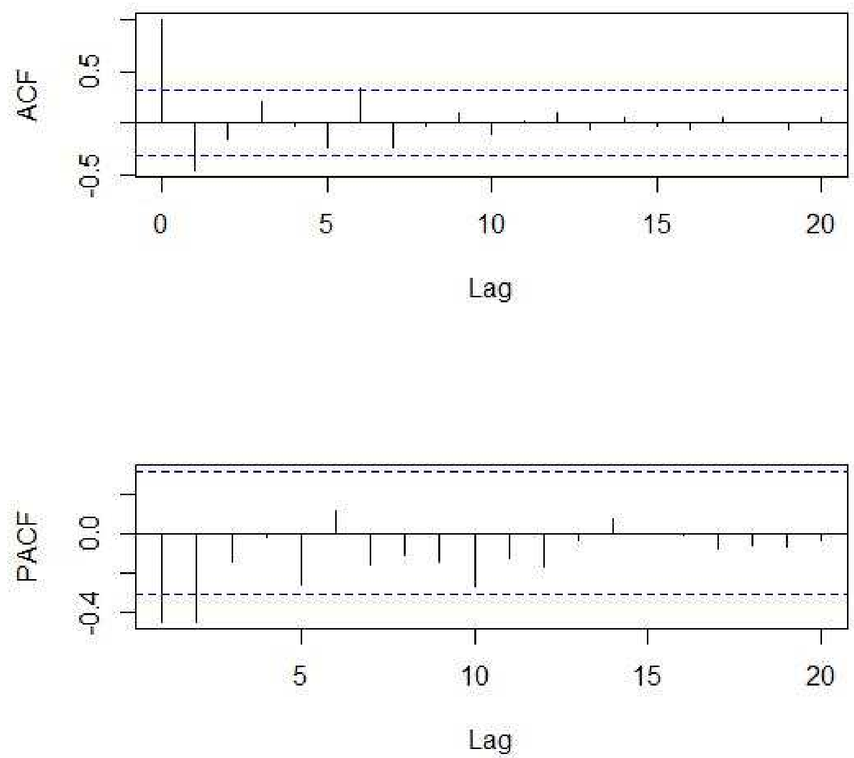

Figure 3: ACF and PACF plots for twice differenced TFR data.

$$BIC = -2\log\hat{L} + k\,\log(T-p), \tag{6}$$

respectively, where k is the number of parameters in the entertained model. The model selection criterion based on AIC(BIC) recommends a model for which the corresponding value of AIC(BIC) is least. The good thing about these two information criteria is that both AIC and BIC penalize the model for its inherent complexities, which is generally indicated by the number of parameters involved in the model.

3.1 Numerical Illustration: Model Selection and Prediction

Considering all the proposed models given in Section 2 and using the corresponding likelihood functions borrowed from Section 3, the ML estimates of resulting model parameters based on twice differenced TFR data are given in Table 2. The table also reports the values of corresponding $\log\,\hat{L}$ for different models. It is to be noted that these values are reported because of their requirement for obtaining AIC and BIC for each of the considered models.

One can easily interpret the results obtained in Table 2. Say, for instance, the effect of intercept terms is not so significant in all the considered models. Also, all the estimates of AR and MA coefficients possess the conditions of stationarity/invertibility and nicely lie in the respective regions. These regions are not reported in the present paper because of their versatile availability in the literature. One can consider, for instance, Box *et al.* [4] and Tripathi *et al.* [21], among others. The effect of error variance is least in all the models; however, the estimates of AR and MA coefficients show the significance of preceding observations and disturbances in all the considered models and thus the significance of error variances especially in MA parts.

As our objective includes predicting TFR for a better understanding of fertility pattern of India, we begin by choosing a most appropriate model among the considered models based on the values of AIC and BIC. The corresponding values

Table 2: ML estimates and the corresponding maximized log likelihood functions for the considered specific cases of ARIMA models

Model	Parameter	ML estimate	$\log \hat{L}$
ARIMA(0, 2, 1)	θ_0	$-0.13e^{-2}$	40.06
	ψ_1	-0.93	
	σ^2	$0.75e^{-2}$	
ARIMA(0, 2, 2)	θ_0	$-0.15e^{-2}$	40.17
	ψ_1	-0.86	
	ψ_2	-0.08	
	σ^2	$0.74e^{-2}$	
ARIMA(1, 2, 0)	θ_0	$0.41e^{-2}$	37.11
	ϕ_1	-0.45	
	σ^2	$0.83e^{-2}$	
ARIMA(2, 2, 0)	θ_0	$0.04e^{-2}$	42.51
	ϕ_1	-0.56	
	ϕ_2	-0.42	
	σ^2	$0.58e^{-2}$	
ARIMA(1, 2, 1)	θ_0	$0.31e^{-2}$	41.33
	ϕ_1	-0.65	
	ψ_1	-0.58	
	σ^2	$0.67e^{-2}$	
ARIMA(2, 2, 1)	θ_0	$-0.56e^{-2}$	48.54
	ϕ_1	-0.19	
	ϕ_2	-0.02	
	ψ_1	-1.24	
	σ^2	$0.43e^{-2}$	
ARIMA(1, 2, 2)	θ_0	$0.32e^{-2}$	41.49
	ϕ_1	-0.56	
	ψ_1	0.43	
	ψ_2	-1.07	
	σ^2	$0.43e^{-2}$	
ARIMA(2, 2, 2)	θ_0	$-0.52e^{-2}$	50.72
	ϕ_1	-0.29	
	ϕ_2	-0.20	
	ψ_1	0.78	
	ψ_2	-0.63	
	σ^2	$0.38e^{-2}$	

of AIC and BIC are reported in Table 3. It is obvious that the evaluated values of AIC and BIC support ARIMA(0, 2, 1) model as the corresponding values of AIC and BIC happen to be the least. We, therefore, consider ARIMA(0, 2, 1) model for the forthcoming analysis and prediction of fertility pattern of India.

Table 3: Values of AIC and BIC for all the considered models

Model	AIC	BIC
ARIMA(0, 2, 1)	-77.58	-72.59
ARIMA(0, 2, 2)	-75.64	-68.99
ARIMA(1, 2, 0)	-63.18	-58.19
ARIMA(2, 2, 0)	-70.94	-64.29
ARIMA(1, 2, 1)	-75.64	-68.98
ARIMA(2, 2, 1)	-73.83	-65.51
ARIMA(1, 2, 2)	-73.61	-65.30
ARIMA(2, 2, 2)	-72.29	-62.31

We next provide the likelihood based retrospective prediction of TFR based on ARIMA(0, 2, 1) model. To start with the retrospective prediction, we initially considered the first 36 observations out of the given 41 observations (see Table 1) and obtained the ML estimates for the parameters of ARIMA(0, 2, 1) model as detailed in Section 3. Based on these ML estimates, we predicted the next 37^{th} observation. This predicted observation was then used to form a sample size 37 and the corresponding ML estimates were obtained using these 37 observations in order to predict the next 38^{th} observation. This process was continued until all the remaining observations were predicted. It is to be noted that for predicting the observation in the original series, one actually works by predicting the observation corresponding to twice differenced data. Thus if we assume the original data size as T, say, at T^{th} stage, we predicted w^{th}_{T-1} observation in the series corresponding to twice differenced data. And, therefore, the estimated predictive values $\hat{y}_{T+1}$ can be obtained from the relation given below in a recursive manner

$$\hat{y}_{T+1}=\hat{w}_{T-1}+2y_T-y_{T-1}. \tag{7}$$

where $\hat{w}_{T-1}$ is the estimated predictive observation corresponding to w_{T-1}.

The actual values and the corresponding estimated predictive values are given in Table 4. The table also provides the associated predictive intervals with 0.95 confidence coefficient. It may be noted that for obtaining the predictive intervals, we once again worked on twice differenced data and obtained the corresponding values for the actual observations. Thus $(1-\alpha)\%$ predictive interval corresponding to w^{th}_{T-1} observation can be obtained using the relationship

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

where $z_{1-\alpha/2}$ is the standard normal percentile and $\hat{\varepsilon}_{T-1}$ is the estimated error term with $Var(\hat{\varepsilon}_{T-1})=\hat{\sigma}^2\hat{\psi}_1^2$. If $\hat{w}L_{T-1}$ and $\hat{w}U_{T-1}$ are the lower and upper limits of the estimated predictive intervals of $\hat{w}_{T-1}$, respectively, the corresponding limits for $\hat{y}_{T+1}$ can be obtained by a similar transformation as given in (7) by replacing $\hat{w}_{T-1}$ by $\hat{w}L_{T-1}$ and $\hat{w}U_{T-1}$, respectively. Obviously, the estimated predictive values obtained in Table 4 are not too far away from the actual values and the values are nicely covered by the corresponding predictive intervals with confidence coefficient 0.95.

Table 4: Likelihood based retrospective predictions of TFR for the period 2007 to 2011

y_t	True value	Estimated predictive value	Estimated predictive interval	
y_{37}	2.7	2.84	2.67	3.01
y_{38}	2.6	2.77	2.59	2.94
y_{39}	2.6	2.67	2.51	2.85
y_{40}	2.5	2.61	2.48	2.81
y_{41}	2.4	2.55	2.37	2.72

4 Bayesian Model Formulation

For the differenced data set, the conditional likelihood function of the chosen model ARIMA(0, 2, 1) can be obtained by ignoring the AR component of (4) and the same can be written as

$$f(\underline{w}|\theta_0,\psi_1,\sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{(T-2)/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{T-2}(w_t-\theta_0-\psi_1\varepsilon_{t-1})^2\right), \tag{9}$$

where w_t is obviously $\Delta^2 y_t$.

In order to perform a Bayes analysis, one is required to begin by specifying prior distributions for the parameters. These prior distributions portray our beliefs about the parameters before the observed data are made available. In case where the experimenter does not have enough information to specify appropriate prior distributions, it is advisable going with non-informative priors and allowing inferences to be data driven. Moreover, since prior distributions play crucial role in any Bayesian analysis, specifying wrong priors may result in poor inferences. For the situation under consideration, we consider the same strategy and consider the use of non-informative or vague priors for the parameters. We, therefore, define the prior distribution for the parameters σ^2, ψ_1 and θ_0 as

$$\pi_1(\sigma^2) \propto \frac{1}{\sigma^2}; \qquad \sigma^2\geq 0, \tag{10}$$

$$\pi_2(\theta_0) \propto U[-M, M]; \qquad M>0, \tag{11}$$

and

$$\pi_3(\psi_1) \propto U[-N_1,\, N_2]; \qquad N_1>0,\, N_2>0, \tag{12}$$

respectively, where M and N_i, i=1,2, are the hyperparameters that may be taken large enough so that the priors remain vague. The prior distribution given in (10) for the scale parameter is obviously Jeffreys' type and it has been widely used in the literature (see, for example, Marriot *et al.* [13] and Kleibergen and Hoek [9]). Thus combining these prior distributions (10) to (12) with the corresponding likelihood function (9) via Bayes theorem, yields the joint posterior distribution that can be written up to proportionality as

$$p(\theta_0,\psi_1,\sigma^2|\underline{w}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{T}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{T-2}(w_t-\theta_0-\psi_1\varepsilon_{t-1})^2\right) I_{[-M,M]}(\theta_0) I_{[-N_1,N_2]}(\psi_1), \tag{13}$$

where $I(.)$ denotes the indicator function defined as

$$I_A(x)=\begin{cases}1 \text{ if } x\in A\\ 0 \text{ otherwise.}\end{cases} \tag{14}$$

In order to obtain the posterior based inferences from (13), one obviously needs to rely on sample based approaches as the resulting posterior is not analytically tractable. We consider Gibbs sampler algorithm among various available alternatives as Gibbs sampler offers conceptually easy solution. We do not describe this Markovian algorithm in detail rather comment simply that it offers a kind of straightforward updating mechanism that proceeds by iterating from various (often) unidimensional full conditionals in a cyclic order. After a large number of iterations, the iterating chain so obtained converges in distribution to a random sample from the actual posterior distribution. For details, one may refer to Upadhyay and Smith [22] (see also Gelfand and Smith [6]), among others.

Once the posterior distribution is simulated to get the samples of desired size, the unobserved future data w_{T-1}, for each of the posterior samples, can be easily simulated from the parent sampling distribution $p(w_{T-1}|\theta_0,\sigma^2,\psi_1,\underline{w})$ where the distribution $p(w_{T-1}|\theta_0,\sigma^2,\psi_1,\underline{w})$ is nothing but a univariate normal with mean $\mu_{T+1}=\theta_0+\psi_1\hat{\varepsilon}_1$ and variance σ^2. Thus having got the corresponding samples from the posterior (13), one can easily obtain the samples of future data w_{T-1} and, as it has been done previously, the predictive observation for the original series y_{T+1} using the transformation in (7). Obviously, the predictive estimates such as point predictions, predictive intervals, etc. can be easily obtained on the basis of the predictive samples corresponding to y_{T+1}.

4.1 Gibbs Sampler Implementation

The full conditionals, specified up to proportionality, for different variates can be specified from the joint posterior (13) (see, for example, Upadhyay *et al.* [23]). The same can be written as

$$p(\theta_0|\sigma^2,\psi_1,\underline{w}) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{T-2}(w_t-\theta_0-\psi_1\varepsilon_{t-1})^2\right), \tag{15}$$

$$p(\psi_1|\sigma^2,\theta_0,\underline{w}) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{T-2}(w_t-\theta_0-\psi_1\varepsilon_{t-1})^2\right), \tag{16}$$

and

$$p(\sigma^2|\theta_0,\psi_1,\underline{w}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{T}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{T-2}(w_t-\theta_0-\psi_1\varepsilon_{t-1})^2\right). \tag{17}$$

Once the full conditionals are obtained, the next strategy is to look on these for their availability from the viewpoint of sample generation. It can be easily verified that (17) can be generated using a gamma generating routine after making a transformation $\tau = 1/\sigma^2$. It can be seen that the transformed variate τ follows gamma density with shape parameter $(\frac{T}{2}+1)$ and scale parameter $[\frac{1}{2}\sum_{t=1}^{T-2}(w_t-\theta_0-\psi_1\varepsilon_{t-1})^2]$. The full conditionals (15) and (16) cannot be reduced to standard family of distributions and, therefore, we propose to simulate both of these via the Metropolis algorithm. Since we are using Metropolis algorithm for generating samples from two full conditionals, the proposed algorithm is not actually the

Gibbs sampler rather one can refer it as hybrid Gibbs-Metropolis algorithm. For details on Metropolis algorithm, one can refer to Smith and Roberts [19], Upadhyay *et al.* [23], among others.

The necessary implementation of the Metropolis algorithm for generating from (15) and (16) separately, we proceed by taking a univariate normal proposal in each case whose location is centred at the corresponding ML estimate and the standard deviation is taken to be c times the Hessian based approximation at the ML estimate where c is a scaling constant whose value may be taken between 0.5 and 1.0 (see, for example, Upadhyay *et al.* [23]). Thus the implementation of Gibbs sampler algorithm can be easily done on the posterior (13) by simulating variate values from the corresponding full conditionals. To extract the corresponding posterior samples, we propose a single long run of the chain after an initial transient behaviour and pick up the variate values at an appropriate interval so as to minimize the serial correlation (see also Upadhyay *et al.* [23]). Since the choice of initial values plays a crucial role in the convergence diagnostic of iterating chain, we consider using ML estimates as the initial values for necessary implemention of the algorithm.

4.2 Numerical Illustration for Bayesian Results

For the full posterior analysis of the ARIMA(0, 2, 1) model, we used the modelling formulation and the corresponding Gibbs sampler implementation as detailed in Section 4. The complete posterior analysis was done using the ML estimates of the parameters and the subsequent Hessian based approximation as the initial values for iterating the chain. It is to be noted that Hessian based approximation was needed for getting samples from the full conditionals (15) and (16) using the Metropolis algorithm. We, however, used a scaling constant c=0.6 that provided a good acceptance probability in the two cases. In order to have vague consideration of priors, the values of hyperparameters M and N_i, i=1,2, were chosen to be 100 in each case. These considerations allowed the inferences to depend on likelihood surface only.

Table 5: Posterior estimates for the parameters of ARIMA(0, 2, 1) model corresponding to twice differenced data

Model	Parameter	MLE	Posterior Mean	Posterior Mode	0.95 HPD Interval	
ARIMA(0, 2, 1)	θ_0	$0.35e^{-2}$	$0.35e^{-2}$	$0.37e^{-2}$	$-1.23e^{-2}$	$2.53e^{-2}$
	ψ_1	-0.59	-0.47	-0.57	-0.79	-0.12
	σ^2	$0.84e^{-2}$	$0.94e^{-2}$	$0.86e^{-2}$	$0.59e^{-2}$	$1.45e^{-2}$

The results are based on a simulated posterior sample of size 1K obtained through a single long run of the chain after an initial transient behaviour noted at about 60K iterations. We, however, picked up observations at a gap of 10 after ensuring that the serial correlation becomes negligibly small. Some important sample based posterior characteristics are shown in Table 5.

Referring to Table 5, it can be seen that the estimated marginal posterior density for $\sigma^2(\psi_1)$ exhibits a slight positive(negative) skeweness, whereas the parameter θ_0 is more or less symmetrical. A natural finding is the closeness of estimated posterior modes with the corresponding ML estimates, which was expected as well because of vague choices of priors. Our Bayesian results based on Gibbs sampler is of course advantageous not only because of its ease of interpretation and associated advantages of Bayesian paradigm but also because of its enormous scope once the samples are made available. Truly speaking, any inferential aspect can be easily drawn once the samples are made available. The complete posterior density estimates and other bivariate or trivariate posterior characteristics are not shown though former can be guessed based on the results given in Table 5. We are leaving other inferential aspects treating them as natural extensions.

Table 6: Retrospective predictions of TFR for the period 2007 to 2011 based on Bayesian tools

y_t	True value	Estimated Bayes predictive value	Estimated highest predictive density interval	
y_{37}	2.7	2.70	2.58	2.82
y_{38}	2.6	2.68	2.55	2.82
y_{39}	2.6	2.67	2.57	2.78
y_{40}	2.5	2.61	2.49	2.76
y_{41}	2.4	2.51	2.39	2.72

Let us finally work for retrospective prediction in a Bayesian framework as it was done earlier in subsection 3.1 using the tools of classical paradigm. We shall focus on the same ARIMA model that was finally recommended and used for the Bayesian analysis. As attempted earlier, we considered only 36 observations as the informative data out of the given 41 observations (see Table 1) though we worked on twice differenced data to go for actual prediction of the given observations. It is to be noted that before going for the actual prediction of these (assumed) unknown 05 entities, the entire Bayes analysis was repeated using the first 36 observations to obtain a posterior sample of size 1K from ARIMA(0, 2, 1) model. For each posterior sample, we next obtained the predictive sample corresponding to the next observation, that is, 37^{th} as discussed in Section 4. We thus have predictive sample of size 1K corresponding to the next unknown observation. Based on these 1K predictive samples, we obtained the estimated Bayes predictive value as the corresponding modal value. These predictive samples were also used to estimate the highest predictive density interval with coverage probability 0.95. The next observation, that is, 38^{th} is predicted only after including the estimated Bayes predictive value for the 37^{th} observation into the informative data set and repeating the whole posterior and predictive analyses as it has been done previously for 37^{th} observation. This entire process was repeated recursively until all the left out observations are predicted.

The results for retrospective point prediction in the form of estimated Bayes predictive values and the corresponding highest predictive density intervals with coverage probability 0.95 are shown in Table 6. The table also shows the true values taken from Table 1 for immediate comparison. It can be seen from the results of Table 6 that the predictive point estimates in the form of estimated Bayes predictive values are, in general, close enough to the corresponding true vales. Also, the estimated highest predictive density intervals with coverage probability 0.95 do cover the corresponding true values in every case.

Let us also compare the results of retrospective prediction obtained using the classical tools (see Table 4) with those obtained using the Bayesian tools as given in Table 6. It can be seen from the comparison of the two tables that estimated Bayes predictive values are, in general, closer to the corresponding true values than the likelihood based estimated predictive values. Besides, we also see that the estimated highest predictive density intervals with coverage probability 0.95 are, in general, narrower than corresponding likelihood based estimated predictive intervals with confidence co-efficient 0.95. These findings are of course striking and convey indisputably in favour of Bayesian results.

Table 7: Future predictions of TFR for the period 2012 to 2020 based on Bayesian tools

y_t	Year	True value	Estimated Bayes predictive value	Estimated highest predictive density interval	
y_{42}	2012	2.4	2.33	2.15	2.44
y_{43}	2013	2.3	2.29	2.16	2.42
y_{44}	2014	2.3	2.25	2.13	2.38
y_{45}	2015	2.3	2.27	2.15	2.41
y_{46}	2016	2.2	2.29	2.16	2.45
y_{47}	2017	–	2.27	2.16	2.39
y_{48}	2018	–	2.28	2.18	2.43
y_{49}	2019	–	2.27	2.15	2.39
y_{50}	2020	–	2.26	2.19	2.45

Before we end the section, let us obtain the Bayesian results of future prediction of TFR beyond 2011. It is to be noted that our considered data set in Table 1 provides the values of TFR only up to 2011 so any value beyond that is being treated as future prediction. For the purpose, we considered all the 41 observations reported in Table 1 and performed the posterior and the predictive analyses of the considered ARIMA(0, 2, 1) model as detailed in the previous paragraphs. The results for point prediction in the form of estimated Bayes predictive values and the corresponding highest predictive density intervals with coverage probability 0.95 are given in Table 7. These results are obtained exactly the way the results of retrospective predictions are obtained in Table 6. The table also provides the actual values of TFR for the years 2012 to 2016. The source of the actual TFR values for the years 2012 to 2015 is same as that of the previous data set (Table 1) and the TFR value for the year 2016 is taken from the NFHS-IV report as available on IIPS's website.

From the results obtained in Table 7, it can be asserted that the average level of TFR for India will remain close to 2.28 for the year 2012 to 2020. It can be seen that the figures obtained in Table 7 are, in a sense, close to the expected value of fertility at its level of replacement, that is, 2.1. Moreover, the prediction results obtained in Table 7 are based on the Bayesian analysis of a simple time series model, being in some sense probabilistic, and surely ignore other important demographic aspects which directly or indirectly affect the dynamics of overall fertility. It appears as if the increasing awareness among people, availability of different contraceptive methods and various other demographic interventions by the government have an important role in governing the future fertility scenario. Overall, our results are somewhat optimistic in the sense that the average value of TFR is expected to remain stationary and close to the theoretical replacement level of fertility in the coming future. The estimated highest predictive density intervals with coverage probability 0.95 further support our conviction.

5 Conclusion

This paper is a successful attempt to predict TFR of India using a simple ARIMA model. Both classical and Bayesian paradigms are successfully employed for obtaining the intended prediction although the latter paradigm appears to have slightly better performance as outlined in the paper. We agree that this approach is certainly not an ultimate approach as it fails to take in to account a number of important demographic considerations that control fertility behaviour of a population. The simplicity of the approach is, however, an apparent advantage that provides very close predicted values of TFR of India. It is expected that such an analysis will help the practitioners to get at least an approximate idea of future fertility trend.

Acknowledgement

The authors are grateful to the anonymous referees for a careful checking of the paper and for the helpful comments that have improved this paper.

References

[1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716-723.
[2] Alkema, L., Raftery, A. E., Gerland, P., Clark, S. J., Pelletier, F., Buettner, T., and Heilig, G. K. (2011). Probabilistic projections of the total fertility rate for all countries. *Demography, 48*(3), 815-839.
[3] Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting, 22*(3), 547-581.
[4] Box, G.E., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons..
[5] Bozik, J. E., and Bell, W. R. (1987). Forecasting age specific fertility using principal components. In *Proceedings of the American Statistical Association, Social Statistics Section* (Vol. 396, p. 401).
[6] 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.
[7] Hyndman, R. J., and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis, 51*(10), 4942-4956.
[8] Keilman, N., and Pham, D. Q. (2000). Predictive intervals for age-specific fertility. *European Journal of Population/Revue europenne de Dmographie, 16*(1), 41-65.
[9] Kleibergen, F., and Hoek, H. (2000). Bayesian Analysis of ARMA Models. *Tinbergen Institute Discussion Paper* Amsterdam, the Netherlands.
[10] Lee, R. D., and Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American statistical association, 87*(419), 659-671.
[11] Lee, R. D. (1993). Modeling and forecasting the time series of US fertility: Age distribution, range, and ultimate level. *International Journal if Forecasting 9*, 187-202.
[12] Lee, R. D., and Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: Beyond high, medium, and low. *Journal of the American Statistical Association, 89*(428), 1175-1189.
[13] 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*, pp.243-256..
[14] McDonald, J. B. (1979). A time series approach to forecasting Australian total live-births. *Demography, 16*(4), 575-601.
[15] Miller, R. B. (1986). A bivariate model for total fertility rate and mean age of childbearing. *Insurance, Mathematics and Economics* 5, 133-140.
[16] Ortega, J. A., and Poncela, P. (2005). Joint forecasts of Southern European fertility rates with non-stationary dynamic factor models. *International Journal of Forecasting, 21*(3), 539-550.
[17] Saboia, J. L. M. (1977). Autoregressive integrated moving average (ARIMA) models for birth forecasting. *Journal of the American Statistical Association, 72* (358), 264-270.
[18] Schwarz, G., (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461-464.
[19] Smith, A.F.M., and G.O. Roberts. (1993). Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society, Series B* 55:3-25.

[20] Thompson, P. A., Bell, W. R., Long, J. F., and Miller, R. B. (1989). Multivariate time series projections of parameterized age-specific fertility rates. *Journal of the American Statistical Association, 84*(407), 689-699.
[21] Tipathi, 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 & Management Systems 20*(3):399-419.
[22] Upadhyay, S.K., and A.F.M. Smith. (1994). Modelling Complexities in Reliability and the Role of Simulation in Bayesian Computation. *International Journal of Continuing Engineering Education* 4:93-104.
[23] Upadhyay, S.K., N. Vasishta, and A.F.M. Smith. (2001). Bayes Inference in Life Testing and Reliability Via Markov Chain Monte Carlo Simulation. *Sankhya, Series A* 63:15-40.

Praveen Kumar Tripathi is a Ph.D. scholar in the Department of Statistics, Banaras Hindu University. He has submitted his thesis for the partial fulfilment of the Ph.D. degree. His research interests are in the area of applied statistics, time series analysis, econometrics, statistical modelling and computing, classical and Bayesian analysis. His Ph.D. research work was supported by the University Grants Commission in the form of Basic Scientific Research fellowship.

Rahul Kumar Mishra is a Ph.D. scholar in the Department of Statistics, Banaras Hindu University. He has received B.Sc. (Hons.) and M.Sc. degrees in Statistics from Banaras Hindu University, Varanasi, India, in 2011 and 2013, respectively. His research interests include Bayesian statistics, demographic modelling, hierarchical modelling and computation. He is awarded with Junior Research Fellowship by the University Grants Commission.

Satyanshu K. Upadhyay received the M.Sc. and Ph.D. degrees from Banaras Hindu University, Varanasi, India, in 1980 and 1984, respectively. He started his teaching career as a Lecturer in statistics in the year 1981. He is currently a Professor in the Department of Statistics, Banaras Hindu University. He is also the Co-ordinator of DST Centre for Interdisciplinary Mathematical Sciences, Banaras Hindu University. He received a number of distinctions in his career, and worked as an EEC Research Bursary at Imperial College, London, during the year 1992−1993. His research interests include preliminary test procedures, life testing and reliability, biostatistics, and Bayes inferences. He has published several research papers, books, etc. Dr. Upadhyay is associated with a number of academic and administrative bodies and also the Fellow of the Royal Statistical Society.