Bayesian Prediction of Moving Average Processes Using Different Types of Priors

Samir M. Shaarawy
Department of Quantitative Methods and Information Systems
Kuwait University
samir@cba.edu.kw

Emad E.A. Soliman[1]
Department of Statistics
Faculty of Science
King Abulaziz University
easalam@hotmail.com

Heba E.A.Shahin
Department of Statistics
Research Sector
Central Bank of Egypt
heba_shahin2002@yahoo.com

ABSTRACT

The current article approaches the Bayesian prediction of moving average processes using three well-known priors; g prior, natural conjugate (NC) prior, and Jeffreys' prior. The main goal of the study is to derive approximate one step-ahead predictive densities for moving average (MA) processes using each of the above mentioned priors. However, the basic contribution is the derivation of the predictive density based upon the g prior. Investigating the performance of the three one step-ahead predictive densities is performed via comprehensive simulation studies using MA(1) and MA(2) processes for illustration. The simulation results show the equivalence of the performance of the three one step-ahead predictive densities based on the three considered priors in the forecasting process.

Keywords: *Forecasting, Prediction, one step-ahead predictive density, Moving Average process, g prior, Jeffreys' prior, natural conjugate prior, Informative prior, Non-informative prior.*

1. INTRODUCTION

Many real time series in the literature are identified to follow moving average (MA) models as found in economics, business, medicine ...etc (Frederick (2000), Gyllenberg and Koski (2002), and Singh (2006)). Analyzing time series using the Bayesian approach is rapidly becoming accepted as a way to solve applied statistical problems. Prediction is usually the ultimate goal of time series analysis and is the

[1] The permanent position of the second author is Department of Statistics, Faculty of Economics and Political Science, Cairo University.

phase of interest in the current article. The Bayesian tool to predict the first future observation is the one step-ahead predictive density.

The earliest standard reference for Bayesian analysis is that of Zellner (1971) in which he has derived both the posterior and predictive densities for the AR(1) model and the posterior density for the AR(2) models using Jeffreys' prior. Monahan (1983) has asserted an important contribution to solve the problem of non-linearity of the errors in the model's coefficients in low order autoregressive moving average (ARMA) models by developing a numerical technique that implements the identification, estimation, and forecasting phases. Broemeling and Land (1984) and Broemeling (1985) have derived the one step-ahead and the more general k step-ahead predictive densities of the AR(p) model using the normal-gamma prior.

Using normal-gamma and Jeffreys' priors, Soliman (1999) has derived the approximate one step-ahead predictive densities for both ARMA and seasonal ARMA (SARMA) models. In the current article, the predictive densities of MA(q) based on a Natural Conjugate (NC) prior and Jeffreys' prior are to be derived as special cases of Soliman (1999) derivations. Soliman (2008) has studied the numerical efficiency of three approximate one step-ahead predictive densities for ARMA models using three different approximations based on Jeffreys' prior.

El-Zayat (2007) has investigated the performance of the posterior densities of AR(1) models via simulation studies. In the derivations of the posterior densities, she considered the NC and the Jeffreys' prior in addition to the g prior. She concluded that informative priors are more efficient than non-informative priors, in terms of a certain efficiency criterion, for estimating the coefficients of a stationary AR(1) model. Moreover, Shaarawy et al. (2010) have employed the above mentioned three priors to derive one step-ahead predictive densities for AR models and studied the numerical effectiveness of each prior. Furthermore, Al-Bassam et al. (2013) have considered the above mentioned three priors in solving the identification problem of moving average processes. More recently, Shaarawy et al. (2015) have investigated the effectiveness of the three priors in solving the identification problem of autoregressive processes.

Importance of the prior selection phase has been an outstanding issue in time series literature. The prior distribution represents initial beliefs about the model parameters. There are two sorts of priors; informative and non-informative priors. A non-informative prior is used when there is little or no information about the model parameters otherwise; an informative prior is used (see DeGroot (1970)).

In 1961, Jeffreys introduced the non-informative Jeffreys' prior to overcome the lack of invariance in the preceding introduced priors. Raiffa and Schlaifer (1961) among others discussed the informative family of conjugate priors, which is a widely used approach. The natural conjugate (NC) prior belongs to this class and its form, by definition, depends on the form of likelihood function (LF). Difficulties in assessment of NC prior lie in the estimation of its hyper-parameters. Estimating the hyper-parameters can be performed using several techniques such as those discussed by Raiffa and Schlaifer (1961), Berger (1985), Broemeling (1985), and others.

In 1986, Zellner introduced the g prior that may be considered as a middle ground of sorts between informative and non-informative priors (Karlsson (2001)). To estimate a g prior, one should determine a value for a related constant g (for more details see Robbins (1956) and Fernández et al (2001)).

In the next section, the moving average models are defined. In section three, the principal tools and concepts used in the study are simply explained. In section four, the predictive densities based on the three proposed priors are derived. Moreover, section five develops wide scale simulation studies. Finally, section six summarizes the study findings and concludes.

2. MOVING AVERAGE MODELS

The moving average model of order q, denoted by MA(q), is a special case of the ARMA(p,q) model. The model can be written as follows (Box and Jenkins (1970)):

$$y_t = \Theta(B)\varepsilon_t \tag{2.1}$$

Where, $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$

Where, B is the backshift operator such that $B^j y_t = y_{t-j}$. $\{y_t, t = \cdots, -1,0,1, \ldots\}$ is the time series. ε_t's are assumed to be i.i.d. normally distributed random errors with mean zero and variance τ^{-1}, where $\tau = 1/\sigma^2 > 0$ is the precision parameter. θ_i's are the model coefficients. The model (2.1) can be presented explicitly as follows:

$$y_t = \varepsilon_t \quad \theta_1 \varepsilon_{t-1} \quad \theta_2 \varepsilon_{t-2} \quad \cdots \quad \theta_q \varepsilon_{t-q} \tag{2.2}$$

Moving average models are always stationary, while they are invertible under some conditions. The invertibility conditions are such that the roots of $\Theta(B)$ lie outside the unit circle. Special cases of (2.2) are the MA(1) model given by

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} \tag{2.3}$$

And the MA(2) model given by

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \tag{2.4}$$

For MA(1) model the invertibility condition is $|\theta_1| < 1$ and for MA(2) model, the invertibility conditions are $\theta_2 - \theta_1 < 1$, $\theta_2 + \theta_1 < 1$, and $|\theta_2| < 1$. (Box and Jenkins (1970)).

3. BASIC CONCEPTS

Derivation of the predictive density depends on both the Likelihood Function (LF) and the prior distribution which are both defined hereafter. Given a time series of n observations $\underline{Y} = [\, y_1 \;\; y_2 \;\; \ldots \;\; y_n \,]'$ and letting the initial error values $\varepsilon_0 = \varepsilon_{-1} = \cdots = \varepsilon_{1-q} = 0$, i.e. they equal their unconditional mean, the conditional LF for the MA(q) model, conditioned on the initial values of the errors, can be written in the form (see Broemeling (1985))

$$L(\underline{\gamma}, \tau \mid \underline{Y}) \propto \tau^{\frac{n}{2}} \exp\left\{ -\frac{\tau}{2} \sum_{t=1}^{n} (y_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j})^2 \right\} \tag{3.1}$$

Where, τ is the precision parameter, $\underline{Y}$ is the vector of observations such that, $\underline{Y} = [\, y_1 \;\; y_2 \;\; \ldots \;\; y_n \,]'$ and $\underline{\gamma}$: is the coefficients' vector defined as

$$\underline{\gamma} = [\, \theta_1 \;\; \theta_2 \;\; \ldots \;\; \theta_q \,]' \tag{3.2}$$

Let X be an n×q matrix of regressors such that the t^{th} row vector $\underline{X}_t$ is represented by

$$\underline{X}_t = [-\varepsilon_{t-1} \; -\varepsilon_{t-2} \ldots -\varepsilon_{t-q}] \qquad t = 1, 2, \ldots, n \tag{3.3}$$

Then, (3.1) can be written in matrix notation as

$$L(\underline{\gamma}, \tau \mid \underline{Y}) \propto \tau^{\frac{n}{2}} \exp\left\{-\frac{\tau}{2}\left(\underline{Y} - X\underline{\gamma}\right)'\left(\underline{Y} - X\underline{\gamma}\right)\right\} \tag{3.4}$$

The conditional LF (3.4) is analytically intractable since the errors of MA models ε_t's are nonlinear functions in the coefficients. Broemeling and Shaarawy (1988) have approximated the errors of the MA model as linear functions in the model's coefficients by their non-linear least squares estimates $\hat{\varepsilon}_t$'s . Thus, the conditional LF (3.4) can be approximated by

$$L^*(\underline{\gamma}, \tau \mid \underline{Y}) \propto \tau^{\frac{n}{2}} \exp\left\{-\frac{\tau}{2}\left(\underline{Y} - \hat{X}\underline{\gamma}\right)'\left(\underline{Y} - \hat{X}\underline{\gamma}\right)\right\} \tag{3.5}$$

Where, $\hat{X}$ is the estimated matrix of regressors for which the t^{th} row $\underline{\hat{X}}_t$ is given by

$$\underline{\hat{X}}_t = [-\hat{\varepsilon}_{t-1} \; -\hat{\varepsilon}_{t-2} \ldots -\hat{\varepsilon}_{t-m}] \tag{3.6}$$

The approximate LF (3.5) can be rewritten as

$$L^*(\underline{\gamma}, \tau \mid \underline{Y}) \propto \tau^{\frac{n}{2}} \exp\left\{-\frac{\tau}{2}\left[\underline{\gamma}' A_1 \underline{\gamma} - 2\underline{\gamma}' \underline{B}_1 + C_1\right]\right\} \tag{3.7}$$

Where,

$$A_1 = \hat{X}'\hat{X}, \qquad \underline{B}_1 = \hat{X}'\underline{Y} \quad \text{and} \quad C_1 = \underline{Y}'\underline{Y} \tag{3.8}$$

In addition, $\hat{X}$ and $\underline{Y}$ are as defined above.

Regarding the g prior, it is considered as a middle ground of sorts between the informative NC prior and non-informative Jeffreys' prior (Karlsson (2001)). It was developed by Zellner (1986) to avoid the evaluation of prior covariates of the hyper-parameters which is found in the NC prior. Zellner's motivation was to derive an appropriate prior for the regression parameters in the General Linear Models (GLM).

Shaarawy et al. (2010) have derived the joint g-prior of $\underline{\gamma}$ and σ for MA(q) model in the form

$$p_g(\underline{\gamma}, \sigma) \propto \sigma^{-(q+1)} \exp\left\{-\frac{g}{2\sigma^2}(\underline{\gamma} - \bar{\underline{\gamma}})' X'X (\underline{\gamma} - \bar{\underline{\gamma}})\right\}. \tag{3.9}$$

Where, $\underline{\gamma}$ is defined in (3.2), σ is the errors variance, $\underline{\bar{\gamma}}$ is an anticipated value of $\underline{\gamma}$ and X is the matrix of regressors of MA(q) model with t^{th} row $\underline{X}_t$ defined in (3.3).

There are many potential values given in the literature for the constant g, mostly g is considered as a function of the sample size n or the number of the parameters k. Fernàndez et al. (2001), based on a simulation study, concluded that the most reasonable choices of the value of g are as follows:

$$\begin{cases} g = \dfrac{1}{n} & \text{for} \quad n > k^2 \\ g = \dfrac{1}{k^2} & \text{for} \quad n \le k^2 \end{cases} \tag{3.10}$$

It should be noted that when the sample under consideration follows a normal distribution, the normal-gamma prior is the NC prior and is written for MA(q) model as follows:

$$p(\underline{\gamma},\tau) \propto \tau^{\alpha+\frac{q}{2}-1} e^{-\frac{\tau}{2}[2\beta+(\underline{\gamma}-\underline{\mu})'V(\underline{\gamma}-\underline{\mu})]}, \quad \tau > 0, \tag{3.11}$$

where α, β, $\underline{\mu}$, and V are the hyper-parameters of the prior distribution.

In addition to the above two priors, the Jeffreys' prior introduced by Jeffreys (1961), is given by:

$$p(\underline{\gamma},\tau) \propto \tau^{-1}. \tag{3.12}$$

The current study is interested in Bayesian prediction of MA models based on g-prior, NC prior, and Jeffreys' prior. The Bayesian tool to predict the first future observation y_{n+1} is the one step-ahead predictive density $p(y_{n+1} \mid \underline{Y})$ which is defined as follows (Shaarawy and Broemeling (1984)):

$$p(y_{n+1} \mid \underline{Y}) \propto \int_{\underline{\gamma}} \int_{\tau} p(y_{n+1}, \underline{Y} \mid \underline{\gamma}, \tau)\ p(\underline{\gamma},\tau)\, d\tau\, d\underline{\gamma} \tag{3.13}$$

4. APPROXIMATE PREDICTIVE DENSITIES OF MA MODELS

The main goal of this section is to derive the approximate one step-ahead predictive density for the first future observation of MA(q) models using the three

above mentioned priors; g prior, NC prior, and Jeffreys' prior. The section is divided into two subsections; the first 4.1 is reserved for the derivation of the approximate one step-ahead predictive density based on g prior, the second 4.2 is reserved for displaying the derivation based on both the NC and Jeffreys' prior.

4.1. Derivation Based On g Prior

In this section the moving average model MA(q) will be considered as a special case of the standard GLM. The model can be written in a matrix notation as

$$\underline{Y} = X\,\underline{\gamma} + \underline{\varepsilon} \tag{4.1}$$

Where $\underline{Y} = [y_1\ y_2\ \dots\ y_n]'$, $\underline{\varepsilon} = [\varepsilon_1\ \varepsilon_2\ \dots\ \varepsilon_n]'$, $\underline{\gamma}$ is the coefficients' vector defined in (3.2), and finally $X_{n\times q}$ is the matrix of regressors such that the t^{th} observation y_t is represented by,

$$y_t = \underline{X}_t\underline{\gamma} + \varepsilon_t\ , \qquad t = 1, 2, \dots, n,$$

where ε_t is the t^{th} error, and the t^{th} row vector $\underline{X}_t$ is defined in (3.3).
Further, define the row vector $\underline{X}_{n+1}$ by letting $t = n+1$ in $\underline{X}_t$ defined in (3.3).

The following theorem gives the approximate one step-ahead predictive density of the first future observation generated from the MA(q) model (2.1) based on the g prior on the form (3.9):

Theorem 4.1:

Based on g prior on the form (3.9), the approximate one step-ahead predictive density of the MA(q) model is a non-central t distribution, with $(n+q-1)$ degrees of freedom, location E and precision P defined as

$$\begin{cases} E(y_{n+1} \mid \underline{Y}) = (g+1)^{-1}\,\underline{\hat{X}}_{n+1}A_1^{-1}(\underline{B}_1 + gA_1\underline{\bar{\gamma}}) \\ P(y_{n+1} \mid \underline{Y}) = \dfrac{n+q-1}{d_1L_1} \end{cases} \tag{4.3}$$

Where $\underline{\bar{\gamma}}$ is an anticipated value of $\underline{\gamma}$, A_1, B_1 and C_1 are as defined in (3.8) and

$$\begin{cases} L_1 = g\bar{\underline{\gamma}}' A_1 \bar{\underline{\gamma}} + C_1 - (g+1)^{-1}(\underline{B}_1 + gA_1\bar{\underline{\gamma}})' A_1^{-1}(\underline{B}_1 + gA_1\bar{\underline{\gamma}}) \\ d_1 = 1 + (g+1)^{-1}\hat{\underline{X}}_{n+1} A_1^{-1} \hat{\underline{X}}'_{n+1} \end{cases} \qquad (4.4)$$

Proof:

Consider the MA(q) model defined earlier in equation (2.1). Also, consider the approximate LF defined in (3.7) and rewrite it in terms of σ^2 instead of τ.

Moreover, it is possible to define the approximate sum of squares $Q_{n+1}^{(1)} = \sum_{t=1}^{n+1} \hat{\varepsilon}_t^2$ as follows

$$Q_{n+1}^{(1)} = \underline{\gamma}' A \underline{\gamma} - 2\underline{\gamma}' \underline{B} + C \qquad (4.5)$$

Where

$$A = A_1 + \hat{\underline{X}}'_{n+1}\hat{\underline{X}}_{n+1}, \qquad \underline{B} = \underline{B}_1 + y_{n+1}\hat{\underline{X}}'_{n+1} \quad \text{and} \quad C = C_1 + y_{n+1}^2. \qquad (4.6)$$

And A_1, B_1 and C_1 are as defined in (3.8).

From (4.5) and the approximate LF (3.7) after replacing τ by $1/\sigma^2$, the joint distribution of the first future observation y_{n+1} and the time series $\underline{Y} = [y_1\ y_2 \ldots y_n]'$ given the model parameters σ and $\underline{\gamma}$ can be written as follows

$$p(y_{n+1}, \underline{Y} | \underline{\gamma}, \sigma) \propto \sigma^{-[n+1]} \exp\left\{ -\frac{1}{2\sigma^2} Q_{n+1}^{(1)} \right\}$$

i.e.

$$p(y_{n+1}, \underline{Y} | \underline{\gamma}, \sigma) \propto \sigma^{-[n+1]} \exp\left\{ -\frac{1}{2\sigma^2} (\underline{\gamma}' A \underline{\gamma} - 2\underline{\gamma}' \underline{B} + C) \right\} \qquad (4.7)$$

Combining the joint g prior of $\underline{\gamma}$ and σ given by (3.9) with the joint distribution (4.7), we get

$$p_g(\underline{\gamma}, \sigma).p(y_{n+1}, \underline{Y} | \underline{\gamma}, \sigma) \propto \sigma^{-[n+q+2]} \exp\{-Z/2\sigma^2\}$$

where

$$Z = g(\underline{\gamma} - \bar{\underline{\gamma}})' A_1 (\underline{\gamma} - \bar{\underline{\gamma}}) + (\underline{\gamma}' A \underline{\gamma} - 2\underline{\gamma}' \underline{B} + C).$$

Thus, the approximate one step-ahead predictive density is given by

$$p(y_{n+1} \mid \underline{Y}) \propto \int_{\underline{\gamma}} \int_{\sigma} \sigma^{-[n+q+2]} \exp\left\{-Z/2\sigma^2\right\} d\sigma \, d\underline{\gamma} \qquad (4.8)$$

In order to extract a perfect square in $\underline{\gamma}$, Z can be rewritten as follows:

$$Z = Z_1 + Z_2$$

Where

$$Z_1 = (\underline{\gamma} - M^{-1}\underline{V})' M (\underline{\gamma} - M^{-1}\underline{V})$$

Which is a perfect square in $\underline{\gamma}$ and

$$Z_2 = g\bar{\underline{\gamma}}' A_1 \bar{\underline{\gamma}} + C - \underline{V}' M^{-1} \underline{V}$$

$$= g\bar{\underline{\gamma}}' (\hat{X}'\hat{X})\bar{\underline{\gamma}} + C_1 + y_{n+1}^2 - \underline{V}' M^{-1} \underline{V},$$

Where

$$M = gA_1 + A = (g+1)A_1 + \underline{X}'_{n+1}\underline{X}_{n+1},$$

And

$$\underline{V} = gA_1\bar{\underline{\gamma}} + \underline{B} = (gA_1\bar{\underline{\gamma}} + \underline{B}_1) + y_{n+1}\underline{X}'_{n+1}.$$

Substituting in (4.8), the approximate one step-ahead predictive density for the first future observation has the following form:

$$p(y_{n+1} \mid \underline{Y}) \propto \int_{\sigma} \sigma^{-[n+q+2]} e^{-z_2/2\sigma^2} \int_{\underline{\gamma}} e^{-z_1/2\sigma^2} d\underline{\gamma} \, d\sigma .$$

The inner integral is a multivariate normal distribution. Thus

$$p(y_{n+1} \mid \underline{Y}) \propto \int_{\sigma} \sigma^{-[n+q+1]} e^{-z_2/2\sigma^2} d\sigma$$

The integrand in the last integral is an inverted gamma density of the second type, which yields

$$p(y_{n+1} \mid \underline{Y}) \propto (Z_2)^{-\frac{n+q}{2}}.$$

Note that,

$$V'M^{-1}V = y_{n+1}^2 \hat{\underline{X}}_{n+1} M^{-1} \underline{X}'_{n+1} + 2y_{n+1}\hat{\underline{X}}_{n+1} M^{-1}(\underline{B}_1 + gA_1\bar{\underline{\gamma}})$$

$$+ (\underline{B}_1 + gA_1\bar{\underline{\gamma}})' M^{-1} (\underline{B}_1 + gA_1\bar{\underline{\gamma}})$$

Note that (see Press (1982) p. 23),

$$M^{-1} = (g+1)^{-1} A_1^{-1} - \frac{(g+1)^{-1} A_1^{-1} \hat{\underline{X}}'_{n+1} \hat{\underline{X}}_{n+1} (g+1)^{-1} A_1^{-1}}{1+(g+1)^{-1} \hat{\underline{X}}_{n+1} A_1^{-1} \hat{\underline{X}}'_{n+1}}$$

$$= \frac{1}{d_1(g+1)} A_1^{-1}$$

Where

$$d_1 = 1 + (g+1)^{-1} \hat{\underline{X}}_{n+1} A_1^{-1} \hat{\underline{X}}'_{n+1},$$

Then

$$Z_2 = \frac{1}{d_1}(y_{n+1} - d_1 k_2)^2 + k_3 - d_1 k_2^2$$

Where

$$k_2 = \hat{\underline{X}}_{n+1} \frac{A_1^{-1}}{d_1(g+1)} (\underline{B}_1 + g A_1 \bar{\underline{\gamma}})$$

$$k_3 = g \bar{\underline{\gamma}}' A_1 \bar{\underline{\gamma}} + C_1 - (\underline{B}_1 + g A_1 \bar{\underline{\gamma}})' \frac{A_1^{-1}}{d_1(g+1)} (\underline{B}_1 + g A_1 \bar{\underline{\gamma}})$$

Thus, the approximate one step-ahead predictive density can easily be shown to follow a non-central t distribution with $(n+q-1)$ degrees of freedom, location E and precision P defined in (4.3).

4.2. Predictive Densities Based On N-C and Jeffreys' Prior

Considering a normal-gamma prior on the form (3.11), the approximate one step-ahead predictive density of the MA(q) model is a non-central t distribution, with $(n+2\alpha+q-1)$ degrees of freedom, location E and precision P defined as

$$\begin{cases} E(y_{n+1}|\underline{Y}) = \hat{\underline{X}}_{n+1}(A_1+V)^{-1}(\underline{B}_1 + V\underline{\mu}) \\ P(y_{n+1}|\underline{Y}) = \dfrac{n+2\alpha+q-1}{d_2 L_2} \end{cases} \tag{4.9}$$

Where A_1, B_1 and C_1 are as defined in (3.8) while α , β, $\underline{\mu}$ and V are the hyper-parameters of the normal-gamma prior defined earlier in (3.11) and

$$\begin{cases} L_2 = C_1 + 2\beta + \underline{\mu}' V \underline{\mu} - (\underline{B}_1 + V\underline{\mu})'(A_1 + V)^{-1}(\underline{B}_1 + V\underline{\mu}) \\ d_2 = 1 + \underline{\hat{X}}_{n+1}(A_1 + V)^{-1}\underline{\hat{X}}'_{n+1} \end{cases}$$

On the other hand, consider Jeffreys' prior on the form (3.12), it can easily be shown that the approximate one step-ahead predictive density of the MA(q) model is a non-central t distribution, with $(n - 1)$ degrees of freedom, location E and precision P defined as

$$\begin{cases} E(y_{n+1} \mid \underline{S}_n) = \underline{\hat{X}}_{n+1} A_1^{-1} \underline{B}_1 \\ P(y_{n+1} \mid \underline{S}_n) = \dfrac{n-1}{d_3 \hat{Q}_n} \end{cases} \qquad (4.10)$$

Where A_1 and B_1 are as defined in (3.8), $d_3 = 1 + \underline{\hat{X}}_{n+1} A_1^{-1} \underline{\hat{X}}'_{n+1}$, and $\hat{Q}_n = \sum_{t=1}^{n} \hat{\varepsilon}_t^2$ is the least errors' sum of squares.

5. EFFECTIVENESS STUDIES

The main objective of this section is to assess and compare the performance of the above mentioned three approximate one step-ahead predictive densities in forecasting future observations of MA models. Some MA(1) and MA(2) models are used for illustration. The investigation is conducted via comprehensive Monte Carlo simulation studies.

The simulation studies have the following general design: first, a time series is generated from a given model with specific parameters. Second, the generated data are used to evaluate each of the three predictive densities. Third, efficiency criteria are calculated for each predictive density. Fourth, 500 replications of the above three steps are done. Finally, the output is tabulated.

Generally, the generation process begins by generating 500 data sets of normal variates, each of size 752, to represent the noise ε_t. Then an MA model with certain coefficients is used recursively to generate 500 realizations, of length 751.

The first 200 observations are ignored to remove the initialization effect. Thus, we get 500 time series each of length 551.

Out of the 551 observations, the first 50 observations are reserved for estimating the priors' hyper-parameters while the next 500 observations are reserved for estimating the approximate predictive density parameters, and finally the last observation, namely, y_{551}, is considered as the first future observation. Estimation of the priors' hyper-parameters is performed using a training sample, from the first 50 observations of each data set (see El-Zayat (2007)). From the first 50 observations a training sample of size either 10 (for short time series) or 10% of the considered time series length - which is used for estimating the predictive density - whatever is larger. For instance, if the time series length is 20 the training sample starts from y_1 to y_{10}, while if the time series length is 200 the training sample starts from y_1 to y_{20} and so on. From the succeeding 500 observations, a time series of the desired size is used for estimating the approximate predictive density. In this simulation study, the chosen time series lengths are 20, 25, 30, 100, 200, and 500.

To investigate and compare the performance of the approximate predictive densities two efficiency criteria are used: the first is the measure P^* and the second is the MAD. The measure P^* checks the goodness of interval forecasts drawn from a specified predictive density. Defining Highest Predictive Density (HPD) region as the interval having probability 0.95 centered at the mean of the predictive density. The percentage P^* of time series for which the HPD region of the predictive density contains the true future observation is defined as

$$P^* = (n^* / 500) * 100 \qquad (5.1)$$

Where n^* is the number of time series where the HPD region includes the true future value (see Soliman (1999)). However, P^* doesn't account for the distance of the future observation from the center of the region or its boundaries. Therefore, another measure, called percentage MAD, is provided. Percentage MAD stand for the mean absolute deviations of the future observation from the location parameter and is defined as

$$\text{MAD} = \left(\sum_{j=1}^{500} \left| y_{n+1,j} - E_j \right| / 500\right) * 100 , \qquad (5.2)$$

Where $y_{n+1,j}$ and E_j are the first future observation and the location parameter, respectively, of the j^{th} simulated time series.

5.1. Results of MA(1) Models

In the current simulation study, three MA(1) models are considered. The values of the coefficient θ are assumed to be 0.2, 0.5, and 0.8. The first choice is close to the center of the invertibility domain of the MA(1) model. The last choice is close to the boundary, whereas the second choice is in between. The results of each model are summarized in one of the tables (5.1) up to (5.3).

The three tables have the same design. Each row of the table displays the results obtained for a certain time series length. Whereas, the columns of the table are divided into two parts: the first part is devoted to the values of P^* for the considered three approximate one step ahead predictive densities and the second part is devoted to the corresponding values of the percentage MAD.

Table (5.1)
The Percentages P^* and MAD for MA(1) with $\theta = 0.2$

	P^*			MAD		
n	g prior	NC prior	Jeffreys' prior	g prior	NC prior	Jeffreys' prior
20	96.6	97.4	97.0	55.85	55.67	56.00
25	94.0	94.8	94.2	58.62	58.62	58.68
30	95.4	95.6	95.6	55.81	55.68	55.83
100	94.6	94.0	94.6	60.43	60.42	60.41
200	95.6	95.8	95.6	52.60	52.63	52.57
500	96.2	95.4	96.2	53.93	53.86	53.87

In view of table (5.1), one may observe the following: first, the values of P^* fluctuate around 95% (the theoretical probability of the HPD region) for the three approximate one step ahead predictive densities for all considered time series lengths. Moreover, the differences in the values of P^* are negligible. Second, the values of the percentage MAD are close to each other for all considered time series lengths and the differences in the MAD values are also negligible. These results

indicate that the considered three approximate one step ahead predictive densities succeeded to forecast the first future observation of the considered MA(1) model. Moreover, the performance of the approximate one step ahead predictive densities based on the considered three priors are equivalent.

Table (5.2)
The Percentages P^* and MAD for MA(1) with $\theta = 0.5$

	P^*			MAD		
n	g prior	NC prior	Jeffreys' prior	g prior	NC prior	Jeffreys' prior
20	95.8	97.0	96.0	57.91	57.54	57.96
25	94.2	94.6	95.0	60.99	60.76	61.06
30	94.0	95.4	95.0	57.33	57.15	57.32
100	94.2	93.8	94.6	61.75	61.73	61.75
200	95.2	95.0	95.4	56.12	56.15	56.07
500	96.2	96.2	96.2	55.09	55.05	54.99

Table (5.3)
The Percentages P^* and MAD for MA(1) with $\theta = 0.8$

	P^*			MAD		
N	g prior	NC prior	Jeffreys' prior	g prior	NC prior	Jeffreys' prior
20	95.2	95.0	95.8	64.22	63.28	64.34
25	95.0	95.2	95.2	67.86	67.60	67.99
30	93.0	94.2	93.4	63.93	63.64	63.97
100	93.8	93.6	94.2	66.84	66.87	66.83
200	95.2	94.8	95.2	64.28	64.24	64.22
500	97.0	97.0	97.0	59.73	59.72	59.74

Regarding tables (5.2) and (5.3), one may obtain similar conclusions as those obtained from table (5.1).

5.2. Results of the MA(2) Model

Three MA(2) models are considered. The selected coefficients (θ_1, θ_2) are (0.2, 0.2), (0.1, -0.8), and (0.8, -0.5) which satisfy the invertibility conditions of the MA(2) model. The results of these model are summarized in tables (5.4) up to (5.6).

Table (5.4)
The Percentages P^* and MAD for MA(2) with $\theta_1 = 0.2$ and $\theta_2 = 0.2$

	P^*			MAD		
n	g prior	NC prior	Jeffreys' prior	g prior	NC prior	Jeffreys' prior
20	93.60	94.40	94.40	58.57	57.29	58.86
25	92.20	92.60	93.40	60.64	60.58	60.74
30	94.40	94.80	94.80	57.05	56.95	57.14
100	93.00	92.80	93.20	60.46	60.48	60.46
200	95.40	95.60	95.40	52.29	52.31	52.29
500	95.80	96.00	95.80	54.08	54.02	54.08

Table (5.5)
The Percentages P^* and MAD for MA(2) with $\theta_1 = 0.1$ and $\theta_2 = -0.8$

	P^*			MAD		
n	g prior	NC prior	Jeffreys' prior	g prior	NC prior	Jeffreys' prior
20	94.00	94.60	94.60	63.74	63.26	64.06
25	92.20	91.60	92.60	65.56	65.70	65.68
30	93.00	93.60	93.60	66.17	65.86	66.32
100	93.00	92.60	93.40	66.92	66.79	66.93
200	95.80	95.80	95.80	57.87	57.84	57.88
500	95.20	95.20	95.20	61.06	61.03	61.06

Table (5.6)
The Percentages P^* and MAD for MA(2) with $\theta_1 = 0.8$ and $\theta_2 = -0.5$

	P^*			MAD		
n	g prior	NC prior	Jeffreys' prior	g prior	NC prior	Jeffreys' prior
20	94.00	94.60	94.60	58.77	58.10	59.10
25	92.20	91.60	92.60	60.59	60.63	60.78
30	93.00	93.60	93.60	59.44	58.96	59.60
100	93.00	92.60	93.40	62.66	62.60	62.66
200	95.80	95.80	95.80	54.16	54.14	54.16
500	95.20	95.20	95.20	55.85	55.83	55.85

The conclusions obtained from tables (5.4), (5.5) and (5.6) are similar to those obtained from table (5.1).

6. Discussion and Conclusion

The current study introduces three approximate one step ahead predictive densities, based on three well known priors, to forecast the first future observation of MA models. The three approximate one step ahead predictive densities are all non-central t with different parameters. The performance of the approximate predictive densities was investigated via comprehensive simulation studies in which a group of time series following MA(1) and MA(2) models are extensively analyzed.

In view of the simulation results one may assert the following conclusions:

1. The three approximate one step ahead predictive densities lead to good interval forecast since they achieve almost the same high values of P^* which is around its theoretical probability of the HPD region, namely 0.95, in all cases.
2. There is no observable difference in the accuracy of the interval forecasts between the three approximate one step-ahead predictive densities in the sense of both P^* and MAD measures defined above.
3. The above conclusions are not affected by changes in the values of the models' coefficients or the time series length.

Generally speaking the g prior, NC prior, and Jeffreys prior performs almost equally in forecasting the first future observation of MA models. It may be recommended to use the Jeffreys' prior since it overcomes the problem of estimating the hyper-parameters.

References

[1] Albassam, M.S., Soliman E.E.A. and Ali, S.S. (2013). Selection of Prior Distribution in the Indirect Bayesian Identification of Moving Average Models. *Canadian Journal on Computing in Mathematics, Natural Sciences, Engineering and Medicine*, Vol.4, No.2, pp(205-212).

[2] Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer Verlag, New York.

[3] Box, G. and Jenkins, G. (1970), *Time Series Analysis: Forecasting and Control*. Holden day, New York.

[4] Broemeling, L. D. (1985). *Bayesian Analysis of Linear Models.* Marcel Dekker, New York.

[5] Broemeling, L. D. and Land, M. (1984). On Forecasting with Univariate Autoregressive Processes: A Bayesian Approach. *Communications in Statistics*, Vol. 13(11), pp. (1305-1320).

[6] Broemeling, L., Shaarawy, S. (1988). Time series: A Bayesian analysis in time domain.. *Bayesian Analysis of Time Series and Dynamic Models edited by Spall, J. pp. 1–22.*

[7] DeGroot, M.H. (1970). *Optimal Statistical Decision*. McGraw-Hill, New York.

[8] El-Zayat, N. (2007). *On The Selection of Prior Distributions in Bayesian Analysis*. Unpublished M.Sc. thesis, Dept. of statistics, Faculty of Economics and Political Sciences. Cairo University.

[9] Fernández, Carmen & Ley, Eduardo & Steel, Mark F. J. (2001). Benchmark Priors for Bayesian Model Averaging. *Journal of Econometrics*, *Elsevier*, vol. 100(2), pages 381-427, February.

[10] Frederick, B. (2000). *Development of a Predictive Bayesian Microbial Dose-Response Function*. Unpublished Ph.D. thesis, University of Miami, Dissertation Abstracts International, Vol. 62-03(B), p. (1491).

[11] Gyllenberg, M. and Koski, T. (2002). Bayesian Predictiveness Exchangeability and Sufficientness in Bacterial Taxonomy. *Mathematical Biosciences Journal*, 177/178, pp. (161-184).

[12] Jeffreys, H. (1961). *Theory of Probability (3rd Ed.)*. Oxford University Press, London.

[13] Karlsson, S. (2001). *Bayesian Methods in Econometrics: Linear Regression.* (A course to a graduate level class), Stockholm School of Economics, Swedish School of Economics and Business Administration (Helsinki).

[14] Monahan, J. F. (1983). Fully Bayesian Analysis of ARIMA Time Series Models. *Journal of Econometrics*, Vol. 21, pp. (307-331).

[15] Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory.* Division of research, graduate school of business administration, Harvard University, Boston.

[16] Robbins, H. (1956). *An Empirical Bayes Approach to Statistics.* Proceeding of the third Berkeley symposium on mathematical statistics, 1, pp. (157-163), University of California Press, Berkeley.

[17] Singh, A. B. M. (2006). *Enrollment Prediction Using Bayesian Multiple Logistic Regression.* Unpublished MS. thesis, Dissertation Abstracts International, Vol. 44-05, p. (2318).

[18] Shaarawy, S. M. and Broemeling, L. (1984). Bayesian Inferences and Forecasts with Moving Averages Processes. *Communications in Statistics*, 13(15), pp. (1871-1888).

[19] Shaarawy, S.M., Soliman, E.E.A. and Shahin, H. (2010). Bayesian prediction of autoregressive models using different types of priors. *The Egyptian Statistical Journal*, Vol.54, No.2, pp(108-126).

[20] Shaarawy, S.M., Soliman, E.E.A. and Sorour, W.W. (2015). On Bayesian identification of autoregressive processes. *Pakistan Journal of Statistics and Operation Research*, Vol.11, No.1, pp(11-28).

[21] Soliman, E. E. A. (1999). *On Bayesian Time Series Analysis*. Unpublished Ph.D. thesis, Dept. of statistics, Faculty of Economics and Political Sciences, Cairo University.

[22] Soliman, E.E.A. (2008). Efficiency of the Approximate Bayesian Prediction of ARMA Models: A Simulation Study. *Egyptian Statistical Journal*, Vol.52, No.2, pp (56-75).

[23] Zellner, A. (1971). *An Introduction to Bayesian in Econometrics.* John Wiley and Sons, New York.

[24] Zellner, A. (1986). *On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions. Bayesian Inference and Decision Techniques*. Essays in Honor of Bruno de Finetti, p.d.f. Goel and A. Zellner, eds., North Holland, pp. (233-243), New York.