

# Bayesian Prediction in Threshold Autoregressive Models With Exponential White Noise

Isabel M. S. Pereira<sup>\*1</sup>

*Mathematics Dep., University of Aveiro, Portugal*

M. Antonia Amaral-Turkman<sup>\*2</sup>

*Departamento de Estadística e Investigación Operativa,  
University of Lisbon, Portugal.*

## Abstract

In this paper, we develop a Bayesian analysis of a threshold autoregressive model with exponential noise. An approximate Bayes methodology, which is introduced here, and the Gibbs sampler are used to compute marginal posterior densities for the parameters of the model, including the threshold parameter, and to compute one-step-ahead predictive density functions.

The proposed methodology is illustrated with a simulation study and a real example.

**Key Words:** Threshold model; Bayesian prediction; Gibbs sampler.

**AMS subject classification:** 62CF15, 62M10, 62M20

## 1 Introduction

In applications we are frequently faced with time series data which, for a variety of different reasons, have characteristics not compatible with the usual assumptions of linearity or/and Gaussian errors. One of the many ways the assumption of linearity may fail is the presence of limit cycle (see, for example, Tong 1990). Threshold models introduced by Tong (1978) are, for instance, very powerful in the analysis of such time series.

Processes with non-Gaussian white noise are useful for modelling a wide range of phenomena that do not allow negative values or have a highly skewed distribution. Many problems such as daily flows of a river, wind speeds, amount of dissolved oxygen in a river, etc, fall well within this category. Since 1980 several time series models with non-Gaussian white

---

M. A. Amaral Turkman acknowledges financial support from INVOTAN, Portugal to carry out some of the research work, during her sabbatical leave, at the University of North Dakota, USA.

\*Correspondence to: \*1 Email: isabelp@mat.ua.pt; \*2 antonia.turkman@fc.ul.pt

noise have been suggested. Some references are Gaver and Lewis (1980), Lawrance and Lewis (1985), Bell and Smith (1986), Andel (1988, 1989), Andel and Garrido (1991), Davis and McCormick (1989), McCormick and Mathew (1993).

In this paper we address the problem of predicting future observations from a two-regime autoregressive threshold model with exponential errors, which is a very special case of the general threshold autoregressive model for non-negative variables. Due to the nature of the model considered, both classical and conventional explicit Bayesian solutions are difficult to obtain. We analyse a simple model to highlight the sort of problems that arise. However, contrary to classical analysis, Bayesian methodology can be applied with success through the use of Markov Chain Monte Carlo (MCMC) methods.<sup>1</sup>

The model we suggest in this paper is motivated from the following pollution problem: A situation which concerns municipalities where pulp and paper mills are very active, is the pollution caused by their wastewater to the rivers and streams, mainly during the summer months, when waters are warmer and the flow is low. Well accepted indicators of water quality in the management of rivers and streams, are besides temperature, colour and smell, maximum concentrations of unfavourable substances like toxic chemicals, and minimum concentrations of a favourable one, dissolved oxygen. Dissolved oxygen is both a direct measure of the quality of water and an indirect indicator that other pollutants are present. It is also known that some external factors, like air temperature, precipitation, tides, and son on, can influence the amount of dissolved oxygen in the water, contributing to different behaviours along the time of the amount of dissolved oxygen. In fact, due to these external factors, the amount of dissolved oxygen in rivers near paper mills can reach, in summer months, values very much below the accepted levels, contrary to what happens in winter months.

The paper is organized as follows; In section 2, we define the two-regime threshold autoregressive model with exponential white noise. In section 3, we study Bayesian inference and prediction for this model. An exact analysis, although possible to perform, becomes quite demanding computationally due to the fact that we are in the presence of a constrained parameter model. Integrations needed to perform Bayesian analysis cannot be calcu-

---

<sup>1</sup>Congdon (2001) presents a wide range of worked examples, drawing on Bayesian literature pre-and post-MCMC methods

lated since an explicit form for the support of the posterior distribution is difficult to obtain, particularly for large sample sizes. Hence, we suggest two different approaches to solve the problem. First we consider an approximation for the support of the posterior distribution. This approach works well for a reasonable sample size and has the advantage of leading to explicit inferences and prediction. Second we suggest using Gibbs sampling to obtain samples from the posterior distribution. Posterior expectations required for inference and prediction are then approximated by sample averages. Contrary to the other methods, generalization of Gibbs methodology to more complicated models presents no further difficulties. In section 4, we conduct a simulation study to compare the performance of the three versions of the Bayesian methodology and the classical approach. In section 5 we apply our model to study the amount of dissolved oxygen in river Vouga, which is located near a paper mill.

## 2 Definition of the model

Here we consider a particular two-regime autoregressive threshold model for non-negative random variables,

$$X_t = \rho X_{t-1} + \epsilon_t + \begin{cases} 0 & \text{if } X_{t-1} < r; \\ \gamma & \text{if } X_{t-1} \geq r \end{cases} \quad (2.1)$$

for  $t \geq 2$ , where  $r$  is the single threshold parameter and  $\epsilon_t$ , are independent exponentially distributed ( $\epsilon_t \sim Exp(\alpha)$ ) with p.d.f.

$$f_{\epsilon_t}(x; \alpha) = \alpha e^{-\alpha x} I_{[0, +\infty)}(x).$$

The parameter  $\gamma$  can be interpreted as a shift in the error caused by external factors.

This model is a very special case of the threshold autoregressive model introduced by Tong (1978) and Tong and Lim (1980). Tong (1990) provides an excellent review of classical analysis and properties of threshold models. Geweke and Terui (1993) and Chen and Lee (1995) proposed Bayesian analysis for two-regime threshold models with normal innovations.

The parameter space for the model is

$$\Theta = \{\theta = (\alpha, r, \gamma, \rho) : \alpha > 0, r > 0, \gamma > 0, 0 < \rho < 1\}.$$

Stationarity of the process is guaranteed by the restriction imposed on the parameter  $\rho$ .

In what follows we assume that the first observation  $X_1$  is known and that at least one observation from the sample in each regime exists.

The likelihood function based on  $\mathbf{x} = (x_1, \dots, x_n)$  is

$$L(\theta|\mathbf{x}) = \alpha^{n-1} \exp\left\{-\alpha \left[ \sum_{t=2}^n x_t - \rho \sum_{t=2}^n x_{t-1} - \gamma N(r) \right]\right\} I_U(\theta) \quad (2.2)$$

where

$$N(r) = \sum_{t=2}^n I_{[r,+\infty)}(x_{t-1})$$

and

$$U = \{\theta \in \Theta : x_t - \rho x_{t-1} - \gamma I_{[r,+\infty)}(x_{t-1}) \geq 0, t = 2, \dots, n\}.$$

The likelihood function is a step function in  $r$ , with breaks at the observed  $x_{t-1}$ .

### 3 Bayesian Inference and Prediction

We assume that, a priori, the parameters  $\rho$  and  $r$  are independent, uniformly distributed in  $(0, 1)$  and  $(0, \beta)$ , respectively. For the parameters of the exponential error we assume a conjugate prior independent of  $(\rho, r)$ , of the form

$$(\alpha, \gamma) \sim \text{Exp}(\gamma|\alpha, f) \text{Ga}(\alpha|g, h), \quad f, g, h > 0.$$

This prior implies that the shift in the error has a conditional mean proportional to the mean of the error. Hence the posterior distribution for  $\theta$  is

$$p(\theta|\mathbf{x}) \propto \alpha^{n_1-1} \exp\left\{-\alpha \left[ S_1 - \rho S_2 - \gamma(N^*(r)) \right]\right\} I_{\Theta_n(\mathbf{x})}(\theta), \quad (3.1)$$

where

$$S_1 = \sum_{t=2}^n x_t + h, S_2 = \sum_{t=2}^n x_{t-1}, n_1 = n + g, N^*(r) = N(r) + f,$$

$$\Theta_n(\mathbf{x}) = \{\theta \in \Theta : x_t - \rho x_{t-1} - \gamma I_{[r, +\infty)}(x_{t-1}) \geq 0, \forall t = 2, \dots, n, r \leq \beta^* = \min(x_{(n-1)}, \beta)\},$$

with  $x_{(r)}$  being the  $r$  order statistic.

Summary inferences and prediction can be obtained in the form of posterior expectations

$$E[g(\theta|\mathbf{x})] = \int_{\Theta_n} g(\theta)p(\theta|\mathbf{x})d\theta, \tag{3.2}$$

for suitable choices of  $g(\cdot)$ .

### 3.1 Exact analysis

We have here a constrained parameter problem, since  $\Theta_n(\mathbf{x})$ , the support of the posterior distribution, depends on the data. This poses problems in the computation of quantities such as (??), since it is not always easy to express  $\Theta_n(\mathbf{x})$  in an explicit form, particularly if  $n$  is large.

It is convenient to notice that  $\Theta_n(\mathbf{x})$  can be written as

$$\Theta_n(\mathbf{x}) = (0, +\infty) \times (0, \beta^*) \times M_n(\mathbf{x}),$$

where

$$M_n(\mathbf{x}) = \{(\gamma, \rho) : \gamma > 0, 0 < \rho < 1, x_t - \rho x_{t-1} - \gamma I_{[r, +\infty)}(x_{t-1}) \geq 0, t = 2, \dots, n\}.$$

$M_n(\mathbf{x})$  is a polyvertex which intersects the coordinated axes at the points  $(\gamma^*, 0)$  and  $(0, \rho^*)$ , where

$$\rho^* = \min_{2 \leq t \leq n} \frac{x_t}{x_{t-1}}$$

and

$$\gamma^* = \min_{\substack{x_{t-1} \geq r, \\ t=2, \dots, n}} x_t,$$

A typical shape for  $M_n(\mathbf{x})$  is represented in figure ??.

In order to carry on with the Bayesian methodology we need to specify the region  $M_n(\mathbf{x})$  explicitly in terms of  $\gamma$  and  $\rho$ , so that the appropriate

Figure 1: Typical shape for  $M_n(\mathbf{x})$ 

integrals can be computed. This requires, in general, a tremendous computational task, since one has to solve simultaneously the  $n - 1$  inequalities which appear in the definition of  $M_n(\mathbf{x})$ . (An algorithm, written in Fortran, which accomplishes this for any given sample is available from the authors). Bayesian inference for the parameters  $\alpha, \gamma, \rho$  is then straightforward.<sup>2</sup>

Let  $nd$  be the number of distinct  $x_t, t = 1, \dots, n-1$  and  $N_k(r)$  the number of observations  $x_t \geq r$  when  $r \in (x_{(k-1)}, x_{(k)}], (x_{(0)} = 0), k = 1, \dots, nd$ . Since

$$\begin{aligned} p(r|\mathbf{x}) &= \int_0^\infty \int_{M_n} p(\theta|\mathbf{x}) d\gamma d\rho d\alpha \\ &\propto \int_0^\infty \int_{M_n} p(\alpha, \gamma, \rho|r, \mathbf{x}) d\gamma d\rho d\alpha, \end{aligned} \quad (3.3)$$

and  $N_k(r)$  is constant in each interval  $(x_{(k-1)}, x_{(k)}]$ , the marginal posterior density for the threshold parameter  $r$  is, in each of these intervals, equal to

$$p_k \propto \int_0^\infty \int_{M_n} \alpha^{n_1-1} \exp\left\{-\alpha \left[S_2 - \rho S_1 - \gamma N_k^*(r)\right]\right\} d\gamma d\rho d\alpha, \quad x_{(k-1)} < r \leq x_{(k)}, \quad (3.4)$$

where  $N_k^*(r) = N_k(r) + f$  and the constant of proportionality is such that  $\sum_{k=1}^{nd} p_k(x_{(k)} - x_{(k-1)}) = 1$ .

An estimate for the threshold parameter based on a quadratic loss function is then,

$$\hat{r}_B = \frac{1}{2} \sum_{k=1}^{nd} p_k [x_{(k)}^2 - x_{(k-1)}^2].$$

Predictive density function for future observations can be computed via Monte Carlo integration based on synthetic random sampling from the posterior density, as in Geweke (1989). However, using the fact that the conditional distribution of  $(X_{n+1}|\mathbf{x}, \alpha, \gamma, \rho, r)$  is a shifted exponential distribution<sup>3</sup> ( $Exp(\alpha, \gamma I_{[r, +\infty)}(x_n) + \rho x_n)$ ) with shift parameter  $\gamma I_{[r, +\infty)}(x_n) + \rho x_n$ ,

<sup>2</sup>In order to avoid heavy notation, we will write, in what follows,  $M_n$  instead of  $M_n(\mathbf{x})$

<sup>3</sup>We say that  $X \sim Exp(a, b)$  if  $X - b \sim Exp(a)$

a Bayesian point predictor for  $X_{n+1}$  is

$$\hat{X}_{n+1} = E[X_{n+1}|\mathbf{x}] = \hat{\alpha}_{1/B} + \hat{\gamma}_B I_{[\hat{r}_B, +\infty)}(x_n) + x_n \hat{\rho}_B,$$

where  $\hat{\alpha}_{1/B}$ ,  $\hat{\gamma}_B$ ,  $\hat{\rho}_B$  and  $\hat{r}_B$  are Bayesian estimates of  $1/\alpha$ ,  $\gamma$ ,  $\rho$  and  $r$ , respectively.

### 3.2 An approximation for $M_n$

The general exact treatment for large values of  $n$  is computationally quite demanding and hence it may be advisable to obtain an approximation  $M_n^*$  for the region  $M_n$ .

Following McCormick and Mathew (1993), it can be shown (see Appendix) that

$$M_n \xrightarrow{a.s.} M, \quad \text{as } n \rightarrow +\infty, \quad (3.5)$$

where

$$M = \{(\gamma, \rho) : 0 < \rho < \rho_0, 0 < \gamma < \frac{\gamma_0}{1 - \rho_0}, \gamma_0(1 - \rho) - (1 - \rho_0)\gamma \geq 0\},$$

and  $\gamma_0$  and  $\rho_0$  are the true parameter values.

Hence, from (??), it can be shown that

$$\begin{aligned} \gamma^* &\xrightarrow{a.s.} \frac{\gamma_0}{1 - \rho_0}, \\ \rho^* &\xrightarrow{a.s.} \rho_0, \\ (\hat{\gamma}, \hat{\rho}) &\xrightarrow{a.s.} (\gamma_0, \rho_0), \end{aligned}$$

where  $(\hat{\gamma}, \hat{\rho})$  is the value of  $(\gamma, \rho)$  that maximizes  $\sum_{t=1}^{n-1} x_t \rho + N(r)\gamma$  subject to  $x_t - I_{[r, +\infty)}(x_{t-1})\gamma - \rho x_{t-1} \geq 0$ ,  $2 \leq t \leq n$ .

Let  $M_n^*$  be the tetragon with vertices  $(0, 0)$ ,  $(\gamma^*, 0)$ ,  $(\hat{\gamma}, \hat{\rho})$  and  $(0, \rho^*)$ . It follows from McCormick and Mathew (1993) that  $M_n^* \subset M_n$ . Simulation studies have shown that, for large  $n$ ,  $M_n^*$  is a good approximation for  $M_n$ . Using  $M_n^*$  instead of  $M_n$ , integrations of the type (??), needed for Bayesian calculations, are then straightforward and approximate Bayesian inference and prediction can easily be performed.

### 3.3 Bayesian Analysis via Gibbs sampling

Geman and Geman (1984) introduced the Gibbs sampler as a way of simulating from high-dimensional complex distributions arising in image restoration. Gelfand and Smith (1990) showed how the algorithm can be used to simulate from posterior distributions, and hence showed how to use it to solve problems in Bayesian Statistics. In this situation, the algorithm is based on the fact that (Besag, 1974), if the joint distribution  $p(\theta|\mathbf{x})$ , where  $\theta = (\theta_1, \dots, \theta_k)$  is positive over its entire domain, then it is uniquely determined by the  $k$  full conditional distributions

$$p(\theta_i|\mathbf{x}, \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k), i = 1, \dots, k.$$

The algorithm is a Markovian updating scheme which requires sampling from these full conditional distributions as follows.

Suppose we are given an arbitrary set of initial values  $\theta^0 = (\theta_1^{(0)}, \dots, \theta_k^{(0)})$ . Then we draw  $\theta_1^{(1)}$  from  $p(\theta_1|\mathbf{x}, \theta_2^{(0)}, \dots, \theta_k^{(0)})$ ,  $\theta_2^{(1)}$  from  $p(\theta_2|\mathbf{x}, \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)})$ ,  $\dots$ ,  $\theta_k^{(1)}$  from  $p(\theta_k|\mathbf{x}, \theta_1^{(1)}, \dots, \theta_{k-1}^{(1)})$ . This completes one iteration of the scheme and a transition from  $\theta^0$  to  $\theta^1 = (\theta_1^{(1)}, \dots, \theta_k^{(1)})$ . Iteration of this cycle of random variate generation from each of the full conditional distributions, produces a sequence  $\theta^0, \dots, \theta^m, \dots$ , which is a realization of a Markov chain with transition probabilities given by (e.g. Smith and Roberts, 1993)

$$\pi(\theta^t, \theta^{t+1}) = \prod_{i=1}^k p(\theta_i^{(t+1)}|\mathbf{x}, \theta_j^{(t)}, j > i, \theta_j^{(t+1)}, j < i). \quad (3.6)$$

It can be shown that, as  $t \rightarrow \infty$ ,  $(\theta_1^{(t)}, \dots, \theta_k^{(t)})$  tends in distribution to a random vector whose joint density is  $p(\theta|\mathbf{x})$ . In particular,  $\theta_i^{(t)}$  tends in distribution to a random quantity whose density is  $p(\theta_i|\mathbf{x})$  and

$$\frac{1}{t} \sum_{i=1}^t g(\theta^i) \xrightarrow{a.s.} E_{\theta|\mathbf{x}}\{g(\theta)\},$$

for any function  $g(\cdot)$ , where  $E_{\theta|\mathbf{x}}\{g(\theta)\}$  represents the expected value of  $g(\theta)$  with respect to the posterior density function  $p(\theta|\mathbf{x})$ . It is very important to notice that  $\theta_i$  is not necessarily an univariate random variate. We can consider groups of random vectors. In order to speed up convergence

McCulloch and Tsay (1994), e.g., suggest to draw groups of parameters jointly which are highly dependent and Gamerman (1997) presents several examples considering linear transformations.

In the model we are considering the parameter  $r$  affects particularly the parameter  $\gamma$  of the error term. Hence, for the problem we have in hand we consider  $k = 3$ , with  $\theta_1 = (r, \gamma)$ ,  $\theta_2 = \alpha$  and  $\theta_3 = \rho$ . The full conditional posterior densities are, respectively

$$p(r, \gamma | \mathbf{x}, \alpha, \rho) = p(\gamma | \mathbf{x}, r, \alpha, \rho) p(r | \mathbf{x}, \alpha, \rho) \quad (3.7)$$

with  $p(r | \mathbf{x}, \alpha, \rho)$  defined by

$$p_k(r | \mathbf{x}, \alpha, \rho) \propto [\alpha N_k^*(r)]^{-1} [\exp\{\alpha N_k^*(r) \gamma_r^*\} - 1] \quad r \in (x_{(k-1)}, x_{(k)}) \quad (3.8)$$

$k = 1, \dots, nd$ , where  $nd$  and  $N_k^*(r)$  are defined in ??,

$$\gamma | \mathbf{x}, r, \alpha, \rho \sim \text{Exp}(\alpha N^*(r), \gamma_r^*),$$

where

$$\gamma_r^* = \min_{x_{t-1} \geq r} (x_t - \rho x_{t-1}), \quad t = 2, \dots, n$$

and  $Y \sim \text{Exp}(a, b)$  means that the distribution of  $Y = \{b - X | X \leq b\}$  is  $\text{Exp}(a)$ , ie, with p.d.f.

$$p(y | a, b) = \frac{a \exp\{-a(b - y)\}}{1 - \exp(-ab)} I_{(0,b)}(y) \quad (3.9)$$

$$\alpha | \mathbf{x}, r, \gamma, \rho \sim \text{Ga}(n_1, S_2 - \rho S_1 - N^*(r)\gamma),$$

i.e.

$$p(\alpha | \mathbf{x}, r, \gamma, \rho) \propto \alpha^{n_1-1} \exp\left\{-\alpha \left[S_2 - \rho S_1 - N^*(r)\gamma\right]\right\} \quad (3.10)$$

$$\rho | \mathbf{x}, r, \gamma, \alpha \sim \text{Exp}(\alpha S_1, \rho_r^*),$$

where

$$\rho_r^* = \min_{t=2, \dots, n} \left[ 1, \frac{x_t - \gamma I_{[r, +\infty)}(x_{t-1})}{x_{t-1}} \right].$$

Full Bayesian inference and prediction is then possible. For instance, the predictive density function of a future observation  $x_{n+1}$  can be computed through

$$p(x_{n+1}|\mathbf{x}) \approx \frac{1}{m} \sum_{i=1}^m p(x_{n+1}|\mathbf{x}, \theta^{(i)}), \quad (3.11)$$

for a conveniently chosen  $m$ .

Also, through Gibbs sampling, one-step-ahead predictor can be calculated as

$$\hat{X}_{n+1} = E[X_{n+1}|\mathbf{x}] \approx \sum_{i=1}^m \left[ \frac{1}{\alpha_i} + \gamma_i I_{[r_i, +\infty)}(x_n) + \rho_i x_n \right]. \quad (3.12)$$

Generalization of Gibbs methodology to more general models, such as

$$X_t = \begin{cases} \rho_1 X_{t-1} + \epsilon_t^{(1)} & X_{t-1} < r; \\ \rho_2 X_{t-1} + \epsilon_t^{(2)} & X_{t-1} \geq r. \end{cases} \quad (3.13)$$

where  $\epsilon_t^{(i)} \sim \text{Exp}(\alpha_i, \gamma_i)$ ,  $i = 1, 2$ , can be done without further difficulties.

#### 4 A simulation study

We illustrate the proposed methodology through a simulation study, considering both Bayesian inference and prediction.

We simulated 5000 samples, of size  $n = 50$  and  $n = 110$ , from the threshold model defined in (2.1), with different values for the parameters. To start with we considered the threshold parameter  $r$  fixed (equal to the known value) and we obtained estimates for the parameters  $\alpha, \gamma$  and  $\rho$ , using least squares, approximate maximum likelihood and Bayesian (exact and approximate) methodology (similar to the study described in ?? and ?? for unknown  $r$ ) developed in Pereira and Amaral Turkman (1995). The result of this simulation study, for two particular sets of parameters, is displayed in tables 1 and 2. For comparison purposes we assumed vague prior knowledge for the parameters.

parameter	$n$	least-square	maximum likelihood	approximate Bayesian	exact Bayesian
$\alpha$	50	0.3600 (0.6343)	0.4008 (0.0584)	0.3943 (0.0575)	0.3953 (0.0575)
$\gamma$	50	1.6402 (1.7114)	1.5668 (0.3260)	1.6609 (0.4811)	1.5559 (0.2445)
$\rho$	50	0.0487 (0.2012)	0.1202 (0.0448)	0.0922 (0.0384)	0.0963 (0.0165)
$\alpha$	110	0.3730 (0.4214)	0.3939 (0.0859)	0.3904 (0.0871)	0.3884 (0.0384)
$\gamma$	110	1.6069 (0.7779)	1.5883 (0.2681)	1.6164 (0.4803)	1.5976 (0.0598)
$\rho$	110	0.0732 (0.1333)	0.1097 (0.0403)	0.0952 (0.0364)	0.0900 (0.0167)

Table 1: Results of a simulation study based on 5000 replications with  $\alpha = 0.38, \gamma = 1.6, \rho = 0.09, r = 3.4$  (standard errors in brackets).

parameter	$n$	least-square	maximum likelihood	approximate Bayesian	exact Bayesian
$\alpha$	50	2.8139 (0.1669)	2.8147 (0.3404)	3.1243 (0.4575)	3.1243 (0.4575)
$\gamma$	50	0.3016 (0.2404)	0.4441 (0.1313)	0.3058 (0.4811)	0.3058 (0.0975)
$\rho$	50	0.7579 (0.0772)	0.8037 (0.0036)	0.7999 (0.0036)	0.7999 (0.0036)
$\alpha$	110	2.9149 (0.1025)	2.8976 (0.1836)	3.0585 (0.3893)	3.0556 (0.3024)
$\gamma$	110	0.2724 (0.2155)	0.4513 (0.1420)	0.3085 (0.1133)	0.3046 (0.0959)
$\rho$	110	0.77602 (0.0538)	0.8025 (0.0098)	0.8003 (0.0174)	0.7999 (0.0018)

Table 2: Results of a simulation study based on 5000 replications with  $\alpha = 3.0, \gamma = 0.3, \rho = 0.80, r = 4.0$  (standard errors in brackets).

These results show, as it was expected, that least-squares estimates are in general poor, with large standard errors, especially when the sample size is small. In table 2 we observe, contrary to what happens in table 1, that the the maximum likelihood estimate of  $\gamma$  is far from the true value.

Actually this happened also in other simulated situations but only for the estimation of this parameter. The unstability of the results may be due to convergence problems of the algorithm used. The Bayesian method offered always very similar situations to those documented on tables 1 and 2.

When the threshold parameter is not known, we suggest the use of the anti-mode as a possible estimate for  $r$  when classical methodology is applied. Tong and Lim (1980) suggest the use of AIC criterion. However, the estimation procedure used for  $r$  is not very relevant since, simulation results also showed, that estimates of the other parameters are not sensitive to reasonable differences in estimates of  $r$ .

A similar simulation study was performed, with the purpose of comparing the different estimation procedures, when  $r$  is unknown. However, since each of the 5000 samples had to be studied separately in order to find the anti-mode, the computational work involved was very heavy and it was not expected that different conclusions, from the ones obtained with fixed  $r$ , would be obtained. Hence, we opted instead to estimate  $r$  from the first simulated sample and use it to estimate, via least square and maximum likelihood methods, the parameters for all the other samples. In tables 3 and 4 we show the results obtained. Gibbs sampling was also used as an alternative method to obtain Bayesian estimates (named here as Gibbs estimates) for the parameters. Gibbs was started using the anti-mode as initial value for  $r$  and least square estimates as initial values for the other parameters. We run a single chain of size 31000, rejecting the first 21000 iterations and using the following  $m = 500$  iterations spaced by 20.

parameter	$n$	least-square	maximum likelihood	approximate Bayesian	exact Bayesian	Gibbs
$r$	50	3.566	3.566	3.8659 (0.5702)	3.8822 (0.5573)	3.9546 (0.8275)
$\alpha$	50	0.3227 (0.6249)	0.3563 (0.0592)	0.3484 (0.0587)	0.3492 (0.0588)	0.3489 (0.0585)
$\gamma$	50	1.3244 (1.2033)	1.5889 (0.3060)	1.6029 (0.3592)	1.5371 (0.3060)	1.5627 (0.3488)
$\rho$	50	0.0464 (0.2014)	0.1010 (0.0411)	0.0742 (0.0368)	0.0778 (0.0354)	0.0772 (0.0372)
$r$	110	3.666	3.666	3.4302 (0.6187)	3.4264 (0.1207)	3.4237 (0.1229)
$\alpha$	110	0.3535 (0.4225)	0.3599 (0.0439)	0.3635 (0.0365)	0.3642 (0.0383)	0.3647 (0.0380)
$\gamma$	110	1.3359 (0.7830)	1.5889 (0.2485)	1.6194 (0.2507)	1.6163 (0.12638)	1.6143 (0.1290)
$\rho$	110	0.2368 (0.1349)	0.1010 (0.0368)	0.0846 (0.0268)	0.0860 (0.0165)	0.0867 (0.0178)

Table 3: Estimates for the parameters of the threshold model with  $\alpha = 0.38$ ,  $\gamma = 1.6$ ,  $\rho = 0.09$  and  $r = 3.5$  (standard errors in brackets).

parameter	$n$	least-square	maximum likelihood	approximate Bayesian	exact Bayesian	Gibbs
$r$	50	3.1600	3.1600	2.9742 (0.1736)	2.9736 (0.1792)	3.0525 (0.1406)
$\alpha$	50	0.9281 (0.2716)	1.0364 (0.1479)	1.0298 (0.1547)	1.0310 (0.1540)	1.0329 (0.1492)
$\gamma$	50	1.5136 (0.6603)	1.5962 (0.1764)	1.4321 (0.2156)	1.4484 (0.2176)	1.4813 (0.1321)
$\rho$	50	0.2519 (0.1517)	0.3162 (0.0166)	0.3007 (0.0217)	0.2994 (0.0206)	0.2996 (0.0192)
$r$	110	3.1400	3.1400	2.9919 (0.2380)	2.9901 (0.0465)	3.0184 (0.0514)
$\alpha$	110	0.9715 (0.1766)	1.0329 (0.1162)	0.9933 (0.2239)	1.0159 (0.0969)	1.0170 (0.0943)
$\gamma$	110	1.4822 (0.4859)	1.5629 (0.17672)	1.6239 (0.1069)	1.4469 (0.0620)	1.5001 (0.0589)
$\rho$	110	0.2816 (0.1074)	0.3146 (0.0317)	0.2947 (0.0423)	0.29990 (0.0094)	0.2999 (0.0086)

Table 4: Estimates for the parameters of the threshold model with  $\alpha = 1.0$ ,  $\gamma = 1.5$ ,  $\rho = 0.3$  and  $r = 3.0$  (standard errors in brackets).

As it can be seen, the estimates using Gibbs sampling are quite good. Gibbs sampling is easy to implement and is a very good alternative to the exact approach, with the advantage of being able to simultaneously handle the prediction problem.

In Fig 2 we represent graphically  $p(x_{n+1}|\mathbf{x})$ , defined in (??), for a single simulation of the model with  $\alpha = 0.38$ ,  $\gamma = 1.6$ ,  $\rho = 0.09$  and  $r = 3.5$ , using the first  $n = 100$  observations. For the corresponding simulation data we have  $x_n = 7.2 > r$ . Using nonlinear regression analysis we fitted an exponential model to the points obtained. For the particular situation the fitted model is

$$p(x_{n+1}|\mathbf{x}) \approx 0.3856 \exp\{-0.3856(x_{n+1} - 2.1786)\} \quad (4.1)$$

Figure 2: Predictive density function  $p(x_{n+1}|\mathbf{x})$ . Dots represent points obtained using Gibbs sampling.

In table 5 we compare the performance of the different methods with respect to the one-step-ahead prediction problem, using the mean square error criterion,

$$RMSE = \left\{ \frac{1}{t^*} \sum_{i=t}^{t+t^*} (\hat{x}_{i+1} - x_{i+1})^2 \right\}^{1/2},$$

where  $t^*$  is the number of predictions,  $\hat{x}_{i+1}$  is the one-step ahead predictor given by the data up to time  $i$ . Prediction performance was investigated for  $t^* = 20$ , with an initial sample size of 90 observations.

RMSE least square	RMSE maximum likelihood	RMSE exact Bayesian	RMSE approximate Bayesian	RMSE Gibbs
1.4748	1.5011	1.0577	1.2279	1.0149
2.7068	2.5918	2.2030	2.3468	2.2180

Table 5: Mean square errors for one-step-ahead predictions (using two different samples).

We observe that, according to this criterion, Bayesian predictors behave better than classical linear predictors.

## 5 Example

Here we analyse a set of pollution data from a river in a site near a bridge, which is close to a paper factory. Figure ?? suggests that data values have different behaviour depending on whether they are above or below a threshold value.

*Figure 3: Pollution data of dissolved oxygen*

In fact there are some factors which may affect these values, namely when another paper factory (located some kilometers above) is working, the river flow is low, etc. So applied to these data the model suggested in section 2. We obtained, using the Bayesian approach the following estimates for the parameters  $\gamma = 0.1604$ ,  $\rho = 0.0968$ ,  $r = 3.4444$  and  $\alpha = 0.3775$ .

## 6 Concluding remarks

We have proposed here a Bayesian analysis of a threshold autoregressive model with exponential errors. The model we assumed is quite simplistic, but it served to show the difficulties associated with the problem. The results obtained also show that Gibbs sampler offers an attractive alternative to the other methods studied, keeping the easiness of the treatment even for a more complicated model. Bayesian analysis based on the approximated region has the advantage of allowing a full analysis of the model, since analytic computation of estimates and credible regions for the parameters, and Bayesian predictive regions for the possible outcomes of future observations, is then possible. However, the ideas behind the approximation used are not easily extended to the analysis of more complicated models.

Computer programs (written in Fortran) are available from the authors on request.

## Appendix

In order to prove that

$$M_n \xrightarrow{a.s.} M \quad \text{as } n \rightarrow +\infty, \quad (6.1)$$

where

$$M = \{(\gamma, \rho) : 0 < \rho < \rho_0, 0 < \gamma < \frac{\gamma_0}{1 - \rho_0}, \gamma_0(1 - \rho) - (1 - \rho_0)\gamma \geq 0\},$$

with  $\gamma_0$  and  $\rho_0$  the true parameter values, we follow the proof of theorem 2.1 given in McCormick and Mathew (1993).

First, it is immediate that

$$\min_{\substack{2 \leq t \leq n, \\ x_{t-1} \geq r}} X_t \xrightarrow{a.s.} \frac{\gamma}{1 - \rho} \quad \text{as } n \rightarrow +\infty,$$

since it corresponds to the model  $X_t = \gamma + \rho X_{t-1} + \epsilon_t$  studied by the authors.

Next we have to prove that

$$\min_{2 \leq t \leq n} \frac{X_t}{X_{t-1}} \xrightarrow{a.s.} \rho, \quad \text{as } n \rightarrow +\infty. \quad (6.2)$$

Consider the models

$$\begin{aligned} X_n^{(1)} &= \rho X_{n-1}^{(1)} + \epsilon_n, \\ X_n^{(2)} &= \gamma I_{[r, +\infty)}(X_{n-1}^{(2)}) + \rho X_{n-1}^{(2)} + \epsilon_n, \\ X_n^{(3)} &= \gamma + \rho X_{n-1}^{(3)} + \epsilon_n, \end{aligned} \quad (6.3)$$

for  $n \geq 2$ , where  $\epsilon_n \sim Ex(\alpha)$  and the initial value  $X_1$  is the same for the three models.

We have that

$$X_n^{(1)} = \rho^{n-1} X_1 + \sum_{i=0}^{n-2} \rho^i \epsilon_{n-i}$$

and

$$X_n^{(3)} = \gamma \frac{1 - \rho^{n-1}}{1 - \rho} + \rho^{n-1} X_1 + \sum_{i=0}^{n-2} \rho^i \epsilon_{n-i} = \gamma \frac{1 - \rho^{n-1}}{1 - \rho} + X_n^{(1)}.$$

For any  $\delta > 0$

$$P \left[ \min_{n \geq 2} \frac{X_n^{(2)}}{X_{n-1}^{(2)}} > \delta \right] < P \left[ \min_{n \geq 2} \frac{X_n^{(3)}}{X_{n-1}^{(3)}} > \delta \right]. \quad (6.4)$$

Let us define

$$M_n^{(2)} = \min_{n \geq 2} \frac{\epsilon_n + \gamma I_{[r, +\infty)}(X_{n-1}^{(2)})}{X_{n-1}^{(2)}}$$

and

$$M_n^{(1)} = \min_{n \geq 2} \frac{\epsilon_n + \frac{\gamma(1 - \rho^{n-1})}{1 - \rho}}{X_{n-1}^{(1)}}.$$

It is easy to see that

$$\min_{n \geq 2} \frac{X_n^{(2)}}{X_{n-1}^{(2)}} = \rho + M_n^{(2)}$$

and

$$\min_{n \geq 2} \frac{X_n^{(3)}}{X_{n-1}^{(3)}} = \rho + M_n^{(1)}.$$

Hence, from (??) we have

$$P[M_n^{(2)} > \delta] < P[M_n^{(1)} > \delta]. \quad (6.5)$$

Therefore, developing (??) and using the fact that  $X_{j-1} \geq \epsilon_{j-1}$ ,

$$\begin{aligned} P[M_n^{(2)} > \delta] &< P \left[ \frac{\epsilon_j + \frac{\gamma(1 - \rho^{j-1})}{1 - \rho}}{X_{j-1}^{(1)}} > \delta, 2 \leq j \leq n \right] \\ &\leq P \left[ (1 - \rho)\epsilon_j + \gamma > (1 - \rho)\delta\epsilon_{j-1}, 2 \leq j \leq n \right] \\ &< P \left[ (1 - \rho)\epsilon_{2j} + \gamma > (1 - \rho)\delta\epsilon_{2j-1}, 1 \leq j \leq [n/2] \right] \\ &= \left\{ P \left[ (1 - \rho)\epsilon_2 + \gamma > (1 - \rho)\delta\epsilon_1 \right] \right\}^{[n/2]} \end{aligned}$$

where  $[n/2]$  is the greatest integer less than or equal to  $n/2$ .

Since

$$P\left[(1-\rho)\epsilon_2 + \gamma > (1-\rho)\delta\epsilon_1\right] = 1 - \frac{\delta}{1+\delta} \exp\left\{-\frac{\alpha\gamma}{\delta(1-\rho)}\right\} < 1,$$

then

$$\sum_{n=2}^{+\infty} P[M_n^{(2)} > \delta] \leq \sum_{n=2}^{+\infty} \left\{ P[(1-\rho)\epsilon_2 + \gamma > (1-\rho)\delta\epsilon_1] \right\}^{[n/2]} < +\infty.$$

Hence, by Borel-Cantelli lemma, for all  $\delta$  sufficiently small, the events  $\{M_n^{(2)}\}$  can occur only finitely often. Consequently

$$M_n^{(2)} \xrightarrow{a.s.} 0 \quad n \rightarrow +\infty,$$

which is equivalent to (??) as desired.

## References

- Andel, J. (1988). On Ar(1) processes with exponential white noise. *Communications in Statistics-Theory and Methods*, **17**, 1481-1495.
- Andel, J. and Garrido, M. (1991). Bayesian analysis of non-negative AR(2) processes. *Statistics*, **22**, 579-588.
- Bell, C. B. and Smith E. P. (1986). Inference for non-negative autoregressive schemes. *Communications in Statistics-Theory and Methods*, **15**, 2267-2293.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society B*, **36**, 192-236 .
- Congdon, P. (2001) *Bayesian Statistical Modelling*. Wiley, New York.
- Gaver, D. P. and Lewis, P. A. W. (1980). First-order autoregressive gamma sequences and point processes. *Advances in Applied Probability*, **12**, 727-745.
- Chen, C. W. S. and Lee, J. C. (1995). Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis*, **16**, 483-492.
- Davis, R. and McCormick, W.P. (1989). Estimation for first-order autoregressive processes with positive or bounded innovations. *Stochastic Processes and Their Applications*, **31**, 237-250.
- Gamerman, D. (1997) *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall, London.

- Geisser, S. (1971). The inferential use of predictive distributions. *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.) Holt, Rinehart and Winston, Toronto, 459-469.
- Gelfand, A. E., Smith, A. F. M. and Lee, T., (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association*, **87**, 523-531.
- Gelfand, A. E. and Smith, A. F. M., (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398-409.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **6**, 712-741.
- Geweke, J. (1989). Exact predictive densities for linear models with ARCH disturbances. *Journal of Econometrics*, **40**, 63-86.
- Geweke, J. and Terui, N. (1993). Bayesian autoregressive models for non-linear time series. *Journal of Time Series Analysis*, **14**, 441-454.
- Lawrance, A. J. and Lewis, P. A. (1985). Modelling and residual analysis of non-linear autoregressive time series in exponential variables. *Journal of the Royal Statistical Society B*, **47**, 165-202.
- McCormick, W. P. and Mathew, G. (1993). Estimation for nonnegative autoregressive processes with an unknown location parameter. *Journal of Time Series Analysis*, **14**, 71-92.
- McCulloch, R. E. and Tsay, R. S. (1994). Bayesian Analysis of autoregressive time series via Gibbs sampler. *Journal of Time Series Analysis*, **15**, 235-250.
- Pereira, I. M. S. and Amaral Turkman, M. A. (1995). A Bayesian analysis of piecewise autoregressive model with exponential errors. *Proceedings of the II Conference of the Portuguese Statistical Society*, 305-318.
- Smith, A. F. M. and Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods (with discussion). *Journal of the Royal Statistical Society B*, **55**, 3-23.
- Tong, H. (1978). On a threshold model. *Pattern Recognition and Signal Processing* (C. H. Chen, ed.) Sijthoff and Noordoff, Amsterdam.
- Tong, H. (1983). Threshold models in non-linear time series analysis. *Lecture Notes in Statistics*, **21**. Springer, Heidelberg.
- Tong, H. (1990) *Non-linear time series model: a dynamical system approach*. Oxford University Press, London.
- Tong, H. and Lim, K. S. (1980). Threshold autoregressive, limit cycles and cyclical

- data (with discussion). *Journal of the Royal Statistical Society* **B**, **42**, 245-292.
- Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, **84**, 231-240.