Bayesian Inference for Threshold ARMA Models

Mohamed A Ismail

Department of Statistics, UAE University
P. O. Box 17555, Al-Ain, United Arab Emirates

Abstract

Bayesian inference for threshold autoregressive moving average models (SETARMA) is developed. Using both informative and non-informative prior, the posterior density of the model coefficients is approximated by a multivariate t distribution. The posterior density of the model precision parameter is approximated by a gamma density. The proposed Bayesian methodology is checked via a simulation study. In addition, the new methodology is illustrated using the series of the US unemployment rates.

Keywords: Posterior distributions, t density, Monte Carlo, US unemployment.

1 Introduction

recently, quite a few nonlinear time series models have been proposed in the literature. Most of the suggested nonlinear models include some type of nonlinearity in the autoregressive part. Among these models, the self exciting threshold autoregressive (SETAR) model is perhaps the most popular one. The SETAR model was introduced by Tong and Lim (1980) and studied steadily in Tong (1983,1990) and Tasy (1989) among others. The dual counterpart of the SETAR model, i.e the self exciting threshold moving average (SETMA) model has received less attention. The SETMA model is considered by Gooijer and Kumar (1992) and studied in detail by Gooijer (1998). Tong (1990) and Brockwell et al. (1992) introduced the self exciting threshold autoregressive moving average (SETARMA) model. Stramer (1996) discussed the estimation of the conditional moments of SETARMA model.

Unfortunately most of the reported work on threshold models in the literature is focused on non Bayesian analysis. However, Kheradmandnia (1991) introduced Bayesian analysis of threshold autoregressive models. Unaware of Kheradmandnia's work, Broemeling and Cook (1992) and Geweki and Terui (1993) proposed Bayesian analysis of threshold autoregressive models. Chen and Lee (1995) employed Gibbs sampling approach to develop a Bayesian analysis for SETAR models. Chen (1998) extended Chen and Lee's (1995) SETAR Bayesian analysis to a generalized SETAR model where exogeneous variables are added and the threshold is allowed to be a function in an exogeneous variable. On the other hand, Bayesian analysis of SETARMA models is not known.

Bayesian analysis of moving average models is difficult even for linear models since the likelihood function is highly nonlinear in the parameters, which cause problems in prior specification and posterior analysis. Thus, the integrations involved in Bayesian analysis must be done numerically. Obviously Bayesian analysis for nonlinear time series models is much harder simply because the likelihood function is more complicated. A possible solution to the problem is Quasi Conjugate Analysis, where any constrains on the model coefficients are ignored and moving average terms are replaced by sensible estimates. The advantage of using the quasi conjugate approach is that it leads to nice standard distribution results for the posterior densities which facilitates the calculations required for inference. Quasi Conjugate Analysis is used by several authors including Broemeling and Shaarawy (1988), Chen (1992) and Ismail (1994) among others and it is going to be employed in this paper. The objective of this paper is to develop an approximate non numerical Bayesian inferential technique for SETARMA models. The approximate posterior densities for the model coefficients and precision in each regime are derived. Both of informative and non informative prior are used. The adequacy of the proposed technique is checked via a Monte Carlo study and a real data set.

This paper is organized as follows. Section 2 introduces and explains the self exciting threshold autoregressive moving average model. The posterior analysis is developed in Section 3. Section 4 illustrates the methodology using a simulation study and a real data set. Section 5 is conclusions.

2 Threshold Autoregressive Moving Average Models

A time series y_t is said to follow a SETARMA of order $(m; p_1, q_1, p_2, q_2, \cdots, p_m, q_m)$, if it satisfies

$$y_t = \mu^{(j)} + \varepsilon_t^{(j)} + \sum_{i=1}^{p_j} \phi_i^{(j)} y_{t-i}^{(j)} + \sum_{i=1}^{q_j} \theta_i^{(j)} \varepsilon_{t-i}^{(j)}, \quad r_{j-1} \le y_{t-d} \le r_j \tag{1}$$

where $j = 1, \cdots, m$, d is a positive integer commonly referred to as the delay of the model, $\mu^{(j)}$, $\theta^{(j)}$ are constants. The m sequences $\varepsilon^{(j)}$ satisfy $\varepsilon_t^{(j)} = \tau_j^{-\frac{1}{2}} \varepsilon_t$, where $\varepsilon_t \sim NID(0,1)$ and where $0 < \tau_j = \frac{1}{\sigma_j^2} < \infty$, $(j = 1, \cdots, m)$. The real numbers r_j (called thresholds) satisfy $-\infty = r_0 < r_1 < \cdots < r_m = \infty$ and form a partition of the space of y_{t-d} and the partition $r_{j-1} \le y_{t-d} \le r_j$ forms the j^{th} regime of the SETARMA model.

Model (1) is said to be self exciting because changes in the model parameters are generated by past values of y_t itself. The model equation (1) reduces to the defining equation for an ARMA(p,q) when m=1. In this paper, we shall take up the case where $m = 2$ in detail. Let $\mu^{(1)} = \mu^{(2)} = 0$ for simplicity, the SETARMA$(2; p_1, q_1, p_2, q_2)$ may be expressed as

$$
y_t = \begin{cases} \sum_{i=1}^{p_1} \phi_i^{(1)} y_{t-i}^{(1)} + \sum_{i=1}^{q_1} \theta_i^{(1)} \varepsilon_{t-i}^{(1)} + \varepsilon_t^{(1)}, \ if\ y_{t-d} \le r_1 & (2.a) \\ \sum_{i=1}^{p_2} \phi_i^{(2)} y_{t-i}^{(2)} + \sum_{i=1}^{q_2} \theta_i^{(2)} \varepsilon_{t-i}^{(2)} + \varepsilon_t^{(2)}, \ if\ y_{t-d} > r_1 & (2.b) \end{cases} \qquad (2)
$$

Equations (2.a) and (2.b) are sometimes called the lower regime and upper regime, respectively.

3 Posterior Analysis

3.1 The Conditional Likelihood Function

Suppose $S_n = (y_1, y_2, \cdots, y_n)$ is a realization of SETARMA$(2;\, p_1, q_1, p_2, q_2)$ process. In this study we employ the concept of arranged autoregression where the cases (observations) are rearranged according to the threshold variable y_{t-d}. The autoregression approach was used by Tsay (1989), Lee and Chen (1996) for SETAR model and by Chen (1998) for generalized SETAR model.

Employing the arranged autoregression approach the SETARMA$(2;\, p_1, q_1, p_2, q_2)$ given by (2) can be written as

$$
y_{\pi_i+d} = \begin{cases} \sum_{j=1}^{p_1} \phi_j^{(1)} y_{\pi_i+d-j}^{(1)} + \sum_{j=1}^{q_1} \theta_j^{(1)} \varepsilon_{\pi_i+d-j}^{(1)} + \varepsilon_{\pi_i+d}^{(1)} & i \le s \\ \sum_{j=1}^{p_2} \phi_j^{(2)} y_{\pi_i+d}^{2} + \sum_{j=1}^{q_2} \theta_j^{(2)} \varepsilon_{\pi_i+d-j}^{(2)} + \varepsilon_{\pi_i+d}^{2} & otherwise \end{cases}
$$

where s satisfies $y_{\pi_s} \le r < y_{\pi_s+1}$, $\gamma^{(1)} = \left(\phi_1^{(1)}, \phi_2^{(1)}, \cdots \phi_{p_1}^{(1)}, \theta_1^{(1)}, \theta_2^{(1)}, \cdots \theta_{q_1}^{(1)}\right)$, $\gamma^{(2)} = \left(\phi_1^{(2)}, \phi_2^{(2)}, \cdots \phi_{p_2}^{(2)}, \theta_1^{(2)}, \theta_2^{(2)}, \cdots \theta_{q_2}^{(2)}\right)$. The initial errors $\varepsilon_{1-q}, \varepsilon_1, \cdots, \varepsilon_0$ where $q = \max(q_1, q_2)$ are assumed zeros. Let π_i be the index of the i-th smallest observation of $\{y_1, y_2, \cdots, y_{n-d}\}$ and $\eta = (d, r)$. The conditional likelihood function of $\left(\gamma^{(1)}, \gamma^{(2)}, \tau_1, \tau_2, \eta\right)$ is given by

$$
\begin{aligned}
L\left(\gamma^{(1)}, \gamma^{(2)}, \tau_1, \tau_2, \eta \mid S_n\right) &\propto \tau_1^{\frac{s}{2}} \tau_2^{\frac{n-s}{2}} \times \exp\left\{-\frac{\tau_1}{2}\sum_{i=1}^{s}\left(\varepsilon_t^{(1)}\right)^2 - \frac{\tau_2}{2}\sum_{i=s+1}^{n-d}\left(\varepsilon_t^{(2)}\right)^2\right\} \\
&= \tau_1^{\frac{s}{2}} \tau_2^{\frac{n-s}{2}} \times \exp\left\{-\frac{\tau_1}{2}\sum_{i=1}^{s}\left(y_{\pi_i+d} - \sum_{j=1}^{p_1}\phi_j^{(1)} y_{\pi_i+d-j}^{(1)} - \sum_{j=1}^{q_1}\theta_j^{(1)}\varepsilon_{\pi_i+d-j}^{(1)}\right)^2\right. \\
&\qquad \left. -\frac{\tau_2}{2}\sum_{i=s+1}^{n-d}\left(y_{\pi_i+d} - \sum_{j=1}^{p_2}\phi_j^{(2)} y_{\pi_i+d-j}^{(2)} - \sum_{j=1}^{q_2}\theta_j^{(2)}\varepsilon_{\pi_i+d-j}^{(2)}\right)^2\right\} \qquad (3)
\end{aligned}
$$

It is clear that the likelihood is a complicated nonlinear function in the parameters $\phi_l^{(i)}, \theta_j^{(i)}$ $i = 1,2$; $l = 1,2,\cdots,p_i$; $j = 1,2,\cdots,q_i$. Suppose the errors $\varepsilon_t^i, i = 1,2$ are estimated by nonlinear least squares $e_t^i, i = 1, 2$ as follows.

$$\begin{aligned}
e^{(1)}_{\pi_t+d} &= y_{\pi_t+d} - \sum_{j=1}^{p_1} \hat{\phi}^{(1)}_j \; y^{(1)}_{\pi_t+d-j} - \sum_{j=1}^{q_1} \hat{\theta}^{(1)}_j \; e^{(1)}_{\pi_t+d-j} \qquad t = 1, 2, \cdots n_1, \\
e^{(1)}_{1-q} &= \cdots = e^{(1)}_{-1} = e^{(1)}_0 = 0 \\
e^{(2)}_{\pi_t+d} &= y_{\pi_t+d} - \sum_{j=1}^{p_2} \hat{\phi}^{(2)}_j \; y^{(2)}_{\pi_t+d-j} - \sum_{j=1}^{q_2} \hat{\theta}^{(2)}_j \; e^{(2)}_{\pi_t+d-j} \quad t = n_1 + 1, n_1 + 2, \cdots n - d - n_1 \\
e^{(2)}_{1-q} &= \cdots = e^{(2)}_{-1} = e^{(2)}_0 = 0
\end{aligned} \tag{4}$$

where $y_{\pi_{n_1}} \leq \hat{r} < y_{\pi_{n_1+1}}$ and $\hat{\theta}_j, \hat{d}, \hat{r}$ are nonlinear least squares estimates found by minimizing

$$S(\gamma^{(1)},\ \gamma^{(2)},\ d,\ r) = \sum_{t=1}^{s} (\epsilon_t^{(1)})^2 + \sum_{i=s+1}^{n-d-s} (\epsilon_t^{(2)})^2$$

with respect to $\gamma^{(1)},\ \gamma^{(2)},\ d,\ r$

Substituting the estimated errors $e_t^{(i)}, i = 1, 2$ in the conditional likelihood function $L\left(\gamma^{(1)}, \gamma^{(2)}, \tau_1, \tau_2, \eta \mid S_n\right)$ results in an approximate conditional likelihood function $L^*\left(\gamma^{(1)},\ \gamma^{(2)}, \tau_1, \tau_2\ , \eta \mid S_n\right)$ which may be written in the form :

$$\begin{aligned}
L^*\left(\gamma^{(1)},\ \gamma^{(2)}, \tau_1, \tau_2\ , \eta \mid S_n\right) &\propto \tau_1^{\frac{s}{2}} \tau_2^{\frac{n-s-d}{2}} \exp\left(-\frac{\tau_1}{2} \sum_{t=1}^{s} (e_t^{(1)})^2 - \frac{\tau_2}{2} \sum_{t=s+1}^{n-s-d} (e_t^{(2)})^2\right) \\
&= \tau_1^{\frac{s}{2}} \tau_2^{\frac{n-s}{2}} \times \exp\left\{-\frac{\tau_1}{2} \sum_{i=1}^{s} \left(y_{\pi_i+d} - \sum_{j=1}^{p_1} \phi_j^{(1)} y^{(1)}_{\pi_i+d-j} - \sum_{j=1}^{q_1} \theta_j^{(1)} e^{(1)}_{\pi_i+d-j}\right)^2\right. \\
&\qquad \left. - \frac{\tau_2}{2} \sum_{i=s+1}^{n-d} \left(y_{\pi_i+d} - \sum_{j=1}^{p_2} \phi_j^{(2)} y^{(2)}_{\pi_i+d-j} - \sum_{j=1}^{q_2} \theta_j^{(2)} e^{(2)}_{\pi_i+d-j}\right)^2\right\} \\
&= \tau_1^{\frac{s}{2}} \tau_2^{\frac{n-s}{2}} \times \prod_{i=1}^{2} \exp\left\{-\frac{\tau_i}{2}\left(y^{(i)^T} y^{(i)} - 2\gamma^{(i)^T} \mathbf{B}^{(i)}\right.\right. \\
&\qquad \left.\left. + \gamma^{(i)^T} \mathbf{A}^{(i)} \gamma^{(i)}\right)\right\}
\end{aligned}$$

where,

$$\begin{aligned}
y^{(1)} &= \left(y_{\pi_1+d}, y_{\pi_2+d}, \cdots, y_{\pi_s+d}\right)^T \textit{ is the observations of the first regime}, \\
y^{(2)} &= \left(y_{\pi_{s+1}+d}, y_{\pi_{s+2}+d}, \cdots, y_{\pi_{n-d}+d}\right)^T \textit{ is the observations of the second regime}, \\
\mathbf{X}^{(1)} &= \left(X_1^{(1)}, \cdots X_s^{(1)}\right)^T, \mathbf{X}_t^{(1)} = \left(y^{(1)}_{\pi_t+d-1}, y^{(1)}_{\pi_t+d-2}, \cdots y^{(1)}_{\pi_t+d-p_1}, e^{(1)}_{\pi_t+d-1}, e^{(1)}_{\pi_t+d-2}, \cdots e^{(1)}_{\pi_t+d-q_1}\right)^T, \\
\mathbf{X}^{(2)} &= \left(X_{s+1}^{(2)}, \cdots X_{n-d}^{(2)}\right)^T, \mathbf{X}_t^{(2)} = \left(y^{(2)}_{\pi_t+d-1}, y^{(2)}_{\pi_t+d-2}, \cdots y^{(2)}_{\pi_t+d-p_2}, e^{(2)}_{\pi_t+d-1}, e^{(2)}_{\pi_t+d-2}, \cdots e^{(2)}_{\pi_t+d-q_2}\right)^T, \\
\mathbf{A}^{(i)} &= X^{(i)^T} X^{(i)}, \mathbf{B}^{(i)} = X^{(i)^T} y^{(i)} \quad , i = 1, 2.
\end{aligned} \tag{6}$$

Note that $\mathbf{X}^{(i)}, \mathbf{A}^{(i)}, \mathbf{B}^{(i)}, i = 1,2$ are functions in $\eta = (d, r)$ however, the arguments are suppressed for simplicity.

3.2 Prior Information

Prior specification is an important step in developing any Bayesian analysis. Both proper and improper prior distributions are used to represent prior information about the parameters. In both cases $\eta = (d, r)$ is treated as independent of the coefficients $\gamma^{(i)}$ (i=1,2) and precision parameters $\tau_i, i = 1,2$. Thus, a suitable choice for the proper prior distribution for the model parameters $\gamma^{(i)}$ and τ_i is a normal gamma distribution.

That is

$$\begin{aligned} \zeta(\gamma^{(i)} \mid \tau_i) &\sim N(\mu^{(i)}, \tau_i^{-1}\, \mathbf{V}^{(i)^{-1}}), \\ \zeta(\tau_i) &\sim \Gamma(a_i, b_i) \end{aligned}$$

The approximate conditional likelihood function (5) for each regime, as a function in the parameters $\gamma^{(i)}, \tau$ given d, r is a normal gamma density.

Hence the joint prior of $\gamma^{(i)}, \tau_i$ may be written as:

$$\zeta\left(\gamma^{(i)}, \tau_i\right) \propto \tau^{\left(\frac{p_i+q_i+2a_i}{2}\right)-1} \exp\left\{-\frac{\tau_i}{2}\left[\,2b_i + (\gamma^{(i)} - \mu^{(i)})^T \mathbf{V}^{(i)} (\gamma^{(i)} - \mu^{(i)})\right]\right\} \qquad (7)$$

where $V^{(i)}$ is a square positive definite matrix of order $p_i + q_i$, $a_i > 0$, $b_i > 0$. The marginal prior for η, namely $\zeta(\eta)$ could be any distribution. So, the joint prior of the model parameters, $\zeta\left(\gamma^{(1)}, \gamma^{(2)}, \tau_1, \tau_2, \eta\right)$ may be written as:

$$\begin{aligned} \zeta\left(\gamma^{(1)}, \gamma^{(2)}, \tau, \eta\right) &= \zeta(\eta) \prod_{i=1}^{2} \zeta\left(\gamma^{(i)} \mid \tau_i\right) \zeta(\tau_i) \\ &\propto \zeta(\eta) \prod_{i=1}^{2} \tau_i^{\left(\frac{p_i+q_i+2a_i}{2}\right)-1} \exp\left\{-\frac{\tau_i}{2}\left[\,2b_i + (\gamma^{(i)} - \mu^{(i)})^T V^{(i)} (\gamma^{(i)} - \mu^{(i)})\right]\right\} \end{aligned} \qquad (8)$$

This class of prior is flexible enough to be used in a lot of applications. It also, facilities the mathematical calculations. It is, at least conditionally, the conjugate prior for the approximate conditional likelihood since the approximate likelihood given by (5) conditioning on $\eta = (d, r)$, is a normal gamma function, which is the same form as $\zeta(\gamma, \tau)$ given by (7). However, the use of the proposed prior distribution (8) is not a necessity and any other distribution could be used provided we are prepared to use numerical integration.

If one has little information about the hyperparameters, or are unwilling to determine them, one may use the improper Jeffreys' prior, namely

$$\zeta\left(\gamma^{(1)}, \gamma^{(2)}, \tau_1, \tau_2\right) \propto \frac{1}{\tau_1 \tau_2} \qquad (9)$$

Jeffreys' prior distribution is a special case of the normal gamma class when $b_i = 0$, $V^{(i)} = 0$, $a_i = -\left(\frac{p_i+q_i}{2}\right)$.

3.3 Posterior Analysis

Multiplying the approximate conditional likelihood function, $L^*(\gamma, \tau, \eta \mid S_n)$, given by (5) by the prior distribution, $\zeta(\gamma^{(1)}, \gamma^{(2)}, \tau_1, \tau_2, \eta)$, given by (8) results in the approximate joint posterior distribution which may be written as:

$$\begin{aligned}\zeta^*(\gamma^{(1)}, \gamma^{(2)}, \tau_1, \tau_2, \eta \mid S_n) \;\propto\; & \zeta(\eta) \prod_{i=1}^{2} \tau_i^{(\frac{n_i+p_i+q_i+2a_i}{2})-1} \exp\left\{-\frac{\tau_i}{2}\left[\, 2b_i + Y^{(i)^T} Y^{(i)} - \right.\right. \\ & \left.\left. 2\gamma^{(i)^T} B^{(i)} (\gamma^{(i)} - \mu^{(i)})^T V^{(i)} (\gamma^{(i)} - \mu^{(i)})\right] \right\}\end{aligned}$$

where n_i is the number of observations in the i^th regime.

In principle, the marginal posterior distribution of $\gamma^{(i)}$ for example can be found from

$$\zeta^*\left(\gamma^{(1)} \mid S_n\right) = \int_{\gamma^{(2)}} \int_{\tau_1} \int_{\tau_2} \int_{\eta} \zeta^*\left(\gamma^{(1)}, \gamma^{(2)}, \tau_1, \tau_2, \eta \mid S_n\right)\right) d\gamma^{(2)} d\tau_1 d\tau_2 d\eta$$

It is not difficult to show that the marginal posteriors of $\gamma^{(1)}, \gamma^{(2)}, \tau_1, \tau_2$ and η are all non-standard. So, all exact Bayesian inferences must be done numerically. However, the following two theorems show that the approximate posterior distributions of $\gamma^{(i)}$ and τ_i given η are standard.

Theorem 3.1 *Given n observations $S_n = (y_1, y_2, \cdots y_n)$ from the SETARMA$(2; p_1, q_1, p_2, q_2)$ model given by (2), a conditional likelihood function approximated by (3) and a prior distribution given by (8), then the approximate conditional marginal posterior distribution of $\gamma^{(i)}$ given η is a $T_{p_i+q_i}$ distribution with $n+2a$ degrees of freedom, location vector $E\left(\gamma^{(i)} \mid S_n, \eta\right)$ and precision matrix $T\left(\gamma^{(i)} \mid S_n, \eta\right)$*
where,

$$\begin{aligned}E\left(\gamma^{(i)} \mid S_n, \eta\right) &= \left(A^{(i)} + V^{(i)}\right)^{-1} \left(B + V^{(i)} \mu^{(i)}\right) \\ T\left(\gamma^{(i)} \mid S_n, \eta\right) &= \left(\frac{n_i + 2a_i}{C_i}\right)\left(A^{(i)} + V^{(i)}\right) \\ C_i &= 2b_i + Y^{(i)^T} Y^{(i)} + \mu^{(i)^T} V^{(i)} \mu^{(i)} - \left(B^{(i)} + V^{(i)} \mu^{(i)}\right)^T \times \\ & \quad (A+V)^{-1}\left(B^{(i)} + V^{(i)} \mu^{(i)}\right)\end{aligned}$$

and $X^{(i)}, A^{(i)}, B^{(i)}$ are defined in (6).

Theorem 3.2 *Under the same conditions of theorem 3.1, the approximate conditional posterior of τ_i given η is a gamma distribution with parameters $\frac{n_i+2a_i}{2}$ and $\frac{C_i}{2}$.*

The approximate conditional posterior distribution of τ_i, $\zeta^*(\tau_i \mid \eta, S_n)$, can be used to make inferences about τ_i. Unlike $\gamma^{(i)}$ which has symmetric T distribution, τ_i has a skewed distribution. Quoting the mean and the variance could be misleading. Instead, an approximate credible interval is constructed as follows. It is not difficult to show that $C_i\tau_i \sim \chi^2_{n_i+2a_i}$. Therefore, a $(1-\alpha)$ equal tailed credible interval for τ_i is given by :

$$\frac{\chi^2_{n_i+2a_i,\ \frac{\alpha}{2}}}{C_i} \leq \tau_i \leq \frac{\chi^2_{n_i+2a_i,\ 1-\frac{\alpha}{2}}}{C_i}$$

where $\chi^2_{n_i,\ \alpha}$ is defined so that $p\left(\chi^2_{n_i} > \chi^2_{n_i,\ \alpha}\right) = \alpha$.

The next two corollaries report the approximate posterior distributions of $\gamma^{(i)}$ and τ_i given η when little information about the parameters are known *apriori*.

Corollary 3.1 *If the parameters $\gamma^{(i)}$, τ_i and η are assumed independent apriori and have Jeffreys' prior distribution (9) and the other conditions of theorem 3.1 are not changed, the approximate posterior distribution of $\gamma^{(i)}$ given η is a T distribution with $(n_i - p_i - q_i)$ degrees of freedom, location vector $(A^{(i)})^{-1}B^{(i)}$ and precision matrix $\left(\frac{n_i-p_i-q_i}{y^{(i)T}y^{(i)} - B^{(i)T}A^{(i)-1}B^{(i)}}\right) A^{(i)}$.*

Corollary 3.2 *Under the same set-up as corollary 3.1, the approximate posterior distribution of τ given η is gamma with parameters $(\frac{n_i-p_i-q_i}{2})$ and $\frac{B^{(i)T}A^{(i)-1}B^{(i)}}{2}$*

3.4 Approximate posterior distribution of η

Integrating out $\gamma^{(i)}$ and τ_i from the approximate joint posterior function results in the marginal posterior function of η. It's not difficult to show that the marginal posterior function of η takes the form:

$$\zeta(\eta \mid S_n) \ \propto \ \zeta(\eta) \prod_{i=1}^{2} \frac{\Gamma\left(\frac{n_i+a_i}{2}\right)}{\sqrt{\mid A(i) + V(i) \mid}} (C_i)^{-\left(\frac{n_i+2a_i}{2}\right)}$$

where, C_i and $A^{(i)}$ are defined as before.

If Jeffreys' prior for γ_i and τ_i is employed, the marginal posterior of η has the form:

$$\zeta(\eta \mid S_n) \ \propto \ \zeta(\eta) \prod_{i=1}^{2} \frac{\Gamma\left(\frac{n_i-p_i-q_i}{2}\right)}{\sqrt{\mid \mathbf{A}^{(i)} \mid}} \left(\mathbf{y}^{(i)T}\mathbf{y}^{(i)} - \mathbf{B}^{(i)T}\mathbf{A}^{(i)-1}\mathbf{B}\right)^{-\left(\frac{n_i-p_i-q_i}{2}\right)}$$

It's clear that the marginal posterior of η is non-standard. Thus all inferences about η must be done numerically.

3.5 Posterior Analysis of SETARMA(2;1,1,1,1) Model

The objective of this section is twofold. Firstly, the proposed Bayesian technique is demonstrated using SETARMA(2;1,1,1,1) model. Secondly, to give the formulae used in calculating results which will be introduced in next section.

Consider the SETARMA(2;1,1,1,1) given by :

$$y_{\pi_i+d} = \begin{cases} \phi_1^{(1)} y^{(1)}_{\pi_i+d-1} + \theta_1^{(1)} \varepsilon^{(1)}_{\pi_i+d-1} + \varepsilon^{(1)}_{\pi_i+d} & i \le s \\ \phi_1^{(2)} y^{(2)}_{\pi_i+d-1} + \theta_1^{(2)} \varepsilon^{(2)}_{\pi_i+d-j} + \varepsilon^{(2)}_{\pi_i+d} & otherwise \end{cases}$$

Assuming $\phi_1^{(1)}, \theta_1^{(1)}, \phi_1^{(2)}, \theta_1^{(2)}, \tau_1, \tau_2$ and η are independent a priori and have vague prior,

$$\zeta\left(\phi_1^{(1)}, \theta_1^{(1)}, \phi_1^{(2)}, \theta_1^{(2)}, \tau_1, \tau_2, \eta\right) \propto \frac{1}{\tau_1\ \tau_2}$$

From results of corollaries (3.1) and (3.2) we get

$$\begin{aligned} \zeta^*(\gamma^{(i)} \mid S_n, \eta) &\sim T_2\left[n_i - 2,\ (A^{(i)})^{-1}B^{(i)},\ \frac{(n_i-2)\ A^{(i)}}{C_i}\right] \\ \zeta^*(\tau_i \mid S_n, \eta) &\sim \Gamma\left[\frac{n_i-2}{2},\ \frac{C_i}{2}\right] \end{aligned}$$

where,

$$\begin{aligned} \gamma^{(i)} &= \left(\phi_1^{(i)}, \theta_1^{(i)}\right) \quad i = 1,2 \\ \mathbf{X}^{(1)} &= \left(X_1^{(1)}, \cdots X_{n_1}^{(1)}\right)^T, \mathbf{X}_t^{(1)} = \left(y^{(1)}_{\pi_t+d-1}, e^{(1)}_{\pi_t+d-1}\right)^T, \\ \mathbf{X}^{(2)} &= \left(X^{(2)}_{n_1+1}, \cdots X^{(2)}_{n_1+n_2}\right)^T, \mathbf{X}_t^{(2)} = \left(y^{(2)}_{\pi_t+d-1}, e^{(2)}_{\pi_t+d-1}\right)^T, \\ y^{(1)} &= \left(y_{\pi_1+d}, y_{\pi_2+d}, \cdots, y_{\pi_{n_1}+d}\right)^T, \\ y^{(2)} &= \left(y_{\pi_{n_1+1}+d}, y_{\pi_{n_1+2}+d}, \cdots, y_{\pi_{n_1+n_2}+d}\right)^T, \\ \mathbf{A}^{(i)} &= X^{(i)^T} X^{(i)}, \mathbf{B}^{(i)} = X^{(i)^T} y^{(i)} \quad , i = 1,2. \\ C_i &= y^{(i)^T} y^{(i)} - B^{(i)^T} A^{(i)^{-1}} B^{(i)} \end{aligned}$$

Therefore,

$$E(\gamma^{(i)} \mid S_n, \eta) = A^{(i)^{-1}} B^{(i)}, \quad i = 1,2$$

$$
\begin{aligned}
Var(\gamma^{(i)} \mid S_n, \eta) &= \left(\frac{C_i}{n_i - 4}\right) A^{(i)^{-1}}, \quad i = 1, 2 \\
E(\tau_i \mid S_n, \eta) &= \frac{n_i - 2}{C_i}, \quad i = 1, 2 \\
Var(\tau_i \mid S_n, \eta) &= \frac{2(n_i - 2)}{C_i^2} \quad i = 1, 2
\end{aligned} \tag{10}
$$

The marginal posterior distribution of η is given by

$$
\zeta(\eta \mid S_n) \propto \prod_{i=1}^{2} \frac{\Gamma\left(\frac{n_i - 2}{2}\right)}{\sqrt{\mid A^{(i)} \mid}} \left(y^{(i)^T} y^{(i)} - B^{(i)^T} A^{(i)^{-1}} B\right)^{-\left(\frac{n_i - 2}{2}\right)} \tag{11}
$$

4 Illustrative Examples

In this section, the proposed methodology is illustrated with a simulation study and a real data set. In both cases, a non informative prior is used.

4.1 Simulation

This study deals with generating 100 sets of n=200 observations from the following SETARMA(2;1,0,0,1) model

$$
y_t = \begin{cases} 0.7\, y_{t-1}^{(1)} + \varepsilon_t^{(1)} & y_{t-1} \leq 0 \\ \\ 0.5\, \varepsilon_{t-1}^{(2)} + \varepsilon_t^{(2)} & y_{t-1} > 0 \end{cases} \tag{12}
$$

where the delay parameter d=1, the threshold parameter r=0, $\varepsilon_t^{(1)} \sim N\left(0, \tau_1^{-1} = 2\right)$, $\varepsilon_t^{(2)} \sim N\left(0, \tau_2^{-1} = 1\right)$ and $\left\{\varepsilon_t^{(1)}\right\}$ and $\left\{\varepsilon_t^{(2)}\right\}$ are independent.

The initial values of the errors are set to zeros and using the equation (12), 400 observations are generated. The first 200 observations are deleted to remove the initialization effect. A lower limit of 1 and upper limit of 4 are used for d. Thus, the candidate domain for d is $1, 2, \cdots, 4$. The candidate domain for the threshold is chosen to be $P_{10}, P_{20}, \cdots P_{90}$ where P_i is the i^{th} percentile of the data.

For each generated series, the least squares estimates are calculated and the residuals are computed. Then, the posterior means of $\phi_1^{(1)}$, $\theta_1^{(2)}$, τ_1 and τ_2 are calculated using formulae (10). The mode of the marginal posterior distribution is chosen as a Bayesian estimator for both of d and r. The simulation results are shown in table(1). Columns 3 through 9 contains the mean, standard deviation, minimum (Min), first quartile (Q_1), median, third quartile (Q_3) and maximum (Max) for 100 posterior modes of d and r and for 100 posterior means of other parameters. The 6^{th} and last rows of Table (1) present the previous descriptive statistics for nonlinear least squares estimates of d and r respectively.

Table (1): Simulation Results for 100 simulated data sets from model(12).

Parameters	True Value	Mean	Standard Deviation	Min	Q_1	Median	Q_3	Max
$\phi_1^{(1)}$	0.7	0.675	0.076	0.434	0.619	0.673	0.734	0.806
$\theta_1^{(2)}$	0.5	0.494	0.292	-0.291	0.338	0.472	0.582	1.63
τ_1	2	1.914	0.392	1.088	1.619	1.908	2.175	3.037
τ_2	1	1.039	0.284	0.233	0.893	1.024	1.159	2.033
d	1	1.71	1.094	1	1	1	2	4
		2.06	1.023	1	1	2	2.75	4
r	0	0.311	0.651	-1.45	-0.118	0.108	0.681	1.75
		0.15	0.866	-2.355	-0.407	0.147	0.794	1.639

Inspection of the results in table (1) shows that the proposed Bayesian methodology gives sound inferences for the parameters.
The comparison of the last two rows of table (1) indicates that Bayesian estimates for r is closer to the true value than the classical nonlinear least squares ones.

Moreover, Relative Frequency distributions for posterior mode and nonlinear least squares estimate of d for the 100 generated data sets are given in table(2). Table (2) shows that the Bayesian technique estimates d correctly in 64 cases while the least squares method estimates d correctly in 34 cases only out of the 100 cases. This may indicate that the Bayesian approach is superior to the classical approach in estimating d in this study.

Table (2): Relative Frequency distribution for posterior mode and nonlinear least squares estimate of d.

d	Posterior mode	Nonlinear least squares estimate
1	64	34
2	15	41
3	7	10
4	14	15

Figures (1) and (2) show the frequency histogram of posterior modes of r and nonlinear least squares estimates for the 100 generated data sets. The comparison of the two figures confirm the superiority of the proposed Bayesian technique in estimating r.

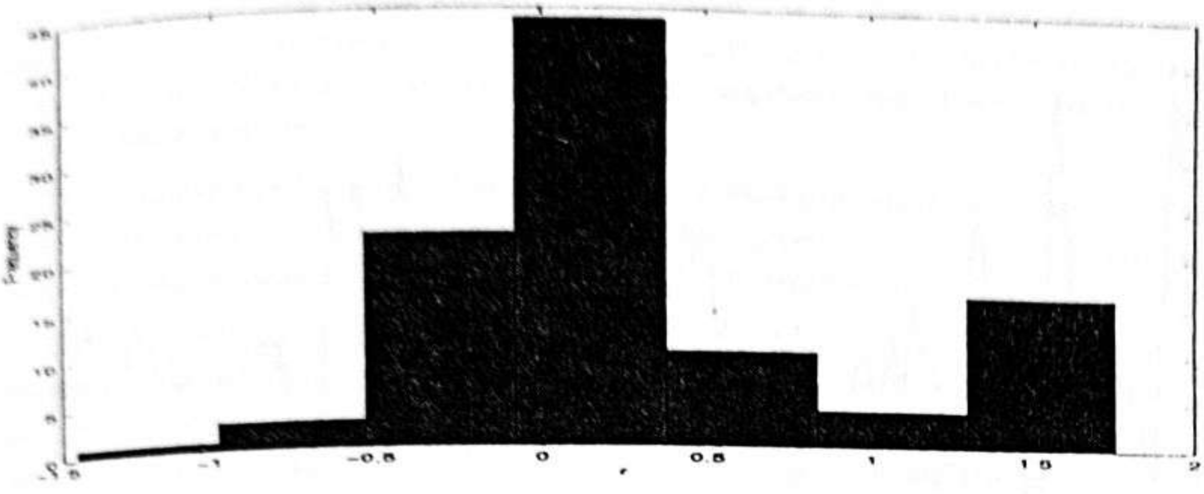

Figure 1: Histogram of posterior modes for r

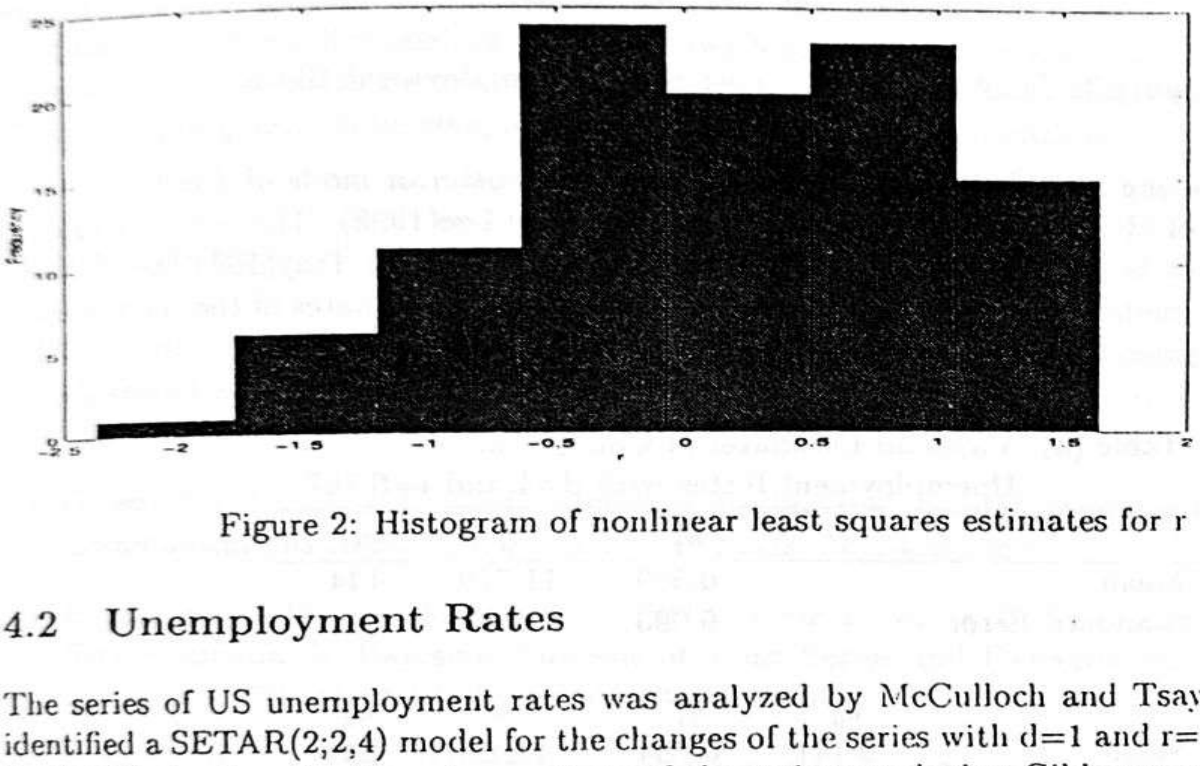

Figure 2: Histogram of nonlinear least squares estimates for r

4.2 Unemployment Rates

The series of US unemployment rates was analyzed by McCulloch and Tsay(1993). They identified a SETAR(2;2,4) model for the changes of the series with d=1 and r=0.3. Chen and Lee(1998) developed a Bayesian analysis of the series employing Gibbs sampling approach to calculate the posterior densities of the parameters. A SETAR(2;2,4) model with d=1 and r=0.299 was used. Figure (3) displays the time Plot for Changes of Us Unemployment Rates.

Assuming a non informative prior, the proposed Bayesian methodology is applied to the changes of the US unemployment rates series from the second quarter of 1948 to the first

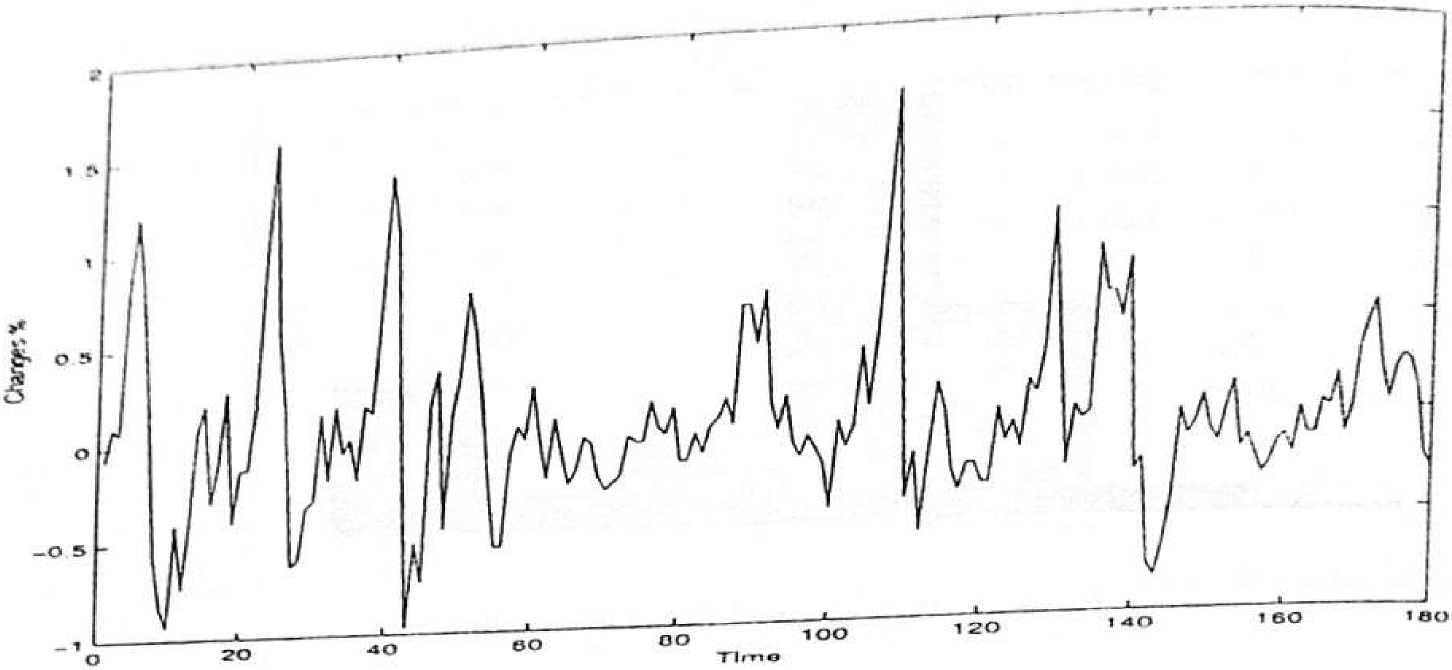

Figure 3: Time Plot for Changes of Us Unemployment Rates

quarter of 1993 using a SETARMA(2;0,1,0,1) model. The posterior mode of d is 1 which is identical to that of McCulloch and Tsay(1993) and Chen and Lee(1998). The posterior mode of r is 0.267 which is close to its estimates found by McCulloch and Tsay(1993) and Chen and Lee(1998). Conditioning on d=1 and r=0.267, the Bayesian estimates of the coefficients and precision parameter in each regime are reported in table (3).

Table (3): Bayesian Estimates of Changes in US Unemployment Rates with d=1 and r=0.267

		$\theta_1^{(1)}$	τ_1	n_1
Mean		0.569	11.779	144
Standard Error		0.093	1.394	
	$\theta_0^{(2)}$	$\theta_1^{(2)}$	τ_2	n_2
Mean	0.141	0.754	4.649	34
Standard Error	0.061	0.104	2.374	

It should be noted that our methodology like Chen and Lee (1998), does not need a subjective determination of the threshold parameter r via scatter plots. In addition, our estimated model is more parsimoneous than SETAR models which are used in McCulloch and Tsay(1993) and Chen and Lee(1998).

5 Conclusion and open problems

We have shown that a complete Bayesian inference for SETARMA model is possible and is no more difficult that doing switching regression. It is shown that, conditioning on the delay and threshold parameters, the conditional posteriors of the model coefficients and precision are approximated by a multivariate t and gamma densities. Therefore, the need for numerical integration is avoided.

The proposed technique is demonstrated by working out details for a SETARMA(2;1,1,1,1) model. Moreover, a Monte Carlo study is conducted and the results indicate the success of the Bayesian approach in analyzing SETARMA time series data.

We assumed that the model is identified, i.e. the model orders are known. However, the identification problem may be investigated by extending the parameter vector to include the orders and then using the marginal posterior for the model orders to identify the model. In addition, the forecasting problem may be studied by deriving the predictive density of future observations.

Other analytic approximations such as Lindley (1980) and Tierney and Kadane (1986) can be tried to implement a Bayesian analysis for SETARMA models. Moreover, the relationship between Broemeling and Shaarawy's approximation and the above mentioned approximations needs to be investigated. In addition, Monte Carlo based methods such as Gibbs Sampling need to be compared with the analytic approximations.

References

[1] Brockell, P. J, Liu, J. and Tweedie, R. L. (1992). "On the Existence of stationary threshold autoregressive moving average Processes." *Journal of Time Series Analysis*, 13, 95-107.

[2] Broemeling, L.D., and Cook, P. (1992). "Bayesian analysis of threshold autoregressions." *Communications in Statistics: Theory Methods*, 21, 2459-2482.

[3] Broemeling, L.D. and Shaarawy, S. (1988). *Time series: A Bayesian Analysis in Time Series Domain*. In Bayesian Analysis of Time Series and Dynamic Models, (ed J. C. Spall), pp. 1-21. Marcel Dekker Inc., New York.

[4] Chen, C.W.S. (1992). "Bayesian inferences and forecasting in bilinear time series models." *Communications in Statistics: Theory Methods.*, 21, 1725-1743.

Chen, C.W.S. (1998). "Bayesian analysis of generalized threshold autoregressive models." *Statistics and Probability Letters*, 40, 15-22.

[5] Chen, C.W.S., and Lee, J.C. (1995). "Bayesian inference of threshold autoregressive models." *Journal of Time Series Analysis*, 16, 483-492.

[6] De Gooijer, J.G. (1998). "On Threshold Moving-Average Models." *Journal of Time Series Analysis*, 19 (1), 1-18.

[7] De Gooijer, J.G., and Kumar, K. (1992). "Some recent developments in non-linear time series modelling, testing, and forecasting." *International Journal of Forecasting*, 8, 135-156.

[8] Geweki, J., and Terui N. (1993). "Bayesian threshold autoregressive models for nonlinear time series." *Journal of Time Series Analysis*, 14, 441-454.

[9] Ismail, M. A. (1994). Bayesian Forcasting for Nonlinear Time Series. Unpublished Ph.D. Thesis, Department of Mathematics, University of Wales, Aberystwyth, U.K.

[10] Kheradmandnia, M. (1991). "Aspects of Bayesian threshold autoregressive Modelling." Ph.D. Department of Statistics, University of Kent.

[11] Lindley, D.V. (1980). "Approximate Bayesian methods." Proc. of the First International Meeting held in Valencia (Spain), 223-237.

[12] MucCulloch, R.E., and Tsay, R. S. (1993). "Bayesian analysis of threshold autoregressive processes with a random number regimes." Proc. 25th Symp. on the Interface, 253-262.

[13] Stramer, O.(1996). "On the approximations of moments for continuous time threshold ARMA Processes." *Journal of Time Series Analysis*, 17, 189-202.

[14] Tierney, L., and Kadane, J.B. (1986). "Accurate approximations for posterior moments and marginal densities." *Journal of the American Statistical Association*, 81, 82-86.

[15] Tong, H. (1983). *Threshold models in non-linear time series analysis*. Vol. 21 of Lecture Notes in Statistics (ed. K. Krickegerg). New York: Springer-Verlag.

[16] Tong, H. (1990). *Non-Linear Time Series Analysis: A Dynamical System Approach*. Oxford: Oxford, University Press.

[17] Tong, H., and Lim, K.S. (1980). "Threshold autoregression, limit cycles and cyclical data (with discussion." *Journal of the Royal Statistical Society: Series B*, 42, 245-292.

[18] Tsay, R.S. (1989). "Testing and modelling threshold autoregressive processes." *Journal of the American Statistical Association*, 84, 231-240.