A Bayes algorithm for model compatibility and comparison of ARMA(p,q) models

Praveen Kumar Tripathi,[1] Rijji Sen,[2] S. K. Upadhyay[3]

ABSTRACT

The paper presents a Bayes analysis of an autoregressive-moving average model and its components based on exact likelihood and weak priors for the parameters where the priors are defined so that they incorporate stationarity and invertibility restrictions naturally. A Gibbs-Metropolis hybrid scheme is used to draw posterior-based inferences for the models under consideration. The compatibility of the models with the data is examined using the Ljung-Box-Pierce chi-square-based statistic. The paper also compares different compatible models through the posterior predictive loss criterion in order to recommend the most appropriate one. For a numerical illustration of the above, data on the Indian gross domestic product growth rate at constant prices are considered. Differencing the data once prior to conducting the analysis ensured their stationarity. Retrospective short-term predictions of the data are provided based on the final recommended model. The considered methodology is expected to offer an easy and precise method for economic data analysis.

Key words: ARMA model, exact likelihood, Gibbs sampler, Metropolis algorithm, posterior predictive loss, model compatibility, Ljung-Box-Pierce statistic, GDP growth rate.

1. Introduction

The general form of an autoregressive-moving average (ARMA) model, of order p and q, is defined as

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

where y_t's are the time series observations, θ_0 is the intercept, ϕ_i's and ψ_j's are autoregressive (AR) and moving average (MA) coefficients, respectively, and ε's are the independent and identically distributed (iid) components of the Gaussian white noise distributed with mean zero and variance σ^2. We shall denote the model (1) by ARMA(p,q). There may be, of course, several choices of p and q but the large choices usually complicate the models and often lead to intractable solutions. This paper, therefore, considers a few arbitrary choices of p and q such that $p + q \leq 2$ and then finally recommends a model that is

---

[1]Department of Mathematics and Statistics, Banasthali Vidyapith, Rajasthan, India.
E-mail: pkt2088@gmail.com. ORCID: https://orcid.org/0000-0002-5652-8301.
[2]Department of Statistics, Behala College, Calcutta University, India. E-mail: rijjisen@gmail.com.
ORCID: https://orcid.org/0000-0002-7805-037X.
[3]Department of Statistics, Banaras Hindu University, Varanasi, India. E-mail: skupadhyay@gmail.com.
ORCID: https://orcid.org/0000-0001-6051-5821.

most appropriate for our intended objective. It is to be noted that the choice of $p + q \leq 2$ simplifies the model considerably.

The literature is abundant with some significant references on classical analyses of ARMA models and its variants. Box and Jenkins (1976) popularized the ARMA model by simplifying the analysis in the classical framework, especially with reference to order identification of the model via autocorrelation function (ACF) and partial autocorrelation function (PACF). Later, Tsay and Tiao (1984, 1985) developed a unified approach for order identification for both stationary and non-stationary ARMA models. They proposed some consistent estimates of the auto-regressive parameters, which in turn were utilized to define an extended sample ACF for determination of the order of an AR model. The method proposed was mostly appropriate for non-seasonal data. However, in the case of seasonal data, a number of studies have used filtering approach. Reilly (1980) used an automatic methodology, similar to that used by Box and Jenkins (1976), to model the macroeconomic variables, like gross domestic product (GDP). A similar method was developed by Reynolds et al. (1995), which is more automatic and well illustrated by utilizing the time-series data for a single variable. Recently, Tripathi et al. (2018) have used Box-Jenkins methodology on the Integrated form of ARMA model.

The Bayesian analysis of ARMA models has a vast literature and the references to this context include Zellner (1971), Monahan (1983), Marriot and Smith (1992), Chib and Greenberg (1994), Marriot et al. (1996), Kleinbergen and Hoek (2000), Fan and Yao (2008) and Tripathi et al. (2017), etc. Among these, a sophisticated numerical integration technique was used by Monahan (1983) and that was later extended by Marriot and Smith (1992). Considered as a breakthrough in the analysis of ARMA models, the study done by Chib and Greenberg (1994) was the first to use the Markov chain Monte Carlo (MCMC) technique. In this paper, the authors relied on state space version of the model. Although based on more realistic assumptions than those used in the preceding works, Chib and Greenberg (1994) do have the disadvantage of carrying out the analysis on only a subset of the parameter set. Marriot et al. (1996) is another significant reference where the authors developed sampling-based approach for the estimation of parameters of ARMA model and its components. They suggested sampling from the conditional densities of AR and MA coefficients subject to the restriction of stationarity and invertibility. The present paper provides the full Bayesian analysis of ARMA models with special focus on model compatibility and comparison. It is based on a more logical and more computation friendly formulation of the ARMA likelihood in comparison with that of Tripathi et al. (2017). More elaborately speaking, the previous work uses an approximate conditional likelihood. The present work brings an improvement in the analysis by considering the joint density of the previous observations in addition to the conditional likelihood (as suggested by Box et al. (2004)). Thus, this is a closer approximation to the exact likelihood, which is given by Newbold et al. (1974). Naturally, the analysis proposed in the present paper appears to be more accurate.

To the best of our belief, none of the papers on ARMA models addresses the problems of verifying model compatibility as well as model comparison using the tools of the Bayesian paradigm. Since the idea of prediction is an integral part to any time series analysis, we use this to verify the compatibility of the considered models. The basic idea is to judge whether the predicted data are in compliance with the observed data. The discrepancy, if any, is

quantified by the calculation of the Bayesian p-value, which essentially uses a discrepancy measure, the Ljung-Box-Pierce statistic in our case. Realising the indispensability of the idea of prediction, we have further used this notion in comparing the different ARMA sub-models considered in the paper. It may be noted that the earlier papers mostly considered the Box-Jenkins methodology to select the models (see, for example, Reilly (1980), Tsay and Tiao (1984, 1985), Reynolds et al. (1995), Tripathi et al. (2018), etc.). This does not appear logical in some sense and, in no way, complies with the Bayesian paradigm. On the other hand, we have used the predictive loss criterion that successfully merges the ideas of prediction and the loss incurred thereof, making it true to Bayesian sensibilities.

Recently, Tripathi et al. (2017) performed an approximate Bayes analysis of ARMA model (1) and used the GDP growth rate data of India to illustrate their procedure. In this paper, the authors resorted to using an approximate form of the likelihood by considering the values of observations and the subsequent error terms, prior to the very first observation, say y_1, to be zero. Although the approximations were made keeping in mind the computational ease, it does have a logical lacuna to some extent. If one considers time series processes such as the GDP growth rate data, one must remember that the dataset considered is actually a part of a large data series and, in reality, there are non-zero observations before y_1. The present treatment of the problem, therefore, adds a reasonable amount of soundness by considering a more logical approximation of the likelihood function for the ARMA model and its components, based on a line of suggestion by Box et al. (2004). As a further extension of the work done by Tripathi et al. (2017), we used a Gibbs-Metropolis hybrid scheme to perform the complete Bayesian analysis of ARMA models instead of pure Gibbs sampler using adaptive-rejection sampling (see Tripathi et al. (2017)).

Undoubtedly, the ARMA model has the capability to model a variety of observations on time series data probably because of its generality and flexibility. The present paper is no more an exception and considers a time series data on GDP growth rate of India at constant prices (considering base year to be 2004-05) collected over a period of 1951-52 to 2013-14 and uses the ARMA model to explain and analyze the data (see also Tripathi et al. (2017)). The use of ARMA models to explain GDP growth rate data is quite prevalent in the literature (see, for example, Morley et al. (2003) and Ludlow and Enders (2000)). We can motivate the model based on two lines of thought. First, GDP growth in one quarter is obviously affected by the same in other quarters. It is similar to the situation when the current value in a time series can be considered to depend upon the lagged observations. Second, the GDP observations at a particular point of time are not only affected by the random shocks at that time, but also by the shocks (such as a natural disaster) that have taken place earlier. Hence, an ARMA model that takes care of these two aspects of modelling is quite plausible. It may be noted that this motivation is general and can be applied to other time series as well.

It is our understanding that the choice of a particular model cannot be completely pre-specified by theoretical consideration alone, rather it should be selected from many competing models using some model selection criteria. In most of the studies using the ARMA model, the particular model is selected using the Box-Jenkins methodology, that is, with the help of autocorrelation and partial autocorrelation functions (see, for example, Tsay and Tiao (1984, 1985) and Pankratz (1983)). Forecasting or predicting is an indispensable element of time series. Gelfand and Ghosh (1998) suggested a criterion based on minimization

of posterior predictive loss (PPL) arising due to a model. Going by this criterion, models are rewarded not only for their predictive capabilities, but also for their fidelity to the observed data. Since in time series analysis prediction is the ultimate objective, the proposed model selection criterion seems to be quite appealing while keeping in mind the criterion of prediction. We finally look into retrospective prediction because of the general belief that a model which predicts well retrospectively is expected to do well at least for the short term prospective prediction (see, for example, Tripathi et al. (2017, 2018)) in majority of cases, although not always.

The outline of the paper is as follows. The next section provides the stationarity and invertibility conditions for the various components of the considered ARMA model. Section 3 provides the Bayesian formulation of the ARMA model for its possible posterior analysis. The analysis is extended for AR and MA components as well, although the developments are routine once the general formulation for the ARMA model is obtained. Vague priors are used for the model parameters and posterior analysis is performed using Gibbs sampler algorithm after imposing the necessary conditions for stationarity and invertibility. The algorithm is actually a hybrid scheme with Metropolis algorithm within the Gibbs sampler as some of the full conditionals are not available for routine sample generation. Finally, an explanation is given for getting predictive samples once the posterior samples are obtained. Section 4 discusses the model compatibility issues based on Ljung-Box-Pierce statistic. The section also comments briefly on the PPL criterion of Gelfand and Ghosh (1998) as a tool for comparison of the compatible models. Section 5 considers the Indian GDP growth rate data and analyses the same using the formulation given in the previous sections. The analysis is done for a number of combinations of p and q such that $p+q \leq 2$ and finally the entertained models are compared using the PPL criterion. Some numeric evidences for the adequacy of the models are given and, also, the short term retrospective prediction based on the final selected model is considered. The last section is a brief conclusion that summarizes our general findings. The paper also has an appendix with some supplementary developments required for the completeness.

2. Stationarity and Invertibility conditions

In order that a time series model becomes reasonable, two vital conditions, namely the stationarity and the invertibility, need to be verified. The stationarity and invertibility conditions in the ARMA(p,q) $(p+q \leq 2)$ model can be defined separately in terms of its AR and MA components, respectively. The exact forms have been extensively worked out in the literature (see, for example, Pankratz (1983) and Box et al. (2004)). Without going into the various aspects of deriving these conditions, we can directly state them as follows. For AR(1) or ARMA(1,0), the condition of stationarity is simply, $|\phi_1| < 1$ whereas for AR(2) or ARMA(2,0), these conditions are

$$|\phi_2| < 1, \tag{2}$$

$$\phi_1 + \phi_2 < 1, \tag{3}$$

and
$$\phi_2 - \phi_1 < 1. \tag{4}$$

Similarly for MA(1) or ARMA(0, 1), the condition of invertibility is simply $|\psi_1| < 1$ whereas for MA(2) or ARMA(0, 2), these conditions are

$$|\psi_2| < 1, \tag{5}$$

$$-\psi_1 - \psi_2 < 1, \tag{6}$$

and
$$\psi_1 - \psi_2 < 1. \tag{7}$$

Also, for the ARMA(1, 1) model the conditions of stationarity as well as invertibility are applied simultaneously and these are simply,

$$|\phi_1| < 1 \qquad \text{and} \qquad |\psi_1| < 1. \tag{8}$$

We shall denote the stationarity and invertibility regions for AR(p) and MA(q) models by C_p and C_q, respectively. Non-stationarity in a time series can occur in several ways and should be handled accordingly. Say, for instance, non-stationarity in the variance can be checked by considering some appropriate transformations on the variates whereas non-stationarity in the location can be checked by using differenced data instead (see, for example, Shumway and Stoffer (2011)). It has been observed that the stationarity and invertibility conditions are very complicated for higher order ($p > 2$ and $q > 2$) ARMA models (see Marriot et al. (1996)) and this is perhaps the reason that higher order is often avoided in a true sense. Exception includes Okereke et al. (2015), where the authors derived and illustrated the consequences of the invertibility conditions on the parameters of MA process of order three. We shall not go into the details of such situations although the interested readers may refer to Fan and Yao (2008), Box et al. (2004), etc.

Several tests have been proposed to check the stationarity of a time series. Among them, two are commonly used in the time series literature. These are augmented Dickey-Fuller (ADF) test (see, for example, Dickey and Fuller (1979)) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test (see, for example, Kwiatkowski et al. (1992)). ADF test is a unit root test and it is based on t-test of the coefficients of a generalized AR process. KPSS test, on the other hand, tests the stationarity of the process and assumes that the time series can be represented as the sum of a deterministic trend, a random walk and a stationarity error. The final conclusion in both the tests is normally drawn on the basis of p-values. For more details on the two tests, one may refer to Dickey and Fuller (1979) and Kwiatkowski et al. (1992).

3. Bayesian Model Formulation and Posterior Simulation

Let $\underline{y} : y_1, y_2, ..., y_T$ be the time series observations from the assumed ARMA(p, q) model. The exact likelihood function corresponding to ARMA(p, q) model is always difficult to write since the observation at any stage depends on its p lagged observations and we may not have lagged observations corresponding to the first p observed time series data sets. Newbold (1974) gave the exact form of likelihood for the general ARMA(p, q) model that

is certainly pragmatic but quite difficult at least computationally except in the case of very small sample sizes. Tools such as those based on Monte Carlo or other sample-based approaches also do not support much if employed directly on the likelihood suggested by Newbold (1974). As an alternative, Marriot et al. (1996) suggested a computational friendly form of the likelihood function for ARMA model by introducing the latent variables, $\mathbf{y_0} = (y_0, y_{-1}, ..., y_{1-p})$ and $\varepsilon_\mathbf{0} = (\varepsilon_0, \varepsilon_{-1}, ..., \varepsilon_{1-q})$ into the existing set of unknowns. This form was certainly easy to implement but resulted in the increase of dimensionality of (unknowns) parameter space.

Tripathi et al. (2017) recently used an approximate form of the likelihood function of a general ARMA (p,q) model. This approximation, although easy to implement, has a limitation in the sense that it assumes all the components of $\mathbf{y_0}$ and $\varepsilon_\mathbf{0}$ as zero, providing no contribution of these components in the likelihood function. An alternative strategy was suggested by Box et al. (2004) although the strategy provided an approximation to the exact likelihood. Since the latent variables $\mathbf{y_0}$ and $\varepsilon_\mathbf{0}$ are unknowns, Box et al. (2004) suggested to consider likelihood as the product of two terms where the first term is the joint density of first p observations $(y_1, y_2, ..., y_p)$ and the second term may be considered as the product of conditional density of remaining observations with each observation conditioned on its p lagged observations. This approach is certainly an approximation but makes sense when T is large but p is comparatively small. Moreover, the components of $\varepsilon_\mathbf{0}$ can be easily taken to be zero. Obviously, the resulting likelihood function can be written as

$$L(y_1, y_2, ..., y_T | \theta_0, \Phi, \Psi) = f(y_1, y_2, ..., y_p | \theta_0, \Phi, \Psi) \times \prod_{t=p+1}^{T} f(y_t | y_{t-1}, y_{t-2}, ..., y_{t-p}; \theta_0, \Phi, \Psi),$$
(9)

where $\Phi = (\phi_1, ..., \phi_p)$ and $\Psi = (\psi_1, ..., \psi_q)$. The likelihood in (9) can be considered as an extension of the likelihood proposed by Tripathi et al. (2017) and it can be obtained, up to proportionality, as (see Appendix)

$$L(y_1, y_2, ..., y_T | \theta_0, \Phi, \Psi) \propto \left(\frac{1}{\sigma^2} \right)^{T/2} \times |V_{\phi, \psi}|^{-1/2} \times$$
$$\exp\left(-\frac{1}{2\sigma^2} \{ (Y_p - \mu_p)' V_{\phi, \psi}^{-1} (Y_p - \mu_p) + \sum_{t=p+1}^{T} (y_t - \theta_0 - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{j=1}^{q} \psi_j \varepsilon_{t-j})^2 \} \right).$$
(10)

It may be noted that for Bayesian implementation, we do not require the exact likelihood, but rather the likelihood defined up to proportionality is sufficient. Moreover, the two-component AR(p) and MA(q) models can be obtained from the general ARMA(p,q) model just by ignoring the MA and AR terms, respectively. The likelihood for AR(p) model

can be obtained, up to proportionality, as

$$
L_{AR}(y_1, y_2, ..., y_T | \theta_0, \Phi) \propto \left(\frac{1}{\sigma^2} \right)^{T/2} \times |V_\phi|^{-1/2} \times
$$

$$
\exp \left(-\frac{1}{2\sigma^2} \{ (Y_p - \mu_p)' V_\phi^{-1} (Y_p - \mu_p) + \sum_{t=p+1}^{T} (y_t - \theta_0 - \sum_{i=1}^{p} \phi_i y_{t-i})^2 \} \right),
$$

$$(11)$$

where V_ϕ denotes a $p \times p$ matrix involving the terms of ϕ's that may be determined by the relationship $\Sigma_p = \sigma^2 V_\phi$. Similarly, the likelihood function for MA(q) model can be obtained as the joint distribution of the observations and it can be obtained, up to proportionality, as

$$
L_{MA}(y_1, y_2, ..., y_T | \theta_0, \Psi) \propto \left(\sigma^2 \right)^{\frac{-T}{2}} \times |V_\psi|^{-1/2} \times \exp \left(-\frac{1}{2\sigma^2} (Y_T - \mu_T)' V_\psi^{-1} (Y_T - \mu_T) \right),
$$

$$(12)$$

where V_ψ is obtained from the relation $\Sigma_T = \sigma^2 V_\psi$ (see also Appendix for other relevant details).

3.1. Priors

Besides the likelihood function, another important component in any Bayesian analysis is the specification of prior distribution. If one has enough *a priori* information, it is advisable to go with the informative prior. The prior distribution in that case will have a dominant role in posterior-based inferences. If enough details on *a priori* evidence is not available to go for informative prior, it is often suggested that one should instead use a vague prior or a weakly informative prior. It is to be noted that a weakly informative prior can be very well described by a proper prior density with large to very large scatteredness. Obviously, if the considered prior is weak, the posterior will be completely dominated by the likelihood function and the inferences can be said to depend exclusively on the data-based information. Such a consideration will of course remove the possibility of any negative impact that inappropriately chosen strong *a priori* belief could have had on the analysis. Keeping this in mind, we have considered the following non-informative priors similar to those proposed by Tripathi et al. (2017) and defined under stationarity and/or invertibility restrictions as detailed in Section 2. The considered priors are

$$
\pi_1(\sigma^2) \propto \frac{1}{\sigma^2}; \qquad \sigma^2 \geq 0, \tag{13}
$$

$$
\pi_2(\theta_0) \propto U[-M, M]; \qquad M > 0, \tag{14}
$$

$$\pi_3(\phi_i) \propto U[-N_1, N_1]; \qquad N_1 > 0, \qquad i = 1, 2, ..., p, \tag{15}$$

and

$$\pi_4(\psi_j) \propto U[-N_2, N_2]; \qquad N_2 > 0, \qquad j = 1, 2, ..., q, \tag{16}$$

where the constants M and N_i, $i = 1, 2$, used in the priors are the hyperparameters and $U[a, b]$ is the uniform distribution over the interval $[a, b]$. The prior distribution (13) is obviously a Jeffreys' type of prior for the scale parameter and it has often been suggested in the literature by a number of authors (see, for example, Marriot et al. (1996) and Kleinbergen and Hoek (2000)). The ranges of uniform distributions can, in general, be recommended large enough in order that the priors remain vague over the corresponding intervals. Moreover, since stationarity and invertibility conditions are essential requirements, it is necessary that we restrict our parameters ϕ_i and ψ_j in (15) and (16) to lie in the regions of stationarity and invertibility, that is, C_p and C_q, respectively, as defined in Section 2. We thus consider the priors for ϕ_i and ψ_j to be uniform distributions defined over the regions C_p and C_q. That is the value of the hyperparameters N_1 and N_2 are so chosen that the stationarity and invertibility conditions are satisfied. We must also keep in mind that the restrictions were calculated only for models satisfying the condition $p + q \leq 2$.

3.2. Posterior Distributions

Updating the prior distributions (13) to (16) with the likelihood (10) via Bayes theorem yields the joint posterior distribution for the parameters of an ARMA(p, q) model which, up to proportionality, can be written as

$$p(\theta_0, \Phi, \Psi, \sigma^2 | \underline{y}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{T}{2}+1} \times |V_{\phi,\psi}|^{-1/2} \times$$

$$\exp\left(-\frac{1}{2\sigma^2}\{(Y_p - \mu_p)'V_{\phi,\psi}^{-1}(Y_p - \mu_p) + \sum_{t=p+1}^{T}(y_t - \theta_0 - \sum_{i=1}^{p}\phi_i y_{t-i} - \sum_{j=1}^{q}\psi_j \varepsilon_{t-j})^2\}\right)$$

$$\times I_{[-M,M]}(\theta_0) \times \prod_{i=1}^{p}I_{[-N_1,N_1]}(\phi_i) \times \prod_{j=1}^{q}I_{[-N_2,N_2]}(\psi_j), \tag{17}$$

where $I_{[v_1, v_2]}(.)$ is the indicator function that takes value unity if (.) belongs to $[v_1, v_2]$ and zero otherwise.

Next, combining the prior distributions (13) to (15) with the likelihood (11) via Bayes theorem yields the joint posterior distribution for the parameters of an AR(p) model, which

can be written, up to proportionality, as

$$p_{AR}(\theta_0, \Phi, \sigma^2 | \underline{y}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{T}{2}+1} \times |V_\phi|^{-1/2} \times$$

$$\exp\left(-\frac{1}{2\sigma^2}\{(Y_p - \mu_p)'V_\phi^{-1}(Y_p - \mu_p) + \sum_{t=p+1}^{T} (y_t - \theta_0 - \sum_{i=1}^{p} \phi_i y_{t-i})^2\}\right)$$

$$\times I_{[-M,M]}(\theta_0) \times \prod_{i=1}^{p} I_{[-N_1,N_1]}(\phi_i).$$

$$(18)$$

Similarly, the posterior distribution for the parameters of a MA(q) model can be obtained up to proportionality as

$$p_{MA}(\theta_0, \Psi, \sigma^2 | \underline{y}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{T}{2}+1} \times |V_\psi|^{-1/2} \times \exp\left(-\frac{1}{2\sigma^2}\{(Y_T - \mu_T)'V_\psi^{-1}(Y_T - \mu_T)\}\right)$$

$$\times I_{[-M,M]}(\theta_0) \times \prod_{i=1}^{q} I_{[-N_2,N_2]}(\psi_i).$$

$$(19)$$

Obviously, this posterior is the result of updating of prior distributions (13), (14) and (16) through the likelihood (12) via the Bayes theorem.

The posteriors (17), (18) and (19) are analytically intractable and cannot be obtained in nice closed forms. We, therefore, do not have many options except going for sample-based approaches and then drawing the corresponding inferences based on the simulated samples from the corresponding posteriors. In the next subsection, we shall discuss briefly the implementation of our proposed MCMC scheme and model estimation.

3.3. MCMC Implementation and Model Estimation

Markov chain Monte Carlo procedures basically construct a Markov chain such that simulating from its stationary distribution renders samples from the target posterior. The Gibbs sampler requires that the target posterior may be reduced into full conditionals corresponding to every variate. The algorithm progresses by simulating from, often unidimensional, full conditionals in a cyclic fashion. Since the implementation of Gibbs sampler requires simulation from various full conditionals, it is indispensable that all the full conditionals are routinely available for sample generation. We may check that this latter requirement may not be easily ensured for all the full conditionals through any standard sample generating schemes. We, therefore, make use of the Metropolis algorithm for such full conditionals although procedures are there where indirect strategies can be always developed. The Metropolis algorithm implemented separately on some of the full conditionals works by generating a probable variate value from a proposal density. The resulting variate value is

accepted if it has the large posterior probability. This is decided with the help of an acceptance probability. Choosing the mean and variance of the proposal density is a vital decision. The mean of the proposal is usually taken as the maximum likelihood (ML) estimate and its variance can be taken as the inverse of observed Fisher's information. Moreover, as far as the variance is concerned, it must be noted that if the variance is too large, some of the generated variate values will be quite far away from the current value leading to rejection. On the other hand, if the variance is too small, the chain will take more time to cover the entire support of the density with slightly low probability regions being under-sampled. Hence, a properly centered and dispersed kernel is highly essential. We often use a tuning constant 'c' to make this adjustment. We skip further discussions on the Gibbs sampler and the Metropolis algorithms for want of space and refer Gelfand and Smith (1991) and Upadhyay and Smith (1994) for details.

For the posterior corresponding to the ARMA(p,q) model given in (17), the full conditionals up to proportionality for different parameters can be written as

$$p_1(\theta_0|\sigma^2,\Phi,\Psi,\underline{y}) \propto$$
$$\exp\left(-\frac{1}{2\sigma^2}\{(Y_p-\mu_p)'V_{\phi,\psi}^{-1}(Y_p-\mu_p) + \sum_{t=p+1}^{T}(y_t-\theta_0-\sum_{i=1}^{p}\phi_i y_{t-i}-\sum_{j=1}^{q}\psi_j\varepsilon_{t-j})^2\}\right),$$
$$(20)$$

$$p_2(\phi_i|\sigma^2,\theta_0,\Psi,\underline{y}) \propto |V_{\phi,\psi}|^{-1/2}\times$$
$$\exp\left(-\frac{1}{2\sigma^2}\{(Y_p-\mu_p)'V_{\phi,\psi}^{-1}(Y_p-\mu_p) + \sum_{t=p+1}^{T}(y_t-\theta_0-\sum_{i=1}^{p}\phi_i y_{t-i}-\sum_{j=1}^{q}\psi_j\varepsilon_{t-j})^2\}\right);$$
$$i=1,2,...,p, \quad (21)$$

$$p_3(\psi_j|\sigma^2,\theta_0,\Phi,\underline{y}) \propto |V_{\phi,\psi}|^{-1/2}\times$$
$$\exp\left(-\frac{1}{2\sigma^2}\{(Y_p-\mu_p)'V_{\phi,\psi}^{-1}(Y_p-\mu_p) + \sum_{t=p+1}^{T}(y_t-\theta_0-\sum_{i=1}^{p}\phi_i y_{t-i}-\sum_{j=1}^{q}\psi_j\varepsilon_{t-j})^2\}\right);$$
$$j=1,2,...,q, \quad (22)$$

and

$$p_4(\sigma^2|\theta_0,\Phi,\Psi,\underline{y}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{T}{2}+1}\times$$
$$\exp\left(-\frac{1}{2\sigma^2}\{(Y_p-\mu_p)'V_{\phi,\psi}^{-1}(Y_p-\mu_p) + \sum_{t=p+1}^{T}(y_t-\theta_0-\sum_{i=1}^{p}\phi_i y_{t-i}-\sum_{j=1}^{q}\psi_j\varepsilon_{t-j})^2\}\right).$$
$$(23)$$

On the other hand, if we consider the posterior (18) corresponding to a particular case of AR(p) model, the corresponding full conditionals are

$$p_5(\theta_0|\sigma^2,\Phi,\underline{y}) \propto$$

$$\exp\left(-\frac{1}{2\sigma^2}\{(Y_p-\mu_p)'V_\phi^{-1}(Y_p-\mu_p) + \sum_{t=p+1}^{T}(y_t-\theta_0-\sum_{i=1}^{p}\phi_i y_{t-i})^2\}\right), \quad (24)$$

$$p_6(\phi_i|\sigma^2,\theta_0,\underline{y}) \propto |V_\phi|^{-1/2} \times$$

$$\exp\left(-\frac{1}{2\sigma^2}\{(Y_p-\mu_p)'V_\phi^{-1}(Y_p-\mu_p) + \sum_{t=p+1}^{T}(y_t-\theta_0-\sum_{i=1}^{p}\phi_i y_{t-i})^2\}\right);$$

$$i = 1,2,...,p \quad (25)$$

and

$$p_7(\sigma^2|\theta_0,\Phi,\underline{y}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{T}{2}+1} \times$$

$$\exp\left(-\frac{1}{2\sigma^2}\{(Y_p-\mu_p)'V_\phi^{-1}(Y_p-\mu_p) + \sum_{t=p+1}^{T}(y_t-\theta_0-\sum_{i=1}^{p}\phi_i y_{t-i})^2\}\right). \quad (26)$$

Similarly, MA(q) model results in the full conditionals up to proportionality obtained from the posterior (19) as

$$p_8\left(\theta_0|\sigma^2,\Psi,\underline{y}\right) \propto \exp\left(-\frac{1}{2\sigma^2}\{(Y_T-\mu_T)'V_\psi^{-1}(Y_T-\mu_T)\}\right), \quad (27)$$

$$p_9(\psi_j|\sigma^2,\theta_0,\underline{y}) \propto |V_\psi|^{-1/2} \times \exp\left(-\frac{1}{2\sigma^2}\{(Y_T-\mu_T)'V_\psi^{-1}(Y_T-\mu_T)\}\right);$$

$$j = 1,2,...,q \quad (28)$$

and

$$p_{10}(\sigma^2|\theta_0,\Psi,\underline{y}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{T}{2}+1} \times \exp\left(-\frac{1}{2\sigma^2}\{(Y_T-\mu_T)'V_\psi^{-1}(Y_T-\mu_T)\}\right). \quad (29)$$

Obviously, we have a total of $(p+q+2)$ full conditionals corresponding to a general ARMA(p,q) model given by (1). Similarly, a total of $(p+2)$ full conditionals corresponding to a general AR(p) model and a total of $(q+2)$ full conditionals corresponding to a general MA(q) model are obtained. Among the various full conditionals, samples from (23), (26) and (29) corresponding to the posterior distributions of σ^2 can be obtained using a gamma generating routine after making the transformation $\tau = 1/\sigma^2$. It can be noted that the gamma

distribution, with shape and scale parameters $\alpha > 0$ and $\beta > 0$, respectively, for a random variable X is defined as;

$$g(x) = \frac{\beta^{\alpha}}{\Gamma \alpha} x^{\alpha-1} e^{-\beta x}; \;\; 0 < x < \infty, \tag{30}$$

where

$$\Gamma \alpha = \int_0^{\infty} e^{-x} x^{\alpha-1} dx.$$

Thus, it can be seen that the transformed variate τ follows a gamma density with shape parameter $\alpha = \frac{T}{2}$ and scale parameter $\beta = \frac{1}{2} \{ (Y_p - \mu_p)' V_{\phi, \psi}^{-1} (Y_p - \mu_p) + \sum_{t=p+1}^{T} (y_t - \theta_0 - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{j=1}^{q} \psi_j \varepsilon_{t-j})^2 \}$ in the case of ARMA(p,q) model. In the case of AR(p) model, τ follows a gamma density with shape parameter $\alpha = \frac{T}{2}$ and scale parameter $\beta = \frac{1}{2} \{ (Y_p - \mu_p)' V_{\phi}^{-1} (Y_p - \mu_p) + \sum_{t=p+1}^{T} (y_t - \theta_0 - \sum_{i=1}^{p} \phi_i y_{t-i})^2 \}$. Similarly, in the case of MA(q) model, τ follows a gamma density with shape parameter $\alpha = \frac{T}{2}$ and the scale parameter $\beta = \frac{1}{2} \{ (Y_T - \mu_T)' V_{\psi}^{-1} (Y_T - \mu_T) \}$.

The full conditionals (20) to (22), for each i and j, are not available from the point of view of sample generation although they may be managed easily by using the Metropolis algorithm. Our final scheme can, therefore, be a kind of hybrid (Metropolis within Gibbs) because of these full conditionals and the same can be referred to as the Gibbs-Metropolis hybrid sampler. Similar hybrid schemes can be thought of for generating samples from (24) to (25), for each i, corresponding to AR(p) model and from (27) to (28), for each j, corresponding to MA(q) model. For Metropolis implementation separately on each of these full conditionals, one can use univariate normal candidate generating density with mean taken as the current realization and standard deviation to be approximated on the basis of the particular element of the Hessian matrix obtained at ML estimates. It is important to mention that although the candidate generating density is univariate and the corresponding full conditionals are each univariate, the ML estimates and the corresponding Hessian-based approximations are obtained for a multi-parameter likelihood function. Moreover, as mentioned earlier, the value of the standard deviation so obtained is adjusted by multiplying with a scaling constant c, taken in the range 0.5 and 1.0 (see, for example, Upadhyay and Smith (1994)) for reducing the number of rejections in the Metropolis step. The other important thing in the implementation of the Metropolis algorithm is the choice of initial values for running the chain. One can, of course, use any standard classical estimates to begin running the chain. We have used the ML estimates in particular. For relevant details on the Metropolis algorithm and issues on its convergence diagnostics, one can refer to Smith and Roberts (1993) and Upadhyay et al. (2001), among others.

Thus, the Gibbs-Metropolis hybrid sampler strategy can be easily implemented on the posteriors given in (17), (18) and (19) corresponding to ARMA, AR and MA models, respectively. The Gibbs-Metropolis hybrid strategy is being referred to simply because some of the full conditionals do not ascribe to any standard form of distributions and, as such, the generations are difficult. We, therefore, implement intermediate Metropolis steps for

generating variate values from the same and base our decision on a single long run of the chain. Obviously, the implementation of the algorithm results in a long sequence of iterating chains. After sufficiently large number of iterations, the generating sequence converges in distribution to random samples from the target joint posterior and the generated components to the corresponding marginal posteriors (see, for example, Smith and Roberts (1993)). In order to get the final samples, one has to discard the initial burn-in samples and then pick up the variate values at equidistant intervals to minimize serial correlation among the generating variates (see, for example, Upadhyay et al. (2012)). These final samples can be used in a variety of ways to draw the desired posterior-based inferences. Say, for instance, one can use sample-based posterior estimates for the model parameters or one can estimate the entire posterior densities by means of some nonparametric density estimates. Other desired features of the posteriors can also be likewise obtained once the posterior samples are made available (see also Upadhyay et al. (2012)).

3.4. Prediction

As mentioned, an important part of our analysis includes predicting the unobserved future data y_{T+1} given the informative data $\underline{y} = y_1, y_2, ..., y_T$. Now, one can easily confirm from (1) that for the given set of observations $\underline{y}$, the future observation y_{T+1} follows normal distribution with mean

$$\mu_{T+1} = \theta_0 + \sum_{i=1}^{p} \phi_i y_{t+1-i} + \sum_{j=1}^{q} \psi_j \varepsilon_{t+1-j} \qquad (31)$$

and variance σ^2. Thus, the future observation y_{T+1} can be easily simulated from this normal distribution after replacing the corresponding parameters by their appropriately chosen posterior estimates (say, for example, estimated posterior modes) obtained on the basis of final posterior simulated samples. The ε_t's in (1) can be simulated from the normal density with mean zero and variance equal to the corresponding posterior estimate of σ^2. This strategy can be easily applied to get the predictive samples of y_{T+1} from which any desired sample-based predictive characteristic can be assessed. It is essential to mention here that the posterior mode is the highest probable value and, therefore, it makes sense if the resulting posterior distribution is non-symmetric. On the other hand, if the resulting posterior distribution is symmetric, it is immaterial whether one uses mean or median or mode. Moreover, in the case of AR and MA models, one can simply ignore the MA and AR components, respectively, in the general form of the ARMA model (1) and proceed to obtain the corresponding predictive samples. It may, however, be noted that one requires the posterior samples corresponding to AR and MA models.

4. Model Compatibility

A model compatibility study can be performed using a variety of tools. An important one among these may utilize the idea of predictive simulation where the predictive observations are obtained from the model under consideration and then compared with the observed

data $\underline{y}$. Obviously, a model may be considered compatible with the observed data if its predictive output compliances with the former. In the case of no or poor resemblance of the two data sets, the considered model creates a suspicion and cannot be recommended for the data in hand. Among other things, this resemblance can often be measured by means of a quantitative summary in the form of p-value that can be obtained using an appropriate model discriminating statistic. Informally, graphical tools are also suggested in the literature where some characteristics based on the two data sets may be shown on the same graphical scale (see, for example, Gelman et al. (1996), Bayarri and Berger (1998), Upadhyay and Peshwani (2003), etc.).

The present study, however, begins with an informal approach where simple time series plots for the observed and the predicted data from the model(s) are shown graphically in a way that the plots corresponding to the latter are superimposed over that corresponding to the former. For p-value based study of model compatibility, we advocate for a Bayesian version of the same based on an appropriately chosen statistic. Obviously, if the calculated p-value happens to be large, the considered model may be regarded compatible with the data.

The Ljung-Box-Pierce statistic that follows a chi-square distribution if the null hypothesis is true, is commonly used in autoregressive integrated moving average (ARIMA) modelling to test the overall randomness behaviour based on a number of lags. It may be noted that the test is applied on the residuals of a fitted ARIMA model and not on the original series. Thus, the hypothesis actually being tested in this case is that the residuals from the ARIMA model have no autocorrelation and that the model is adequate. When testing the residuals of an estimated ARIMA model, the degrees of freedom need to be adjusted to reflect the parameter estimation. For example, for an ARIMA$(p,0,q)$ or ARMA(p,q) model, the degrees of freedom should be set to $(l - p - q)$, where l is the number of lags, that is, the order of autocorrelation being tested and the correction due to p and q is because of the fact that the degrees of freedom must account for the estimated model parameters. The statistic, used to test for uncorrelated residuals, is then calculated by

$$Q(l) = T(T+2) \sum_{j=1}^{l} \frac{r_j^2(\hat{\varepsilon}_t)}{T - j}, \tag{32}$$

where

$$r_j^2(\hat{\varepsilon}_t) = \frac{\sum_{t=1}^{T-j} \hat{\varepsilon}_t \hat{\varepsilon}_{t+j}}{\sum_{t=1}^{T} \hat{\varepsilon}_t^2}, \tag{33}$$

$\hat{\varepsilon}_t = y_t - \hat{y}_t$ and the statistic $Q(l)$ follows chi-square distribution with $(l - p - q)$ degrees of freedom. Here $\hat{y}_t$ is the predicted value of y_t that can be obtained from (1) in a way described in subsection 3.4, T is the number of residuals computed for the model and $\hat{\varepsilon}_t$ is the residual at time t. This statistic was advocated by Ljung and Box (1978) and is often referred to as the Ljung-Box statistic or the Ljung-Box-Pierce statistic. The Bayesian p-value can then be calculated by the probability P^{post} (based on the posterior samples) of acceptance of the

null hypothesis and it is given by

$$pval = P^{post}[Q(l) < \chi^2_{(l-p-q),(1-\alpha)}],$$ (34)

where $\chi^2_{(l-p-q),(1-\alpha)}$ is the tabulated value of chi-square at $(l-p-q)$ degrees of freedom and the level of significance α. The model is adequate if this probability is large, that is, at least larger than the assumed significance level (usually 0.05). There is, however, very little practical advice on how to choose the number of lags for the test. A recommendation based on power considerations is to consider $l = 10$ for non-seasonal data and $l = 2k$ for seasonal data where k is the period of seasonality (see, for example, Hyndman and Athanasopoulos (2018)). Needless to mention that we want to ensure that l is large enough to capture any meaningful and troublesome correlations. So for our data set, we prefer to choose $l = 10$ to increase the power of the test.

4.1. Model Comparison: A PPL Approach

After the compatibility of a model is ascertained, we go for its comparison with other compatible models. Model comparison is intuitively based on two major criteria - its fitting to the observed data and its inherent complexity. Most of the model comparison tools, Bayesian or otherwise, are based on a weighted trade-off between the two criteria, the weights being decided according to some specific needs. One such criterion, known as PPL criterion, was initially given by Gelfand and Ghosh (1998). Based on predictive simulation, this criterion parallels to standard utility ideas and partitions the total loss into loss due to fit and loss due to complexity (see also Upadhyay and Mukherjee (2008)). A simplified version of this criterion was given by Sahu and Dey (2000) (see also Upadhyay et al. (2012)). This criterion recommends a model m that minimizes the joint effect of two measures, namely, the closeness of observed and predictive data sets and variability of the prediction. The criterion can be defined as

$$D(m) = G(m) + P(m)$$ (35)

where $G(m) = \sum_{t=1}^{T} (\mu_t^{(m)} - y_t)^2$, $P(m) = \sum_{t=1}^{T} \sigma_t^{2(m)}$, $\mu_t^{(m)} = E(z_t|y_t,m)$, $\sigma_t^{2(m)} = \text{Var}(z_t|y_t,m)$ and z_t denotes the tth component of predictive data, $t = 1,2,3,...,T$. Obviously, the term $\mu_t^{(m)}$ represents the predictive mean and the term $\sigma_t^{2(m)}$ represents the predictive variance of tth component under the model m.

In (35), $G(m)$ represents the goodness of fit term and it increases when the entertained model provides poor fitting at the observed data points. Similarly, the second term $P(m)$ represents the penalty term and it increases with the increasing complexity in the model. A model m that provides least value of $D(m)$ when compared with all other models, is finally recommended. We are not going into details of its formulation rather refer to Sahu and Dey (2000) (see also Upadhyay et al. (2012) and Tripathi et al. (2017)).

5. Real Data Illustration

For numerical illustration, we consider a real data set on GDP growth rate of India at constant prices. The data set given in Table 5 (see Appendix) is collected over a period of sixty three years, 1951-52 to 2013-14, and is taken from the publication of Central Statistical Organization (CSO) (2014) (see `http://planningcommission.nic.in/data/datatable/0814/comp_databook.pdf`). This data set has been used by a number of authors, a recent reference being Tripathi et al. (2017).
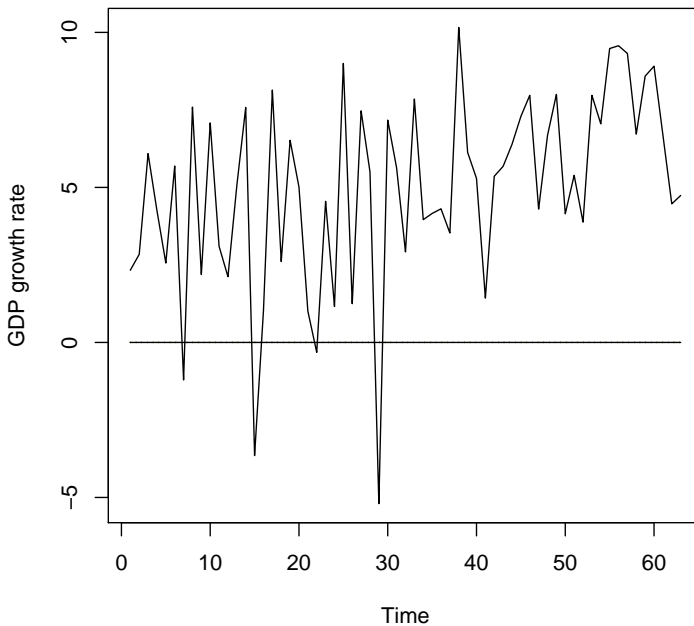

Figure 1: Time series plot showing the GDP growth rate of India (straight line shows the mean level).

Before we begin the intended numerical illustration with GDP growth rate data, let us draw the simple time series plot corresponding to the given GDP observations to see if there is stationarity behaviour in the series. The corresponding time series plot is shown in Figure 1. It can be seen that the series exhibits non-stationarity behaviour indicated by its growth (see Figure 1) and, therefore, it is essential to perform an appropriate transformation (see also Clement (2014)) on the data to remove its non-stationarity behaviour. To resolve this issue, we took first difference of the data and the corresponding time series plot for the transformed data, as shown in Figure 2. Obviously, the figure shows stationarity behaviour. Some outliers at intermediate stages can also be seen, which cause some abrupt hikes in

the first half of the series. Since the study of outliers is beyond the scope of our work, we continue our study up to this level of stationarity only. We further strengthen our claim using some numerical evidences. We assess stationarity of data (or its absence) using the ADF and KPSS tests, first on non-differenced data and then on differenced data. The two tests for stationarity are found to be significant at 5% level and, therefore, provide enough evidence against stationarity of actual data. The p-value in the ADF test is found to be 0.1 and that in the KPSS test is 0.02, which evidently refuse the presence of stationarity in the data. Moreover, after differencing the data, the p-values are found to be 0.01 for the ADF test and 0.1 for the KPSS test, which now ensure the stationarity in the data. Hence, this transformed data can be used for the final analysis.
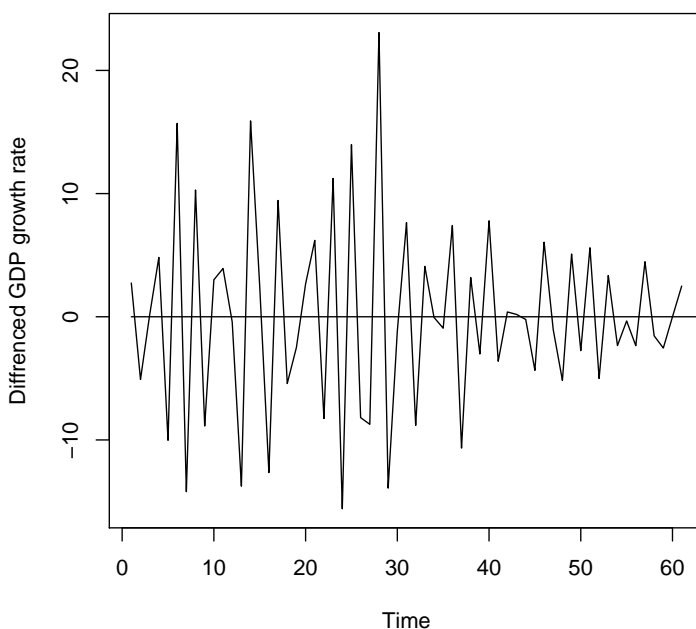

Figure 2: Time series plot corresponding to the first difference of GDP growth rate data (straight line shows the mean level).

Our analysis of ARMA(p,q) was restricted to $p+q \leq 2$ and, therefore, we considered five different ARMA modelling combinations by taking (p,q) as (0, 1), (0, 2), (1, 0), (1, 1) and (2, 0), respectively. These restrictions on the values of p and q were imposed for the ease of performing the analysis and also for the ease of implementation of invertibility conditions as mentioned in Section 2. It is to be noted that for higher values of p and q, as mentioned in Section 2, the invertibility conditions are not available in analytically closed forms and the situation might be difficult to ensure these numerically too.

To perform the complete posterior analysis, we first obtained the ML estimates of the corresponding parameters of the considered models by maximizing the log likelihood function in each case. These ML estimates were utilized as the initial values for the necessary MCMC implementation. The complete posterior analysis was done for each considered model as discussed in Section 3. In order to nullify the prior effect and hence to draw exclusively data dependent inferences, the values of prior hyperparameters M and N_i, $i = 1, 2$, were chosen to be 100 in each case. The priors for AR and/or MA coefficients were chosen under the restrictions of stationarity and invertibility as discussed in Section 2 and subsection 3.1.

Under the prior assumptions as mentioned above, we analyzed the posteriors (17), (18) and (19) using the Gibbs-Metropolis hybrid scheme as discussed in subsection 3.3. It is important to mention that for the full conditional distributions corresponding to the intercept, the AR coefficients and/or the MA coefficients, the corresponding generating algorithm was Metropolis with properly centred and dispersed normal kernel as the proposal density. The mean value of the kernel was approximated by the ML estimate of the corresponding parameter (see Table 1). As far as the standard deviation is concerned, the exact value was difficult to obtain either analytically or numerically. We, therefore, considered its numerical approximation by evaluating the second derivative numerically at the corresponding ML estimate. The need for some adjustments occurred probably because of the approximations that we considered. This adjustment was carried out with the help of a scaling constant, $c = 0.6$, that provided a good acceptance probability in each case.

To get the posterior sample by implementing the Gibbs-Metropolis hybrid sampler, we considered a single long run of the chain. After an initial transient behaviour, convergence of the chain, based on ergodic averages, was observed at about 40K iterations. This cannot be considered as a deterrent issue considering the development in high speed computing that took approximately 3.1423 minutes in 40K iterations. Thus, it can be said that the algorithm works reasonably well. After successfully monitoring the convergence, we picked up equally spaced generated values to form the samples from the corresponding posteriors (see also Upadhyay et al. (2001)). These gaps were taken to be 20 to make the serial correlation negligibly small.

Some of the important estimated posterior characteristics based on the posterior samples of size 1K from each model are shown in Table 1. These estimates are shown in the form of estimated posterior means, medians, modes and 0.95 highest posterior density (HPD) intervals (HPD interval with coverage probability 0.95). Besides, we have also shown the ML estimates of different parameters for each of the considered models, which were actually obtained for defining the initial values for running the proposed Gibbs-Metropolis hybrid algorithm but can also be used for the purpose of comparison with the corresponding Bayes estimates. It can be seen that the ML estimates are, in general, close enough to the corresponding Bayes estimates, a conclusion that is expected for the considered modelling combination. From the posterior estimates shown in Table 1, one can draw several conclusions, the one among these may be to get an overall idea of various estimated posterior densities. It can be seen that the estimated marginal posterior of σ^2 reveals slight positive skewness. The corresponding estimates for other parameters, however, exhibit almost symmetrical behaviour for their posterior densities except for the parameter θ_0 for the first two

models, which reveal negatively skewed posterior densities. It may also be noted that the degree of skewness, in general, is less for ARMA$(0, 1)$ model as compared to all other models. As a final remark, it can be said that these estimates are obtained under the restrictions of stationarity and invertibility and, as mentioned earlier, the results are reasonable from that point of view as well. One can also see the significance of preceding observations in our reported results. It can be seen that the estimates of ϕ's and ψ's are, in general, appreciably larger than the corresponding estimated values for the intercepts. This of course advocates the usability of the considered model in some sense. Moreover, the estimates of σ^2's also convey an important message that increasing complexity in the model, in general, decreases the variability inherent in the model.

Table 1. Posterior summaries for different variates of considered ARMA models based on first difference data

Model	Parameter	ML Estimates	Posterior Means	Posterior Medians	Posterior Modes	0.95 HPD Interval	
ARMA$(1,0)$	θ_0	0.054	0.039	0.057	0.099	-0.829	1.084
	ϕ_1	-0.558	-0.557	-0.555	-0.550	-0.770	-0.346
	σ^2	13.079	14.076	13.777	13.667	9.297	19.491
ARMA$(2,0)$	θ_0	0.065	0.039	0.059	0.124	-0.701	0.924
	ϕ_1	-0.808	-0.791	-0.796	-0.794	-0.981	-0.603
	ϕ_2	-0.435	-0.421	-0.424	-0.444	-0.654	-0.218
	σ^2	10.501	11.374	11.005	10.563	7.178	15.689
ARMA$(0,1)$	θ_0	0.071	0.064	0.067	0.068	-0.043	0.201
	ψ_1	-0.900	-0.884	-0.896	-0.913	-0.998	-0.744
	σ^2	8.015	9.197	9.023	8.779	6.011	12.838
ARMA$(0,2)$	θ_0	0.073	0.053	0.059	0.065	-0.093	0.228
	ψ_1	-1.230	-0.903	-0.925	-0.969	-0.999	-0.743
	ψ_2	0.230	0.074	0.072	0.066	-0.106	0.250
	σ^2	7.595	9.482	9.253	8.745	6.051	13.175
ARMA$(1,1)$	θ_0	0.072	0.083	0.086	0.087	-0.044	0.183
	ϕ_1	-0.195	-0.226	-0.228	-0.251	-0.476	0.019
	ψ_1	-0.990	-0.901	-0.916	-0.970	-0.999	-0.738
	σ^2	7.665	8.738	8.533	8.068	5.876	12.291

We next consider the issue of examining compatibility of the assumed models for the given time series data. Since the stationarity behaviour is established on the basis of first difference data, we shall consider the same set of observations for examining compatibility. Our study is based on some graphical as well as quantitative summaries. Graphically, we have studied it by plotting the time series of observed first difference data along with the corresponding predictive differenced data. For this purpose, 10 predictive samples were generated from each of the considered models using the final posterior samples of size 10 generated using the Gibbs-Metropolis hybrid sampler algorithm (see subsection 3.3). Thus, each posterior sample resulted in one predictive sample of size equal to that of the observed data. We next considered obtaining the first difference from each of the predictive samples. Now, the differenced form of 10 predictive samples are plotted along with the corresponding observed (differenced) data in the form of time series. One such plot corresponding to ARMA$(0, 1)$ model is shown in Figure 3, where the bold line corresponds to first difference of observed time series data. The time series plots corresponding to the first differenced predictive samples are shown by means of dotted lines. It can be seen that the predictive sample plots and the observed data plot exhibit more or less similar overlapping behaviour and, therefore, the ARMA$(0, 1)$ model can be considered compatible for the observed first

differenced data. We had a similar message from the plots corresponding to all other models although the plots are not shown due to space restriction. Thus, all the models can be regarded compatible with the observed time series data.
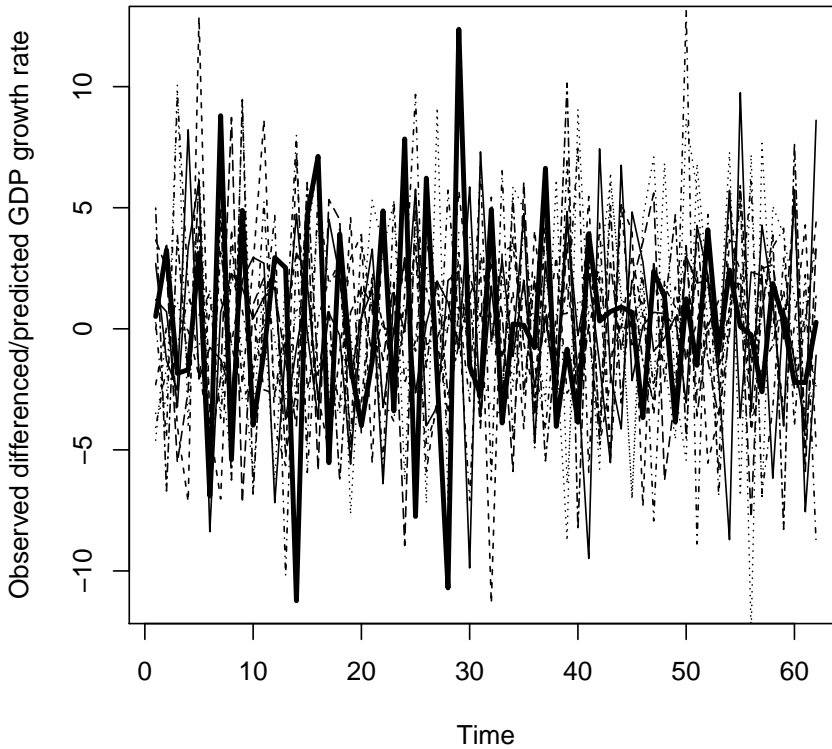

Figure 3: Time series plots for the first difference of observed and predictive data sets from ARMA(0, 1) model (the bold line corresponds to observed data).

For the study of model compatibility based on quantitative evidence, we considered evaluating the Bayesian p-value based on Ljung-Box-Pierce chi-square statistic (see Section 4). It is to be noted that the values of residuals can be obtained once the predictive observations corresponding to each of the original observations are made available. To clarify the computational stages, we first simulated 1K posterior samples as discussed in Section 3.3 and then obtained 1K predictive samples, each predictive sample of size equal to that of the observed time series data. Based on these simulated samples, we can have 1K samples of residuals where each sample of residuals is of size equal to that of the observed data. Finally, each set of residuals is used to get the predicted value of $Q(l)$ by substituting the values of residuals in (32). Hence, we calculate a total of 1K values of $Q(l)$. These values can then be used to

obtain the estimated p-values (see (34)) by counting the number of values of $Q(l)$ which are less than the tabulated value of $\chi^2_{(l-p-q),(1-\alpha)}$ and calculating the corresponding fraction. The p-values have been calculated for the considered ARMA models and are presented in Table 2.

Table 2. p-values for the considered ARMA models

Model	pval
ARMA(1,0)	0.758
ARMA(2,0)	0.723
ARMA(0,1)	0.739
ARMA(0,2)	0.743
ARMA(1,1)	0.716

Obviously, the values are large enough to support the adequacy of the models. Thus, our model compatibility study conveys that each of the considered models may be a good candidate to describe the data in hand. One can, of course, use parsimony principle and recommend a model that happens to be the simplest among these but we shall compare these models using the PPL criterion described in subsection 4.1 before recommending a model.

The values of $P(m)$, $G(m)$ and hence $D(m)$ for all the considered models are given in Table 3. These values are based on 1K posterior and correspondingly 1K predictive samples from each of the considered models. It can be seen that the value of $D(m)$ is least for ARMA(0,1) model mainly because it has the smallest value of loss due to complexity although it provides poor fitting when compared with ARMA(1,0), ARMA(2,0), ARMA(0,2) and ARMA(1,1) models (see Table 3). Thus, our final recommended model is ARMA(0,1), a model that has no autoregressive component.

Table 3. Results based on PPL criterion for the considered ARMA models

Model	$P(m)$	$G(m)$	$D(m)$
ARMA(1,0)	1345.309	1206.245	2551.554
ARMA(2,0)	1280.723	1168.402	2449.125
ARMA(0,1)	1022.513	1194.609	**2217.123**
ARMA(0,2)	1045.388	1182.324	2227.711
ARMA(1,1)	1188.329	1185.996	2374.325

Before we end the section, let us consider the problem of predicting the future observations based on the finally selected ARMA(0,1) model. We, however, confine ourselves to short term retrospective prediction so that the scope of predicting the GDP values through the considered model can be verified based on the comparison of predicted values with the observed data points. To proceed with the task of retrospective prediction, let us consider the first 55 observations as the informative data out of a total of 63 entertained observations (see Table 5) and obtain the predictive estimates for the next 56th observation. We may then include this predicted observation in the informative data and proceed with 56 informative observations to develop the prediction for 57th observation. The process may be continued until all the 63 observations are predicted.

To explain the implementation, let us consider the informative data size as r. Thus, we begin with the complete posterior analysis based on these r observations before going for the actual prediction. The details of running the hybrid strategy on the posterior corresponding to ARMA$(0, 1)$ model and hence obtaining the posterior and corresponding predictive samples are provided in subsections 3.3 and 3.4. The results are shown in terms of point prediction as well as the corresponding predictive interval in Table 4. These results are based on a posterior sample of size 1K and corresponding predictive observation for the next unobserved future data $((r+1)^{th})$ obtained for each value of the simulated posterior sample. The predictive estimates are then given as the corresponding predictive modes based on such 1K predictive samples. Similarly, the predictive intervals correspond to highest predictive density intervals each with coverage probability 0.95 obtained on the basis of such 1K predictive samples.

Table 4. The retrospective predictions of GDP growth rate data for the period 2006-07 to 2013-14

Year	y_t	True value	Predictive point estimate	0.95 predictive interval	
2006-07	y_{56}	9.57	10.694	1.133	17.665
2007-08	y_{57}	9.32	9.844	0.967	16.924
2008-09	y_{58}	6.72	9.051	1.105	17.099
2009-10	y_{59}	8.59	8.915	1.427	17.770
2010-11	y_{60}	8.91	8.092	0.551	16.684
2011-12	y_{61}	6.69	9.194	1.015	16.721
2012-13	y_{62}	4.47	10.358	0.993	16.773
2013-14	y_{63}	4.74	9.633	2.714	18.467

It can be seen from the results that the predictive point estimates in the form of modal values are, in general, not too far away from the actual observed data points except for the situations where there is a high fluctuation in the values from those in the previous years. Say, for instance, the values corresponding to the years 2011-12 (y_{61}) to 2013-14 (y_{63}) where the predictive point estimates are too far away from the actual values. It might be possible that there are structural breaks in the GDP data for these years perhaps because of global economic recession and, as such, our model fails to reflect the same. The situation is, however, not too susceptible when we see the estimated predictive intervals with coverage probability 0.95. All these estimated intervals not only cover the true values but also indicate that the true values fall in the high probability central regions of the corresponding predictive density estimates. Although such predictive density estimates are not shown, an idea can be derived based on the values of estimated predictive intervals. As a word of final remark- in no way we claim that our model is most appropriate for the situation rather it appears as if there is always a scope for its improvement. The simplicity of our model and its analysis are certainly the important features in its favour.

6. Conclusions

The paper emphasizes the analysis of a general ARMA model along with its components in a fully Bayesian framework, although a few classical tools, such as evaluation of the ML estimates and the observed information, are employed to support our analysis. The analysis finally proceeds for five particular cases of the considered general form under stationarity and invertibility restrictions. A sample-based approach based on the Gibbs-Metropolis hybrid sampler appears to provide routine posterior implementation on the considered ARMA model and its particular forms. The paper then considers model compatibility and comparison, the former using predictive simulation ideas and the latter using predictive loss criterion. A real data illustration of GDP growth rate data of India at constant prices conveys that ARMA$(0, 1)$ model appears to be the most appropriate, although other components also provide good compatibility with the data in hand. A short-term retrospective prediction based on the final chosen model conveys that the proposed model can be used, in general, except when there is abrupt fluctuation in the data from those of previous years.

Acknowledgements

References

Bayarri, M. J., Berger, J. O., (1998). Quantifying Surprise in the Data and Model Verification, In *Bayesian Statistics*, 6, ed. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, pp. 53–82. London: Oxford University Press.

Box, G. E. P., Jenkins, G. M., (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., (2004). *Time Series Analysis: Forecasting and Control*. Second Indian Reprint, Pearson Education.

Chib, S., Greenberg, E., (1994). Bayes Inference in Regression Models with ARMA(p, q) Errors. *Journal of Econometrics*, 64, pp. 183–206.

Clement, E. P., (2014). Using Normalized Bayesian Information Criterion (BIC) to Improve Box-Jenkins Model Building. *American Journal of Mathematics and Statistics*, 4(5), pp. 214–221.

Dickey, D. A., Fuller, W. A., (1979). Distribution of the Estimators for the Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74, pp. 427–431.

Fan, C., Yao. S., (2008). Bayesian Approach for ARMA Process and Its Application. *International Business Research*, 1(4), pp. 49–55.

Gelfand, A. E., Smith, A. F. M., (1991). Gibbs Sampling for Marginal Posterior Expectations. *Communications in Statistics, Theory and Methods*, 20(5–6), pp. 1747–1766.

Gelfand, A. E., Ghosh, S. K., (1998). Model Choice: A Minimum Posterior Predictive Loss Approach. *Biometrika*, 85, pp. 1–11.

Gelman, A., Meng, X. L., Stern, H. S., (1996). Posterior Predictive Assessment of Model Fitness via Realized Discrepancies. *Statistica Sinica*, 6, pp. 733–807.

Hyndman, R. J., Athanasopoulos, G., (2018). *Forecasting: principles and practice.* OTexts.

Kleibergen, F., Hoek, H., (2000). Bayesian Analysis of ARMA Models. *Tinbergen Institute Discussion Paper* Amsterdam, the Netherlands.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., Shin, P., (1992). Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root. *Journal of Econometrics*, 54, pp. 159–178.

Ljung, G. M., Box, G. E. P., (1978). On A Measure of Lack of Fit in Time Series Models. *Biometrika*, 65, 297–303.

Ludlow, J., Enders, W., (2000). Estimating non-linear ARMA models using Fourier coefficients. *International Journal of Forecasting*, 16(3), pp. 333–347.

Morley, J. C., Nelson, C. R., Zivot, E., (2003). Why are the Beveridge-Nelson and unobserved-components decompositions of GDP so different? *Review of Economics and Statistics*, 85(2), pp. 235–243.

Marriot, J. M., Smith A. F. M., (1992). Reparameterization Aspects of Numerical Bayesian Methodology for Autoregressive Moving Average Models. *Journal of Time Series Analysis*, 13(4), pp. 327–343.

Marriot, J. M., Ravishanker, N., Gelfand, A., PAI, J., (1996). Bayesian Analysis of ARMA Processes: Complete Sampling Based Inference Under Exact Likelihoods. In: Berry, D. A., K. M. Chaloner, and J. Geweke (eds.): *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner* John Wiley, New York.

Monahan, J. F., (1983). Fully Bayesian Analysis of ARMA Time Series Models. *Journal of Econometrics*, 21, pp. 307–331.

Newbold, P., (1974). The Exact Likelihood Function for a Mixed Autoregressive-Moving average Process. *Biometrika*, 61(3), pp. 423–426.

Okereke, O. E., Iwueze, I. S., Johnson, O., (2015). Necessary Conditions for the Application of Moving Average Process of Order Three. *Applied Mathematics*, 6(01), pp. 173–181.

Pankratz, A., (1983). *Forecasting with Univariate Box-Jenkins Models: Concepts and Cases.* John Wiley & Sons, New York.

Reilly, D. P., (1980). Experiences with An Automatic Box-Jenkins Modeling Algorithm. *Time Series Analysis, Proceedings of Houston Meeting on Time Series Analysis*, North-Holland publishing, Amsterdam.

Reynolds, S. B., Mellichamp, J. M., Smith, R. E., (1995). Box-Jenkins Forecast Model Identification. *AI Expert* June, pp. 15–28.

Sahu, S. K., Dey D. K., (2000). A Comparison of Frailty and Other Models for Bivariate Survival Data. *Lifetime Data Analysis*, 6, pp. 207–228.

Shumway, R. H., Stoffer, D. S., (2011). *Time Series Analysis and Its Applications*, Springer-Verlag, New York.

Smith, A. F. M., Roberts, G. O., (1993). Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society, Series B*, 55, pp. 3–25.

Tripathi, P. K., Ranjan, R., Pant, R., 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), pp. 399–419.

Tripathi, P. K., Mishra, R. K., Upadhyay, S. K., (2018). Bayes and Classical Prediction of Total Fertility Rate of India Using Autoregressive Integrated Moving Average Model. *Journal of Statistics Applications & Probability*, 7(2), pp. 233–244.

Tsay, R. S., Tiao, G. C., (1984). Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Non-Stationary ARMA Models. *Journal of the American Statistical Association*, 79, pp. 84–96.

Tsay, R. S., Tiao, G. C., (1985). Use of Canonical Analysis in Time Series Model Identification. *Biometrika*, 72(2), pp. 299–315.

Upadhyay, S. K., Gupta, A., Dey, D. K., (2012). Bayesian Modeling of Bathtub Shaped Hazard Rate Using Various Weibull Extensions and Related Issues of Model Selection. *Sankhya, Series B*, 74, pp. 15–43.

Upadhyay, S. K., Mukherjee, B., (2008). Assessing the Value of the Threshold Parameter in the Weibull Distribution Using Bayes Paradigm. *IEEE Transactions on Reliability*, 57(3), pp. 489–497.

Upadhyay, S. K., Peshwani, M., (2003). Choice Between Weibull and Lognormal Models: A Simulation Based Bayesian Study. *Communications in Statistics, Theory and Methods*, 32, pp. 381–405.

Upadhyay, S. K., Smith, A. F. M., (1994). Modelling Complexities in Reliability and the Role of Simulation in Bayesian Computation. *International Journal of Continuing Engineering Education*, 4, pp. 93–104.

Upadhyay, S. K., Vasishta, N., Smith, A .F. M., (2001). Bayes Inference in Life Testing and Reliability via Markov Chain Monte Carlo Simulation. *Sankhya, Series A*, 63, pp. 15–40.

Zellner, A., (1971). *An Introduction to Bayesian Inference in Econometrics.* John Wiley & Sons, New York.

Appendix

Table 5. GDP growth rate of India at constant prices for the period 1951-52 to 2013-14
(from left to right)

2.33	2.84	6.09	4.25	2.56	5.69	-1.21	7.59	2.19
7.08	3.10	2.12	5.06	7.58	-3.65	1.02	8.14	2.61
6.52	5.01	1.01	-0.32	4.55	1.16	9.00	1.25	7.47
5.5	-5.2	7.17	5.63	2.92	7.85	3.96	4.16	4.31
3.53	10.16	6.13	5.29	1.43	5.36	5.68	6.39	7.29
7.97	4.30	6.68	8.00	4.15	5.39	3.88	7.97	7.05
9.48	9.57	9.32	6.72	8.59	8.91	6.69	4.47	4.74

Likelihood for ARMA(p,q) model:

Let us consider the likelihood of ARMA(p,q) model as given in (9) and let $Y_p = (y_1, y_2, ..., y_p)'_{p \times 1}$ be the vector of first p observations in the sample of size T. Then Y_p follows a p-variate normal distribution with mean vector $\mu_p = (\mu, \mu, ..., \mu)'_{p \times 1}$ of dimension $p \times 1$ and variance-covariance matrix Σ_p of dimension $p \times p$. It may be noted that $\mu = \theta_0/(1 - \phi_1 - \phi_2 - ... - \phi_p)$ and the elements of Σ_p can be obtained by solving the Yule-Walker equations of autocorrelation in terms of AR parameters (see, for example, Box et al. (2004)). Thus, the joint distribution of Y_p can be written as

$$f(y_1, y_2, ..., y_p | \theta_0, \Phi, \Psi) \propto \left(\sigma^2\right)^{\frac{-p}{2}} \times |V_{\phi,\psi}|^{-1/2} \times$$
$$\exp\left(-\frac{1}{2\sigma^2}(Y_p - \mu_p)' V_{\phi,\psi}^{-1}(Y_p - \mu_p)\right), \quad (36)$$

where $V_{\phi,\psi}$ is a $p \times p$ matrix involving terms of ϕ's and ψ's and it may be determined by using the relationship $\Sigma_p = \sigma^2 V_{\phi,\psi}$.

The second term on the right-hand side of (9) is the product of conditional densities of y_t with conditioning variates $(y_{t-1}, y_{t-2}, ..., y_{t-p})$, $t = (p+1), ..., T$. This conditional density can be shown to follow $N((\theta_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \psi_j \varepsilon_{t-j}), \sigma^2)$ and the same can be written as,

$$f(y_t | y_{t-1}, y_{t-2}, ..., y_{t-p}; \theta_0, \Phi, \Psi) \propto \left(\frac{1}{\sigma^2}\right) \times$$
$$\exp\left(-\frac{1}{2\sigma^2}(y_t - \theta_0 - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{j=1}^{q} \psi_j \varepsilon_{t-j})^2\right). \quad (37)$$

Obviously, the conditional likelihood function (the second term in the right-hand side of (9)) reduces to,

$$f_c \propto \left(\frac{1}{\sigma^2}\right)^{(T-p)/2} \times \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p+1}^{T}(y_t - \theta_0 - \sum_{i=1}^{p}\phi_i y_{t-i} - \sum_{j=1}^{q}\psi_j \varepsilon_{t-j})^2\right). \qquad (38)$$

Now, combining (9) and (38), the likelihood for ARMA(p,q) model in (9) can be written as in (10).

The ARMA(p,q) model is generally difficult to implement in practice due to the fact that stationarity and invertibility conditions are not easily encountered for higher values of p and q beyond, say, 2 in each case. This is perhaps the reason that we restrict to $p+q \leq 2$ in our present study. Moreover, for ARMA$(1,1)$ model, the things are comparatively easier. In this case, we have $Y_1 = y_1 \sim N(\mu_1, \Sigma_1)$, where $\mu_1 = \frac{\theta_0}{1-\phi_1}$ and $\Sigma_1 = \frac{\sigma^2(1+\psi_1^2+2\psi_1\phi_1)}{1-\phi_1^2}$.

Exact likelihood for AR(p) model:

The AR(p) model corresponds to ARMA$(p,0)$ and, therefore, the corresponding likelihood function can be obtained by putting zero in the place of MA component q in the likelihood (10) and the same can be written as in (11).

Moreover, as mentioned earlier for ARMA process, the AR(p) process is also not easy to implement in practice for values of p beyond 2. For AR(1) model, we have $Y_1 = y_1 \sim N(\mu_1, \Sigma_1)$ where $\mu_1 = \frac{\theta_0}{1-\phi_1}$ and $\Sigma_1 = \frac{\sigma^2}{1-\phi_1^2}$ and for AR(2) model, we have $Y_2 = (y_1, y_2)' \sim N(\mu_2, \Sigma_2)$ with $\mu = \frac{\theta_0}{1-\phi_1-\phi_2}$ and

$$\Sigma_2 = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

where $\Sigma_{11} = \Sigma_{22} = \frac{\sigma^2(\phi_2-1)}{(\phi_2-1)(1-\phi_1^2-\phi_2^2)+2\phi_1^2\phi_2}$ and $\Sigma_{12} = \Sigma_{21} = \frac{-\sigma^2\phi_1}{(\phi_2-1)(1-\phi_1^2-\phi_2^2)+2\phi_1^2\phi_2}$.

Exact likelihood for MA(q) model:

The MA(q) model corresponds to ARMA$(0,q)$ and, therefore, MA(q) process can be easily written from (1) as

$$y_t = \theta_0 + \sum_{j=1}^{q}\psi_j \varepsilon_{t-j} + \varepsilon_t. \qquad (39)$$

The likelihood function corresponding to MA(q) model cannot be obtained directly by putting zero in the place of AR component p in the likelihood (10). This may be because of the fact that the process (1) will not have any component of lagged time series observation in its right-hand side and the components of ε_0 are already taken to be zero while writing the likelihood corresponding to ARMA(p,q) model. One can, however, write a computationally friendly form of the exact likelihood for MA(q) model by an alternative argument.

Let the vector of complete sample observations be given by $Y_T = (y_1, y_2, ..., y_T)'_{T \times 1}$ and correspondingly the mean vector be given by $\mu_T = (\mu, \mu, ..., \mu)'_{T \times 1}$ where $\mu = \theta_0$. The variance-covariance matrix Σ_T is a $T \times T$ matrix and the same can be obtained as

$$
\Sigma_T = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & ... & \Sigma_{1T} \\ \Sigma_{21} & \Sigma_{22} & ... & \Sigma_{2T} \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ \Sigma_{T1} & \Sigma_{T2} & ... & \Sigma_{TT} \end{bmatrix},
$$

where

$$
\Sigma_{11} = \Sigma_{22} = ... = \Sigma_{TT} = \sigma^2 (1 + \psi_1^2 + \psi_2^2 + ... + \psi_q^2); \qquad \text{for} \quad t = 1, 2, ..., T \quad (40)
$$

and

$$
\Sigma_{t,t-k} = \Sigma_{t-k,t} = \begin{cases} \sigma^2 (\psi_k + \psi_{k+1} \psi_1 + ... + \psi_q \psi_{q-k}), & \text{for} \quad k = 1, 2, ..., q \\ 0, & \text{for} \quad k > q. \end{cases} \quad (41)
$$

The exact likelihood for the MA(q) model is, therefore, given by the T-variate normal distribution with mean vector μ_T and variance-covariance matrix Σ_T and it can be written up to proportionality as in (12).