On Bayes Estimation of the First Order Moving Average Model

Mahmoud M. Smadi and Ahmed A. Abu-Taleb
Department of Mathematics and Statistics, Jordan University of Science and Technology Jordan
P.O. BOX 3030, 22110 Irbid, Jordan

Abstract: In this work, Bayes estimation of the first order moving average model (MA(1)) were studied. Theoretical justification of the Bayes estimates based on the estimated innovations is given. The convergence of Bayes and maximum likelihood estimates are examined via simulation using different parameter values. Also, Bayes estimates were determined when the model is invertible using the estimated innovations. For long series lengths, it has been noted that the Bayes estimate of \( \theta \) of invertible MA(1) model assuming uniform prior on \( \theta \) and inverted gamma prior on \( \sigma^2 \), equals the Bayes estimate of \( \theta \) for noninvertible MA(1) model. Generally, the simulation results showed that the performance of the Bayes estimates using estimated innovations depends on the values of \( \theta \) within the invertibility region. As expected, we note that the performance of the maximum likelihood and Bayes estimates are equally likely for long series lengths.

Keywords: Bayes Estimation; Estimated Innovations; Moving Average Model.

INTRODUCTION

Bayesian inferences on autoregressive moving average (ARMA) models are limited due to the complicated form of the likelihood function which makes it difficult to get analytical tractable results. In addition, the Bayesian estimation of ARMA models is based on the causality and inevitability conditions of the model coefficients. The Bayesian inference on time series goes back to Zellner\(^{[14]}\), when the posterior and predictive distributions were derived for the first and the second order AR models using vague priors and analysis of regression models with autocorrelated errors. Box and Jenkins\(^{[3]}\) introduced Bayesian analysis for ARMA models without any restrictions on causality and inevitability conditions. Monahan\(^{[12]}\), used numerical integration technique to implement Bayesian time series. Bromeling and Shaarawy\(^{[6]}\) proposed an approximate Bayes estimates based on the estimated innovations for the analysis of MA and ARMA models without any restrictions on the causality and inevitability conditions. The estimated innovations were used later by Chen\(^{[9]}\) in Bayesian inference of the bilinear models. Marriot et. al.\(^{[11]}\) used Monte Carlo Markov Chain (MCMC) methods to implement the Bayesian inference on the ARMA models. Chen et. al\(^{[10]}\) used MCMC methods for exploration of the joint posterior distribution of threshold autoregressive (TAR) models. Smadi\(^{[3]}\) used MCMC methods for the Bayesian inference on the threshold autoregression moving average (TARMA) models using the estimated innovations.

In this work, the Bayes estimation of the first order moving average model (MA(1)) has been investigated using the estimated innovations. Asymptotic justification of Bayes estimation of MA(1) model is provided. The convergence of the Bayes and the maximum likelihood estimates are examined via simulation using different parameter values. Moreover, Bayes estimates of the first order invertible moving average model have been considered. Bayes theorem of estimation is described below\(^{[7]}\).

\textbf{Theorem (1)}

\( \pi(\theta) \) be a prior distribution of \( \theta \). A sample \( X_1, X_2, \ldots, X_n \) is then drawn from a population indexed by \( \theta \). Let \( f(x | \theta) \) be the sampling distribution of \( X_1, X_2, \ldots, X_n \). The prior distribution is updated by the sample information. The updated prior is called the posterior distribution which is a conditional distribution of \( \theta \) given a sample can be described as:

\[ \pi(\theta | x) = f(x | \theta) \pi(\theta) / m(x), \]

where \( m(x) \) is the marginal distribution of \( X \) which is given by:

\[ m(x) = \int f(x | \theta) \pi(\theta) d\theta \]

The mean of the posterior distribution can be used as a point estimate of \( \theta \) by means of squared loss function.

The first order moving average model (MA(1)) is given by:

\[ X_t = \theta Z_{t-1} + Z_t, \quad t = 0, \pm1, \pm2, \ldots \]

where \( \{Z_t\} \) are \( I.I.D \ N(0, \sigma^2) \) random variables.
The first order moving average model given in (3) is always stationary. According to Brockwell and Davis[4] a model is judged to be invertible if 
\[ -1 < \theta < 1, \]
recursively.
\[
Z_t = \sum_{j=0}^{\infty} \theta^j X_{t-j}
\]  
\[ (4) \]
For first order moving average model (MA(1)) given in (3), the joint probability density function of \( X_1, X_2, ..., X_n \) can be written as[3]:
\[
f(X|\theta, \sigma^2, Z_0) = (2\pi)^{\frac{n}{2}} (\tau)^{\frac{n}{2}} \exp(\frac{-\tau}{2} \hat{\Sigma} (X_t - \theta Z_{t-1})^2).
\]  
\[ (5) \]
where \( Z_0 \) is non-observable random variable; assuming \( Z_0 \) equals its unconditional expectation zero. The innovations \( Z_t (t = 1,2,...,n) \) can be calculated recursively.

The conditional likelihood function is given by
\[
L(X; \theta, \sigma^2 | Z_0 = 0) = (2\pi)^{\frac{n}{2}} (\tau)^{\frac{n}{2}} \exp(\frac{-\tau}{2} \sum_{t=1}^{n} (Z_t - \theta Z_{t-1})^2)
\]  
\[ (6) \]
Bromeling[5] derived approximate marginal posterior distributions and posterior means for the parameters of the MA(q) model. Accordingly, the true innovations \( Z_t \) is estimated as \( \hat{Z}_t = X_t - \hat{\theta} Z_{t-1} \) \( (t = 1,...,n) \), where \( \hat{Z}_0 = 0 \), and \( \hat{\theta} \) is the least squares estimate of \( \theta \). Thus, the approximate conditional likelihood function is given by:
\[
\hat{L}(X; \theta, \sigma^2 | \hat{Z}_0 = 0) \propto (\tau)^{\frac{n}{2}} \exp(\frac{-\tau}{2} \sum_{t=1}^{n} (Z_t - \hat{\theta} Z_{t-1})^2)
\]  
\[ (7) \]
Bromeling[5] assumed multi-normal-gamma prior density on the model parameters of the MA(q) model. For the MA(1) model, the normal-gamma prior density on \( \theta \) and \( \tau \) is:
\[
g(\theta, \tau) = \tau^{n-1} e^{-\frac{n\theta}{2}} e^{\frac{\mu}{2\theta}} \tau^{\frac{n}{2}} \exp(\frac{\mu}{2\theta}), \quad \theta \in \mathbb{R}, \tau > 0, \alpha > 0, \beta > 0, \mu \in \mathbb{R}
\]  
\[ (8) \]
Based on the above, Bromeling[5] concluded the following results:
\begin{itemize}
  \item \( i \) The marginal posterior density of \( \theta \) has a \( t \)-distribution with \( n + 2\alpha - 1 \) degrees of freedom. The mean and precision are given by
    \[
    E(\theta | X) = \frac{p \mu - \sum_{t=1}^{n} x_t \hat{z}_{t-1}}{p + \sum_{t=1}^{n} \hat{z}_{t-1}^2}
    \]  
    \[
    (9)
    \]
    Precision = \[
    \frac{\mu^2 p + 2\mu + \sum_{t=1}^{n} x_t^2 - (\sum_{t=1}^{n} x_t \hat{z}_{t-1})^2 / \sum_{t=1}^{n} \hat{z}_{t-1}^2}
    \]  
    \[
    (10)
    \]
  \item \( ii \) The marginal posterior density of \( \tau \) is a gamma distribution, where:
    \[
    \alpha' = \frac{n + 2\alpha}{2}
    \]  
    \[
    (11)
    \]
    \[2\beta' = \mu^2 p + 2\beta + \sum_{t=1}^{n} x_t^2 - (\sum_{t=1}^{n} x_t \hat{z}_{t-1})^2 / \sum_{t=1}^{n} \hat{z}_{t-1}^2
    \]  
    \[
    (12)
    \]
    Then, if \( \alpha \to -1/2, \beta \to 0, \) and \( p \to 0, \) the normal-gamma prior density given in (8) reduces to Jeffrey’s prior described by:
    \[
g(\theta, \tau) \propto \tau^{-1}, \quad \text{where} \quad \tau > 0, \mu \in \mathbb{R}
\]  
    \[
    (13)
    \]
    Therefore, the Bayes estimate of the precision parameter \( \tau \) is given by:
    \[
    \frac{1}{n - 1} \left[ \sum_{t=1}^{n} x_t^2 - (\sum_{t=1}^{n} x_t \hat{z}_{t-1})^2 / \sum_{t=1}^{n} \hat{z}_{t-1}^2 \right]
    \]  
    \[
    (14)
    \]  
\end{itemize}

Asymptotic justification of the estimated innovations in Bayes estimation of the MA(1) model is described below using basic convergence results of the probability theory. The following theorem will be used[2].

**Theorem (2)**
\begin{itemize}
  \item \( i \) If \( Z_n \xrightarrow{P} Z \), then \( Z_n \xrightarrow{D} Z \).
  \item \( ii \) If \( Z = Z_0 \) (a constant), \( Z_n \xrightarrow{D} Z \Rightarrow Z_n \xrightarrow{P} Z \).
  \item \( iii \) If \( Z_n \xrightarrow{D} Z_0 \) and \( g \) is continuous at \( Z_0 \), then \( g(Z_n) \xrightarrow{P} g(Z_0) \).
  \item \( iv \) If \( Z_n \xrightarrow{D} Z \) and \( g \) is continuous, then \( g(Z_n) \xrightarrow{D} g(Z) \).
  \item \( v \) If \( Z_n \xrightarrow{D} Z \) and \( U_n \xrightarrow{P} u_0 \) (a constant), then
    \[
    (a) \quad Z_n + U_n \xrightarrow{D} Z + u_0
    \]  
    \[
    (b) \quad U_n Z_n \xrightarrow{D} u_0 Z
    \]  
\end{itemize}

**Theorem (3)**
Consider the first order MA (1) model
\[
X_t = \theta Z_{t-1} + Z_t (t = 0,...,n)
\]  
\[ (15) \]
Let \( \{\hat{Z}_t\} \) be the estimated innovations which are calculated recursively, i.e.
\[
\hat{Z}_t = X_t - \hat{\theta} Z_{t-1} (t = 1,...,n), \quad \text{where} \quad \hat{Z}_0 = 0 \quad \text{and}
\]
\( \hat{\theta} \) is the least squares estimate of \( \theta \). Then, the approximate conditional joint distribution
\[
f(X | \theta, \sigma^2, \hat{Z}_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^2} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (X_i - \theta \hat{Z}_{i-1})^2\right)
\]
converges in distribution to
\[
f(X | \theta, \sigma^2, Z_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^2} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (X_i - \theta Z_{i-1})^2\right)
\]

Proof:
Let \( \hat{\theta} \) be the least squares estimate of \( \theta \), it has been shown that\(^3\)
\[
\hat{\theta} \to_{p} \theta
\]
Then, according to Theorem (2), part (v),
\[
\hat{Z}_t = X_t - \hat{\theta} \hat{Z}_{t-1} \to_{p} Z_t = X_t - \theta Z_{t-1}
\]
Consequently,
\[
\hat{Z}_t \to_{p} Z_t.
\]
Also, according to Theorem (2), part (v),
\[
\sum_{i=1}^{n} (X_i - \theta \hat{Z}_{i-1})^2 \to_{p} \sum_{i=1}^{n} (X_i - \theta Z_{i-1})^2,
\]
Finally, because of the continuity of the exponential function and Theorem (2), part (iv), the following approximate conditional joint distribution function
\[
f(X | \theta, \sigma^2, \hat{Z}_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^2} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (X_i - \theta \hat{Z}_{i-1})^2\right)
\]
converges in distribution to
\[
f(X | \theta, \sigma^2, Z_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^2} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (X_i - \theta Z_{i-1})^2\right)
\]
Similarly this result can be generalized to the MA(q) model.

The Bayes Estimation of the Invertible MA(1) Model
The first order invertible moving average model, where \(-1 < \theta < 1\), will be considered. Abu-Salih and Abd-alla\(^1\) obtained in a closed form the Bayes estimates of the parameters of the stationary AR(1) model using different informative and non-informative priors. The AR(1) model is given by:
\[
X_t = \phi X_{t-1} + Z_t \quad (t = 0, \ldots, n)
\]
where \(\{Z_t\} \sim IID N(0, \sigma^2)\), \(-1 < \phi < 1\).

The following joint prior on \(\phi\) and \(\sigma^2\) is used in the Bayesian estimation:
\[
g(\phi, \sigma^2) = \frac{\alpha^d e^{\frac{-\alpha}{\sigma^2}}}{(\sigma^2)^{d+1} \Gamma(d)} \cdot \frac{1}{\sigma^2 - (\phi^2 - 1)} \quad \sigma^2 > 0, -1 < \phi < 1,
\]
where \(\phi\) and \(\sigma^2\) are independent, with inverted gamma prior on \(\sigma^2\) with hyper parameters \(d\) and \(\alpha\) and uniform prior on \(\phi\).

Assuming squared loss function, Abu-Salih and Abd-alla\(^1\) derived in a closed form the marginal posteriors of \(\phi\) and \(\sigma^2\) for the stationary AR(1) model. The derivation for the above invertible MA(1) model is similar to the stationary AR(1) model; where \(X_i\) is regressed on \(\hat{Z}_{i-1}\) rather than \(X_{i-1}\). The final expressions of the marginal posterior means will be followed, details of the derivations and integration results can be found in Abu-Salih and Abd-alla\(^1\).

The Bayes estimates of \(\theta\) is given by
\[
\hat{\theta} = r + \frac{1}{\sqrt{a}} \left(\frac{1 - r}{\sqrt{a}} + \frac{1 + r}{\sqrt{a}}\right),
\]
where
\[
k = \frac{1}{\sqrt{a}} \left(\frac{1 - r}{\sqrt{a}} + \frac{1 + r}{\sqrt{a}}\right),
\]
\[
b(j) = \frac{j(j-1)(4a)^2}{(2j)!} \left(\frac{1 - r}{(1-r)^2 + a} + \frac{1 + r}{(1-r)^2 - a}\right),
\]
\[
r = \frac{\sum_{i=1}^{n} X_i \hat{Z}_{i-1}}{\sum_{i=1}^{n} \hat{Z}_{i-1}^2}, \quad a = \left(\frac{2\alpha + \sum_{i=1}^{n} \hat{Z}_{i-1}^2}{\sum_{i=1}^{n} \hat{Z}_{i-1}^2}\right) - r^2,
\]
for large \(n\),
\[
\hat{\theta} \sim r + \frac{\sum_{i=1}^{n} X_i \hat{Z}_{i-1}}{\sum_{i=1}^{n} \hat{Z}_{i-1}^2},
\]
The Bayes estimator of \(\sigma^2\) is given by:
\[
\hat{\sigma}^2 = \frac{a \sum_{i=1}^{n} \hat{Z}_{i-1}^2}{2m - 1} (1 - \frac{b(m)}{\sum_{j=1}^{m} b(j) + 2\sqrt{ak}}),
\]
For large \(n\),
\[
\hat{\sigma}^2 = \frac{a \sum_{i=1}^{n} \hat{Z}_{i-1}^2}{2m - 1},
\]
It is noted that for long series lengths, the approximate Bayes estimate (Eq. 23) of $\theta$ in the invertible MA(1) model, assuming uniform prior on $\theta$ and inverted gamma prior on $\sigma$, equals the approximate Bayes estimate of $\theta$ (Eq. 9) in the noninvertible MA(1) model, assuming Jeffrey's prior. Thus, it can be concluded that both estimates of invertible and non-invertible models are the same. In addition, there is no effect of the inverted gamma prior on $\sigma$ when the Bayes estimate of $\theta$ for long record lengths. On the other hand, as one expects, there is an effect of the inverted gamma prior on $\sigma$ when comparing the mean of the marginal posterior density of $\tau$ derived by Bromeling (Eq. 14) and the asymptotic Bayes estimate of the invertible model (Eq. 25).

RESULTS AND DISCUSSION

A simulation was carried out to examine the convergence of the Bayes estimates based on the estimated innovations. The results were compared with the maximum likelihood estimates. The means and root mean square errors (RMSE) of the maximum likelihood estimates $\hat{q}$ and $\hat{s}$ and the Bayes estimates $\hat{\theta}$ and $\hat{\sigma}^2$ were calculated using different combinations of parameter values ($\theta$, $\sigma^2$); namely, ($\theta$=-0.3, $\sigma^2$=1), ($\theta$=0.3, $\sigma^2$=1), ($\theta$=-0.6, $\sigma^2$=1), ($\theta$=0.6, $\sigma^2$=1) and ($\sigma^2$=1, $\theta$=-0.3). The results are displayed in Tables 1-5. The simulation results were based on 1000 replicates for series lengths of 30, 50, 100, 200, 400, 700 and 1000.

| Table 1: MLE and Bayes estimates with RMSE of $\theta$ ($\theta$ = 0.3, $\sigma^2$ =1) |
|---|---|---|---|---|
| $n$ | $\hat{\theta}$ | RMSE($\hat{\theta}$) | $\hat{\theta}_B$ | RMSE($\hat{\theta}_B$) |
| 30 | 0.3978 | 0.3192 | 0.2248 | 0.2849 |
| 50 | 0.3490 | 0.1938 | 0.2934 | 0.2162 |
| 100 | 0.3148 | 0.1092 | 0.2943 | 0.1487 |
| 200 | 0.3089 | 0.0714 | 0.2993 | 0.1051 |
| 400 | 0.3056 | 0.0497 | 0.3006 | 0.0510 |
| 700 | 0.3027 | 0.0362 | 0.2996 | 0.0371 |
| 1000 | 0.3019 | 0.0313 | 0.3001 | 0.0322 |

| Table 2: MLE and Bayes estimates with RMSE of $\theta$ ($\theta$ = 0.6, $\sigma^2$ =1) |
|---|---|---|---|---|
| $n$ | $\hat{\theta}$ | RMSE($\hat{\theta}$) | $\hat{\theta}_B$ | RMSE($\hat{\theta}_B$) |
| 30 | 0.7902 | 0.3254 | 0.5696 | 0.2223 |
| 50 | 0.7202 | 0.2362 | 0.5875 | 0.1552 |
| 100 | 0.6737 | 0.1700 | 0.5956 | 0.1035 |
| 200 | 0.6373 | 0.1210 | 0.5971 | 0.0734 |
| 400 | 0.6176 | 0.0807 | 0.5977 | 0.0505 |
| 700 | 0.6070 | 0.0478 | 0.5988 | 0.0373 |
| 1000 | 0.6026 | 0.0311 | 0.5983 | 0.0322 |

| Table 3: MLE and Bayes estimates with RMSE of $\theta$ ($\theta$ = -0.3, $\sigma^2$ =1) |
|---|---|---|---|---|
| $n$ | $\hat{\theta}$ | RMSE($\hat{\theta}$) | $\hat{\theta}_B$ | RMSE($\hat{\theta}_B$) |
| 30 | -0.2832 | 0.2512 | -0.2903 | 0.2249 |
| 50 | -0.2906 | 0.1639 | -0.2927 | 0.1599 |
| 100 | -0.2976 | 0.1078 | -0.2993 | 0.1079 |
| 200 | -0.2970 | 0.0725 | -0.2975 | 0.0730 |
| 400 | -0.2974 | 0.0481 | -0.2979 | 0.0511 |
| 700 | -0.2995 | 0.0355 | -0.2995 | 0.0371 |
| 1000 | -0.2991 | 0.0305 | -0.2993 | 0.0323 |
Table 4: MLE and Bayes estimates with RMSE of $\theta$ ($\theta = -0.6, \sigma^2 = 1$)

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\hat{\theta}$</th>
<th>RMSE($\hat{\theta}$)</th>
<th>$\hat{\theta}_B$</th>
<th>RMSE($\hat{\theta}_B$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>-0.6089</td>
<td>0.2216</td>
<td>-0.5679</td>
<td>0.2152</td>
</tr>
<tr>
<td>50</td>
<td>-0.6220</td>
<td>0.1820</td>
<td>-0.5787</td>
<td>0.1610</td>
</tr>
<tr>
<td>100</td>
<td>-0.6221</td>
<td>0.1340</td>
<td>-0.5890</td>
<td>0.1045</td>
</tr>
<tr>
<td>200</td>
<td>-0.6217</td>
<td>0.1150</td>
<td>-0.5940</td>
<td>0.0735</td>
</tr>
<tr>
<td>400</td>
<td>-0.6110</td>
<td>0.0800</td>
<td>-0.5970</td>
<td>0.0504</td>
</tr>
<tr>
<td>700</td>
<td>-0.6036</td>
<td>0.0509</td>
<td>-0.5977</td>
<td>0.0379</td>
</tr>
<tr>
<td>1000</td>
<td>-0.6005</td>
<td>0.0287</td>
<td>-0.5985</td>
<td>0.0322</td>
</tr>
</tbody>
</table>

Table 5: MLE and Bayes estimates with RMSE of $\sigma^2$ ($\sigma^2 = 1, \theta = 0.3$)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\hat{\sigma}^2$</th>
<th>RMSE($\hat{\sigma}^2$)</th>
<th>$\hat{\sigma}^2_B$</th>
<th>RMSE($\hat{\sigma}^2_B$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.9145</td>
<td>0.2835</td>
<td>0.8420</td>
<td>0.2849</td>
</tr>
<tr>
<td>50</td>
<td>0.9600</td>
<td>0.2041</td>
<td>0.9103</td>
<td>0.2162</td>
</tr>
<tr>
<td>100</td>
<td>0.9821</td>
<td>0.1092</td>
<td>0.9574</td>
<td>0.1487</td>
</tr>
<tr>
<td>200</td>
<td>0.9866</td>
<td>0.1022</td>
<td>0.9750</td>
<td>0.1051</td>
</tr>
<tr>
<td>400</td>
<td>0.9948</td>
<td>0.0720</td>
<td>0.9891</td>
<td>0.0738</td>
</tr>
<tr>
<td>700</td>
<td>0.9978</td>
<td>0.0552</td>
<td>0.9947</td>
<td>0.0556</td>
</tr>
<tr>
<td>1000</td>
<td>0.9990</td>
<td>0.0463</td>
<td>0.9966</td>
<td>0.0464</td>
</tr>
</tbody>
</table>

Generally, the simulation results showed that the performance of the Bayes estimates using the estimated innovations depends on the values of $\theta$ within the invertibility region. As we expect, we note that the performance of the maximum likelihood and the Bayes estimates are equally likely for long series lengths.

**CONCLUSION**

It can be concluded that for long series lengths, the approximate Bayes estimate of $\theta$ of invertible MA(1) model assuming uniform prior on $\theta$ and inverted gamma prior on $\sigma$ equals the Bayes estimate of $\theta$ of noninvertible MA(1) model, assuming Jeffrey's prior. The simulation results showed that the performance of the Bayes estimates using the estimated innovations depends on the values of $\theta$ within the invertibility region. For long series lengths, we note that both maximum likelihood and Bayes estimates performs equally likely.

**REFERENCES**