Sankhyā: The Indian Journal of Statistics
2000, Volume 62, Series B, Pt. 3, pp. 353–371

BAYESIAN ANALYSIS OF THRESHOLD AUTOREGRESSIVE MOVING AVERAGE MODELS

By THELMA SÁFADI
Federal University of Lavras, Brazil
and
PEDRO A. MORETTIN
University of São Paulo, Brazil

SUMMARY. In this paper we consider a Bayesian analysis for threshold autoregressive moving average models. Two different methods are used for the special case of two regimes. In the first we consider a hierarchical prior and obtain posterior distributions in closed form. In the second, a rearranged procedure due Tsay (1989) is used, in conjunction with the Gibbs sampler and Metropolis-Hastings algorithm. Applications are given for a simulated series and for the sunspot data.

1. **Introduction**

In recent years non-linear models have been studied thoroughly and their analysis is facilitated due to increasing developments in computational methodologies. The classical Bayesian linear models are unable to reproduce some of the features frequently found in observed time series, e.g. non-linear processes exhibit such interesting properties as amplitude frequency dependence, limit cycle behavior and jump phenomena.

In order to obtain more appropriate models for such data, various discrete-time non-linear models have been proposed. Among the more successful ones we mention the threshold models, bilinear models and random coefficient autoregressive models. There are others models, but they have received less attention. See Tong (1990) for a more detailed description.

The TAR (threshold autoregressive) class was first studied by Tong and Lim (1980). The Bayesian analysis of TAR process has many characteristics that are similar to shift point problems in statistics. Broemeling (1985) and Broemeling and Tsurumi (1987) give accounts of the Bayesian analysis of shift point problems. Pole and Smith (1985) noted this similarity and developed a non-asymptotic Bayesian analysis of Gaussian threshold non-linear models with regimes determined by

Paper received March 1999; revised July 2000.
AMS (1991) *subject classification.* 62F15, 62M10.
Key words and phrases. ARMA models, Bayesian analysis, Gibbs sampler, Metropolis-Hastings algorithm, predictive distribution, rearranged autoregression, TARMA model, threshold.

external or exogenous variables. This analysis was done by numerical methods. In this paper we are interested in a Bayesian analysis for the TARMA (threshold autoregressive moving average) models and this analysis is done through two different procedures. In the first we extend the works of Broemeling and Shaarawy (1988) and Broemeling and Cook (1992) on the Bayesian analysis of ARMA(autoregressive moving average) and TAR models, respectively.

As in the usual ARMA models, the consideration of mixed models relies on the issue of parsimony, meaning that we may have less parameters in an ARMA model than in a pure AR or MA model. In both models, TAR and TARMA, care should be taken in order to have enough observations in each regime. A hierarchical prior distribution was considered and inference for the parameters was done. In the second procedure we extend the work of Chen and Lee (1995) on the TAR model. We consider the rearranged TARMA model and independent priors. Inference on the parameters was obtained through the Gibbs sampler and the Metropolis-Hastings algorithm.

The paper is organized as follows. The TARMA model is introduced in Section 2 and a posterior analysis with a proper prior is given in Section 3. In Section 4 an improper (Jeffreys) prior is considered. The rearranged model and corresponding analysis are considered next in Sections 5 and 6. The predictive distribution is entertained in Section 7. In Section 8 we present a simulation study and an application to a real series and finally, in Section 9, some further comments are made.

2. The TARMA Model

Suppose that $\{Y_t, t = 0, \pm 1, \ldots\}$ is a discrete parameter stochastic process satisfying

$$Y_t = \phi_{l,0} + \sum_{i=1}^{p_l} \phi_{l,i} Y_{t-i} + a_t^{(l)} + \sum_{j=1}^{q_l} \theta_{l,j} a_{t-j}^{(l)}, \quad \text{if} \quad Y_{t-d} \in R_l \tag{2.1}$$

for $l = 1, \ldots, k$ and:

(i) $\{a_t^{(l)}\}$ is a sequence of iid $\sim N(0, \tau_l^{-1})$, where the precision $\tau_l > 0$, $\tau_l^{-1} = \sigma_l^2 = Var\{a_t^{(l)}\}$ is unknown, for each l;

(ii) R_l is an interval of the real line R such that $R_i \cap R_j = \phi$, if $i \neq j$ and $\bigcup_{l=1}^{k} R_l = R$; that is, $\{R_1, \ldots, R_k\}$ forms a partition of the real line with $k-1$ thresholds;

(iii) d is the delay parameter and we assume $d \in \{1, \ldots, d_0\}$.

Then (2.1) tells us that Y_t follows one of k different ARMA processes, with different orders $(p_l, q_l), l = 1, \ldots, k$, depending on the value of Y_{t-d}. We will denote such a process by TARMA$(k; p_1, \ldots, p_k; q_1, \ldots, q_k)$, a threshold ARMA process with k regimes.

In what follows we restrict attention to the TARMA$(2; p_1, p_2; q_1, q_2)$ model, that is,

$$Y_t = \begin{cases} \phi_{10} + \sum_{i=1}^{p_1} \phi_{1i} Y_{t-i} + a_t^{(1)} + \sum_{j=1}^{q_1} \theta_{1j} a_{t-j}^{(1)}, & \text{if } Y_{t-d} \le r; \\ \phi_{20} + \sum_{i=1}^{p_2} \phi_{2i} Y_{t-i} + a_t^{(2)} + \sum_{j=1}^{q_2} \theta_{2j} a_{t-j}^{(2)}, & \text{if } Y_{t-d} > r. \end{cases} \tag{2.2}$$

and we call r the threshold parameter. The parameters in (2.2) are $\gamma_1 = (\phi_{10}, \dots, \phi_{1p_1}, \theta_{11}, \dots, \theta_{1q_1})'$, $\gamma_2 = (\phi_{20}, \dots, \phi_{2p_2}, \theta_{21}, \dots, \theta_{2q_2})'$, τ_1, τ_2, r and d.

We also assume that the orders p_1, p_2, q_1 and q_2 are known. Otherwise they may be selected by considering the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of Y_t or some information criterion such as AIC or BIC. As the ACF and the PACF often provide guidance for reasonable values of p and q, we can start with them and refine the order later, if necessary.

3. The Posterior Analysis with a Proper Prior

Given a sample $Y = \{Y_1, \dots, Y_n\}$, let $p = \max(p_1, p_2), q = \max(q_1, q_2)$. Conditioning on the first p observations and assuming $a_1^{(i)} = a_2^{(i)} = \dots = a_p^{(i)} = a_0^{(i)} = \dots = a_{p-q-1}^{(i)} = 0$, with $q > p+1$ and $i = 1, 2$, the conditional likelihood may be approximated by

$$\mathcal{L}(\gamma_1, \gamma_2, \tau_1, \tau_2, r, d | Y_{p+1}, \dots, Y_n) \propto \tau_1^{n_1/2} \tau_2^{n_2/2} \exp\left\{ -\frac{\tau_1}{2} \sum_1 (a_t^{(1)})^2 - \frac{\tau_2}{2} \sum_2 (a_t^{(2)})^2 \right\}, \tag{3.1}$$

where $\sum_1$ and $\sum_2$ denote summations over $\{t = p+1, ..., n, Y_{t-d} \le r\}$ and $\{t = p+1, ..., n, Y_{t-d} > r\}$, respectively, and n_1 and n_2 are the numbers of observations in the first and second regimes, respectively.

The conditional likelihood is not linear in γ_i, since the error $a_t^{(i)}$ in (2.2) is not a linear function of the parameters. In order to make this function approximately linear we use the residuals

$$\hat{a_t}^{(i)} = Y_t - \hat{\phi}_{i0} - \sum_{j=1}^{p_i} \hat{\phi}_{ij} Y_{t-j} - \sum_{j=1}^{q_i} \hat{\theta}_{ij} \hat{a}_{t-j}^{(i)}, \tag{3.2}$$

where $t = p+1, \dots, n$, $\hat{a}_{p-1}^{(i)} = \dots = \hat{a}_{p-q-1}^{(i)} = 0$, and $\hat{\phi}_i$ and $\hat{\theta}_i$ are the(nonlinear) least squares estimators of the ϕ_i and θ_i, respectively, obtained by minimizing the sum of squares

$$S(\phi_i, \theta_i) = \sum_{t=p+1}^{n} (a_t^{(i)})^2 \tag{3.3}$$

with respect to the ϕ_i and θ_i, using a nonlinear regression algorithm. See Harvey (1981) for details. The approximate likelihood function (3.1) becomes

$$\mathcal{L}^*(\gamma_1, \gamma_2, \tau_1, \tau_2, r, d | Y_{p+1}, \dots, Y_n)$$

$$\propto \tau_1^{n_1/2}\tau_2^{n_2/2}\exp\left\{-\frac{\tau_1}{2}\left[\sum_1\left(Y_t-\left(\phi_{10}+\sum_{i=1}^{p_1}\phi_{1i}Y_{t-i}+\sum_{j=1}^{q_1}\theta_{1j}\hat{a}_{t-j}^{(1)}\right)\right)^2\right.\right.$$
$$\left.\left.-\frac{\tau_2}{2}\sum_2\left(Y_t-\left(\phi_{20}+\sum_{i=1}^{p_2}\phi_{2i}Y_{t-i}+\sum_{j=1}^{q_2}\theta_{2j}\hat{a}_{t-j}^{(2)}\right)\right)^2\right]\right\}. \tag{3.4}$$

It is difficult to find a family of conjugate densities for the parameters, however the form of the likelihood function suggest a quasi-conjugate type prior density, namely, let $P(\gamma_1,\gamma_2,\tau_1,\tau_2,r,d)$ be the proper prior given by

$$P(\gamma_1,\gamma_2,\tau_1,\tau_2,r,d)=P(\gamma_1|\tau_1)P(\gamma_2|\tau_2)P(\tau_1)P(\tau_2)P(r)P(d), \tag{3.5}$$

where

(1) γ_1 and γ_2 are independent $\sim N(\gamma_{0i},\tau_i^{-1}\mathbf{Q}_i^{-1})$, $i=1,2$,

(2) τ_1 and τ_2 are independent $\sim \text{Gamma}(\alpha_i/2,\beta_i/2)$, $i=1,2$,

(3) $d\sim$ Uniform $\{1,2,...,d_0\}$,

(4) $r\sim$ Uniform(a,b), where a and b are suitably chosen.

The hyperparameters $\gamma_{0i},\mathbf{Q}_i,\alpha_i$ and β_i, $i=1,2$, are supposed to be known, with $\alpha_i>0$, $\beta_i>0$, $\gamma_{0i}\in R^{p_i+q_i+1}$ and $\mathbf{Q_i}$ positive definite of order $(p_i+q_i+1)\times(p_i+q_i+1)$.

It seems imperative that the conditional prior density of $\gamma_i, i=1,2$ depend on r, namely through the hiperparameters γ_{0i} and $\mathbf{Q}_i$, but it is difficult to devise such a density which conveniently combines with the likelihood function.

The prior may be written

$$P(\gamma_1,\gamma_2,\tau_1,\tau_2,r,d)\propto\tau_1^{\alpha_1/2-1}\tau_2^{\alpha_2/2-1}\exp\left\{-\frac{1}{2}\sum_{i=1}^{2}[\tau_i(\beta_i+(\gamma_i-\gamma_{0i})'\mathbf{Q}_i(\gamma_i-\gamma_{0i}))]\right\}. \tag{3.6}$$

Let Y_i be the vector $(n_i\times 1)$ of observations in the i-th regime, $i=1,2$.

Use of the conditional likelihood function will allow derivation of the posterior distributions, condition on the threshold r and on d, without numerical integration. Let $Y_{(i)}$ denote the i-th order statistics. Also, let $D=\{Y_{(1)},\ldots,Y_{(n)}\}$. Note that $r\in D_r=\{Y_{(2)},\ldots,Y_{(n-1)}\}$. If $r=Y_{(1)}$ or $r=Y_{(n)}$ we would have a single regime. We suppose then that the distribution of r is uniform over $\{Y_{(2)},...,Y_{(n-1)}\}$.

We obtain the following theorem.

THEOREM 1. *If the likelihood (3.4) is combined with the prior (3.6) we obtain a normal-gamma posterior distribution*

$$P(\gamma_1,\gamma_2,\tau_1,\tau_2,r,d|D)\propto\tau_1^{\frac{n_1+\alpha_1}{2}-1}\tau_2^{\frac{n_2+\alpha_2}{2}-1}\exp\Big\{-\frac{1}{2}\sum_{i=1}^{2}\tau_i[F_i(r)$$

$$+(\gamma_i-(\mathbf{A}_i(r)+\mathbf{Q}_i)^{-1}(\mathbf{B}_i(r)+\mathbf{Q}_i\gamma_{0i}))'(\mathbf{A}_i(r)+\mathbf{Q}_i)(\gamma_i-(\mathbf{A}_i(r)+\mathbf{Q}_i)^{-1}(\mathbf{B}_i(r)+\mathbf{Q}_i\gamma_{0i}))]\Big\}, \tag{3.7}$$

where

$$F_i(r)=\beta_i+Y_i'Y_i+\gamma_{0i}'\mathbf{Q}_i\gamma_{0i}-(\mathbf{B}_i(r)+\mathbf{Q}_i\gamma_{0i})'(\mathbf{A}_i(r)+\mathbf{Q}_i)^{-1}(\mathbf{B}_i(r)+\mathbf{Q}_i\gamma_{0i}),$$

$$\mathbf{B}_i(r)=(\mathbf{B}_{1i},\mathbf{B}_{2i})', \quad \textit{a vector} \quad (p_i+q_i+1)\times 1, \quad i=1,2,$$

$$\mathbf{B}_{1i}=\begin{pmatrix} \sum_i Y_t \\ \sum_i Y_tY_{t-1} \\ \vdots \\ \sum_i Y_tY_{t-p_i} \end{pmatrix}, \quad \mathbf{B}_{2i}=\begin{pmatrix} \sum_i Y_t\hat{a}^{(i)}_{t-1} \\ \vdots \\ \sum_i Y_t\hat{a}^{(i)}_{t-q_i} \end{pmatrix} \quad i=1,2,$$

$$\mathbf{A}_i(r)=\begin{pmatrix} \mathbf{A}^{(i)}_{11} & \mathbf{A}^{(i)}_{12} \\ \mathbf{A}^{(i)}_{21} & \mathbf{A}^{(i)}_{22} \end{pmatrix}, \quad \textit{symmetric}, \quad (p_i+q_i+1)\times(p_i+q_i+1), \quad i=1,2,$$

$$\mathbf{A}^{(i)}_{11}=\begin{pmatrix} \sum_i 1 & \sum_i Y_{t-1} & \dots & \sum_i Y_{t-p_i} \\ \sum_i Y_{t-1} & \sum_i Y^2_{t-1} & \dots & \sum_i Y_{t-1}Y_{t-p_i} \\ \vdots & \vdots & \dots & \vdots \\ \sum_i Y_{t-p_i} & \sum_i Y_{t-1}Y_{t-p_i} & \dots & \sum_i Y^2_{t-p_i} \end{pmatrix},$$

$$\mathbf{A}^{(i)}_{12}=\begin{pmatrix} \sum_i \hat{a}^{(i)}_{t-1} & \sum_i \hat{a}^{(i)}_{t-2} & \dots & \sum_i \hat{a}^{(i)}_{t-q_i} \\ \sum_i \hat{a}^{(i)}_{t-1}Y_{t-1} & \sum_i \hat{a}^{(i)}_{t-2}Y_{t-1} & \dots & \sum_i \hat{a}^{(i)}_{t-q_i}Y_{t-1} \\ \vdots & \vdots & \dots & \vdots \\ \sum_i \hat{a}^{(i)}_{t-1}Y_{t-p_i} & \sum_i \hat{a}^{(i)}_{t-2}Y_{t-p_i} & \dots & \sum_i \hat{a}^{(i)}_{t-q_i}Y_{t-p_i} \end{pmatrix},$$

$$\mathbf{A}^{(i)}_{22} \quad : \quad (q_i\times q_i) \textit{ matrix whose (j,k)-th element is } \textstyle\sum_i \hat{a}^{(i)}_{t-j}\hat{a}^{(i)}_{t-k}.$$

We first determine the marginals of γ_i and τ_i conditional on the threshold variable r and on d and afterwards the marginals of γ_i and τ_i, $i=1,2$.

From (3.7) the joint posterior distribution of γ_i and τ_i, conditional on r and on d, $P(\gamma_i,\tau_i|r,d,D)$, is

$$\begin{gathered} P(\gamma_i,\tau_i|r.d,D)\propto P(\gamma_i,\tau_i,r,d|D)\propto|\tau_i(\mathbf{A}_i(r)+\mathbf{Q}_i)|^{1/2}\times \\ \exp\{-\frac{\tau_i}{2}(\gamma_i-(\mathbf{A}_i(r)+\mathbf{Q}_i)^{-1}(\mathbf{B}_i(r)+\mathbf{Q}_i\gamma_{0i}))'(\mathbf{A}_i(r)+\mathbf{Q}_i) \\ \times(\gamma_i-(\mathbf{A}_i(r)+\mathbf{Q}_i)^{-1}(\mathbf{B}_i(r)+\mathbf{Q}_i\gamma_{0i}))\} \\ \times|\mathbf{A}_i(r)+\mathbf{Q}_i|^{-1/2}\tau_i^{(n_i+\alpha_i-p_i-q_i-1)/2-1}\exp\{-\frac{\tau_i}{2}[\beta_i+Y_i'Y_i \\ -(\mathbf{B}_i(r)+\mathbf{Q}_i\gamma_{0i})'(\mathbf{A}_i(r)+\mathbf{Q}_i)^{-1}(\mathbf{B}_i(r)+\mathbf{Q}_i\gamma_{0i})+\gamma_{0i}'\mathbf{Q}_i\gamma_{0i}]\}\,. \end{gathered} \tag{3.8}$$

Corollary 1. *From (3.8) we obtain the marginal distributions*

$$\gamma_i|\tau_i,r,d,D \quad \sim \quad N((\mathbf{A}_i(r)+\mathbf{Q}_i)^{-1}(\mathbf{B}_i(r)+\mathbf{Q}_i\gamma_{0i}),(\tau_i(\mathbf{A}_i(r)+\mathbf{Q}_i))^{-1})\,, \tag{3.9}$$

$$\tau_i|r,d,D \quad \sim \quad \text{Gamma}(\nu_i/2, F_i(r)/2)\,, \nu_i = n_i + \alpha_i - p_i - q_i - 1\,. \tag{3.10}$$

Moreover, integrating (3.8) with respect to τ_i *we obtain*

$$\gamma_i|r,d,D \quad \sim \quad t_{p_i+q_i+1}(\nu_i, (\mathbf{A}_i(r)+\mathbf{Q}_i)^{-1}(\mathbf{B}_i(r)+\mathbf{Q}_i\gamma_{0i}), \mathbf{W_i(r)})\,, \tag{3.11}$$

where $\mathbf{W_i}(r) = \nu_i(\mathbf{A}_i(r)+\mathbf{Q}_i)F_i(r)^{-1}$.

To find the exact distributions of γ_i and τ_i we need the posterior distribution of r and d, which is obtained integrating (3.8) with respect to γ_i and τ_i, $i = 1, 2$. That is,

$$P(r,d|D) \propto \prod_{i=1}^{2} |\mathbf{A}_i(r)+\mathbf{Q}_i|^{-1/2}[F_i(r)]^{-\nu_i/2}, \quad r \in D_r \text{ and } d \in \{1, 2, ..., d_0\}, \tag{3.12}$$

where the normalizing constant of (3.12) and the moments of the distribution can be found numerically. Values of the threshold for which at least one observation is not assigned to each of the two regimes are values where $P(r,d|D)$ should be zero.

Therefore, the exact marginals of τ and γ are

$$P(\tau_i|D) = \sum_{d\in\{1,\ldots,d_0\}} \sum_{r\in D_r} P(\tau_i|r,d,D)P(r,d|D), \quad i = 1, 2, \tag{3.13}$$

and

$$P(\gamma_i|D) = \sum_{d\in\{1,\ldots,d_0\}} \sum_{r\in D_r} P(\gamma_i|r,d,D)P(r,d|D), \quad i = 1, 2, \tag{3.14}$$

respectively. We see that (3.13) is a mixture of gamma distributions, with

$$E(\tau_i|D) = E_{r,d|D}(E(\tau_i|r,d,D)) = \sum_{d\in\{1,\ldots,d_0\}} \sum_{r\in D_r} \frac{\nu_i}{F_i(r)} P(r,d|D),$$

and

$$Var(\tau_i|D) = E_{r,d|D}(Var(\tau_i|r,d,D)) + Var_{r,d|D}(E(\tau_i|r,d,D))$$

$$= \sum_{d\in\{1,\ldots,d_0\}} \sum_{r\in D_r} \frac{2\nu_i}{[F_i(r)]^2} P(r,d|D) + \sum_{d\in\{1,\ldots,d_0\}} \sum_{r\in D_r} [E(\tau_i|r,d,D) - E(\tau_i|D)]^2 P(r,d|D).$$

Similarly, (3.14) is a mixture of multivariate t-distributions, with means and variances given respectively by

$$E(\gamma_i|D) = \sum_{d\in\{1,\ldots,d_0\}} \sum_{r\in D_r} (\mathbf{A}_i(r)+\mathbf{Q}_i)^{-1}(\mathbf{B}_i(r)+\mathbf{Q}_i\gamma_{0i})P(r,d|D)$$

and

$$Var(\gamma_i|D) = \sum_{d\in\{1,\ldots,d_0\}} \sum_{r\in D_r} \frac{F_i(r)(\mathbf{A}_i(r)+\mathbf{Q}_i)^{-1}}{\nu_i - 2} P(r,d|D)$$

$$+\sum_{d\in\{1,\dots,d_0\}}\sum_{r\in D_r}[(\mathbf{A}_i(r)+\mathbf{Q}_i)^{-1}(\mathbf{B}_i(r)+\mathbf{Q}_i\gamma_{0i})-E(\gamma_i|D)][(\mathbf{A}_i(r)+\mathbf{Q}_i)^{-1}$$
$$\times(\mathbf{B}_i(r)+\mathbf{Q}_i\gamma_{0i})-E(\gamma_i|D)]'P(r,d|D).$$

4. Posterior Analysis for the Improper Jeffreys Prior

If we have few or no information on the parameters we can use Jeffreys prior,

$$P(\gamma_1,\gamma_2,\tau_1,\tau_2,r,d)\propto\tau_1^{-1}\tau_2^{-1}. \tag{4.1}$$

For this prior the posterior will be also a normal-gamma distribution. The results obtained above are modified making $\beta_i\to 0$, $\mathbf{Q}_i\to\mathbf{0}$ and $\alpha_i\to-(p_i+q_i+1)$, $i=1,2$, in the joint posterior distribution.

Corollary 2. *If the likelihood function (3.4) is combined with (4.1), the marginals are:*

$$\gamma_i|\tau_i,r,d,D \quad\sim\quad N(\mathbf{A}_i(r)^{-1}\mathbf{B}_i(r),\tau_i^{-1}\mathbf{A}_i(r)^{-1}),\quad i=1,2, \tag{4.2}$$

$$\tau_i|r,d,D \quad\sim\quad \text{Gamma}(\nu_i'/2,h_i(r)/2), \tag{4.3}$$

where $\nu_i'=n_i-2(p_i+q_i+1)$ *and* $h_i(r)=Y_i'Y_i-\mathbf{B}_i(r)'\mathbf{A}_i(r)^{-1}\mathbf{B}_i(r)$,

$$\gamma_i|r,d,D \quad\sim\quad t_{p_i+q_i+1}(\nu_i',\mathbf{A}_i(r)^{-1}\mathbf{B}_i(r),\nu_i'\mathbf{A}_i(r)h_i(r)^{-1}). \tag{4.4}$$

From (3.12) the distribution of r and d is given by

$$P(r,d|D)\propto\prod_{i=1}^{2}|\mathbf{A}_i(r)|^{-1/2}[h_i(r)]^{-\nu_i'/2},\quad r\in D_r,\quad d\in\{1,2,...,d_0\}. \tag{4.5}$$

The posterior marginals for τ_i and γ_i, $i=1,2$, are as in (3.13) and (3.14), respectively, with ν_i and $F_i(r)$ replaced by ν_i' and $h_i(r)$ in the expressions for $E(\tau_i|D)$ and $Var(\tau_i|D)$. For the distribution of $\gamma_i|D$ we have

$$E(\gamma_i|D)=\sum_{d\in\{1,\dots,d_0\}}\sum_{r\in D_r}\mathbf{A}_i(r)^{-1}\mathbf{B}_i(r)P(r,d|D) \tag{4.6}$$

and

$$Var(\gamma_i|D)=\sum_{d\in\{1,\dots,d_0\}}\sum_{r\in D_r}\frac{h_i(r)\mathbf{A}_i(r)^{-1}}{\nu_i'-2}P(r,d|D)$$
$$+\sum_{d\in\{1,\dots,d_0\}}\sum_{r\in D_r}[(\mathbf{A}_i(r)^{-1}\mathbf{B}_i(r)-E(\gamma_i|D))(\mathbf{A}_i(r)^{-1}\mathbf{B}_i(r)-E(\gamma_i|D))']P(r,d|D). \tag{4.7}$$

Remark 1. A different approach to the formal representation of ignorance is based on entropy. Entropy is a very familiar concept in physics as a measure of the amount of disorder and unpredictability in a system. Jeffreys was primarily a physicist, and the maximum entropy approach has been advocated forcefully by another physicist, Jaynes (1968, 1981). The primary criticism of this approach is that it is not invariant under change of parametrization, a problem for which the Jeffreys prior was designed to avoid.

Although the objetive view of probability leads to the use of Bayes' theorem, with all the characteristics of the Bayesian method for posterior inference, the way in which the prior distribution is derived implies some important differences.

(a) If the Jeffreys prior is used, the posterior distribution will depend on the form of the experiment in a way that violates the Likelihood Principle.

(b) If the prior distribution is defined by the maximizing entropy (reference prior), then the posterior inference will not be coherent with respect to the reparametrization.

(c) If subjective prior information exists then it may not be reflected in the posterior distribution.

Reference prior distributions were proposed by Bernardo(1979), and indicates another possible use of "ignorance priors" in a subjective Bayesian analysis. See O'Hagan(1994) for more details. We will not pursue this further in this paper.

Remark 2. Note that the posterior distributions are known and simulation algorithms are not necessary. In spite of this we would have a lot of work to fit the models for each d and r. The problem is that the orders p_l and q_l change from regime to regime. In order to solve this problem we use next the rearranged autoregression method proposed by Tsay (1989).

5. **Rearranged TARMA Model**

A rearranged TARMA model is useful because effectively separates the regimes. Specifically, the observations are assembled in groups in such way that the observations in each group follow an ARMA model with the same order. Also the separation does not require knowing the precise value of the threshold r. Only the number of observations in each regime depends on r.

An ARMA(p,q) process with n observations can be seen as

$$Y_t = (1, Y_{t-1}, Y_{t-2}, ..., Y_{t-p}, a_{t-1}, a_{t-2}, ..., a_{t-q})\gamma + a_t, \quad t = 1, ..., n,$$

where $\{a_t\}$ is white noise and $\gamma = (\phi_0, \phi_1, ...\phi_p, \theta_1, ..., \theta_q)'$ is the $(p+q+1)$-dimension vector of coefficients.

The threshold variable Y_{t-d} may assume values $\{Y_{p+1-d}, Y_{p+2-d}, ..., Y_{n-d}\}$ for $t = p+1, ..., n$. Observations and threshold values are compared and indices π_i are associated to the observations in order of occurrence. Then, the s observations of the first regime are Y_{π_i+d} for $i = 1, ..., s$ and, for $i > s$ we have observations to the second regime.

We rewrite the model (2.2) as:

$$Y_{\pi_i+d} = \begin{cases} \phi_{10} + \sum_{j=1}^{p_1} \phi_{1j} Y_{\pi_i+d-j} + a^{(1)}_{\pi_i+d} + \sum_{k=1}^{q_1} \theta_{1k} a^{(1)}_{\pi_i+d-k}, & \text{if } i \le s \\ \phi_{20} + \sum_{j=1}^{p_2} \phi_{2j} Y_{\pi_i+d-j} + a^{(2)}_{\pi_i+d} + \sum_{k=1}^{q_2} \theta_{2k} a^{(2)}_{\pi_i+d-k}, & \text{if } i > s \end{cases} \tag{5.1}$$

where s satisfies $Y_{\pi_s} \le r < Y_{\pi_{s+1}}$.

The model (5.1) is called "Rearranged TARMA Model". Note that the values $(Y_t, 1, Y_{t-1}, ..., Y_{t-p_i}, a^{(i)}_{t-1}, ..., a^{(i)}_{t-q_i})$, $i = 1, 2$, were rearranged based on the threshold variable and there are s observations in the first regime and $(n - p - s)$ in the second. Specifically,$\{Y_{\pi_1+d}, ..., Y_{\pi_s+d}\}$ and $\{Y_{\pi_{s+1}+d}, ..., Y_{\pi_{n-p}+d}\}$ are observations in the first and the second regimes, respectively.

Let $Y_1^* = (Y_{\pi_1+d}, ..., Y_{\pi_s+d})'$ and $Y_2^* = (Y_{\pi_{s+1}+d}, ..., Y_{\pi_{n-p}+d})'$. Let also $X_1^* = (x_{1,\pi_1+d}, x_{1,\pi_2+d}, ..., x_{1,\pi_s+d})$, $X_2^* = (x_{2,\pi_{s+1}+d}, ..., x_{2,\pi_{n-p}+d})$ where,

$x_{1,k} = (1, Y_{\pi_k+d-1}, Y_{\pi_k+d-2}, ..., Y_{\pi_k+d-p_1}, a^{(1)}_{\pi_k+d-1}, ..., a^{(1)}_{\pi_k+d-q_1})'$, $k = 1, ..., s$ and

$x_{2,k} = (1, Y_{\pi_k+d-1}, Y_{\pi_k+d-2}, ..., Y_{\pi_k+p_2}, a^{(2)}_{\pi_k+d-1}, ..., a^{(2)}_{\pi_k+d-q_2})'$, $k = s+1, ..., n-p$.

Then,

$$(X_1^*)' = \begin{pmatrix} x_{1,\pi_1+d} \\ x_{1,\pi_2+d} \\ \vdots \\ x_{1,\pi_s+d} \end{pmatrix} = \begin{pmatrix} 1 & Y_{\pi_1+d-1} & \cdots & Y_{\pi_1+d-p_1} & a^{(1)}_{\pi_1+d-1} & \cdots & a^{(1)}_{\pi_1+d-q_1} \\ 1 & Y_{\pi_2+d-1} & \cdots & Y_{\pi_2+d-p_1} & a^{(1)}_{\pi_2+d-1} & \cdots & a^{(1)}_{\pi_2+d-q_1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & Y_{\pi_s+d-1} & \cdots & Y_{\pi_s+d-p_1} & a^{(1)}_{\pi_s+d-1} & \cdots & a^{(1)}_{\pi_s+d-q_1} \end{pmatrix}$$

is a matrix $s \times (1 + p_1 + q_1)$. Define similarly, the matrix $(X_2^*)'$, which is $(n - p - s) \times (1 + p_2 + q_2)$.

6. Posterior Analysis

We assume the prior distributions given as follows:

(1) γ_1 and γ_2 are independent $\sim N(\gamma_{0i}, \mathbf{Q}_i^{-1})$, $i = 1, 2$,

(2) τ_1 and τ_2 are independent $\sim \text{Gamma}(\alpha_i/2, \beta_i/2)$, $i = 1, 2$,

(3) $r \sim$ Uniform (a, b) , where a and b are suitably chosen,

(4) $d \sim$ Uniform $\{1, 2, ..., d_0\}$.

The hyperparameters $\gamma_{0i}, \mathbf{Q}_i, \alpha_i, \beta_i, d_0, i = 1, 2$ are supposed to be known.

The prior may be written

$$P(\gamma_1, \gamma_2, \tau_1, \tau_2, r, d) \propto \left(\tau_1^{(\alpha_1/2)-1} \tau_2^{(\alpha_2/2)-1}\right) \exp\left\{-\frac{1}{2} \sum_{i=1}^{2} (\beta_i \tau_i + (\gamma_i - \gamma_{0i})' \mathbf{Q}_i (\gamma_i - \gamma_{0i}))\right\}. \tag{6.1}$$

The conditional likelihood is given approximately by

$$\mathcal{L}(\gamma_1, \gamma_2, \tau_1, \tau_2, r, d|Y) \propto (\tau_1^{s/2})(\tau_2^{(n-p-s)/2})$$

$$\times \exp\left\{-\frac{\tau_1}{2}\sum_{i=1}^{s}\left(Y_{\pi_i+d}-\phi_{10}-\sum_{j=1}^{p_1}\phi_{1j}Y_{\pi_i+d-j}-\sum_{k=1}^{q_1}\theta_{1k}\hat{a}^{(1)}_{\pi_i+d-k}\right)^2\right.$$
$$\left.-\frac{\tau_2}{2}\sum_{i=s+1}^{n-p-s}\left(Y_{\pi_i+d}-\phi_{20}-\sum_{j=1}^{p_2}\phi_{2j}Y_{\pi_i+d-j}-\sum_{k=1}^{q_2}\theta_{2k}\hat{a}^{(2)}_{\pi_i+d-k}\right)^2\right\} \tag{6.2}$$

where $\hat{a}_t^{(i)}$ is the estimator of $a_t^{(i)}$ obtained as in preceding sections and $Y = \{Y_1, Y_2, ..., Y_n\}$ is the set of observed sample values. Replacing the $a_t^{(i)}$ by the $\hat{a}_t^{(i)}$ in the $x_{i,t}$, $i = 1, 2$, the likelihood function may be rewritten as

$$\mathcal{L}(\gamma_1,\gamma_2,\tau_1,\tau_2,r,d|Y) \propto \tau_1^{n_1/2}\tau_2^{n_2/2}\exp\left\{-\sum_{i=1}^{2}\frac{\tau_i}{2}(Y_i^*-X_i^{*\prime}\gamma_i)'(Y_i^*-X_i^{*\prime}\gamma_i)\right\} \tag{6.3}$$

where n_1, and n_2 are the numbers of observations in the first and second regimes, respectively.

Theorem 2. *If the prior (6.1) is combined with the likelihood (6.3) we obtain a normal-gamma posterior distribution*

$$P(\gamma_1,\gamma_2,\tau_1,\tau_2,r,d|Y) \quad \alpha \quad (\tau_1)^{\frac{(n_1+\alpha_1)}{2}-1}(\tau_2)^{\frac{(n_2+\alpha_2)}{2}-1}$$
$$\times\exp\Big\{-\frac{1}{2}\sum_{i=1}^{2}\Big[\tau_i(\beta_i+Y_i^{*\prime}Y_i^*)-(\tau_iX_i^*Y_i^*+\mathbf{Q}_i\gamma_{0i})'(\tau_iX_i^*X_i^{*\prime}+\mathbf{Q}_i)^{-1}(\tau_iX_i^*Y_i^*+\mathbf{Q}_i\gamma_{0i})$$
$$+[\gamma_i-(\tau_iX_i^*X_i^{*\prime}+\mathbf{Q}_i)^{-1}(\tau_iX_i^*Y_i^*+\mathbf{Q}_i\gamma_{0i})]'(\tau_iX_i^*X_i^{*\prime}+\mathbf{Q}_i)$$
$$\times[\gamma_i-(\tau_iX_i^*X_i^{*\prime}+\mathbf{Q}_i)^{-1}(\tau_iX_i^*Y_i^*+\mathbf{Q}_i\gamma_{0i})]\Big]\Big\}. \tag{6.4}$$

To make inferences for the parameters we need to derive the marginal posterior distributions. For each parameter, the marginal distribution is obtained integrating (6.4) with respect to the other parameters. In this particular case, these distributions are not easily obtained. We can solve this problem using the Gibbs sampler. To implement the Gibbs sampler, we need to derive the conditional posterior distribution of each parameter given all the others (full conditional distributions).

Corollary 3. *From (6.4) we get the conditional posterior distributions*

$$P(\gamma_i|\tau_1,\tau_2,r,d,Y)\propto\exp\{(-1/2)[\gamma_i-(\tau_iX_i^*X_i^{*\prime}+\mathbf{Q}_i)^{-1}(\tau_iX_i^*Y_i^*+\mathbf{Q}_i\gamma_{0i})]'$$
$$(\tau_iX_i^*X_i^{*\prime}+\mathbf{Q}_i)[\gamma_i-(\tau_iX_i^*X_i^{*\prime}+\mathbf{Q}_i)^{-1}(\tau_iX_i^*Y_i^*+\mathbf{Q}_i\gamma_{0i})]\},$$

that is,

$$\gamma_i|\tau_1,\tau_2,r,d,Y\sim N_{p_i+q_i+1}\Big((\tau_iX_i^*X_i^{*\prime}+\mathbf{Q}_i)^{-1}(\tau_iX_i^*Y_i^*+\mathbf{Q}_i\gamma_{0i}),(\tau_iX_i^*X_i^{*\prime}+\mathbf{Q}_i)^{-1}\Big)$$

for $i = 1, 2$, *and*

$$P(\tau_i|\gamma_1, \gamma_2, r, d, Y) \propto \tau_i^{(n_i+\alpha_i)/2-1} \exp\Big\{-\frac{\tau_i}{2}[\beta_i + (Y_i^* - X_i^{*\prime}\gamma_i)'(Y_i^* - X_i^{*\prime}\gamma_i)]\Big\},$$

that is,

$$\tau_i|\gamma_1, \gamma_2, r, d, Y \sim \text{Gamma}((n_i+\alpha_i)/2, (\beta_i+(Y_i^* - X_i^{*\prime}\gamma_i)'(Y_i^*-X_i^{*\prime}\gamma_i))/2), \quad i = 1, 2.$$

Corollary 4. *The conditional posterior probability function of* r *is*

$$P(r|\gamma_1, \gamma_2, \tau_1, \tau_2, d, Y) \propto \tau_1^{n_1/2}\tau_2^{n_2/2} \exp\left\{(-1/2)\sum_{i=1}^{2}\tau_i(Y_i^* - X_i^{*\prime}\gamma_i)'(Y_i^* - X_i^{*\prime}\gamma_i)\right\},$$

where, n_1 *and* n_2 *are functions of* r.

Corollary 5. *The conditional posterior probability function of* d *is a multinomial distribution with probability*

$$P(d|\gamma_1, \gamma_2, \tau_1, \tau_2, r, Y) = \frac{\mathcal{L}(\gamma_1, \gamma_2, \tau_1, \tau_2, r, d|Y)}{\sum_{d=1}^{d_0}\mathcal{L}(\gamma_1, \gamma_2, \tau_1, \tau_2, r, d|Y)},$$

where $\mathcal{L}(\gamma_1, \gamma_2, \tau_1, \tau_2, r, d|Y)$ *is the likelihood function given in* (6.3).

Note that all conditional densities are identified with exception of that of r. The Metropolis-Hastings algorithm is then employed with regard to r, and we get the conditional posterior of r so the Gibbs sampler can then be used.

7. Predictive Distribution

In this section, we determine the predictive density for the one-step ahead prediction, Y_{n+1} given that we have observations $\{Y_1, Y_2, Y_3, ..., Y_n\}$.

We have that Y_{n+1} belongs to the first regime if $Y_{n+1-d} \le r$ and Y_{n+1} belongs to the second regime if $Y_{n+1-d} \le r$. Let us suppose $Y_{n+1-d} \le r$ (if $Y_{n+1-d} > r$ the procedure is similar).

We can see that the last observation in the first regime is Y_{π_s-d}, hence

$$Y_{n+1} = \phi_{10} + \sum_{j=1}^{p_1}\phi_{1j}Y_{\pi_s+d-j} + a^{(1)}_{\pi_s+d} + \sum_{k=1}^{q_1}\theta_{1k}a^{(1)}_{\pi_s+d-k}.$$

The density of Y_{n+1} is given by

$$P(Y_{n+1}|\gamma_1, \gamma_2, \tau_1, \tau_2, r, d) \propto \tau_1^{1/2} \exp\Big\{-\frac{\tau_1}{2}[Y_{n+1}-$$
$$-(1, Y_{\pi_s+d-1}, Y_{\pi_s+d-2}, ..., Y_{\pi_s+d-p_1}, a^{(1)}_{\pi_s+d-1}, a^{(1)}_{\pi_s+d-2}, ..., a^{(1)}_{\pi_s+d-q_1})\gamma_1]^2\Big\}$$

where the vector $(1, Y_{\pi_s+d-1}, Y_{\pi_s+d-2}, ..., Y_{\pi_s+d-p_1}, a^{(1)}_{\pi_s+d-1}, a^{(1)}_{\pi_s+d-2}, ..., a^{(1)}_{\pi_s+d-q_1})$ is the s-th matrix line $(X_1^*)'$. Hence, if we consider $I = (0, 0, 0, ..., 0, 1)'$, a $(p_1 + q_1 + 1) \times 1$ vector, the density of Y_{n+1} is rewritten as

$$P(Y_{n+1}|\gamma_1, \gamma_2, \tau_1, \tau_2, r, d) \propto \tau_1^{1/2} \exp\{-\frac{\tau_1}{2}(Y_{n+1} - I'(X_1^*)'\gamma_1)^2\}. \tag{7.1}$$

The joint distribution of $Y_{n+1}, \gamma_1, \gamma_2, \tau_1, \tau_2, r, d$ and Y is obtained from equations (6.4) and (7.1).

The inference for Y_{n+1} will be done through the Gibbs sampler, so the distribution of Y_{n+1} given all the other parameters and past observations is necessary. We have

$$Y_{n+1}|\gamma_1, \gamma_2, \tau_1, \tau_2, r, d, Y \propto N(I'(X_1^*)'\gamma_1, \tau_1^{-1}).$$

Remark 1. If we have few or no information about the parameters, we could use improper Jeffreys prior,

$$p(\gamma_1, \gamma_2, \tau_1, \tau_2, r, d) \propto \tau_1^{-1}\tau_2^{-1}. \tag{7.2}$$

If prior (7.2) is combined with the likelihood (6.3) the posterior still will be a normal-gamma. The results obtained for the full conditional distributions are modified letting $\beta_i \to 0, \mathbf{Q} \to \mathbf{0}$ and $\alpha_i \to -(p_i + q_i + 1)$, $i = 1, 2$, in the joint posterior distribution (6.4).

Remark 2. Details of the Gibbs sampler and Metropolis-Hastings algorithm can be found in Casella and George (1992) and Chib and Greenberg (1995), respectively.

8. Applications

Example 1. Let us illustrate the method of rearranged TARMA through an example where the data are simulated.

We generated 200 observations for the $TARMA(2; 1, 1, 1, 1)$ model, given by:

$$Y_t = \begin{cases} \phi_{11}Y_{t-1} + a_t^{(1)} + \theta_{11}a_{t-1}^{(1)}, & \text{if } Y_{t-d} \le r \\ \phi_{21}Y_{t-1} + a_t^{(2)} + \theta_{21}a_{t-1}^{(2)}, & \text{if } Y_{t-d} > r \end{cases} \tag{8.1}$$

where $\{a_t^{(i)}\}$ is a sequence of random variables i.i.d. $\sim N(0, \tau_i^{-1}), i = 1, 2$ and $\{a_t^{(1)}\}$ and $\{a_t^{(2)}\}$ are independent. The values chosen for the parameters are $\phi_{11} = 0.8, \theta_{11} = -0.5, \tau_1 = 0.5$, $\phi_{21} = -0.3, \theta_{21} = 0.7, \tau_2 = 1.0$, $r = 0.4$ and $d = 1$.

For the hyperparameters we consider $\gamma_{0i} = (0.0, 0.0)', \alpha_i = 2.0, \beta_i = 1$ and

$$\mathbf{Q}_i = \begin{pmatrix} 0.1 & 0.0 \\ 0.0 & 0.2 \end{pmatrix}, \quad i = 1, 2.$$

Suppose also that $a = \rho_{25}$ and $b = \rho_{75}$, where ρ_k is the k-th percentile for the data.

The residuals were estimated from an AR(2) model fitted to the the data and were recursively obtained from $\hat{a}_t = Y_t - \hat{\phi}_1 Y_{t-1} - \hat{\phi}_2 Y_{t-2}$, where $\hat{\phi}_i$ is the least squares estimator of $\phi_i,\ i = 1, 2$.

The Metropolis-Hastings (MH) algorithm and the Gibbs sampler (GS) were used with five chains and five different initial values. For the MH we considered initially the normal distribution as the transition kernel. Since the convergence was slow we changed to a logistic transition distribution. We decided for an acceptance ratio around 50%. We started with $1,000$ iterations for each chain. Figures 1 and 2 show plots and histograms for the variable r, considering different numbers of iterations. The histograms show that the convergence was reached eventually.

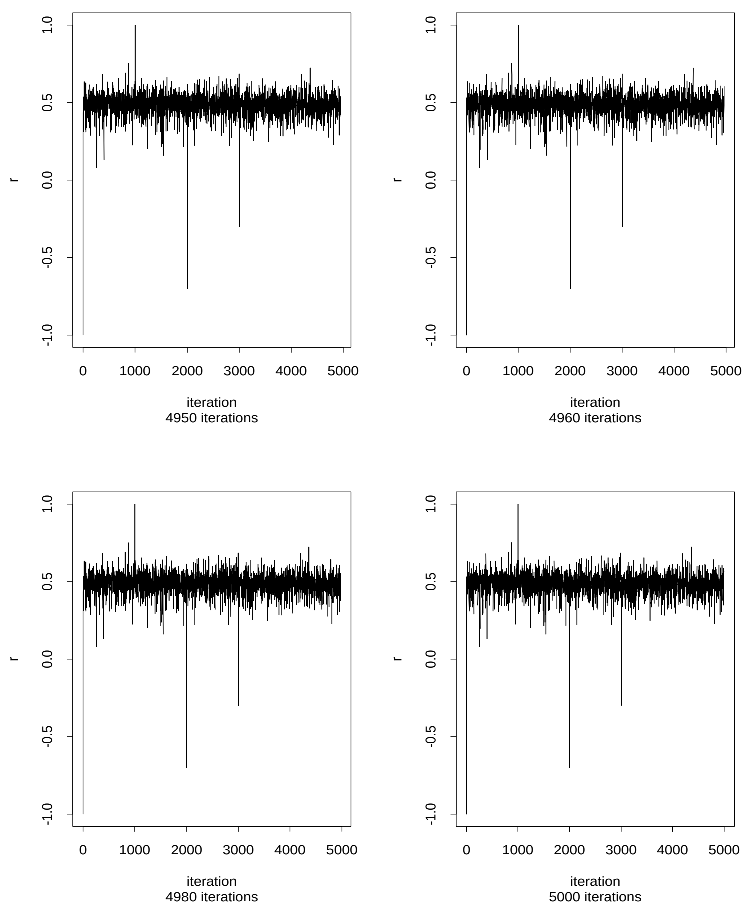

Figure 1. Plots for the variable r, considering different numbers of iterations

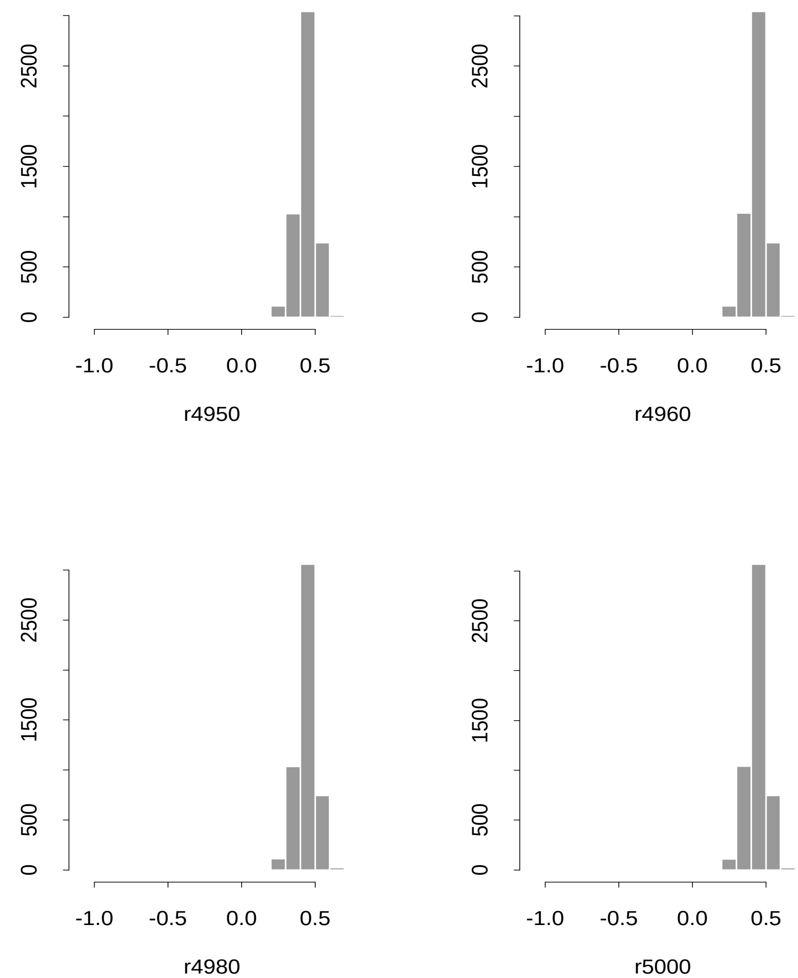

Figure 2. Histograms for the variable r, considering different numbers of iterations

The inferences for the parameters and for the one-step ahead forecast were done considering the $5,000$ iterations. In order to make small the effect of initial values we discarded the first 400 values and took, from the remaining ones, one observation for each group of fifteen, to get approximate independence. The results so obtained are presented in Table 1. For the parameter d we got 1 for the median and 0 for the dispersion measure. We notice that the factor $\sqrt{\hat{R}} \approx 1$ for all parameters, indicating convergence.

Table 1. Simulation results from 5,000 iterations (1,000 in each chain), final 200 values

parameters	**real value**	**mean**	**s.d.**	$\sqrt{\hat{R}}$
ϕ_{11}	0.8	0.776	0.246	1.0147
θ_{11}	-0.5	-0.422	0.263	0.9954
ϕ_{21}	-0.3	-0.276	0.216	1.0000
θ_{21}	0.7	0.738	0.241	1.0001
τ_1	0.5	0.431	0.066	1.0006
τ_2	1.0	0.891	0.208	1.0236
r	0.4	0.443	0.066	1.0101
Y_{n+1}	0.9067	0.9297	1.1224	1.0045

The results in Table 1 tell us that the MH algorithm and the GS performed well. The estimators are "robust," since similar results were obtained for different initial values. To further check, we simulated another five chains, with $5,000$ iterations each and discarded the first $3,500$ values and sampled at each fifteen observations to get the final 500 values(100 values for each chain) for inference. The results are shown in Table 2 and in Figure 3 we have the histograms for the variable r, for iterations between $24,000$ and $25,000$. We can see that the plots in Figures 2 and 3 are similar, indicating that convergence in fact already had occurred. In Figure 4 we present the plots for the simulated data and for the parameters.

Table 2. Simulation results from 25,000 iterations (5,000 in each chain), final 500 values

parameters	**real value**	**mean**	**s.d.**	$\sqrt{\hat{R}}$
ϕ_{11}	0.8	0.754	0.236	1.0019
θ_{11}	-0.5	-0.458	0.254	1.0028
ϕ_{21}	-0.3	-0.274	0.238	1.0029
θ_{21}	0.7	0.731	0.260	0.9985
τ_1	0.5	0.432	0.060	1.0053
τ_2	1.0	0.885	0.216	1.0011
r	0.4	0.448	0.063	1.0065
Y_{n+1}	0.9067	0.8673	1.1406	0.9998

Example 2. One of the most studied series is the sunspot data, first treated by Schuster (1906). Several types of models, linear and non-linear, were tried to model this series. Tong and Lim (1980) have shown that a threshold model is capable of reproducing the asymmetric and periodic behavior present in the sunspots. Among the threshold models we mention in particular: Tong and Lim (1980) fitted a TAR(2; 3,11) model, considering data from 1,700 to 1,920, with $d = 3$ and $r = 36.6$. Ghaddar and Tong (1980) used the TAR(2; 10, 2) model, with $d = 9$ and $r = 11.93$ and the method of rearranged autoregression. Tsay (1989) fitted a TAR model with 3 regimes, $d = 2$ and two threshold values, $r_1 = 34.8$, $r_2 = 70.7$ and orders 11, 10 and 11, with data in the period 1,700 – 1,979.

With the purpose of showing that we can get a more parsimonious model fitting a TARMA model, we consider as in Tsay (1989) the rearranged autoregression procedure and the values $d = 2$, $r_1 = 34.8$ and $r_2 = 70.7$.

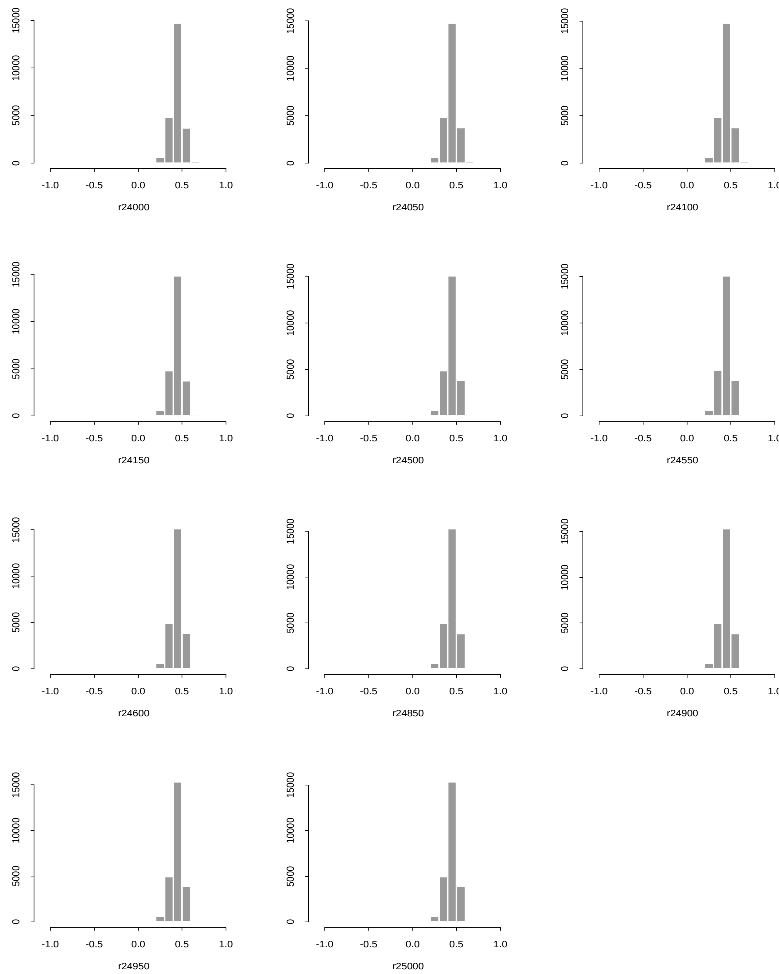

Figure 3. Histograms for the variable r, for iterations between $24,000$ and $25,000$.

For the analysis we assume a proper prior with $\alpha_i = 2, \beta_i = 1, \gamma_{0i} = 0, i = 1, 2, 3$, and the matrix $\mathbf{Q_i}$ such that

$$q_{ij} = \begin{cases} 0, & \text{if } i \neq j \\ 0.i, & \text{if } i = j,\ i = 1, 2, 3. \end{cases}$$

From the autocorrelation and partial autocorrelation functions of the observations in each regime we obtained the orders of the models.

For the Gibbs sampler we considered two chains with different initial values for each parameter and 4,000 iterations, 2,000 for each chain. The first 400 iterations were removed for the computations.

We fitted a TARMA(3; 1, 1, 1; 0, 2, 1) model, and the estimated parameters are shown in Table 3. We see that for the first regime we have an AR(1) model and for the other two regimes we have ARMA models with a total number of parameters smaller than that used by Tsay (1989).

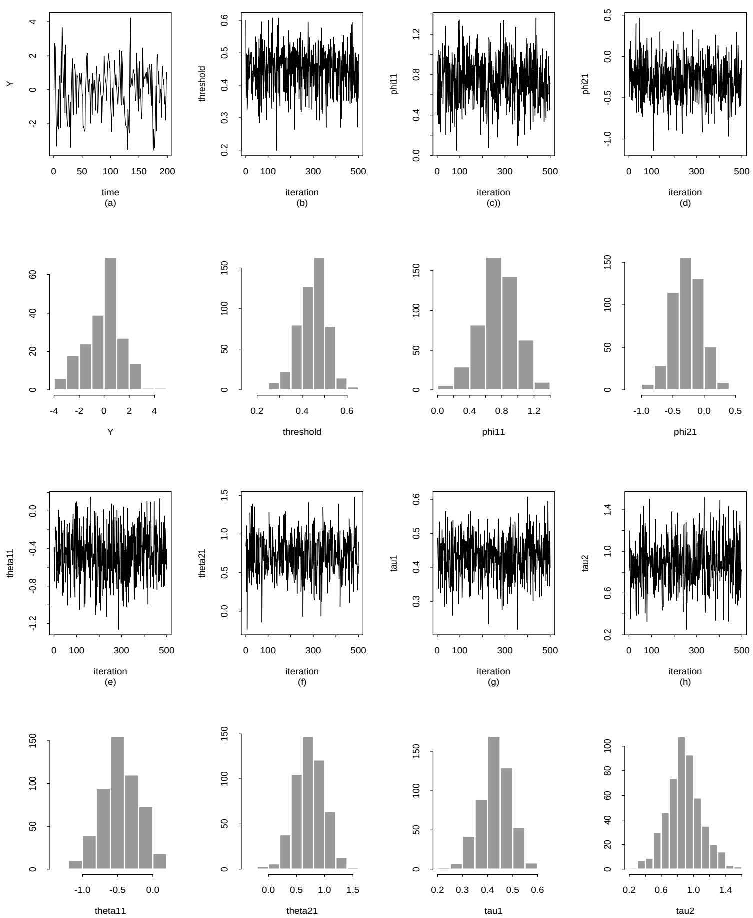

Figure 4. Plots for the simulated data and for the parameters.

9. **Some Remarks**

The aim of this paper was to introduce the analysis of the TARMA model under a Bayesian approach.

The generalization to k regimes shoud bring no further difficulties. The only modification is the choice of the prior for the threshold variables. To illustrate, consider the case $k = 3$. We have three ARMA processes, with orders $(p_i, q_i), i = 1, 2, 3$ and two thresholds r_1, r_2, in such a way that $Y_{t-d} \leq r_1$ for the first regime, $r_1 < Y_{t-d} \leq r_2$ for the second regime and $Y_{t-d} > r_2$ for the third regime, where $-\infty < r_1 < r_2 < \infty$.

Table 3. Model TARMA(3;1,1,1;0,2) fitted to the sunspots series, 1,700 – 1,979, with $d = 2, r_1 = 34.8$ and $r_2 = 70.7$

parameters	mean	s.d.	$\sqrt{\hat{R}}$
ϕ_{10}	6.5169	1.0942	1.0001
ϕ_{11}	0.4086	0.0448	0.9999
ϕ_{20}	15.8547	2.7384	1.0008
ϕ_{21}	0.5834	0.0488	1.0027
θ_{21}	0.4360	0.0936	1.0048
θ_{22}	0.3042	0.0867	1.0017
ϕ_{31}	0.8974	0.0399	1.0003
θ_{31}	0.8156	0.1318	0.9996
τ_1	0.0168	0.0023	1.0022
τ_2	0.0066	0.0011	1.0028
τ_3	0.0012	0.0002	0.9993

Given the sample $Y_1, \ldots, Y_n$, let $D = \{Y_{(1)}, \ldots, Y_{(n)}\}$. Since $r_1 \in (Y_{(1)}, Y_{(n)})$ and $r_2 \in (Y_{(1)}, Y_{(n)})$, with the condition $r_1 < r_2$, we have that $\mathbf{r} = (r_1, r_2)'$ is uniform in the triangular region $\{Y_{(1)} < r_1 < Y_{(n)}, Y_{(1)} < r_2 < Y_{(n)}, r_1 < r_2\}$ and hence the posterior distribution of r, $p(r|D)$ is such that

$$\int_{Y_{(1)}}^{Y_{(n)}} \int_0^{r_2} p(r|D) dr_1 dr_2 = 1.$$

For k regimes, consider the $k-1$ threshold variables $r_1, \ldots, r_{k-1}$, under the condition $r_1 < r_2 < \ldots < r_{k-1}$.

Tests for non-linearity and existence of thresholds are given by Tong (1990). If the threshold is known, the analysis using the first method is simple, since we need to fit only two ARMA models, one for each regime, and inference is obtained directly from the known posterior distributions.

Some comments on the use of (3.6) to approximate (3.3) are in order. First, it has been used elsewhere(see Broemeling and Shaarawy, 1988 and Chen, 1992). Second, it is needed to obtain the theoretical results. It may not be necessary if one chooses to use MCMC algorithms. This will be pursued elsewhere. We believe that point estimators behave nicely under the approximation. A topic for further study would be to check if posterior distributions are robust under such approximation.

Acknowledgments. The authours would like to thank Marinho G. Andrade for computational support and to a referee for comments that led to the improvement of the presentation.

References

Bernardo, J.M. (1979). Reference posterior distributions for Bayesian inference (with discussion), *Journal of the Royal Statistical Society Series B*, **41**, 113–147.

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

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

BROEMELING, L.D. AND SHAARAWY S. (1988). Time series: A Bayesian analysis in the time domain. In *Bayesian Analysis of Time Series and Dynamic Models*, J.C. Spall, Ed., Marcel Dekker, New York, 1–21.

BROEMELING, L.D. AND TSURUMI, H. (1987). *Econometrics and Structural Change.* Marcel Dekker, New York.

CASELLA, G. AND GEORGE, E.I. (1992). Explaining the Gibbs sampler, *The American Statistician*, **46**, 167–174.

CHEN, C.W.S. (1992). Bayesian inference and forecasting in bilinear time series models, *Communications in Statistics – Theory and Methods*, **21**, 1725–1743.

CHEN, C.W.S. AND LEE J.C. (1995). Bayesian inference of threshold autoregressive models, *Journal of Time Series Analysis*, **16**, 483–392.

CHIB, S. AND GREENBERG, E. (1995). Understanding the Metropolis-Hastings algorithm, *The American Statistician*, **49**, 327–335.

GHADDAR, D.K. AND TONG, H. (1981). Data transformation and self-exciting threshold autoregression, *Journal of the Royal Statistical Society, Series C*, **30**, 238–48.

GELMAN, A. AND RUBIN, D.B. (1992). Inference from iterative simulation using multiple sequence, *Statistical Science*, **7**, 457–511.

HARVEY, A.C. (1981). *Time Series Models.* Wiley, New York.

JAYNES, E.T. (1968). Prior probabilities, *IEEE Transactions on Systems Science and Cybernetics*, **4**, 227–241.

– – –– (1981). Marginalization and prior probabilities. In *Bayesian Analysis in Econometrics and Statistics.* A. Zellner (ed.) North-Holland, Amsterdam.

O'HAGAN, A. (1994) *Bayesian Inference: Volume 2B, Kendall's Advanced Theory of Statistics.* Arnold, London.

POLE, A.M. AND SMITH, A.F.M. (1989). A Bayesian analysis of some threshold switching models, *Journal of Econometrics*, **29**, 97–119.

SCHUSTER, A. (1906). On the periodicities of sunspots, *Philosophical Transactions of the Royal Society, Series A*, **206**, 69–100.

TONG, H. (1983). *Threshold Models in Non-linear Time Series.* Springer Verlag, New York.

– – –– (1990). *Non-linear Time Series: A Dynamical Approach.* Oxford University Press.

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.

TSAY, R.S. (1989). Testing and modeling threshold autoregressive process, *Journal of the American Statistical Association*, **84**, 231–240.

THELMA SÁFADI
DEPARTAMENTO DE CIÊNCIAS EXATAS
UNIVERSIDADE FEDERAL DE LAVRAS
37200-000 LAVRAS, MG
BRASIL
E-mail: safadi@ufla.br

PEDRO A. MORETTIN
DEPARTAMENTO DE ESTATISTICA
INSTITUTO DE MATEMÁTICA E ESTATISTICA
UNIVERSITY OF SÃO PAULO, BRAZIL
CAIXA POSTAL 66.281
05315 – 970 SÃO PAULO, SP
BRASIL
E-mail: pam@ime.usp.br