Communications in Statistics - Simulation and Computation

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lssp20

Using the Reversible Jump MCMC Procedure for Identifying and Estimating Univariate TAR Models

Fabio H. Nieto [a] , Hanwen Zhang [b] & Wen Li [c]

[a] Departamento de Estadística , Universidad Nacional de Colombia , Bogotá, D.C. , Colombia

[b] Facultad de Estadística , Universidad Santo Tomás , Bogotá, D.C. , Colombia

[c] Singapore Clinical Research Institute , Singapore

Published online: 21 Dec 2012.

To cite this article: Fabio H. Nieto , Hanwen Zhang & Wen Li (2013) Using the Reversible Jump MCMC Procedure for Identifying and Estimating Univariate TAR Models, Communications in Statistics - Simulation and Computation, 42:4, 814-840, DOI: 10.1080/03610918.2012.655827

To link to this article: http://dx.doi.org/10.1080/03610918.2012.655827

Communications in Statistics—Simulation and Computation®, 42: 814–840, 2013

ISSN: 0361-0918 print / 1532-4141 online
DOI: 10.1080/03610918.2012.655827

Using the Reversible Jump MCMC Procedure for Identifying and Estimating Univariate TAR Models

FABIO H. NIETO,[1] HANWEN ZHANG,[2] AND WEN LI[3]

[1]Departamento de Estadística, Universidad Nacional de Colombia, Bogotá, D.C., Colombia
[2]Facultad de Estadística, Universidad Santo Tomás, Bogotá, D.C., Colombia
[3]Singapore Clinical Research Institute, Singapore

One way that has been used for identifying and estimating threshold autoregressive (TAR) models for nonlinear time series follows the Markov chain Monte Carlo (MCMC) approach via the Gibbs sampler. This route has major computational difficulties, specifically, in getting convergence to the parameter distributions. In this article, a new procedure for identifying a TAR model and for estimating its parameters is developed by following the reversible jump MCMC procedure. It is found that the proposed procedure conveys a Markov chain with convergence properties.

Keywords Bayesian model choice; Nonlinear time series; Regime-switching models; RJMCMC; Threshold autoregressive (TAR) models

Mathematics Subject Classification Primary 62; Secondary 60

1. Introduction

Nieto (2005) developed a procedure for identifying univariate threshold autoregressive (TAR) models, which are nonlinear, based on that of Carlin and Chib (1995) for Bayesian model choice. As quoted by these authors, the so-called *link* distributions are crucial for obtaining efficient mixing in the Gibbs sampler defined in their article. This aspect was noted by Nieto (2005) and he found major difficulties in implementing the method in practice, especially, because of the very small values of the model likelihood function, a fact which signals that the likelihood present in the data is not being extracted optimally.

In order to have an alternative to Nieto's (2005) approach, we present in this article a procedure based on Green's (1995) RJMCMC (reversible jump Markov chain Monte Carlo) sampler. Some authors have addressed problems that are close to the one presented in this article, for example, Campbell (2004) and Vermaak et al. (2004). The first work developed an RJMCMC sampler in the context of the so-called SETAR (self-exciting threshold autoregressive) models, where, a priori, the number of regimes is assumed to be known and the focus is basically on estimating the autoregressive (AR) orders. In the second work, linear AR processes are considered and the interest is to estimate the unknown model order. For choosing different types of time series models, Vrontos et al. (2000) used

Received December 2, 2010; Accepted December 28, 2011
Address correspondence to Fabio H. Nieto, Departamento de Estadística, Universidad Nacional de Colombia, Carrera 30 No. 45-03, Bogotá, Colombia; E-mail: fhnietos@unal.edu.co

the RJMCMC procedure for making full Bayesian inference of generalized autoregressive conditional heteroscedasticity (GARCH) and exponential GARCH (EGARCH) models.

Our main concern in this article is to identify (estimate) the structural parameters, that is, the number of regimes and the AR orders in each regime and, simultaneously, to estimate the nonstructural parameters of the TAR model. A crucial complexity in the design of the RJMCMC sampler is implied by jumps in *both* the number of regimes and the AR orders in the regimes, a situation not addressed in literature. To handle this problem, we propose a two-step procedure, which consists of, first, to identify the number of regimes and, second, conditional on this parameter, to estimate the remaining model parameters.

In Sec. 2, we present a brief summary of the TAR model and the basic steps of the RJMCMC sampler. The two-step algorithm developed for the TAR model is presented in Sec. 3 and in Sec. 4, we include a procedure for checking the convergence of the proposed sampler. Two simulations and one real-data application are provided in Sec. 5. Finally, main conclusions and recommendations for future research are given in Sec. 6.

2. Theoretical Background

2.1. *TAR Models*

Let $\{X_t\}$ and $\{Z_t\}$ be stochastic processes defined on the same probability space and related by the equation (TAR model)

$$X_t \;=\; a_0^{(j)} + \textstyle\sum_{i=1}^{k_j} a_i^{(j)} X_{t-i} + h^{(j)}\varepsilon_t \;, \tag{1}$$

if $Z_t \in B_j = (r_{j-1}, r_j]$, for some j, $j = 1, \ldots, l$, where $r_0 = -\infty$, $r_l = \infty$, and l is a positive integer number. Here, $a_i^{(j)}$ and $h^{(j)}$; $j = 1, \ldots, l; i = 0, 1, \ldots, k_j$; are real numbers and $\{\varepsilon_t\}$ is a Gaussian zero-mean white-noise process with variance 1. The real numbers r_j, $j = 1, \ldots, l-1$, are called the threshold values of the process $\{Z_t\}$ and they define l *regimes* for the bivariate process $\{X_t, Z_t\}$. Additionally, the nonnegative integer numbers $k_1, \ldots, k_l$ denote, respectively, the AR orders of $\{X_t\}$ in each regime. We use the symbol TAR$(l; k_1, \ldots, k_l)$ to denote this model and call $l, r_1, \ldots, r_{l-1}$, and $k_1, \ldots, k_l$ as the model structural parameters.

This kind of models was introduced by Tong (1978) and Tong and Lim (1980), in which the threshold variable is the lagged variable X_{t-d} for some positive integer d. This specific model is called as the SETAR model and there is a rich literature on how to analyze these models, with the assumption that the number l of regimes and the AR orders $k_1, \ldots, k_l$ are known.

Under the open-loop systems (Tong, 1990), we assume that $\{Z_t\}$ is exogenous in the sense that there is no feedback of $\{X_t\}$ toward it and that $\{Z_t\}$ is a homogeneous pth-order Markov chain with initial distribution $F_0(z, \boldsymbol{\theta}_z)$ and kernel distribution $F_p(z_t|z_{t-1}, \ldots, z_{t-p}, \boldsymbol{\theta}_z)$, where $\boldsymbol{\theta}_z$ is a parameter vector in an appropriate numerical space. Furthermore, we assume that these distributions have densities in the Lebesgue measure sense. Let $f_0(z, \boldsymbol{\theta}_z)$ and $f_p(z_t|z_{t-1}, \ldots, z_{t-p}, \boldsymbol{\theta}_z)$ be, respectively, the initial and kernel density functions of the distributions above. In this article, we assume that the p-dimensional Markov chain $\{Z_t\}$ has an invariant or stationary distribution $f_p(z, \boldsymbol{\theta}_z)$.

Tong (1990) considered open-loop models and under this concept, the TAR model means that the nonlinear dynamical behavior of variable X is explained by the exogenous variable Z, according to its values in the sets of a partition of its sample space. Tsay (1998) developed the multivariate version of the TAR model for complete time series,

with the threshold variable Z_{t-d}, where $d > 1$, while in Nieto's (2005) model $d = 0$ is allowed. Nieto (2005) developed the methodology for analyzing the TAR model in the presence of missing data and a procedure for forecasting variable X with the TAR model (Nieto, 2008). If the time series are complete, Nieto's methods are applicable with minor modifications. The typical TAR model can explain certain stylized facts of financial and hydrological/meteorological time series, for example, the presence of extreme-value clusters. The model can also explain the regime-switching characteristic of Markov chains, with the advantage that the state space can be general, that is, not necessarily discrete.

2.2. *The RJMCMC Procedure*

This section refers to Chen et al. (2000). For a more specific presentation, the reader can consult Green (1995).

Suppose that we have a countable collection of candidate models $\{\mathcal{M}_k | k \in M\}$, where $M \subseteq \mathbb{N}$, the set of natural numbers. Model $\mathcal{M}_k$ is parameterized by the unknown vector $\boldsymbol{\theta}^{(k)}$ of dimension p_k, which may vary from model to model. Under model $\mathcal{M}_k$, the posterior distribution of $\boldsymbol{\theta}^{(k)}$ takes the form

$$\pi(\boldsymbol{\theta}^{(k)}|\mathcal{D}, \mathcal{M}_k) \propto \pi^*(\boldsymbol{\theta}^{(k)}|\mathcal{D}, \mathcal{M}_k)p(k)/\,p(\mathcal{M}_k|\mathcal{D}),$$

with $\pi^*(\boldsymbol{\theta}^{(k)}|\mathcal{D}, \mathcal{M}_k) = L(\boldsymbol{\theta}^{(k)}, \mathcal{M}_k|\mathcal{D})\pi(\boldsymbol{\theta}^{(k)}|\mathcal{M}_k)$, where $\mathcal{D}$ denotes the dataset, $L(\boldsymbol{\theta}^{(k)}, \mathcal{M}_k|\mathcal{D})$ is the likelihood function, $\pi(\boldsymbol{\theta}^{(k)}|\mathcal{M}_k)$ is the prior distribution for the parameter vector under model $\mathcal{M}_k$, and $p(k)$ is a prior distribution for k (or $\mathcal{M}_k$). Note that $\pi^*(\boldsymbol{\theta}^{(k)}|\mathcal{D}, \mathcal{M}_k)$ is the unnormalized posterior density, given $\mathcal{M}_k$. Then, the joint distribution of $(k, \boldsymbol{\theta}^{(k)})$ given the data $\mathcal{D}$, the distribution of interest, is given by

$$\pi(k, \boldsymbol{\theta}^{(k)}|\mathcal{D}) = \pi(\boldsymbol{\theta}^{(k)}|\mathcal{D}, \mathcal{M}_k)p(\mathcal{M}_k|\mathcal{D}) \propto \pi^*(\boldsymbol{\theta}^{(k)}|\mathcal{D}, \mathcal{M}_k)p(k).$$

The idea is to obtain joint samples for the model number k and the parameter vector $\boldsymbol{\theta}^{(k)}$ from this joint posterior distribution. Note that, strictly speaking, k is a realization of a discrete random variable K, whose sample space is M.

The main property of the RJMCMC algorithm is that the underlying Markov chain is designed in a way that it can *jump* between models with different dimensional parameter spaces, while retaining a detailed balance that ensures the correct limiting distribution. This last property is accomplished if the chain is irreducible and aperiodic.

The Algorithm

Let us assume that the current state of the chain is $(k, \boldsymbol{\theta}^{(k)})$, then:

Step 1. Propose a new model $\mathcal{M}_{k^*}$ with jump probability $j(k^*|k)$ (this means, given k, one needs to put a probability distribution for choosing moves among the models).

Step 2. Generate $\mathbf{u}$, of certain dimension, from a specified proposal density $q_k(\mathbf{u}|\boldsymbol{\theta}^{(k)}, k, k^*)$.

Step 3. Set $(\boldsymbol{\theta}^{(k^*)}, \mathbf{u}^*) = g_{k,k^*}(\boldsymbol{\theta}^{(k)}, \mathbf{u})$, where g_{k,k^*} is a bijective transformation between Euclidian subspaces of dimension $p_k + \dim(\mathbf{u}) = p_{k^*} + \dim(\mathbf{u}^*)$ (thus, one needs some kind of dimension matching and for that purpose one uses the proposal distribution $q_k(\cdot|\cdot)$).

Step 4. Let

$$r_n = p(k^*)\pi^*(\boldsymbol{\theta}^{(k^*)}|\mathcal{D}, \mathcal{M}_{k^*})j(k^*|k)q_{k^*}(\mathbf{u}^*|\boldsymbol{\theta}^{(k^*)}, k^*, k),$$
$$r_d = p(k)\pi^*(\boldsymbol{\theta}^{(k)}|\mathcal{D}, \mathcal{M}_k)j(k|k^*)q_k(\mathbf{u}|\boldsymbol{\theta}^{(k)}, k, k^*),$$

and J be the Jacobian of the transformation g_{k,k^*}. Accept the proposed move to $(k^*, \boldsymbol{\theta}^{(k^*)})$ with probability $\alpha = \min\{1, r_n J / r_d\}$.

As quoted by Green (1995), the probability distribution for the moves among models, the q distributions, the g functions, etc., strongly depend on the data. The idea is to assure that the defined chain is aperiodic and irreducible.

3. Identifying and Estimating the TAR Model

Following Green's (1995) example about estimating a step function in a multidimensional change-point problem, our interest is on the posterior distribution of the random vector $(l, \mathbf{k}_l, \boldsymbol{\theta}_l)$, where $l = 2, \ldots, l_0$, for some known l_0, $\mathbf{k}_l = (k_{1l}, \ldots, k_{ll})$, and $\boldsymbol{\theta}_l = (\theta_{1l}, \ldots, \theta_{ll})$, with $\theta_{jl} = (a_0^{(j)}, a_1^{(j)}, \ldots, a_{k_{jl}}^{(j)}, h^{(j)})$, for $j = 1, \ldots, l$. Note that $\boldsymbol{\theta}_l$ has a dimension of $k_{1l} + \cdots + k_{ll} + 2l$; if l or k_{jl} changes, for some $j = 1, \ldots, l$, this dimension changes and, consequently, the dimension of the whole vector $(l, \mathbf{k}_l, \boldsymbol{\theta}_l)$ changes. This is our motivation for proposing to use the RJMCMC procedure.

In this article, we consider a modification of the usual RJMCMC sampler, which consists of a two-step approach. In the first step, we design an RJMCMC sampler for the whole parameter vector $(l, \mathbf{k}_l, \boldsymbol{\theta}_l)$, and in the second step, we fix the estimated value of l, $\hat{l}$ say, and reestimate the remaining parameters $(\mathbf{k}_{\hat{l}}, \boldsymbol{\theta}_{\hat{l}})$ via a simplification of the first sampler, which coincides with Campbell's (2004) approach. From this point of view, our approach is an extension of Campbell's (2004) procedure. Our two-step approach was motivated by the fact that there is a double nested dimension movement among models, that is, a move for l and then, a move for $\mathbf{k}_l$, a problem not considered in the current literature to our knowledge. This double nested movement can cause a low mixing for l and even for $\mathbf{k}_l$, as will be seen in the examples below. Consequently, an important problem for future research would be to design an RJMCMC sampler for the whole parameter vector, as one would desire at the first place, that takes into account this complex characteristic.

In the RJMCMC sampler of $(l, \mathbf{k}_l, \boldsymbol{\theta}_l)$, initially we consider three types of transitions in the underlying Markov chain to be developed. They are (1) a change in the number of regimes l (R), (2) a change in some AR order k_{jl}, $j = 1, \ldots, l$, (O), and (3) a change in the total nonstructural parameter vector $\boldsymbol{\theta}_l$ (V). When transition R or O happens, we also need to consider *intrinsic* changes in some of the subvectors of $\boldsymbol{\theta}_l$.

Following Green's (1995) article, the changes in the number of regimes are circumscribed as birth and death moves. That is, a birth, if the number of regimes moves from l to $l + 1$ and a death, if there is a move from $l + 1$ to l. For these moves, a subprobability distribution is needed, which is described as follows.

Let $b_l = \Pr(\text{Birth}) = c \min_l\{1, p(l+1)/p(l)\}$ and $d_{l+1} = \Pr(\text{Death}) = c \min_l\{1, p(l)/p(l+1)\}$, where $p(\cdot)$ denotes a prior distribution for L and c is a constant such that $b_l + d_l \leq 0.9$ for all $l = 2, \ldots, l_0$. For the boundaries, we put $d_2 = b_{l_0} = 0$. At the $(i-1)$th iteration, the above definitions give probabilities for the regime move types, conditioned at the number of regimes for that iteration, $l^{(i-1)}$ say, c is a constant such that $c \max_l\{b'_l + d'_l\} \leq 0.9$, where $b'_l = \min\{1, p(l+1)/p(l)\}$ and $d'_l = \min\{1, p(l-1)/p(l)\}$. In this way, $c \leq 0.9/\max_l\{b'_l + d'_l\}$. Note that given $p(\cdot)$ and the heuristic value 0.9, the upper bound above for c is computable.

Because of the birth/death moves described above for the number of regimes, it is convenient to say that R splits into the moves B (birth) and D (death), that is, R = (B,D), and thus, the full set of transitions for the TAR model is {B, D, O, V}. To choose among these

transitions, we design a probability distribution $q_l = (b_l, d_l, o_l, v_l)$, where $o_l = \Pr(O|l)$ and $v_l = \Pr(V|l)$, and $b_l + d_l + o_l + v_l = 1$. We can put $o_l = v_l$. In summary, so far, we have described the move (discrete) set and a probability mass function on it.

Now, we describe the mathematical form of each one of the jump and proposal ratios, in order to compute the acceptance probabilities. We keep in mind that the chain is going to move from the $(i-1)$th iteration to the ith one. As mentioned previously, a key aspect of this article is the consideration of intrinsic AR-order moves. To give a general idea of the proposed algorithm, we assume that move B is chosen. The number of regimes then satisfies $l^{(i)} = l^{(i-1)} + 1$. For ease of notation, we shall write l instead of $l^{(i)}$, where necessary. To allow for the updates of an AR order k_l and the total nonstructural parameter vector $\boldsymbol{\theta}_l$, from $l^{(i-1)}$ regimes to $l^{(i)}$ regimes, we denote the temporal updated value of k_l and $\boldsymbol{\theta}_l$ as $k_l^{(i^*)}$ and $\boldsymbol{\theta}_l^{(i^*)}$, and set $k_l^{(i^*)} = k_l^{(m)}$ and $\boldsymbol{\theta}_l^{(i^*)} = \boldsymbol{\theta}_l^{(m)}$, where m is the latest iteration among all the previous iterations such that $l^{(m)} = l^{(i)}$. If no such m exists, we put the corresponding priors suggested by Nieto (2005) for each AR order and nonstructural parameter. Next, we do a death/birth move to update the AR order $k_{jl}^{(i^*)}$ in the regime j, for all $j = 1, \ldots, l^{(i)}$. We denote the intrinsic birth and death probabilities as b_k and d_k, respectively, which are defined similarly to b_l and d_l. In the update of AR orders, the probability of a *remain* move is $1 - b_k - d_k$. Note that these intrinsic probabilities are embedded in each regime. If the intrinsic move in a regime is a birth, we update the corresponding AR coefficient using a proposal density $q(\cdot|\boldsymbol{\theta}_j^{(i^*)})$ as given by Campbell (2004). For an intrinsic death move, we simply delete the last AR coefficient and the proposal density of this deletion is 1, as Campbell (2004) did in his approach. In summary, the product of the jump and the proposal ratios become the following:

(1) For transition B, the proposal ratio is $(b_l/d_{l+1}) \times \prod_{j=1}^{l} r_j^{\mathrm{AR}} \times \prod_{j=1}^{l} r_j^{P}$, where r_j^{AR}, $j = 1, \ldots, l$, denotes, in turn, a ratio of jump probabilities for doing a move in the AR order of regime j. That is, if we get a birth, the associated ratio is b_k/d_{k+1}; if a death happens, the ratio is d_k/b_{k-1}; and if the move is remain, the ratio is 1. Here, r_j^P, $j = 1, \ldots, l$, denotes a ratio of proposal densities and its value is $1/q$ if there is a birth in an intrinsic AR-order move and is 1 otherwise.
(2) For transition D, we have a similar product of the jump and the proposal ratios and the only difference is to replace d_l/b_{l-1} with b_l/d_{l+1} in the above expression.
(3) For transition O, the ratio becomes $\prod_{j=1}^{l} r_j^{\mathrm{AR}} \times \prod_{j=1}^{l} r_j^{P}$. Note that this scheme is contrary to Campbell's (2004) procedure.
(4) If transition V is chosen, the ratio is trivially 1.

To end the design of our RJMCMC sampler, we need to compute the Jacobian for transitions B, R, and O. In the first two, the Jacobian is only needed when an *intrinsic* move in the AR orders is birth. The same situation happens if the result of O is birth. It is easy to show that the Jacobian is always 1. In the Appendix, we present the synthesized algorithm that can be implemented for fitting TAR models.

We remark that in the case of remain moves in the AR orders, we can do Gibbs sampling to update the nonstructural parameters in the corresponding regime, using Nieto's (2005, sec. 3.2) results. Another alternative to this updating of nonstructural parameters might use a single-site Metropolis–Hastings random walk algorithm, as in Campbell (2004). It is also important to note here that, although our goal is the whole vector $(l, \mathbf{k}_l, \boldsymbol{\theta}_l)$, given l in an iteration, we *do* intrinsic estimates for the AR orders and the nonstructural parameters, in a manner similar to the Akaike information criterion (AIC) used for identifying a model. This fact is our motivation for the proposed two-step approach.

Once l is estimated, the first step is done. The second step in our approach focuses on the estimates of AR orders and nonstructural parameters in all the regimes, and can be thought as a simplified version of the first step, that is, just containing the transition O and the transition V. Similarly, the product of the jump ratios and the proposal ratios can be calculated as 3 for the transition O and 4 for the transition V in the first step. In this way, our two-step approach extends Campbell's (2004) algorithm.

We now describe how q is computed. Following Campbell's (2004, p. 472) article, in the case of an AR-order birth, from k to $k+1$ say, we propose to use a univariate normal distribution from which the AR coefficient $a_{k+1}^{(j)}$ is drawn, corresponding to the chosen regime j in the model with $l^{(i)}$ regimes. The proposed mean of q is

$$\hat{a}_{k+1} = \sum_{t=k+2}^{T} x_{t-k-1} e_t \Bigg/ \sum_{t=k+2}^{T} x_{t-k-1}^2 ,$$

where $e_t = x_t - \sum_{s=1}^{k} a_s^{(i)} x_{t-s}$, $t = k+2, \ldots, T$, and $i = i^*$ for B and D and $i = i-1$ for O, V. The proposed variance is $\epsilon \mathrm{Var}(a_{k+1}|\mathbf{x}, \mathbf{z}, l^{(i)}, \mathbf{k}_{l^{(i)}})$ for some $\epsilon > 0$, where

$$\mathrm{Var}(a_{k+1}|\mathbf{x}, \mathbf{z}, l^{(i)}, \mathbf{k}_{l^{(i)}}) = \sum_{t=k+2}^{T} e_t^2 \Big/ \left[(T-k-1) \sum_{t=k+2}^{T} x_{t-k-1}^2 \right],$$

with $\mathbf{x}$ and $\mathbf{z}$ denoting the data vectors for variables X and Z, respectively, in the entertained sample period. That is, we seek to set the proposal variance to be a proportion of the marginal posterior variance.

It is important to note that the numerator (denominator) in the product likelihood ratio $\times$ prior ratio is given, in general, for a fixed value of l, by

$$\begin{aligned} L(\mathbf{y}|l, \mathbf{k}_l, \boldsymbol{\theta}_l)\pi(l, \mathbf{k}_l, \boldsymbol{\theta}_l) &= L(\mathbf{y}|l, \mathbf{k}_l, \boldsymbol{\theta}_l)\pi(l)\pi(\mathbf{k}_l|l)\pi(\boldsymbol{\theta}_l|l, \mathbf{k}_l) \\ &= L(\mathbf{y}|l, \mathbf{k}_l, \boldsymbol{\theta}_l)\pi(l) \prod_{j=1}^{l} \pi(k_{jl}|l) \prod_{j=1}^{l} \pi(\theta_{jl}|l, \mathbf{k}_l), \end{aligned}$$

where $L(\cdot|\cdots)$ denotes the likelihood function for the whole parameter vector, which is based on the sample $\mathbf{y} = (\mathbf{x}, \mathbf{z})$, $\pi(l)$ is the unconditional prior for l, and $\pi(\cdot|\cdot)$ denotes appropriate conditional priors. The priors for the nonstructural parameters are obtained from Nieto's (2005) article and the priors for the number of regimes and AR orders are truncated Poisson distributions (see Sec. 5). It is worth noting that

$$L(\mathbf{y}|l, \mathbf{k}_l, \boldsymbol{\theta}_l) = L(\mathbf{x}|\mathbf{z}, l, \mathbf{k}_l, \boldsymbol{\theta}_l) L(\mathbf{z}|l, \mathbf{k}_l, \boldsymbol{\theta}_l),$$

where, given the data $\mathbf{z}$, $L(\mathbf{z}|l, \mathbf{k}_l, \boldsymbol{\theta}_l)$ is constant for all l, $\mathbf{k}_l$, and $\boldsymbol{\theta}_l$. This result is due to the assumption that the process $\{Z_t\}$ is exogenous. In this way, our strict likelihood function becomes $L(\mathbf{x}|\mathbf{z}, l, \mathbf{k}_l, \boldsymbol{\theta}_l)$, which was named *conditional* likelihood function by Nieto (2005).

4. Convergence Diagnostics

Following the unpublished article of Castelloe and Zimmerman (2002) named "Convergence Assessment for Reversible Jump MCMC Samplers" of SAS Institute, Inc., we implement a procedure for checking the convergence of our proposed RJMCMC sampler.

As the parameter of interest, we choose $\lambda = \sum_{j=1}^{l} p_j [h^{(j)}]^2$, where $p_j = \mathrm{P}(Z_t \in B_j)$. The conditional distribution of X_t given $x_{t-1}, \ldots, x_1$ and z_t is Gaussian with mean given by $a_0^{(j)} + \sum_{i=1}^{k_j} a_i^{(j)} x_{t-i}$ and standard deviation $h^{(j)}$ if $z_t \in B_j$, for some $j = 1, \ldots, l$. Nieto (2008) showed that this conditional distribution denoted by $f_j(x_t|x_{t-1}, \ldots, x_1, z_t)$ is a mixture distribution with density function

$$f(x_t|x_{t-1}, \ldots, x_1) = \sum_{j=1}^{l} p_j f_j(x_t|x_{t-1}, \ldots, x_1, z_t \in B_j). \tag{2}$$

We remark that one of the two types of conditioning considered above depends on $x_{t-1}, \ldots, x_1$ and $z_t \in B_j$ and, the other one depends on $x_{t-1}, \ldots, x_1$ only. We call *regime-based conditioning* as the first one and *conditioning* as the second one. From expression (2) and following Nieto's (2008) TAR model characteristics, the conditional variance of X_t, $\sigma^2_{t|t-1}$ say, is given by

$$\sigma^2_{t|t-1} = \sum_{j=1}^{l} p_j [h^{(j)}]^2 + \sum_{j=1}^{l} p_j \mu^2_{j,t} - \left[\sum_{j=1}^{l} p_j \mu_{j,t} \right]^2, \tag{3}$$

where $\mu_{j,t} = a_0^{(j)} + \sum_{i=1}^{k_j} a_i^{(j)} x_{t-i}$. The term $\sum_{j=1}^{l} p_j [h^{(j)}]^2$ is like a "communality" in the conditional variances above because it depends neither on t nor on $x_{t-1}, \ldots, x_1$. Moreover, this quantity is a weighted average of the regime-based conditional variances of the process $\{X_t\}$. Note that $\{\sigma^2_{t|t-1} : t = 2, \ldots\}$ defines the conditional variance function of process $\{X_t\}$, a key element in the GARCH model family.

We use λ as the monitoring parameter and run $C \geq 2$ parallel chains of length $2T$ each and discard the first T iterations. Let M be the number of distinct models, λ^r_{cm} be the value of λ for the rth occurrence of model m in chain c, and R_{cm} be the number of times model m occurred in chain c. According to Castelloe and Zimmerman (2002), the convergence diagnostic is based on the following estimated variations:

$$\hat{V}(\lambda) = \frac{1}{CT - 1} \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} \left(\lambda^r_{cm} - \bar{\lambda}\right)^2, \tag{4}$$

$$W_c(\lambda) = \frac{1}{C(T-1)} \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} \left(\lambda^r_{cm} - \bar{\lambda}_c\right)^2, \tag{5}$$

$$W_m(\lambda) = \frac{1}{CT - M} \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} \left(\lambda^r_{cm} - \bar{\lambda}_m\right)^2, \tag{6}$$

$$W_m W_c(\lambda) = \frac{1}{C(T-M)} \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} \left(\lambda^r_{cm} - \bar{\lambda}_{cm}\right)^2, \tag{7}$$

where

$$\bar{\lambda} = \frac{1}{CT} \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} \lambda^r_{cm},$$

$$\bar{\lambda}_c = \frac{1}{T}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\lambda_{cm}^r,$$

$$\bar{\lambda}_m = \frac{1}{R_{\cdot m}}\sum_{c=1}^{C}\sum_{r=1}^{R_{cm}}\lambda_{cm}^r,$$

$$\bar{\lambda}_{cm} = \frac{1}{R_{cm}}\sum_{r=1}^{R_{cm}}\lambda_{cm}^r,$$

and

$$R_{\cdot m} = \sum_{c=1}^{C} R_{cm}.$$

These quantities could be defined for any parameter λ and interpreted as total variation ($\hat{V}$), variation within chains (W_c), variation within models (W_m), and variation due to interaction between models and chains ($W_m W_c$). Based on a two-way analysis of variance (ANOVA) with interactions and unbalanced data, Castelloe and Zimmerman (2002) concluded some properties for the ratios $\frac{\mathrm{E}(\hat{V}(\lambda))}{\mathrm{E}(W_c(\lambda))}$ and $\frac{\mathrm{E}(W_m(\lambda))}{\mathrm{E}(W_m W_c(\lambda))}$, which could be used in practice to check any violation of convergence. More precisely, they defined the so-called *potential scale reduction factors* (PSRFs) as

$$\mathrm{PSRF}_1(\lambda) = \frac{\hat{V}(\lambda)}{W_c(\lambda)} \tag{8}$$

and

$$\mathrm{PSRF}_2(\lambda) = \frac{W_m(\lambda)}{W_m W_c(\lambda)}, \tag{9}$$

and recommended to use $\mathrm{PSRF}_1(\lambda)$ and $\mathrm{PSRF}_2(\lambda)$ in the following way:

(1) For simulating each chain use overdispersed starting values.
(2) Following Brooks and Gelman (1998), choose the base batch size $b \approx T/20$.
(3) Plot $\mathrm{PSRF}_1^{(q)}(\lambda)$ versus q and $\mathrm{PSRF}_2^{(q)}(\lambda)$ versus q, where the $\mathrm{PSRF}_i^{(q)}(\lambda)$, $i = 1, 2$, is computed using the $(qb + 1)$th iteration to the $2qb$th iteration for batches $q = 1, \ldots, T/b$.
(4) Plot together the numerator and the denominator of $\mathrm{PSRF}_1^{(q)}(\theta)$ versus q.
(5) Plot together the numerator and the denominator of $\mathrm{PSRF}_2^{(q)}(\theta)$ versus q.
(6) Determine q_0 such that for $q \geq q_0$, the plots in step 3 have settled close to 1 and the plots in each of steps 4 and 5 have settled approximately to a common value.
(7) Discard the first $q_0 b$ sweeps from each chain and then pool the remaining ones together to use for inference.

5. Simulations and Applications

The RJMCMC sampler in Sec. 3 was designed for obtaining draws of the whole parameter vector $(l, \mathbf{k}_l, \boldsymbol{\theta}_l)$ by a two-step approach. We use two simulations and one empirical

application to illustrate the adequacy and implementation of our proposed approach. We took prior distributions of the nonstructural parameters as suggested in Nieto (2005), that is, the multivariate normal for the AR coefficients vector and inverse Gamma distributions for the white-noise weights. The likelihood function is also referred to Nieto (2005). In order to check the convergence of the sampler using the procedure proposed in Sec. 4, we run three parallel chains. It seems that this number of chains is enough for our simulated and empirical data.

5.1. *Two Simulations*

Here we consider two simulated models. The first one is a TAR(2; 1, 1) model given by

$$X_t = \begin{cases} -0.5 + 0.6X_{t-1} + \varepsilon_t, & Z_t \leq 0, \\ 0.9 - 0.7X_{t-1} + 2.0\varepsilon_t, & Z_t > 0, \end{cases}$$

where $\{Z_t\}$ is an AR(1) process given by the model $Z_t = 0.5Z_{t-1} + a_t$, with $\{a_t\}$ a Gaussian zero-mean white-noise process with variance $\sigma_a^2 = 1$. The length of the simulated time series was 200. We fixed the simulated parameters for process $\{Z_t\}$ and used our proposed two-step procedure to identify and estimate the TAR model, by using the following prior distributions of the structural parameters: we used a truncated Poisson distribution on the set $\{2, 3, 4\}$ with parameter 3 for the number of regimes l and a truncated Poisson distribution for the AR orders on the set $\{0, 1, 2, 3\}$ with parameter 1, for any given number of regimes. We note that the number of observations in the first regime is $n_1 = 102$ and in the second, it is $n_2 = 97$.

Each chain was run with 16,000 iterations and we took the first 8,000 as the burn-in period, thus $T = 8{,}000$. To check convergence, we used the procedure described in Sec. 4 and the results are presented in Figs. 1 and 2. As we can see, the convergence of the chain is guaranteed. The posterior distribution of the number of regimes is given in Table 1. Clearly, $l = 2$ regimes is the mode of the distribution.

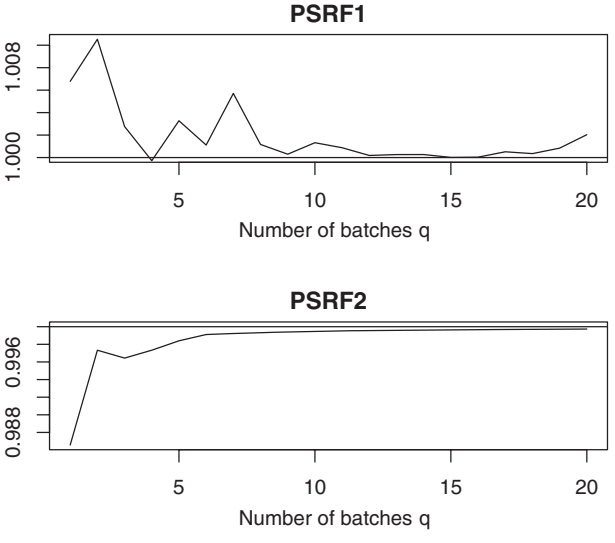

Figure 1. Convergence diagnostics for the first example: PSRFs.

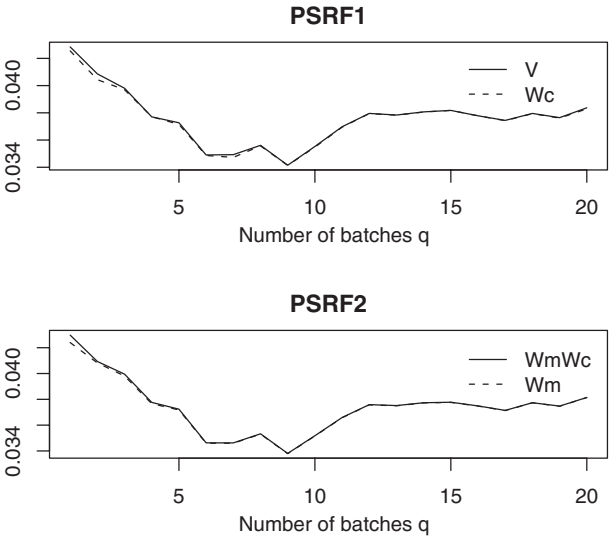

Figure 2. Numerators and denominators of PSRFs in the first example.

Fixing $l = 2$, we estimated the remaining parameters using chains with the same length as before. In Figs. 3 and 4, the chain convergence is guaranteed in the second step of the approach. In Table 2, we can see that the posterior distribution of the AR orders signals that $k_1 = k_2 = 1$. To study the sensitivity of the sampler to prior distributions, we used different priors for the AR orders and looked at the posterior distributions for l and the nonstructural parameters. We tried the values 3, 4, . . . , 14 as the maximum order of these parameters and several parameter values of a Poisson distribution truncated at a certain maximum value, for example, 4, 5, 6 as the parameter value of a Poisson distribution truncated at 8. We found that the method is robust against changes in the prior distribution of the AR orders since the sampler always identified the correct value of $l = 2$ and no drastic changes were detected in the posteriors of the nonstructural parameters. This "sensitivity"check exercise was also carried out in the other simulation and the two applications.

Table 3 presents the estimated nonstructural parameters, with their corresponding 95% and 99% credible intervals. All the simulated parameters lie within their credible intervals. Additionally, we increased the iterations number and the results were practically the same.

In order to check slightly the goodness-of-fit property of the proposed estimation procedure, we examined the cumulative sum (CUSUM) and cumulative sum of squares

Table 1
Posterior probabilities for l in the first example

l	Probability
2	0.998
3	0.002
4	0.000

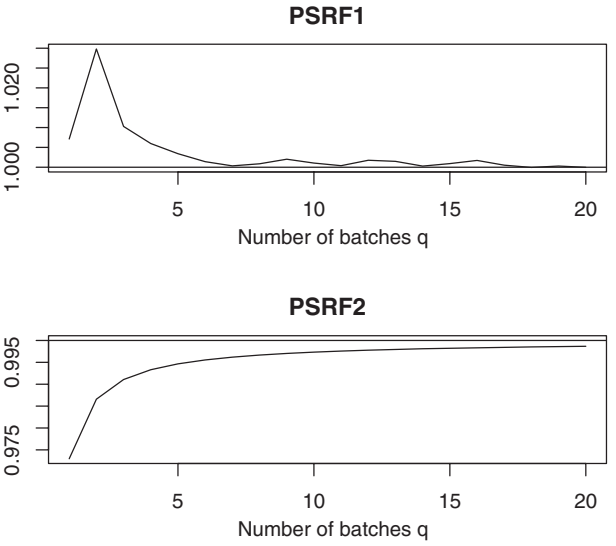

Figure 3. PSRFs with fixed l in the first example.

(CUSUMSQ) charts of the standardized model residuals, as suggested by Nieto (2005). These charts are plotted in Fig. 5 with its corresponding 95% confidence bands and, as we can see there, they signal an adequate fitting of the model.

Next, we simulate a TAR(3;2,1,1) model given by

$$X_t = \begin{cases} -0.5 + 0.6X_{t-1} - 0.7X_{t-2} + \varepsilon_t, & Z_t \leq -1, \\ 0.9 + 0.2X_{t-1} + 2.0\varepsilon_t, & -1 < Z_t \leq 1, \\ -1.0 - 0.7X_{t-1} + 3.0\varepsilon_t, & 1 < Z_t, \end{cases}$$

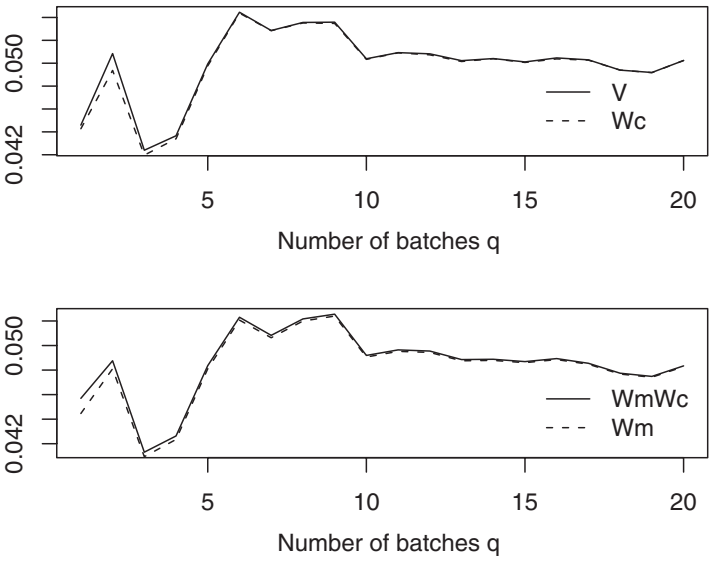

Figure 4. Numerators and denominators with fixed l in the first example.

Table 2
Posterior probabilities for AR orders in the first example

	Regime	
AR order	1	2
0	0.00	0.00
1	0.63	0.78
2	0.36	0.20
3	0.01	0.02

where process $\{Z_t\}$ was simulated as in the first simulation study, with white-noise process $\{a_t\}$ having variance $\sigma_a^2 = 1$. The sample size was 200 and we fix the parameters of process $\{Z_t\}$ and the thresholds $r_1 = -1.0$ and $r_2 = 1.0$ at their simulated values, in order to identify and estimate the TAR model. The whole set of prior distributions and the likelihood function were similar to those of the first simulation study. The chains convergence in the first stage of the procedure can be assessed in Figs. 6 and 7. Given convergence, we obtained the posterior distribution of l in Table 4, which implies that the distribution mode is $l = 3$. Fixing $l = 3$, we obtained the convergence of the chains in the second stage of the approach as seen in Figs. 8 and 9. The posterior distributions of the AR orders and the nonstructural parameters are presented in Table 5. The identified AR orders are $k_1 = 2, k_2 = 1$, and $k_3 = 1$, consistent with the simulated values. The estimated nonstructural parameters are presented in Table 6, with their 95% and 99% credible intervals and all of the simulated values lie within those intervals. Figure 10 shows the CUSUM and CUSUMSQ functions and only CUSUMSQ has a small piece of segment that is slightly outside the 95% confidence band. Overall, we have a reasonable fitting of the TAR model to the simulated data.

Chib (1995) developed a procedure based on the marginal likelihood approach that is helpful for solving model choice problems. Consequently, we use his procedure to do

Table 3
Nonstructural parameter estimates' credible intervals at levels 95% and 99% in the first example

	Regime	
Parameter	1	2
	−0.49	0.70
$a_0^{(j)}$	95% (−0.66, −0.32)	95% (0.63, 0.76)
	99% (−0.72, −0.27)	99% (0.61, 0.78)
	0.94	−0.69
$a_1^{(j)}$	95% (0.52, 1.40)	95% (−0.87, −0.51)
	99% (0.37, 1.50)	99% (−0.91, −0.47)
	0.80	2.40
$h^{(j)}$	95% (0.70, 0.92)	95% (2.11, 2.75)
	99% (0.67, 0.97)	99% (2.01, 2.89)

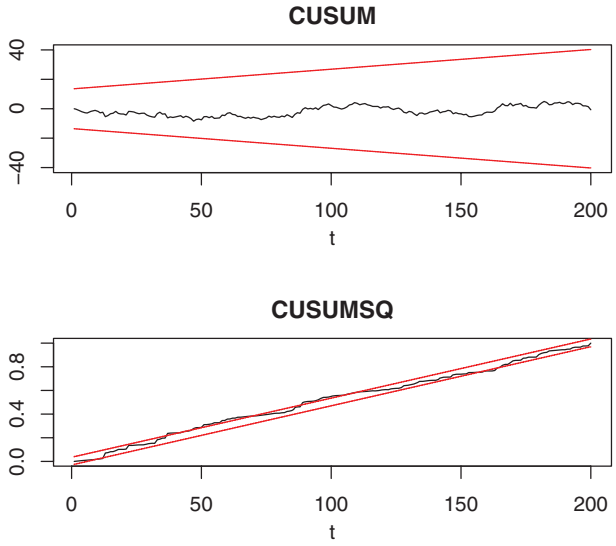

Figure 5. CUSUM and CUSUMSQ charts for the first simulated example.

a comparison with our sampler in this simulated example. Since $l_0 = 4$ and $k_{\max} = 3$, we get 336 possible TAR models (!); but thinking in the practice, we only consider six models for the comparison purpose, being the true model in one of them. We note here that, in general, the number of possible TAR models is $\sum_{j=2}^{l_0}(k_{\max}+1)^j$. For a given $l \in \{2, 3, 4\}$,

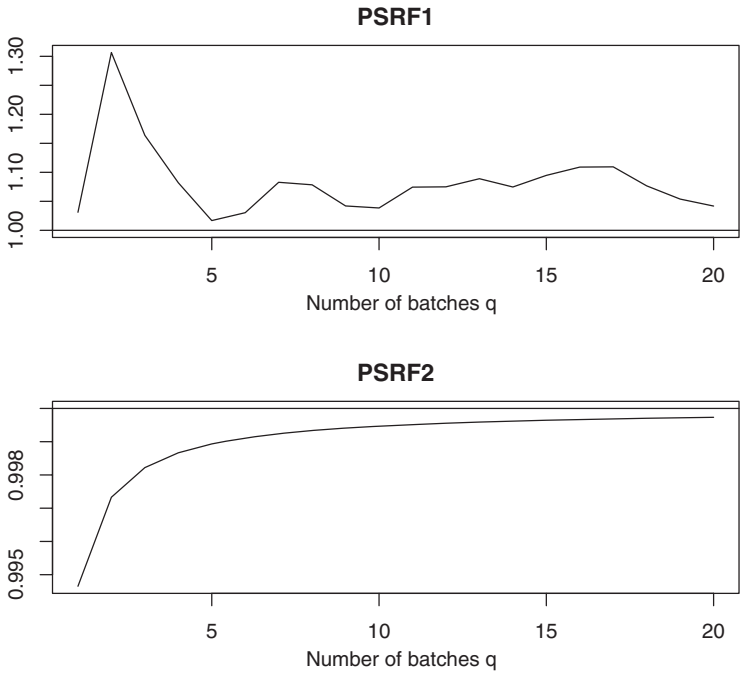

Figure 6. Convergence results for the second example: PSRFs.

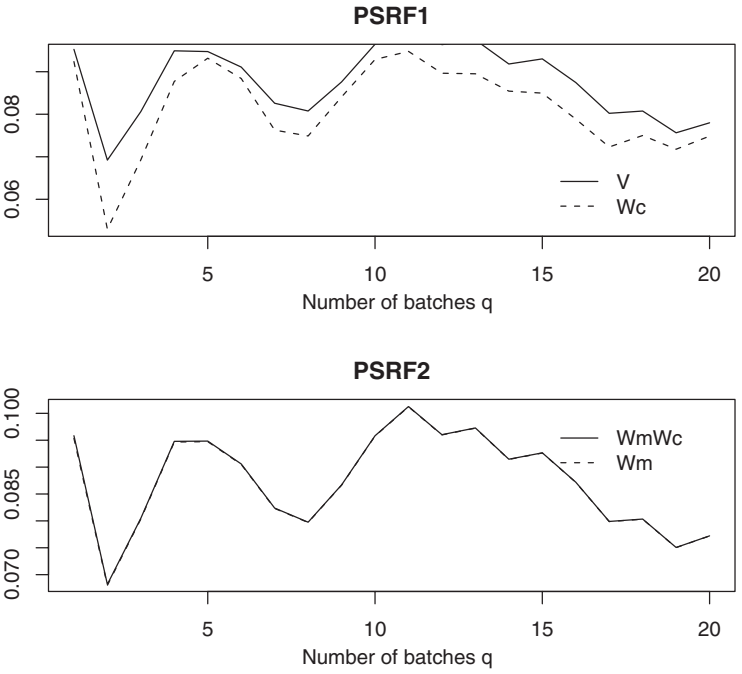

Figure 7. Numerators and denominators of PSRFs in the second example.

our parameter block vector $\boldsymbol{\theta}$ is given by $\boldsymbol{\theta} = (\nu_1, ..., \nu_l, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_l)$, with $\nu_j = [h^{(j)}]^2$ and $\boldsymbol{\theta}_j = (a_0^{(j)}, a_1^{(j)}, \ldots, a_{k_j}^{(j)})$ for each $j = 1, \ldots, l$. Note that we have $2l$ blocks and that their full conditional distributions are given by propositions 3.3 and 3.4 in Nieto's (2005) article.

Chib's (1995) algorithm for computing an estimate of the marginal likelihood $m(\mathbf{y}|l)$ for a TAR model with l regimes, where $\mathbf{y} = (\mathbf{x}, \mathbf{z})$, becomes as follows:

1. *Selection of* $\boldsymbol{\theta}^*$.
   We use a slight modification of Chib's suggestions, namely, we take the median of the Gibbs-sampling drawn samples for the nonstructural parameters $a_i^{(j)}$ and ν_j; $i = 0, 1, \ldots, k_j$; $j = 1, \ldots, l$. With these sample statistics, we conform $\boldsymbol{\theta}^*$.
2. *Computation of* $f(\mathbf{x}|\mathbf{z}, \boldsymbol{\theta}^*, l)$ and $\pi(\boldsymbol{\theta}^*|l)$.

Table 4
Posterior probabilities for l in the second model

l	Probability
2	0.01
3	0.97
4	0.02

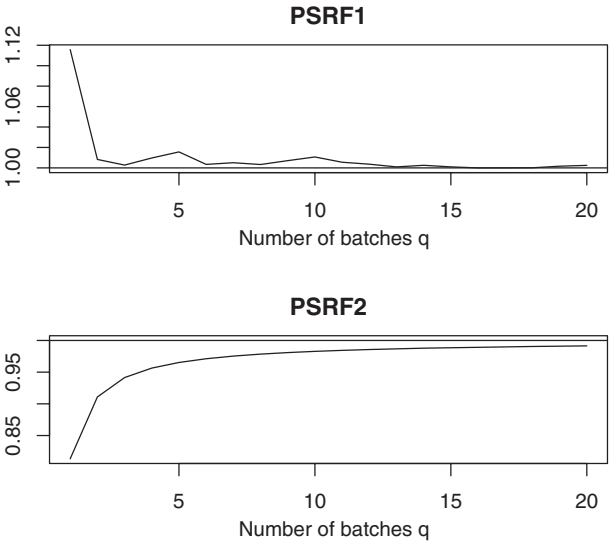

Figure 8. PSRFs with fixed l in the second simulated example.

The prior distribution $\pi(\boldsymbol{\theta}|l)$ is obtained assuming a priori independence among the blocks of $\boldsymbol{\theta}$, as Nieto (2005) did. That is to say,

$$\pi(\boldsymbol{\theta}|l) = \prod_{j=1}^{l} \pi(\boldsymbol{\theta}_j) \prod_{j=1}^{l} \pi(\nu_j),$$

where $\pi(\boldsymbol{\theta}_j)$ is the probability density function (pdf) of a multinormal distribution of order $k_j + 1$ with mean $\mathbf{0}$ and variance matrix $100\boldsymbol{I}$, with $\boldsymbol{I}$ the identity matrix of order $k_j + 1$, and $\pi(\nu_j)$ is the pdf of an inverse Gamma distribution with shape parameter $\alpha_0 = 1.5$ and scale parameter $\beta_0 = 0.5\hat{\sigma}_r^2$, for all $j = 1, \ldots, l$, with $\hat{\sigma}_r^2$ the residual variance after fitting a linear AR model to the time series $\{x_t\}$.
We can note that, with $\boldsymbol{\theta} = \boldsymbol{\theta}^*$,

$$\ln\pi(\boldsymbol{\theta}^*|l) = \sum_{j=1}^{l} \ln\pi(\boldsymbol{\theta}_j^*) + \sum_{j=1}^{l} \ln\pi(\nu_j^*),$$

Table 5
Posterior probabilities for the AR orders in the second example

	Regime		
AR order	1	2	3
0	0	0.24	0.43
1	0	0.73	0.44
2	0.98	0.03	0.11
3	0.02	0.00	0.02

Table 6
Nonstructural parameter estimates with credible intervals at 95% and 99% levels in the second example

	Regime		
Parameter	1	2	3
$a_0^{(j)}$	−0.80	1.20	−0.58
95%	(−1.08, −0.57)	(0.83, 1.56)	(−1.77, 0.64)
99%	(−1.12, −0.37)	(0.72, 1.67)	(−2.11, 1.11)
$a_1^{(j)}$	0.74	0.20	−0.35
95%	(0.57, 0.88)	(0.06, 0.34)	(−0.86, 0.12)
99%	(0.56, 0.98)	(0.01, 0.39)	(−1.12, 0.29)
$a_2^{(j)}$	−0.78		
95%	(−0.91, −0.64)		
99%	(−0.94, −0.61)		
$h^{(j)}$	0.83	2.03	3.82
95%	(0.67, 1.04)	(1.82, 2.27)	(2.84, 5.31)
99%	(0.64, 1.12)	(1.77, 2.38)	(2.57, 6.11)

where, for each $j = 1, \ldots, l$,

$$\ln \pi\left(\boldsymbol{\theta}_j^*\right) = -0.5\left[(k_j + 1)\ln(2\pi) + (k_j + 1)\ln 100 + 0.01\left(\boldsymbol{\theta}_j^*\right)'\boldsymbol{\theta}_j^*\right]$$

and

$$\ln \pi\left(v_j^*\right) = \alpha_0 \ln \beta_0 - \ln \Gamma(\alpha_0) - (\alpha_0 + 1)\ln v_j^* - \beta_0 (v_j^*)^{-1},$$

with Γ denoting the Gamma function.

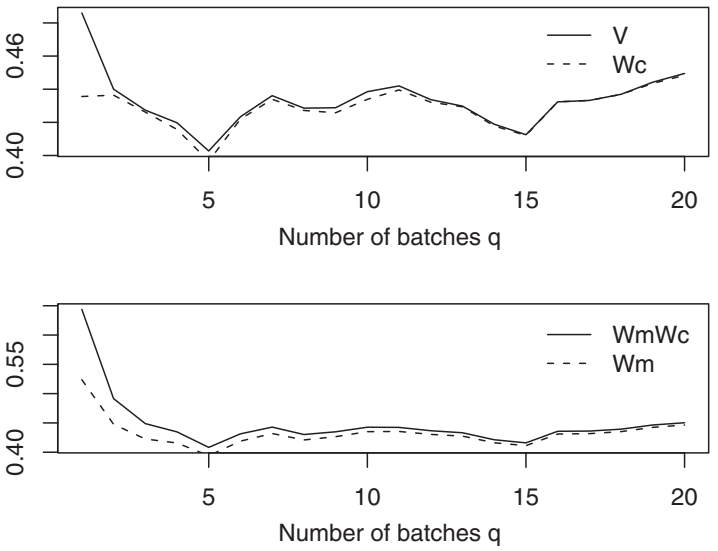

Figure 9. Numerators and denominators of PSRFs in the second example with fixed l.

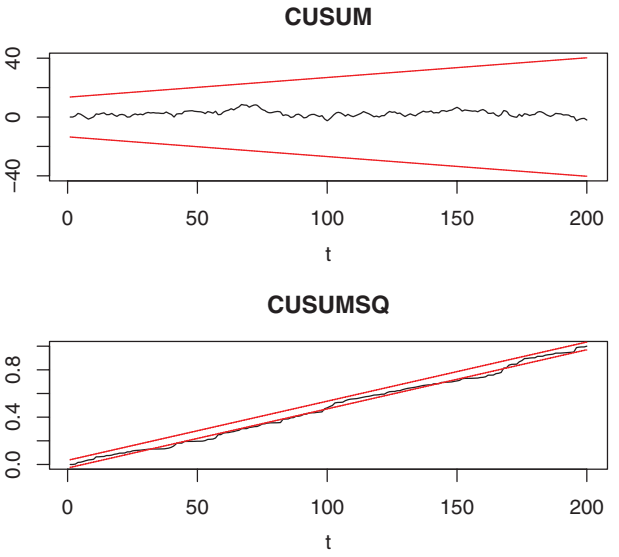

Figure 10. CUSUM and CUSUMSQ charts for the second example.

The log-likelihood is given by

$$\ln f(\mathbf{x}|\mathbf{z}, \boldsymbol{\theta}^*, l) = -0.5\left[(T-k)\ln(2\pi) + \sum_{t=k+1}^{T} \ln \nu_{j_t}^* + \sum_{t=k+1}^{T} \left(e_t^*\right)^2\right],$$

[see Nieto's (2005) article] where $e_t^* = [x_t - a_0^{(j_t)*} - \sum_{i=1}^{k_{j_t}} a_i^{(j_t)*} x_{t-i}]/(\nu_{j_t}^*)^{1/2}$, $j_t = j \in \{1, \ldots, l\}$ if z_t is in regime j, and T is the sample size. In what follows, we omit writing l in the set of conditioning values of each density under consideration.

3. *Computation of* $\hat{\pi}(\nu_1^*|\mathbf{y})$ and $\hat{\pi}(\nu_r^*|\nu_s^*(s < r), \mathbf{y})$, $r = 2, \ldots, l$.
When $r = 1$, we need to simulate from the full conditional distribution $\pi(\nu_1|\nu_s(s = 2, \ldots, l), \boldsymbol{\theta}_s(s = 1, \ldots, l))$. However, this distribution is an inverse Gamma with shape parameter that depends on the number of x observations in the first regime, n_1 say, and scale parameter that depends on $\boldsymbol{\theta}_1$ and the x observations in this regime (see Nieto, 2005). It is worth noting that for all $j = 1, \ldots, l$, n_j depends on k, the maximum of the l AR orders. For example, for the true model TAR(3;2,1,1), $n_1 = 40$, $n_2 = 124$, and $n_3 = 34$ (effective sample size is 198). Thus, we only need draws from the full distribution of $\boldsymbol{\theta}_1$ and they are taken up from the first G iterates of the Gibbs sampler. Hence, we get

$$\hat{\pi}\left(\nu_1^*|\mathbf{y}\right) = \frac{1}{G}\sum_{i=1}^{G} \pi\left(\nu_1^*|n_1, \boldsymbol{\theta}_1^{(i)}\right).$$

For $r = 2, \ldots, l$, we use arguments similar to those given above for obtaining $\hat{\pi}(\nu_1^*|\mathbf{y})$ and, then, we get

$$\hat{\pi}\left(\nu_r^*\middle|\nu_s^*(s < r), \mathbf{y}\right) = \frac{1}{G}\sum_{i=1}^{G} \pi\left(\nu_r^*|n_r, \boldsymbol{\theta}_r^{(i)}\right),$$

where the draws $\boldsymbol{\theta}_r^{(i)}$, $i = 1, \ldots, G$, also come from the first G iterates of the Gibbs sampler and n_r denotes the number of observations in the rth regime. For the true model, we have that $n_2 = 124$ and $n_3 = 34$. Of course, the reduced conditional distribution $\pi(\nu_r|\nu_s^*(s < r), \boldsymbol{\theta}_s(s = 1, \ldots, l), \mathbf{y})$ is an inverse Gamma with shape and scale parameters that depend on n_r, $\boldsymbol{\theta}_r$, and the x observations in the rth regime.

4. *Computation of* $\hat{\pi}(\boldsymbol{\theta}_r^*|\nu_s^*(s = 1, \ldots, l), \boldsymbol{\theta}_s^*(s < r), \mathbf{y})$ for $r = 1, \ldots, l$.

Following Nieto (2005), we can show that for all $r = 1, \ldots, l$, the reduced conditional density function $\pi(\boldsymbol{\theta}_r|\nu_s^*(s = 1, \ldots, l), \boldsymbol{\theta}_s^*(s < r), \boldsymbol{\theta}_s(s > r), \mathbf{y})$ is that of a multinormal distribution of order $k_r + 1$, with mean and variance matrix that depend on ν_r^* and the x observations in the regime r. In this way, the value of this density at $\boldsymbol{\theta}_r^*$ is known exactly and, consequently, $\hat{\pi}(\boldsymbol{\theta}_r^*|\nu_s^*(s = 1, \ldots, l), \boldsymbol{\theta}_s^*(s < r), \mathbf{y}) = \pi(\boldsymbol{\theta}_r^*|\nu_r^*, \mathbf{y})$.

Finally,

$$\ln \hat{m}(\mathbf{y}|l) = \ln f(\mathbf{x}|\mathbf{z}, l, \boldsymbol{\theta}^*) + \ln \pi(\boldsymbol{\theta}^*|l) - \sum_{i=1}^{2l} \ln \hat{\pi}\left(\boldsymbol{\theta}_i^*|\boldsymbol{\theta}_j^*(j \leq i-1), \mathbf{y}\right),$$

where ν_i^* is the ith block of $\boldsymbol{\theta}^*$, $i = 1, \ldots, l$. We then search for the model (number of regimes, AR orders, and nonstructural parameter estimates) for which its marginal likelihood is the maximum among the set of all possible models.

The six models are the following:

Model 1 (M1): the true model TAR(3;2,1,1),
Model 2 (M2): TAR(3;3,3,3) with the same thresholds as M1,
Model 3 (M3): TAR(3;0,1,0) with the same thresholds as M1,
Model 4 (M4): TAR(2;1,1) with threshold $r_1 = -1.0$,
Model 5 (M5): TAR(2;3,2) with threshold $r_1 = -1.0$,
Model 6 (M6): TAR(4;2,2,2,2) with thresholds $r_1 = -1.0$, $r_2 = 1.0$, and $r_3 = 1.8$.

The results are presented in Table 7 and we can see that the marginal likelihood for the true model is the maximum. Nevertheless, it is worth noting that once the values of l and $\mathbf{k}_l$ are fixed, as happens in models M1–M6 above, the RJMCMC sampler is very complex to implement because of the design of (1) the mass probability function for the moves among models and (2) the densities q for completing dimensions, when this situation is necessary. In view of this fact, we do not use our proposed RJMCMC sampler for choosing a TAR model in the set {M1, ..., M6} and, in doing so, for comparing with the results in Table 7.

Table 7
Log of marginal likelihoods for the chosen TAR models

Model	log(marginal)
M1	−434.14
M2	−447.64
M3	−465.75
M4	−477.42
M5	−455.24
M6	−448.64

For future research, however, it would be interesting to explore a two-step strategy based on the two approaches, for identifying and estimating a TAR model. In the first stage, we might use our RJMCMC sampler to find appropriate values for the structural parameters, and in the second, we could use the marginal likelihood approach to choose among the relatively few proposed models. Recall that in this example we have 336 candidate TAR models.

5.2. *An Application in the Hydrology/Meteorology Field*

Nieto (2005) presented a real-data application for illustrating his methodology. The time series considered were the daily rainfall (in mm), as the threshold variable, and a daily river flow (in m^3/s), as the response variable, in a certain Colombian geographical region. The rainfall was measured at the San Rafael Lagoon's meteorological station, with an altitude of 3,420 m and geographical coordinates 2.23 north (latitude) and 76.23 west (longitude). The flow corresponds to the Bedon river, a small one in hydrological terms, and was measured at the San Rafael Lagoon's hydrological station, with an altitude of 3,300 m and coordinates 2.19 north and 76.15 west. These stations are located close to the earth's equator and in a very dry geographical zone. This last characteristic permits to control for hydrological/meteorological factors, which may distort the kind of dynamical relationship explained by the TAR model. The dataset corresponds to the sample period from January 1, 1992, up to November 30,2000 (3,256 data), and it was assembled by IDEAM, the official Colombian agency for hydrological and meteorological studies.

The two time series have missing data but, in this article, we interpolate the time series using Nieto's (2005) procedure. Nieto (2005) did a double transformation of the river flow data with the square root function and an adjustment for conditional heteroscedasticity via an ARCH(1) model. In this section, $\{X_t\}$ always denotes the transformed river flow process and $\{Z_t\}$ the precipitation process lagged one day (because of measurement conventions for this variable). Figure 11 shows the dynamical relationship between the two time series: the more the precipitation is, the more the (transformed) river flow is. Additionally, one

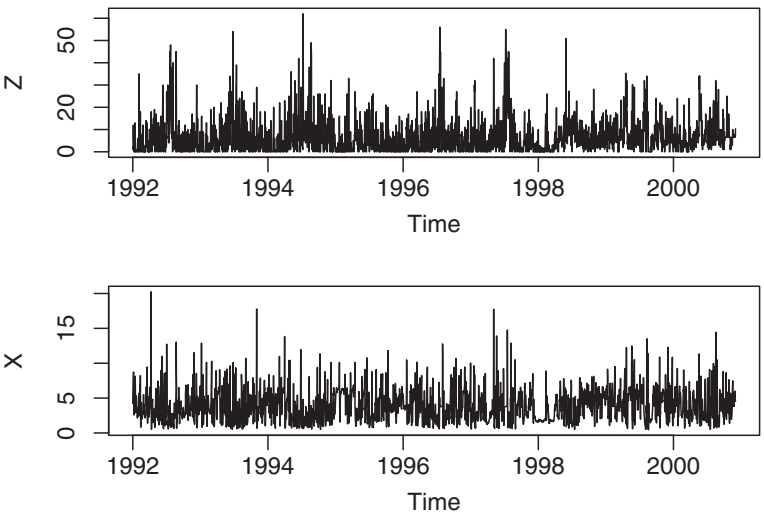

Figure 11. (a) Precipitation. (b) Flow.

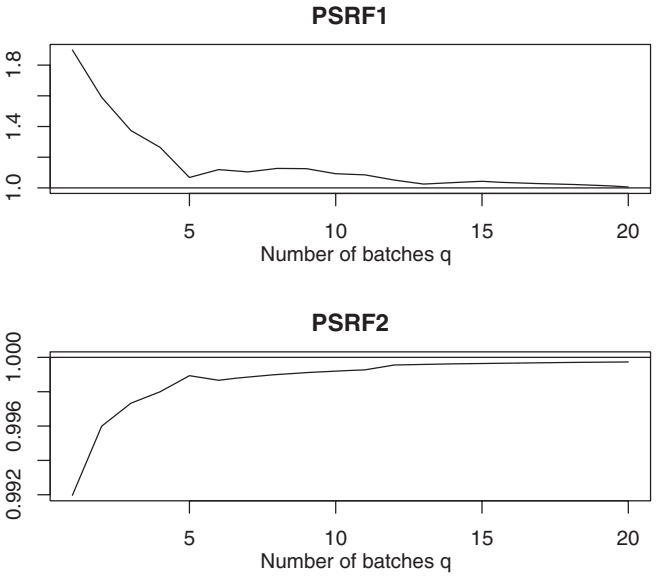

Figure 12. PSRFs for the hydrological/meteorological example.

can see certain stable path in both variables, although there are bursts of large values in the river flow. This fact is a major characteristic of data that are generated by a TAR model.

To implement our two-step approach, we empirically obtained the thresholds values. Following Tong (1990), these values were computed via the minimum normalized AIC (NAIC) criterion in the following way: we fixed a value for l and then we chose the thresholds values for which the NAIC is minimum among the empirical quantiles of the rainfall variable. For these data, the potential thresholds are 6.0 mm for $l = 2$, 6.0 mm and

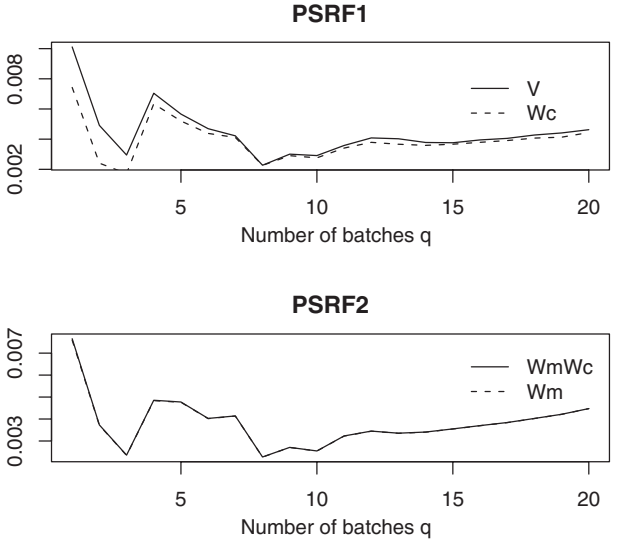

Figure 13. Numerators and denominators of PSRFs for the river flow/rainfall example.

Table 8
Posterior probabilities for the number of regimes in the precipitation/flow example

l	Probability
2	0.00
3	0.32
4	0.68

10.3 mm for $l = 3$, and 6.0 mm, 10.3 mm, and 17.18 mm for $l = 4$. It is important to note that to fit a Markov chain model to process $\{Z_t\}$ is not necessary because we do not estimate missing data or forecast the involved stochastic processes.

As to prior distributions of the structural parameters, we use a truncated Poisson distribution with parameter 3 on the set $\{2, 3, 4\}$ (here $l_0 = 4$) for the number of regimes and truncated Poisson distributions with parameter 2 on the set $\{0, 1, 2, 3\}$ ($k_{\max} = 3$) for the AR orders. In Figs. 12 and 13, we present the convergence diagnostics for the first stage of the procedure and see that convergence is clearly guaranteed. In Table 8, we present the posterior distribution for the number of regimes that was obtained in this first phase of the procedure and find $l = 4$, contrary to Nieto's (2005) result that was $l = 2$. This finding is in accordance with the known fact that in that Colombian geographical zone, there are two periods of rains and two periods of dry meteorological conditions, which alternate trough the whole year in this way: first dry season from middle December to middle March, first rain season from middle March to middle July, second dry period from middle July to middle September, and second rain season from middle September to middle December.

We set $\epsilon = 10$, a large quantity, when it is compared with the simulated examples and, consequently, contrary to the suggestion in Campbell (2004). This fact signals some dependence degree of this parameter on the data. In Figs. 14 and 15, we also see that in the second step of our approach, the convergence of the chains is obtained. Table 9 displays the posterior distributions for the AR orders and Table 10 shows the estimates for the nonstructural parameters. To do a raw checking on the model specification, we plot the CUSUM and CUSUMSQ charts in Fig. 16, which implies an adequate fitting of the model.

Table 9
Posterior probabilities for the AR orders in the hydrological/meteorological example

	Regime			
AR order	1	2	3	4
0	0	0	0	0
1	0	0	0	0.05
2	0	0.46	0.77	0.63
3	0.81	0.45	0.19	0.28
4	0.19	0.09	0.04	0.04

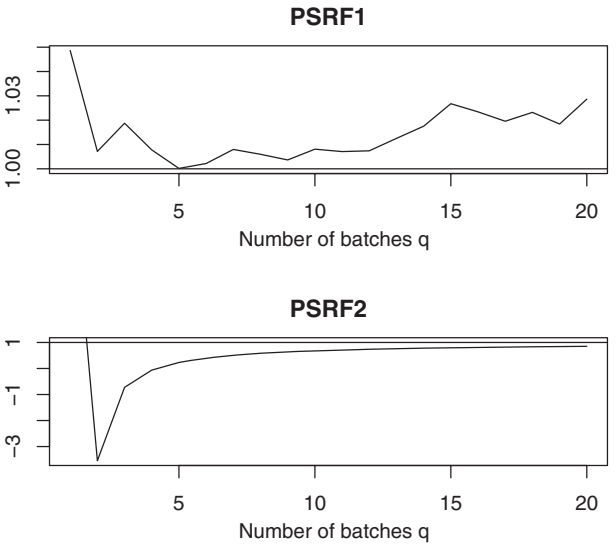

Figure 14. PSRFs for the hydrological/meteorological example with fixed L.

The fitted model is given by

$$X_t = \begin{cases} 1.35 + 0.74X_{t-1} - 0.27X_{t-2} \\ \quad +0.12X_{t-3} + 1.30\varepsilon_t, & Z_t \le 6.0, \\ 1.96 + 0.75X_{t-1} - 0.30X_{t-2} \\ \quad +1.66\varepsilon_t, & 6.0 < Z_t \le 10.3, \\ 2.11 + 0.80X_{t-1} - 0.28X_{t-2} \\ \quad +2.15\varepsilon_t, & 10.3 < Z_t \le 17.18, \\ 2.96 + 0.62X_{t-1} - 0.35X_{t-2} \\ \quad +3.18\varepsilon_t, & 17.18 < Z_t. \end{cases}$$

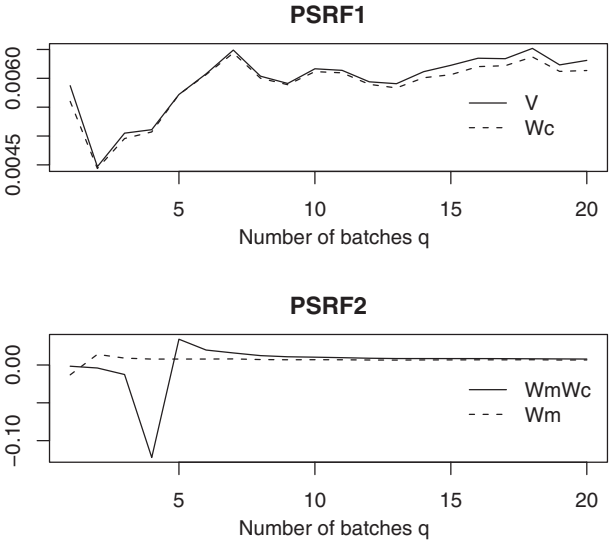

Figure 15. Numerators and denominators of PSRFs for the river flow/rainfall example with fixed L.

Table 10
Nonstructural parameter estimates: 95% credible intervals for the hydrological/meteorological time series

	Regime			
Parameter	1	2	3	4
$a_0^{(j)}$	1.35 (1.2, 1.5)	1.96 (1.6, 2.3)	2.11 (1.6, 2.5)	2.96 (2.0, 3.7)
$a_1^{(j)}$	0.74 (0.7, 0.8)	0.75 (0.7, 0.8)	0.80 (0.7, 0.9)	0.62 (0.5, 0.8)
$a_2^{(j)}$	−0.27 (−0.3, −0.2)	−0.3 (−0.4, −0.2)	−0.28 (−0.4, −0.2)	−0.35 (−0.6, −0.2)
$a_3^{(j)}$	0.12 (0.1, 0.2)			
$h^{(j)}$	1.30 (1.2, 1.3)	1.66 (1.6, 1.7)	2.15 (2.0, 2.3)	3.18 (3.0, 3.4)

It is interesting to note that (1) the river flow has an AR dynamic of lag 2, except in the first regime where it is of lag 3 and (2) the numeric signs of the coefficients of the same lag are the same and their absolute values are close in these regimes. This fact suggests that the AR dynamics of the river flow is almost the same through the different regimes for the precipitation. This empirical characteristic was also detected by Nieto's (2005) TAR(2;1,1) model. As a first impression, one might say that the influence of the regimes on the actual value of the river flow is through the model intercepts and the regime-based

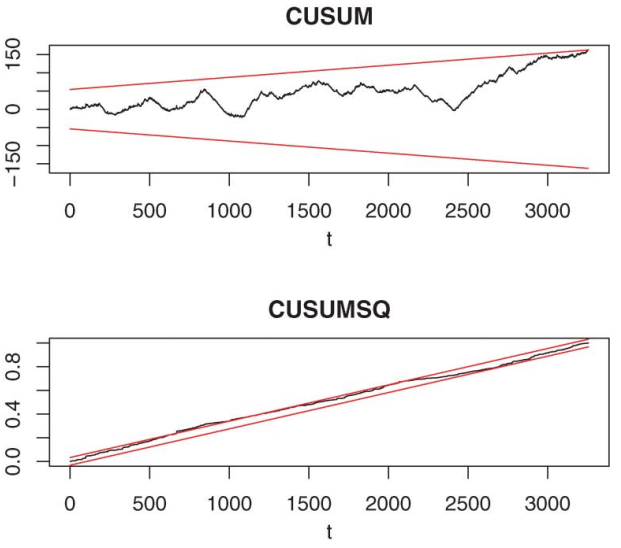

Figure 16. CUSUM and CUSUMSQ charts for the hydrological/meteorological example.

variability. However, model interpretations other than this impression are provided in the next paragraph.

According to TAR model characteristics (Nieto, 2008), we can derive the following important facts. First, with the distribution of X_t, the contribution of each regime to the mean of X at time t, computed as $\mu_{j,t,1} = a_0^{(j)}/\phi_j(1)$ with $j = 1, 2, 3, 4$, is 3.30, 3.56, 4.40, and 4.05 for the first, second, third, and fourth regimes, respectively. In general, $\phi_j(B) = 1 - a_1^{(j)}B - \cdots - a_{k_j}^{(j)}B^{k_j}$, $j = 1, \ldots, l$. Note that these contributions do not depend on t and that these figures are almost in agreement with the fact that the more the precipitation is, the more the river flow is (except in the fourth regime). This last observation holds for the regime-conditional variability of the river flow, which is reported as 1.30^2, 1.66^2, 2.15^2, and 3.18^2. Note that the AR polynomials in each regime have their roots outside the unit circle. Second, we found that $\mathrm{E}(X_t|x_{t-1}, \ldots, x_1) = 1.71 + 0.75x_{t-1} - 0.29x_{t-2} + 0.07x_{t-3}$, where $p_1 = 0.61$, $p_2 = 0.20$, $p_3 = 0.12$, and $p_4 = 0.08$, for

$$\mathrm{E}(X_t|x_{t-1}, \ldots, x_1) = \sum_{j=1}^{4} p_j a_0^{(j)} + \sum_{j=1}^{4} p_j a_1^{(j)} x_{t-1} + \sum_{j=1}^{4} p_j a_2^{(j)} x_{t-2} + p_1 a_3^{(1)} x_{t-3}.$$

The values of p_1, p_2, p_3, and p_4 indicate that it is more possible to have low-intensity rainfall than either medium- or high-intensity rainfall, a result that agrees strongly with the fact that the Colombian geographical region for which the analysis is performed is very dry in the whole solar year. Third, the marginal mean of the river flow is 3.58 for any day t, indicating that the mean function of the stochastic process $\{X_t\}$ is constant in the analyzed sample period—an empirical characteristic observed in the data.

Another real-data application was carried out using US quarterly macroeconomic data in the sample period 1970:01–2004:02. These time series correspond to the seasonally adjusted gross domestic product (GDP) growth rate (the output variable) and the spread between 3-year constant maturity yield (R_1) and 3-month treasury bill discount yield (R_2) (the input variable). The results can be obtained from the authors upon request.

6. Conclusions

An RJMCMC sampler has been developed for identifying and estimating a TAR model (except the thresholds), which constitutes an alternative to the Gibbs-sampling-based approach that uses link priors. Basically, the new approach consists of two steps. In the first step, the number of regimes is identified using samplers that take into account the varying dimension of the whole parameter space, when the chains move from one state to another. In the second, the identified number of regimes is fixed and, conditional on it, the remaining model parameters are estimated, including the AR orders.

The proposed approach improves the three-step Gibbs sampling (Nieto, 2005), in the sense of (1) only two steps involved, (2) quickly obtaining the convergence condition or, equivalently, the stationary distribution of the underlying Markov chain, and (3) better numerical behavior of the likelihood function, avoiding that, approximately, posterior distributions equal prior distributions. The second property is monitored via a convergence diagnostic that is based on two one-way ANOVAs and a two-way ANOVA with interaction and unbalanced data. Also, in the second step, we have proposed a modification to Campbell's (2004) approach, in the sense that we update the nonstructural parameters in all the regimes without randomly choosing a regime to do the updating.

Nevertheless, we observed in several simulated examples and two empirical examples that the posterior distribution of the number of regimes and some AR orders tends to be very concentrated on one or two values. We feel that this low mixing is due to the fact that the varying dimension of the whole parameter space involves a nested movement, that is, first, the underlying chain moves to change the number of regimes and then, it moves intrinsically to change the AR order in each regime. This nested movement causes the complexity for designing a *global* RJMCMC sampler. Additional research must be conducted to study this low mixing problem in the current two-step approach. However, the results obtained in the simulated examples and the real-data applications provide coherent interpretations of the fitted models and, thus, this proposed two-step RJMCMC sampler is adequate for analyzing a TAR model.

A small exercise that uses the marginal likelihood approach of Chib (1995) was conducted, in order to check the performance of this procedure in choosing a TAR model. We used the second simulated example and six competing models (among a large number of possible models) and the results were satisfactory. We feel that some research can be done in the future, in order to combine these two approaches for identifying and estimating TAR models. Specifically, our RJMCMC sampler could be used as a first step to propose a relatively few number of appropriate models.

Appendix

We present here the summarized algorithm that was obtained for implementing our RJMCMC sampler in the fitting of TAR models. We recall that this is a Metropolis–Hastings algorithm and, from this perspective, we organize sequentially their steps. Our parameter vector of interest is $\Lambda = (l, \mathbf{k}_l, \boldsymbol{\theta}_l)$.

Stage 1. *The global procedure to identify the number of regimes* l.

Step 1. Obtain an initial value for Λ, $\Lambda^{(0)}$ say. To do this, we can obtain a draw for l from a truncated Poisson distribution on the set $\{2, \ldots, l_0\}$ and set each component of $\mathbf{k}_l$ equal to $k_{\max}$, where l_0 and $k_{\max}$ are appropriate upper bounds for the number of regimes and the AR orders, respectively. Given the initial values for l and $\mathbf{k}_l$, we can use the prior distributions for the subvectors of $\boldsymbol{\theta}_l$ given by Nieto (2005).

Step 2. For each $i = 1, \ldots, n$:

(1) Draw Λ from a proposal distribution. To do this:
   (a) Use the probability mass function defined on the set $\{B, D, O, V\}$ to draw a move for l. These probability values constitute the jump probabilities $j(\cdot|\cdot)$, which in our case do not depend on the current state of the chain.
   (b) If the draw is B set $l^{(i)} = l^{(i-1)} + 1$. If it is D, put $l^{(i)} = l^{(i-1)} - 1$. If it is either O or V set $l^{(i)} = l^{(i-1)}$. If the move is V go to (d).
   (c) Do *intrinsic* moves for the AR orders in each regime, using the proposed probability mass function on the set $\{B_k, D_k, R_k\}$, where R_k stands for *remain*. If the move at regime j is B_k, $j = 1, \ldots, l$, put $k_{jl}^{(i)} = k_{jl}^{(m)} + 1$, where the superscript (m) was defined on page 8, and consider the proposal distribution q given on page 10 to complete the dimension of the AR coefficient vector at the mth iteration. This is needed to compute α below. If it is D_k, set $k_{jl}^{(i)} = k_{jl}^{(m)} - 1$ without doing dimension completion. If the drawn move for an AR order is R_k, put $k_{jl}^{(i)} = k_{jl}^{(m)}$.
   (d) Update the nonstructural parameters in each regime using their posterior distributions given by Nieto (2005) and let $\boldsymbol{\theta}_l^{(i)}$ be the updated parameter vector.

(2) Compute the acceptance probability α according to the expressions given on page 9.
(3) Draw u from the uniform distribution with parameters 0 and 1. If $u \leq \alpha$, accept the whole move, that is, set $\Lambda^{(i)} = (l^{(i)}, \mathbf{k}_l^{(i)}, \boldsymbol{\theta}_l^{(i)})$. If $u > \alpha$, set $\Lambda^{(i)} = \Lambda^{(i-1)}$.
(4) Go to (1) until convergence.

Stage 2. *Identifying AR orders and estimating nonstructural parameters.*

Once the chains in Stage 1 have converged, we take as the identified value for l the mode of its marginal posterior distribution. We fix this value and concentrate our attention on the reduced parameter vector $\Lambda_r = (\mathbf{k}, \boldsymbol{\theta})$. For notational convenience, we suppress the subscript l. Now, the Metropolis–Hastings algorithm is the following.

Step 1. Obtain an initial value for Λ_r, $\Lambda_r^{(0)}$ say. To do this, we set each component of $\mathbf{k}$ equal to $k_{\max}$, where $k_{\max}$ is the upper bound that was defined in Step 1 of Stage 1. Given the initial values for the components of $\mathbf{k}$, we can use as prior distributions for the subvectors of $\boldsymbol{\theta}$ those proposed by Nieto (2005).

Step 2. For each $i = 1, \ldots, n$:

(1) Draw Λ_r from a proposal distribution. To do this:
   (a) Do a move for each AR order using the probability mass function defined in Stage 1. These function values are the new jump probabilities. If the AR-order move at regime j is B_k, $j = 1, \ldots, l$, put $k_j^{(i)} = k_j^{(i-1)} + 1$ and consider the proposal distribution q given on page 10 to complete the dimension of the AR coefficient vector at the $(i-1)$th iteration. This is needed to compute α below. If it is D_k, set $k_j^{(i)} = k_j^{(i-1)} - 1$ without doing dimension completion. If the drawn move for an AR order is R_k, put $k_j^{(i)} = k_j^{(i-1)}$.
   (b) Update the nonstructural parameters in each regime using their posterior distributions given by Nieto (2005) and let $\boldsymbol{\theta}^{(i)}$ be this updated parameter vector.
(2) Compute the acceptance probability α as it was indicated on page 9.
(3) Draw u from the uniform distribution with parameters 0 and 1. If $u \leq \alpha$, accept the whole move, that is, set $\Lambda_r^{(i)} = (\mathbf{k}^{(i)}, \boldsymbol{\theta}^{(i)})$. If $u > \alpha$, set $\Lambda_r^{(i)} = \Lambda_r^{(i-1)}$.
(4) Go to (1) until convergence.

Acknowledgments

The authors wish to thank DIB, the investigation division of Universidad Nacional de Colombia at Bogotá, the sponsorship that was provided to this research under contract COL0022265-2007. Also, they gratefully acknowledge the very useful comments and suggestions of an anonymous referee, which helped to improve substantially the article.

References

Brooks, S. P., Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7:434–455.

Campbell, E. P. (2004). Bayesian selection of threshold autoregressive models. *Journal of Time Series Analysis* 25:467–482.

Carlin, B. P., Chib, S. (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B* 57:473–484.

Castelloe, J. M., Zimmerman, D. L. (2002). Convergence Assessment for Reversible Jump MCMC Samplers, SAS Institute Inc. (unpublished manuscript).

Chen, M., Shao, Q., Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation.* New York: Springer-Verlag.

Chib, S. (1995). Marginal likelihood from the the Gibbs output. *Journal of the American Statistical Association* 90:1313–1321.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82:711–732.

Nieto, F. H. (2005). Modeling bivariate threshold autoregressive processes in the presence of missing data. *Communications in Statistics—Theory and Methods* 34:905–930.

Nieto, F. H. (2008). Forecasting with univariate TAR models. *Statistical Methodology* 5:263–276.

Tong, H. (1978). On a threshold model. In: Chen, C.ed. *Pattern Recognition and Signal Processing. Volume 29 of NATO ASI Series E: Applied Sciences*. Amsterdam: Sijthoff and Noordhoff, pp. 575–586.

Tong, H. (1990). *Nonlinear Time Series: A Dynamical System Approach.* Oxford: Oxford University Press.

Tong, H., Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society, Series B* 42:245–292.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* 93:1188–1202.

Vermaak, J., Andrieu, C., Doucet, A., Godsill, S. J. (2004). Reversible jump Markov chain Monte Carlo strategies for Bayesian model selection in autoregressive processes. *Journal of Time Series Analysis* 25:785–809.

Vrontos, I. D., Dellaportas, P., Politis, D. N. (2000). Full Bayesian inference for GARCH and EGARCH models. *Journal of Business and Economic Statistics* 18:187–198.