

Time Series Analysis

Modeling Bivariate Threshold Autoregressive Processes in the Presence of Missing Data

FABIO H. NIETO

National University of Colombia, Bogotá D.C., Colombia

In this article, a methodology for analyzing bivariate time series with missing data is presented. It is assumed that there is a dynamical nonlinear relationship between the two time series, which is described by a threshold autoregressive (TAR) model. The time series analysis consists in the identification and estimation of the model in the presence of missing data. As a main result, the model parameters and the missing observations are estimated jointly. The TAR model analysis is accomplished by means of Markov Chain Monte Carlo (MCMC) and Bayesian approaches.

Keywords Missing data; Nonlinear time series; TAR models.

1. Introduction

Sometimes, one is faced with the problem of analyzing multivariate time series with missing data. The missing observations may belong to one or more of the time series. Here the missing data case being considered is that in which the variable(s) of interest had realizations in the sample period considered but were not physically observed. This point differs from those in which the observations are by nature unequally spaced or there is a mechanism that produces the missing data at random. A real example consists of the data basis of meteorological/hydrological official agencies, where, in some countries, is very common that some of their variables (diary rainfall, diary riverflow, etc.) have many missing observations. Here, it is clear that the variables had realizations but, for some reason, those were not measured or registered. This becomes a crucial practical problem when the climatological information is intended to be related to other non climatological variables, because missing data in the climatological variables can induce missing observations in the other variables and potential statistical models could not be fitted. Then, a procedure for estimating missing data in an optimal way is necessary.

Received April 7, 2004; Accepted September 3, 2004

Address correspondence to Fabio H. Nieto, National University of Colombia, A.A. 72157, Bogotá D.C., Colombia; E-mail: fhnetos@unal.edu.co

It is well known that, usually, the time series analysis consists of three basic steps: (1) identification of a statistical model in the presence of missing data, (2) estimation of the unknown model parameters and, when needed, of the missing data, and (3) forecasting. In the case of univariate nonlinear time series models, Tong (1990), Tong and Yeung (1991a,b), and Brockwell (1994) have worked on the problem, but only partially. Their works have the main characteristic that the underlying model is assumed to be known but the missing data estimation is not an issue.

For analyzing nonlinear multivariate time series with missing data, no results are yet available to the best knowledge of the author. For the case of consecutive or complete time series, Tong's (1990) TARSO models or Tsay's (1998) multivariate SETAR (MSETAR) models are adequate tools for their analysis. A related specification to the TARSO and MSETAR models is the so-called TAR model where one only has a variable as the input and one as the output in the underlying dynamical system. The input variable is the threshold variable and depending on the location of its values in its sample space, one has different types of responses of the output variable. This is something like a (nonlinear) transfer function model and its analysis, taking into account missing data, is the main goal of this article. For example, in the case of the real problem quoted above, a reasonable approach is to consider rainfall as an input variable and river flow as an output.

Markov Chain Monte Carlo (MCMC) methods and the Bayesian approach are used to design a methodology for analyzing the nonlinear bivariate model described above. Note that the forecasting part of the problem will not be considered in this article. The article is organized as follows. In Sec. 2, the TAR model specification and its main properties are presented, emphasizing on the likelihood function characteristics. In Sec. 3, the results about the estimation of both the model unknown parameters and the missing data are included and some highlights for model checking are given when there are missing observations. Section 4 presents the findings for model identification when the time series are complete; then, a practical strategy is sketched for the model identification in the presence of missing data. In Sec. 5, a real-life application is shown where the variables involved are the rainfall and a river flow in a certain Colombian geographical region. Finally, the last section concludes the article.

2. Model Specification

Let $\{X_t\}$ and $\{Z_t\}$ be stochastic processes such that the first one is the output and the second is the input of a dynamical system described by the equation (TAR model),

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

if Z_t belongs to the real interval $R_j = (r_{j-1}, r_j]$ for some j ($j = 1, \dots, l$), where $r_0 = -\infty$ and $r_l = \infty$. The real numbers r_j ($j = 1, \dots, l-1$) are called the threshold values of the process $\{Z_t\}$ and they define l regimes for the process $\{Z_t\}$. This equation means that the dynamic response of the variable X depends on the location of the values of variable Z in its sample space.

Additionally, $\{\varepsilon_t\}$ is a Gaussian zero-mean white noise process with variance 1 that is mutually independent of $\{Z_t\}$. The coefficients $a_i^{(j)}$ and $h^{(j)}$ ($j = 1, \dots, l$;

$i = 0, 1, \dots, k_j$) are real numbers and the non negative integer numbers k_1, \dots, k_l denote, respectively, the autoregressive orders of $\{X_t\}$ in each regime. The symbol $\text{TAR}(l; k_1, \dots, k_l)$ is used to denote this model and called l, r_1, \dots, r_{l-1} , and k_1, \dots, k_l the model structural parameters.

These kinds of models were introduced by Tong (1978) and Tong and Lim (1980), specifically in the case where the threshold variable is the lagged variable $Z_t = X_{t-d}$, for some positive integer d . In this case, the model is known as the self-exciting TAR (SETAR) model and, at present, there is a lot of literature about the analysis of these models, under the frequent assumption that we know the number l of regimes and the autoregressive orders k_1, \dots, k_l .

In this article, it is assumed that $\{Z_t\}$ is exogenous in the sense that there is no feedback of $\{X_t\}$ towards it and that the stochastic behavior of $\{Z_t\}$ is described by a homogeneous p th order Markov chain with initial distribution $F_1(z, \theta_z)$ and kernel distribution $F(z_t|z_{t-1}, \dots, z_{t-p}, \theta_z)$, where θ_z is a parameter vector in an appropriate numerical space. Furthermore, it is assumed that these distributions have densities in the Lebesgue-measure sense. Let $f_1(z, \theta_z)$ and $f(z_t|z_{t-1}, \dots, z_{t-p}, \theta_z)$ be, respectively, the initial and kernel density functions of the distributions above. Of course, all the properties of the chain $\{Z_t\}$ can be deduced from the multivariate homogeneous first-order Markov chain $\{\mathbf{Z}_t\}$ where $\mathbf{Z}_t = (Z_t, \dots, Z_{t-p+1})'$ for all $t > p - 1$. For example, if $\{Z_t\}$ has a stationary distribution, $\{\mathbf{Z}_t\}$ also has a stationary one and if $K(\cdot, \cdot)$ and $K^p(\cdot, \cdot)$ denote, respectively, the univariate and the multivariate transition kernels, then $K^p(\mathbf{z}_t, A) = K(\mathbf{z}_t, A_{p-1})$ where $\mathbf{z}_t = (z_t, \dots, z_{t-p+1}) \in \mathcal{X}^p$ and $A_{p-1} = \{y \in \mathcal{X} : (y, z_{t-1}, \dots, z_{t-p+1}) \in A\}$, with \mathcal{X} the state space of the chain $\{Z_t\}$, \mathcal{X}^p its p -dimensional cartesian product, and A an event in the σ -algebra for \mathcal{X}^p . In what follows, it is assumed that the p -dimensional Markov chain $\{\mathbf{Z}_t\}$ has an invariant or stationary distribution and that $f_p(\cdot)$ and $f_p(\cdot|\cdot)$ denote, respectively, its stationary and transition kernel densities. It is remarked here that a stationary distribution indicates that sample paths from $\{Z_t\}$ are long-term stable.

The TAR model has, additionally, the following properties or characteristics.

- (1) It describes a nonlinear relationship between the variables X and Z , which by no means one can assure a priori that it is piece-wise linear as does happen with a SETAR model. In terms of the associated stochastic processes, what one has is a dynamical system without feedback with input process $\{Z_t\}$ and output $\{X_t\}$. The piece-wise linearity is between X_t and the set of variables X_{t-1}, \dots, X_{t-k} and Z_t , where $k = \max\{k_1, \dots, k_l\}$.
- (2) With this kind of model, one can also explain certain types of heteroscedasticity in $\{X_t\}$ because a typical path from it may show bursts of large values. In this way, the TAR model is an alternative to known models for handling this kind of stylized facts, as is the case of the GARCH models.
- (3) It is an alternative to the regime-switching model of Hamilton (1994), in the sense that one can have more than two regimes in the underlying hidden Markov chain, which is represented here for the process $\{Z_t\}$, or that it can has a general state space (Meyn and Tweedie, 1993).

In order to facilitate the inference part of the TAR model analysis, it will be transformed it into a state space model with regime switching. To do that, let $\alpha_t = (X_t, X_{t-1}, \dots, X_{t-k+1})'$ be the state vector, $\omega_t = (\varepsilon_t, 0, \dots, 0)'$ be the system noise, and $\{J_t\}$ be a sequence of indicator variables such that $J_t = j$ if and only if $Z_t \in B_j$

for some j ($j = 1, \dots, l$). Also, let $H = (1, 0, \dots, 0)'$ and for each j ($j = 1, \dots, l$), let $C_j = (a_0^{(j)}, 0, \dots, 0)'$,

$$R_j = \begin{pmatrix} h^{(j)} & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

and

$$A_j = \left(\begin{array}{cccc|c} a_1^{(j)} & a_2^{(j)} & \cdots & a_{k-1}^{(j)} & a_k^{(j)} \\ & I_{k-1} & & & \mathbf{0} \end{array} \right),$$

where $a_i^{(j)} = 0$ for $i > k_j$ and I_{k-1} denotes the identity matrix of order $k - 1$. Then, the state space form for the TAR($l; k_1, \dots, k_l$) model becomes

$$X_t = H\alpha_t \tag{2}$$

as the observation equation, and

$$\alpha_t = C_{J_t} + A_{J_t}\alpha_{t-1} + R_{J_t}\omega_t \tag{3}$$

as the system or state equation, where it is understood that $C_{J_t} = C_j$, $A_{J_t} = A_j$, and $R_{J_t} = R_j$ if at time t , $J_t = j$. Note that one can consider the matrices C_{J_t} , A_{J_t} , and R_{J_t} as stochastic matrices with a finite range of possible values. This kind of nonlinear state space models, where additional to the observation and system equations there is an underlying indicator process that defines the structure of these equations and the probability distributions of the error terms, have been studied in the literature by, among others, Shumway and Stoffer (1991), Carter and Kohn (1994, 1996), and Kim and Nelson (1999). Nowadays, they are termed state space models with regime switching and can be analyzed efficiently by means of MCMC simulation procedures. Other nonlinear and non Gaussian state space models, which do not take into account the presence of a parallel indicator process as described above, have also been studied by means of MCMC methods as is the case of, among others, Carlin et al. (1992) and Geweke and Tanizaki (1999), with the latter paper being a generalization of the first one.

A main tool for the TAR model analysis to be developed below is its likelihood function; consequently, we proceed now to characterize and understand it. Let $\mathbf{y} = (\mathbf{x}, \mathbf{z})$ with \mathbf{x} and \mathbf{z} being, respectively, the observed data vectors for processes $\{X_t\}$ and $\{Z_t\}$ in the consecutive sample period from $t = 1$ up to $t = T$. Conditional on $l, r_1, \dots, r_{l-1}, k_1, \dots, k_l$, and $\mathbf{x}_k = (x_1, \dots, x_k)$, this function is given by the joint density $f(\mathbf{y}|\boldsymbol{\theta}_x, \boldsymbol{\theta}_z)$, where $\boldsymbol{\theta}_x$ denotes the vector of all the non structural parameters a 's and h 's. Now,

$$f(\mathbf{y}|\boldsymbol{\theta}_x, \boldsymbol{\theta}_z) = f(\mathbf{x}|\mathbf{z}, \boldsymbol{\theta}_x, \boldsymbol{\theta}_z)f(\mathbf{z}|\boldsymbol{\theta}_x, \boldsymbol{\theta}_z), \tag{4}$$

where

$$f(\mathbf{z}|\boldsymbol{\theta}_x, \boldsymbol{\theta}_z) = f(z_p|\boldsymbol{\theta}_z)f(z_{p+1}|\mathbf{z}_p; \boldsymbol{\theta}_z) \cdots f(z_T|\mathbf{z}_{T-1}; \boldsymbol{\theta}_z) \tag{5}$$

and

$$f(\mathbf{x}|\mathbf{z}, \boldsymbol{\theta}_x, \boldsymbol{\theta}_z) = f(x_{k+1}|\mathbf{x}_k, \mathbf{z}, \boldsymbol{\theta}_x, \boldsymbol{\theta}_z) \cdots f(x_T|x_{T-1}, \dots, x_1; \mathbf{z}, \boldsymbol{\theta}_x, \boldsymbol{\theta}_z).$$

Since the white noise process $\{\varepsilon_t\}$ is Gaussian, one gets that

$$f(\mathbf{x}|\mathbf{z}, \boldsymbol{\theta}_x, \boldsymbol{\theta}_z) = (2\pi)^{-(T-k)/2} \left[\prod_{t=k+1}^T \{h^{(j_t)}\}^{-1} \right] \exp \left(-\frac{1}{2} \sum_{t=k+1}^T e_t^2 \right), \tag{6}$$

where

$$e_t = \frac{x_t - a_0^{(j_t)} - \sum_{i=1}^{k_{j_t}} a_i^{(j_t)} x_{t-i}}{h^{(j_t)}},$$

and the sequence $\{j_t\}$ is the observed time series for the stochastic process $\{J_t\}$.

It is important to note here that, given the number of regimes l and the data \mathbf{x} and \mathbf{z} , the values of the likelihood function change if one changes the thresholds r_1, \dots, r_{l-1} , in turn $\{j_t\}$, or the autoregressive orders k_1, \dots, k_l . This points out the clear dependence of the likelihood function on the thresholds of $\{Z_t\}$ and the autoregressive orders of $\{X_t\}$. Note that the marginal likelihood function for data \mathbf{z} is determined only by the initial and kernel densities of the assumed Markov chain and this probabilistic structure has to be developed on a case-by-case basis. In expression (5), it is assumed that there is no relation between the parameters $\boldsymbol{\theta}_x$ and $\boldsymbol{\theta}_z$, in the sense that the probability mechanism generating \mathbf{z} does not depend on $\boldsymbol{\theta}_x$. Additionally, it is assumed in what follows that the marginal likelihood function for \mathbf{x} , i.e., $f(\mathbf{x}|\mathbf{z}, \boldsymbol{\theta}_x, \boldsymbol{\theta}_z)$, does not depend on $\boldsymbol{\theta}_z$.

3. Model Estimation

Because estimation of missing data is a main concern in this article, as was motivated in the Introduction, I assume now that there are missing observations in the two time series, in such a way that the observed data are located at the unequally spaced time points t_1, \dots, t_N , with $1 \leq t_1 \leq \dots \leq t_N \leq T$, for $\{X_t\}$ and at s_1, \dots, s_M , $1 \leq s_1 \leq \dots \leq s_M \leq T$, for $\{Z_t\}$. Here, the statistical problem is to estimate both the model non structural parameters and the missing observations. Of course, in other missing data situations, it can be unnecessary to estimate the missing observations, only the fixed unknown model parameters. In this case, some kind of adjustment would be necessary to do to the results in Subsec. 3.2 and the work developed in Subsec. 3.1 below must be skipped.

For the time being, it is assumed that the structural parameters l, r_1, \dots, r_{l-1} , and k_1, \dots, k_l are known, that is to say, one has identified the model. From a methodological point of view, we consider initially the problem of estimating the missing data assuming that the model non structural parameters are also known; then, we will work on the joint estimation of both the unknown parameters and the missing data.

3.1. Missing Data Estimation

One can modify the state space model of Eqs. (2)–(3) by taking into account this unequally spaced time series, defining the new observation equation

$$X_t = H_t \alpha_t + \delta_t W, \tag{7}$$

where $H_t = H$ and $\delta_t = 0$ if $t \in \{t_1, \dots, t_N\}$ and $H_t = \mathbf{0}'$ and $\delta_t = 1$ otherwise; W is a discrete random variable with $\Pr(W = w_0) = 1$ for some w_0 in the support of

the variable X . This idea of including artificial data at the missing observation time points has been used in handling linear time series with missing observations (see, e.g., Brockwell and Davis, 1991 and Peña and Maravall, 1991) and, as can be seen, this has the technical advantage of using the equally spaced Kalman filter. Since the matrix H_t for the artificial data is the zero matrix, there is no influence of these observations on the Kalman filter recursions and consequently on the missing data estimates. That idea is also useful for putting reasonable initial estimates of the missing data as would be required in this section.

As noted by Carter and Kohn (1994, 1996) and Kim and Nelson (1999), the model given by Eqs. (7) and (3) can be studied by means of the Gibbs sampler; thus, one can do inferences about unobserved states or indicator variables. For this reason, we will follow the Gibbs sampling approach to address this missing data estimation problem. Strictly, what one needs is to draw from the joint conditional density $p(\mathbf{x}_m, \mathbf{z}_m | \mathbf{x}_o, \mathbf{z}_o)$, where \mathbf{x}_m and \mathbf{z}_m denote the vectors of missing data in the time series $\{x_t\}$ and $\{z_t\}$, respectively, and \mathbf{x}_o and \mathbf{z}_o denote the vectors of corresponding observed data. Then, the idea is to compute the conditional expectation of the missing data variables given the observed ones, using draws of the above conditional density. This is the optimal estimate of the missing observations in the sense of the minimum mean square error (MMSE) criterion.

In order to accomplish the previous described goal, we briefly recall Carter and Kohn's (1994) approach. They gave a way for computing the so-called full conditional densities $p(\boldsymbol{\alpha} | \mathbf{j}, \mathbf{x})$ and $p(\mathbf{j} | \boldsymbol{\alpha}, \mathbf{x})$ for generating samples from the joint posterior density $p(\boldsymbol{\alpha}, \mathbf{j} | \mathbf{x})$, where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_T)$ is the so-called total state vector and $\mathbf{j} = (j_1, \dots, j_T)$ is the total vector of indicators (discrete variables with finite sample space). Their main results establish that

$$p(\boldsymbol{\alpha} | \mathbf{j}, \mathbf{x}) = p(\alpha_T | \mathbf{j}, \mathbf{x}) \prod_{t=1}^{T-1} p(\alpha_t | \alpha_{t+1}, \mathbf{j}, \mathbf{x}_t) \quad (8)$$

with $\mathbf{x}_t = (x_1, \dots, x_t)$, and that

$$p(\mathbf{j} | \boldsymbol{\alpha}, \mathbf{x}) = p(j_T | \boldsymbol{\alpha}, \mathbf{x}) \prod_{t=1}^{T-1} p(j_t | j_{t+1}, \mathbf{x}_t, \boldsymbol{\alpha}_t) \quad (9)$$

with $\boldsymbol{\alpha}_t = (\alpha_1, \dots, \alpha_t)$. Thus, to generate $\boldsymbol{\alpha}$ from $p(\boldsymbol{\alpha} | \mathbf{j}, \mathbf{x})$, one first generates α_T from $p(\alpha_T | \mathbf{j}, \mathbf{x})$ and then, for $t = T - 1, \dots, 1$, one generates α_t from $p(\alpha_t | \alpha_{t+1}, \mathbf{x}_t, \mathbf{j})$. The same idea applies for generating from $p(\mathbf{j} | \boldsymbol{\alpha}, \mathbf{x})$. In the first case, extensive use of Kalman-filter-based algorithms is carried out, while in the second a discrete filter is considered, which depends strongly on the distribution of the process $\{J_t\}$ (a first-order Markov chain). We note that

$$p(j_t | j_{t+1}, \mathbf{x}_t, \boldsymbol{\alpha}_t) = \Pr(J_t = j_t | J_{t+1} = j_{t+1}, \mathbf{x}_t, \boldsymbol{\alpha}_t)$$

for each $j_t = 1, \dots, l$, where l is the number of states of the chain. The same remark applies for understanding the meaning of the functions $p(\mathbf{j} | \boldsymbol{\alpha}, \mathbf{x})$ and $p(j_T | \boldsymbol{\alpha}, \mathbf{x})$.

A major point emerges here in the context of TAR models. The resulting indicator process $\{J_t\}$ is a transformation of the input process $\{Z_t\}$. While in the state-space regime-switching models literature the interest is on the indicator process itself, the concern here is instead on the process $\{Z_t\}$, which is a Markov

chain with general state space. Notice that missing observations in this process cause missing ones in the indicator process and these values can be estimated using directly expressions (8) and (9) (Carter and Kohn’s 1994, 1996, algorithms). However, the crucial problem here is how to obtain estimates of the missing values in $\{Z_t\}$ given estimates of the corresponding missing ones in $\{J_t\}$. In order to avoid this route, we shall modify Carter and Kohn’s approach by directly taking into account the process $\{Z_t\}$. That is to say, we shall concentrate the attention on the joint conditional density $p(\boldsymbol{\alpha}, \mathbf{z}|\mathbf{x})$, because the required marginal joint density $p(\mathbf{x}_m, \mathbf{z}_m|\mathbf{x}_o, \mathbf{z}_o)$ can be obtained from it, where here \mathbf{x} is constituted by $\mathbf{x}_o, \mathbf{z}_o$, and the artificial data inserted to complete the time series. The idea is that at time t , where there is not observation for X_t , one considers a draw of α_t and takes up its first component as an estimate of x_t (small letters denote data and capital ones denote random variables). The same idea applies for obtaining draws for the components of \mathbf{z}_m .

It is easy to see that expressions (8) and (9) are still valid with \mathbf{z} in place of \mathbf{j} . Indeed, the modified expression (8) is obtained following Carter and Kohn’s (1994) paper, line by line, while the modified expression (9) is given in Proposition 3.1 below.

Proposition 3.1. *Assume that the process $\{X_t\}$ obeys the state-space regime-switching model given by*

$$\begin{aligned} X_t &= H_{J_t}\alpha_t + e_t, \\ \alpha_t &= A_{J_t}\alpha_{t-1} + u_t, \end{aligned}$$

where $J_t = j$ if $Z_t \in B_j$ ($j = 1, \dots, l$), with the set family $\{B_j|j = 1, \dots, l\}$ determining a partition of the real line as indicated in Sec. 2; the noise processes $\{e_t\}$ and $\{u_t\}$ are each one serially uncorrelated and the distribution of e_t and u_t may depend on Z_t through the indicator variable J_t , $t = 1, 2, \dots$; and conditional on $\{Z_t\}$ the remaining standard assumptions on state space models hold (Harvey, 1989). Additionally, assume that (i) Z_t is independent of e_s and u_s for all $s > t$, and (ii) \mathbf{x}_t and $\boldsymbol{\alpha}_t$ are independent of Z_s for all $s > t$. Then, the full conditional density $p(\mathbf{z}|\boldsymbol{\alpha}, \mathbf{x})$ is given by

$$p(\mathbf{z}|\boldsymbol{\alpha}, \mathbf{x}) = p(\mathbf{z}_T|\boldsymbol{\alpha}, \mathbf{x}) \prod_{t=1}^{T-p} p(z_t|z_{t+p}, \mathbf{x}_t, \boldsymbol{\alpha}_t).$$

where, in general, $\mathbf{z}_t = (z_{t-p+1}, \dots, z_t)$.

Proof. One knows that

$$p(\mathbf{z}|\boldsymbol{\alpha}, \mathbf{x}) = p(\mathbf{z}_T|\boldsymbol{\alpha}, \mathbf{x})p(z_{T-p}|z_{T-p+1}, \dots, z_T, \boldsymbol{\alpha}, \mathbf{x}) \dots p(z_1|z_2, \dots, z_T, \boldsymbol{\alpha}, \mathbf{x}).$$

Now, using the state-space model equations and assumption (i) one obtains, for all $t = T - p, \dots, 1$, that

$$p(z_t|z_{t+1}, \dots, z_T, \boldsymbol{\alpha}, \mathbf{x}) = p(z_t|z_{t+1}, \dots, z_T, \boldsymbol{\alpha}_t, \mathbf{x}_t),$$

and, using assumption (ii) and that the process $\{Z_t\}$ is a homogeneous p th-order Markov chain, one has that

$$p(z_t|z_{t+1}, \dots, z_T, \boldsymbol{\alpha}_t, \mathbf{x}_t) = p(z_t|\mathbf{z}_{t+p}, \boldsymbol{\alpha}_t, \mathbf{x}_t).$$

This completes the proof.

As outlined previously, the importance of Proposition 3.1 is in that it signals how to draw random samples from $p(\mathbf{z}|\mathbf{x}, \boldsymbol{\alpha})$ via the conditional marginal densities $p(\mathbf{z}_T|\mathbf{x}, \boldsymbol{\alpha})$ and $p(z_t|\mathbf{z}_{t+p}, \mathbf{x}_t, \boldsymbol{\alpha}_t)$ ($t = T - p, \dots, 1$). The key part now is how to draw samples from these conditional densities and Proposition 3.2 below shows how to do that.

Proposition 3.2. *With the notation and assumptions in Proposition 3.1, one has that*

$$p(\mathbf{z}_T|\boldsymbol{\alpha}, \mathbf{x}) \propto \prod_{j=T-p+1}^T p(\alpha_j|\mathbf{z}_T, \alpha_{j-1})f_p(\mathbf{z}_T)$$

and that for $t = T - p, \dots, 1$,

$$p(z_t|\mathbf{z}_{t+p}, \boldsymbol{\alpha}_t, \mathbf{x}_t) \propto p(\alpha_t|\mathbf{z}_{t+p-1}, \alpha_{t-1})f_p(z_{t+p}|\mathbf{z}_{t+p-1})f_p(\mathbf{z}_{t+p-1}).$$

Proof. Initially, we note that

$$p(\mathbf{z}_T|\boldsymbol{\alpha}, \mathbf{x}) \propto p(x_{T-p+1}, \dots, x_T|\boldsymbol{\alpha}_p, \mathbf{z}_T)p(\boldsymbol{\alpha}_p|\alpha_{T-p}, \mathbf{z}_T)f_p(\mathbf{z}_T),$$

where $\boldsymbol{\alpha}_p = (\alpha_{T-p+1}, \dots, \alpha_T)$, and that

$$p(\boldsymbol{\alpha}_p|\alpha_{T-p}, \mathbf{z}_T) = \prod_{j=T-p+1}^T p(\alpha_j|\mathbf{z}_T, \alpha_{j-1}).$$

For each $t = T - p + 1, \dots, 1$, if X_t is observed,

$$\begin{aligned} p(\mathbf{z}_{t+p-1}|\boldsymbol{\alpha}_t, \mathbf{x}_t) &\propto p(\mathbf{z}_{t+p-1}, x_t, \alpha_t|\boldsymbol{\alpha}_{t-1}, \mathbf{x}_{t-1}) \\ &= p(x_t|\alpha_t, \mathbf{z}_{t+p-1})p(\alpha_t|\mathbf{z}_{t+p-1}, \alpha_{t-1})f_p(\mathbf{z}_{t+p-1}), \end{aligned}$$

and if X_t is not observed,

$$\begin{aligned} p(\mathbf{z}_{t+p-1}|\boldsymbol{\alpha}_t, \mathbf{x}_t) &= p(\mathbf{z}_{t+p-1}|\boldsymbol{\alpha}_t, \mathbf{x}_{t-1}) \\ &\propto p(\mathbf{z}_{t+p-1}, \alpha_t|\boldsymbol{\alpha}_{t-1}, \mathbf{x}_{t-1}) \\ &= p(\alpha_t|\alpha_{t-1}, \mathbf{z}_{t+p-1})f_p(\mathbf{z}_{t+p-1}). \end{aligned}$$

Now, for $t = T - p, \dots, 1$,

$$\begin{aligned} p(z_t|\mathbf{z}_{t+p}, \boldsymbol{\alpha}_t, \mathbf{x}_t) &\propto p(z_{t+p}|\mathbf{z}_{t+p-1}, \boldsymbol{\alpha}_t, \mathbf{x}_t)p(\mathbf{z}_{t+p-1}|\boldsymbol{\alpha}_t, \mathbf{x}_t) \\ &= f_p(z_{t+p}|\mathbf{z}_{t+p-1})p(\mathbf{z}_{t+p-1}|\boldsymbol{\alpha}_t, \mathbf{x}_t). \end{aligned}$$

Finally, if in the observation equation of the state-space regime-switching model there is no error term, i.e., $\Pr(e_t = 0) = 1$ for all t , as is the case of TAR processes, one has that conditional on \mathbf{z}_{t+p-1} and α_t , the distribution of X_t is degenerate, that is $p(x_t|\alpha_t, \mathbf{z}_{t+p-1}) = 1$ ($t = T - p + 1, \dots, 1$), and also $p(x_{T-p+1}, \dots, x_T|\alpha_p, \mathbf{z}_T) = 1$. This ends the proof.

Jointly, Propositions 3.1 and 3.2 generalize Carter and Kohn's (1994) result given by expression (9), in the sense that the underlying Markov chain is of order $p \geq 1$ and with general state space. Of course, they are also extensions of the classical smoothing algorithms.

The results given in the above Propositions suggest that one can use accept reject or Metropolis-Hastings algorithms (Chib and Greenberg, 1995; Robert and Casella, 1999) for drawing from $p(\mathbf{z}_T|\mathbf{x}, \boldsymbol{\alpha})$ and $p(z_t|\mathbf{z}_{t+p}, \mathbf{x}_t, \boldsymbol{\alpha}_t)$ ($t = T - p, \dots, 1$), taking $f_p(\cdot)$ as a reasonable proposal density. It is important to note here that if at time t , Z_t was observed, one does not need to generate a simulated value for this variable. Consequently, in the backward recursive procedure described above for generating \mathbf{z} one can skip the time point t .

To draw from $p(\boldsymbol{\alpha}|\mathbf{z}, \mathbf{x})$, one must take into account the following points. First, for running the conditional Kalman filter, we propose to take as initial conditions $\alpha_0 = m\mathbf{1}$ and $\text{Var}(\alpha_0) = s^2I_k$, where m denotes the median of the time series $\{x_t\}$, s is the median of the absolute deviations with respect to this data median, $\mathbf{1}$ is a column of ones, and I_k is the identity matrix of order k . Additionally, for completing the x - and z -data in a preliminary way and so to run the first iteration of the Gibbs sampler, we propose to take the medians of the corresponding observed data. Using the median is recommended because this is a non parametric robust estimate of the central tendency of variable X . The problem at hand may suggest another reasonable initial estimates. As was noted previously, these initial missing data estimates have no impact on the final estimates. Second, as can be seen in Proposition 3.2, one needs to compute the distribution of $\alpha_t|\alpha_{t-1}, \mathbf{z}_{t+p-1}$. For the state space model given by Eqs. (7) and (3), this is a singular multivariate normal distribution (Anderson, 1984) with mean $C_{j_t} + A_{j_t}\alpha_{t-1}$ and variance matrix $R_{j_t}\Sigma R'_{j_t}$, where

$$\Sigma = \begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Consequently, the density $p(\alpha_t|\alpha_{t-1}, \mathbf{z}_{t+p-1})$ collapses to $\phi(e_t)/R_{j_t}(1, 1)$, with $\phi(\cdot)$ denoting the standard normal density, e_t is the residual in (6), which has the equivalent definition given by

$$e_t = (x_t - C_{j_t}(1) - A_{j_t,1}\alpha_{t-1})/R_{j_t}(1, 1),$$

with $C_{j_t}(1)$ the first component of C_{j_t} , $R_{j_t}(1, 1)$ the entry (1, 1) of matrix R_{j_t} , and $A_{j_t,1}$ the first row of matrix A_{j_t} .

Knowing the full conditional densities $p(\boldsymbol{\alpha}|\mathbf{z}, \mathbf{x})$ and $p(\mathbf{z}|\boldsymbol{\alpha}, \mathbf{x})$ and the way for drawing from them, one obtains a sample of size G from $p(\boldsymbol{\alpha}, \mathbf{z}|\mathbf{x})$, which is denoted for future reference as $(\boldsymbol{\alpha}^{(1)}, \mathbf{z}^{(1)}), \dots, (\boldsymbol{\alpha}^{(G)}, \mathbf{z}^{(G)})$. Note that the set $\{\boldsymbol{\alpha}^{(1)}, \dots, \boldsymbol{\alpha}^{(G)}\}$ can be considered as a sample of size G from the marginal posterior density $p(\boldsymbol{\alpha}|\mathbf{x})$ and the set $\{\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(G)}\}$ as one of $p(\mathbf{z}|\mathbf{x})$. Successively, one can obtain samples

for the required missing-data marginal posterior densities, extracting the values that correspond to the missing data components. The main goal is to compute the conditional means and variances of the distributions for the missing data variables. Additionally, with the generated samples one can obtain $100(1-c)\%$ credible intervals for the missing observations, for some c , $0 < c < 1$, where such intervals are given by $[q_{c/2}, q_{1-c/2}]$, with q_c the c th sample quantile.

One important point is mentioned here. For $t \in \{t_1, \dots, t_N\}$, the drawing of α_t always gives x_t in its first component, which is the actually observed data. This fact can be shown noting that the corresponding MMSE updating matrix in the backward recursive procedure is the zero matrix. This means, in practice, that it is not necessary to draw from the conditional distribution for X_t or, equivalently, that one can skip this step in the aforementioned recursive algorithm. This result is imitated in the procedure for drawing from the conditional distributions $p(z_t | \mathbf{z}_{t+p}, \mathbf{x}_t, \alpha_t)$ when at time t one has an observation for Z_t .

In summary, the procedure for estimating missing observations in time series that obey a TAR model, when the model parameters are known, can be implemented as follows.

- Step 1.** Complete the x - and z -time series with initial values for the missing data.
- Step 2.** Based on these two completed time series, run the Gibbs sampler as indicated in Propositions 3.1 and 3.2 for generating random draws from the missing data posterior distributions.
- Step 3.** With those samples, compute the means, variances, and credible intervals at the level desired. The sample means are the estimates of the missing data and the sample variances are their MMSEs.

3.2. Parameters Estimation

The results obtained in the previous subsection are extended now to deal with both the parameters and missing data estimation, which is a real practical problem. The methodology is also based on the Gibbs sampler for state space models with regime switching. As will be noted, the estimation of the missing observations is an intrinsic step as happens with the classical Expectation Maximization (EM) algorithm.

Let θ be the vector of total unknown parameters in the TAR model, including those for the threshold process $\{Z_t\}$. Assuming that θ is known, the so-called full conditional densities $p(\alpha | \mathbf{z}, \mathbf{x}, \alpha_0, \theta)$ and $p(\mathbf{z} | \alpha, \mathbf{x}, \alpha_0, \theta)$ were obtained in the previous subsection, in order to generate samples from the joint posterior density $p(\alpha, \mathbf{z} | \mathbf{x}, \theta, \alpha_0)$, where α_0 is the initial state vector. From now on, and unless the contrary is indicated, we omit writing α_0 in the conditioning sets. The statistical problem here is to estimate both θ and the missing observations \mathbf{x}_m and \mathbf{z}_m ; consequently, our specific interest is in obtaining the joint conditional density $p(\theta, \mathbf{x}_m, \mathbf{z}_m | \mathbf{x}_o, \mathbf{z}_o)$. With this end, we consider again the state space form of the TAR model and concentrate in obtaining the joint conditional density $p(\theta, \alpha, \mathbf{z} | \mathbf{x})$, which $p(\theta, \mathbf{x}_m, \mathbf{z}_m | \mathbf{x}_o, \mathbf{z}_o)$ can be obtained from.

Usually, the full conditional density $p(\theta | \alpha, \mathbf{y})$ is obtained according to the particular case at hand. In the case of TAR models, we must compute the full conditional densities for the unknown model parameters $a_i^{(j)}$ and $h^{(j)}$ ($j = 1, \dots, l$; $i = 0, 1, \dots, k_j$) and those of the distribution of $\{Z_t\}$. With this goal, let $\theta_j = (a_0^{(j)}, a_1^{(j)}, \dots, a_{k_j}^{(j)})'$ ($j = 1, \dots, l$) and $\mathbf{h} = (h^{(1)}, \dots, h^{(l)})'$. Then $\theta_x = (\theta_1, \dots, \theta_l, \mathbf{h})$

and $\theta = (\theta_x, \theta_z)$. Notice that if $\{Z_t\}$ is a discrete Markov chain, the case reduces to that of Carter and Kohn (1994).

Instead of considering the conditional density $p(\theta|\alpha, \mathbf{y})$, we shall compute, equivalently, the conditional densities

$$p(\theta_j|\theta_i; i \neq j; \mathbf{h}, \theta_z, \alpha, \mathbf{y}),$$

$p(h^{(j)}|h^{(i)}; i \neq j; \theta_z, \theta_1, \dots, \theta_l, \alpha, \mathbf{y})$ ($j = 1, \dots, l$) and $p(\theta_z|\theta_x, \alpha, \mathbf{y})$. Following Geweke and Terui (1993) and Chen and Lee (1995), in a similar problem for SETAR models, we assume that *a priori* the parameters among regimes are independent, that θ_j and $h^{(j)}$ are also independent, as are θ_x and θ_z . These are simplifying assumptions that allow us to obtain reasonable results. In the case of the assumed independence of the parameters θ_j 's, what one is saying is that the dynamic behavior of the process $\{X_t\}$ in a regime depends only on the values of $\{Z_t\}$. Letting some kind of dependence among the θ_j 's parameters might involve potential identification problems. Nevertheless, this is an interesting point for future research.

We take as prior densities for θ_x the following

$$\theta_j \sim N(\theta_{0,j}, V_{0,j}^{-1})$$

and $[h^{(j)}]^2 \sim IG(\gamma_{0,j}, \beta_{0,j})$ ($j = 1, \dots, l$), where "IG" stands for inverse Gamma distribution. Choosing a Gaussian prior for θ_j is very reasonable because it can reflect either some precise knowledge about the actual value of this parameter or a large uncertainty about it, which is explained by a large variance matrix. Even more, the Bayesian theory about conjugate distributions helps to support this choice.

Under the above considerations, we have the following results.

Proposition 3.3. For each $j = 1, \dots, l$, let $\{t : j_t = j\} = \{t_{1,j}, \dots, t_{n_j,j}\}$, with n_j denoting the number of observations in regime j . With the above assumptions and prior distribution specified for θ_j , its full conditional distribution given $\theta_i; i \neq j; \mathbf{h}, \theta_z, \alpha$, and \mathbf{y} is multinormal with mean

$$\theta_j^* = V_j^{-1}[(1/\{h^{(j)}\}^2)\mathbf{W}'_j\mathbf{X}_j + V_{0,j}\theta_{0,j}], \tag{10}$$

and variance

$$V_j = [1/\{h^{(j)}\}^2]\mathbf{W}'_j\mathbf{W}_j + V_{0,j}, \tag{11}$$

where $\mathbf{w}'_t = (1, x_{t-1}, \dots, x_{t-k_j})$, $\mathbf{W}_j = (\mathbf{w}_{t_{1,j}}, \dots, \mathbf{w}_{t_{n_j,j}})'$, and $\mathbf{X}_j = (x_{t_{1,j}}, \dots, x_{t_{n_j,j}})'$.

Proof. For each $j = 1, \dots, l$,

$$\begin{aligned} p(\theta_j|\theta_i; i \neq j; \mathbf{h}, \theta_z, \alpha, \mathbf{y}) &\propto L(\mathbf{x}|\theta, \alpha, \mathbf{z})p(\alpha|\theta, \mathbf{z})\pi(\theta_j) \\ &\propto p(\alpha|\theta, \mathbf{z})\pi(\theta_j), \end{aligned}$$

where $\pi(\cdot)$ is the corresponding prior density. Notice that $L(\cdot|\cdot)$ becomes a product of degenerate densities that do not depend on the model parameters. Since

$$p(\alpha|\theta, \mathbf{z}) = p(\alpha_1|\alpha_0, \theta, \mathbf{z})p(\alpha_2|\alpha_1, \theta, \mathbf{z})\dots p(\alpha_T|\alpha_{T-1}, \theta, \mathbf{z})$$

and each factor in this right-side product is a singular multivariate normal density, which in turn is concentrated on a non singular univariate density given before in Subsec. 3.1, we obtain that

$$p(\boldsymbol{\alpha}|\boldsymbol{\theta}, \mathbf{z}) \propto \exp\left[-\frac{1}{2} \sum_{\{t:j_t=j\}} \left\{x_t - a_0^{(j)} - \sum_{i=1}^{k_j} a_i^{(j)} x_{t-i}\right\}^2 / \{h^{(j)}\}^2\right].$$

Now, the sum in the above exponent is equal to the quadratic form

$$(\mathbf{X}_j - \mathbf{W}_j \boldsymbol{\theta}_j)' [\{h^{(j)}\}^2 \mathbf{I}_{n_j}] (\mathbf{X}_j - \mathbf{W}_j \boldsymbol{\theta}_j).$$

Since

$$\pi(\boldsymbol{\theta}_j) \propto \exp\left[-\frac{1}{2} (\boldsymbol{\theta}_j - \boldsymbol{\theta}_{0,j})' V_{0,j} (\boldsymbol{\theta}_j - \boldsymbol{\theta}_{0,j})\right],$$

we obtain, after some algebraic operations in the exponents, that

$$p(\boldsymbol{\theta}_j | \boldsymbol{\theta}_i; i \neq j; \mathbf{h}, \boldsymbol{\theta}_z, \boldsymbol{\alpha}, \mathbf{y}) \propto \exp\left[-\frac{1}{2} (\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*)' V_j (\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*)\right], \tag{12}$$

where

$$V_j = [1/\{h^{(j)}\}^2] \mathbf{W}'_j \mathbf{W}_j + V_{0,j}$$

and

$$\boldsymbol{\theta}_j^* = V_j^{-1} [(1/\{h^{(j)}\}^2) \mathbf{W}'_j \mathbf{X}_j + V_{0,j} \boldsymbol{\theta}_{0,j}],$$

which corresponds to the kernel of a multivariate normal density with mean $\boldsymbol{\theta}_j^*$ and variance V_j^{-1} . This completes the proof.

Proposition 3.4. *With the same assumptions in Proposition 3.3 and the prior distribution established previously for $[h^{(j)}]^{-2}$, it is found that its full conditional density given $\boldsymbol{\theta}_j$ ($j = 1, \dots, l$), $\boldsymbol{\theta}_z$, $\boldsymbol{\alpha}$, and \mathbf{y} , corresponds to that of a Gamma distribution with shape parameter $\gamma_{0,j} + n_j/2$ and scale parameter*

$$\beta_{0,j} + (1/2)(\mathbf{Y}_j - \mathbf{W}_j \boldsymbol{\theta}_j)' (\mathbf{Y}_j - \mathbf{W}_j \boldsymbol{\theta}_j).$$

Proof. Since

$$p([h^{(j)}]^{-2} | \boldsymbol{\theta}_{-j,h}, \boldsymbol{\alpha}, \mathbf{y}) \propto p(\boldsymbol{\alpha} | \boldsymbol{\theta}, \mathbf{z}) \pi([h^{(j)}]^{-2}),$$

where $\boldsymbol{\theta}_{-j,h}$ denotes the vector $\boldsymbol{\theta}$ without $h^{(j)}$, and

$$\pi([h^{(j)}]^{-2}) \propto \frac{1}{[h^{(j)}]^{2(\gamma_{0,j}-1)}} \exp[-\beta_{0,j}/\{h^{(j)}\}^2],$$

then $p([h^{(j)}]^{-2}|\theta_{-j,h}, \alpha, \mathbf{y})$ is directly proportional to the kernel of a Gamma distribution with shape parameter $\gamma_{0,j} + n_j/2$ and scale parameter

$$\beta_{0,j} + (1/2)(\mathbf{Y}_j - \mathbf{W}_j\theta_j)'(\mathbf{Y}_j - \mathbf{W}_j\theta_j).$$

This ends the proof.

One can see that for the case of SETAR models, similar expressions to (10)–(11) and to the parameters given by Proposition 3.4 were obtained by Geweke and Terui (1993) and Chen and Lee (1995), via the so-called arranged autoregression approach.

In this article, and with simplifying purposes only, we recommend estimating θ_z in accordance with the problem at hand and following the lines of the empirical Bayesian approach. Consequently, this estimation has to be done before the estimation of the remaining parameters. In this way, one has the set of full conditional densities for running the Gibbs sampler for both parameter and missing data estimation in the TAR model. This set is given by $p(\alpha|\theta_x, \mathbf{y})$, $p(\mathbf{z}|\theta_x, \alpha, \mathbf{x})$, and $p(\theta_x|\alpha, \mathbf{y})$, and from the first two, one obtains the marginal densities $p(\mathbf{x}_m|\theta_x, \mathbf{x}_o, \mathbf{z}_o)$ and $p(\mathbf{z}_m|\theta_x, \mathbf{x}_o, \mathbf{z}_o)$.

3.3. Model Adequacy

For model checking, we follow some ideas developed by Tong (1990) and Tong and Yeung (1991a) in the SETAR models setting. For each $t = 1, \dots, T$, let $\hat{e}_t = (X_t - X_{t|t-1})/h^{(j)}$ if $Z_t \in B_j$ for some j ($j = 1, \dots, l$), where $X_{t|t-1}$ is the one-step-ahead predictor of X_t . It can be seen that

$$X_{t|t-1} = a_0^{(j)} + \sum_{i=1}^{k_j} a_i^{(j)} X_{t-i|t-1},$$

where $X_{t-i|t-1} = X_{t-i}$ if at $t - i$ there is an observation for process $\{X_t\}$ and we set $X_{t-i|t-1} = \widehat{X}_{t-i}$, with \widehat{X}_{t-i} the estimator of x_{t-i} , otherwise. Additionally, let $O_t = \{i|i = 1, \dots, k_j, \text{ and } t - i \notin \{t_1, \dots, t_N\}\}$, then

$$\hat{e}_t = \begin{cases} \sum_{i \in O_t} a_i^{(j)} \frac{(X_{t-i} - \widehat{X}_{t-i})}{h^{(j)}} + \varepsilon_t, & \text{if } O_t \neq \Phi \\ \varepsilon_t, & \text{if } O_t = \Phi, \end{cases}$$

with Φ the empty set. For future reference, the process $\{\hat{e}_t\}$ is called standardized pseudo residuals (SPR), and as can be seen, is not necessarily a white noise process as $\{\varepsilon_t\}$ is. For now, we propose to use this process in an exploratory way for checking heteroscedasticity in $\{\varepsilon_t\}$ and model specification. To this aim, we recommend using CUSUM and CUSUMSQ charts. Obviously, the distribution of the usual statistical tests in this crucial case deserves to be investigated in future research.

4. Identification of the TAR Model

The results presented in Secs. 2 and 3 were obtained under the assumption that one knows the model structural parameters. In practice, the first step in the model analysis consists in identifying these parameters. This is the aim of this section, where, for the time being, I assume that there are no missing observations in the time series. In Sec. 5, we shall describe a practical strategy to deal with the missing data case.

Given the threshold process, the key problems in the context of identifying TAR models are (1) the determination of the number of regimes and (2) the specification of the autoregressive orders in each regime. In the literature, these parameters are almost always assumed to be known. In the line of problem (1), McCulloch and Tsay (1994) have proposed a procedure that uses the concept of multiple structural changes for SETAR processes, being flexible, in the authors words, about both the location and the number of possible thresholds, i.e., without estimating them.

Although in this article we are concerned only with the TAR-model philosophy described above, it is convenient to say that the TAR model can be cast within the context of the so-called open-loop dynamical systems, in Tong's (1990) sense, or the multivariate threshold models, in Tsay's (1998) sense, but without covariates in both cases. In these two settings, we find that the problems quoted above are addressed basically under empirical considerations and strong use of the Akaike Information Criterion (AIC) information criterion. Indeed, Tong (1990) proposes to use a non parametric method to find some relation between the process of interest and the chosen threshold process and, based on this observation and expertise knowledge, one may postulate the number of regimes and threshold locations, these last ones being reestimated via a conditional least-squares procedure combined with the information criterion. The information criterion is also used for choosing the autoregressive orders. Tsay (1998) proposes a building model procedure for identifying a nonlinear model, which is similar to that developed by himself (Tsay, 1989) for univariate SETAR models. His approach begins with the use of a nonlinearity test and follows with the use of the information criterion for adjusting initial entertained values for the thresholds. As Tong (1990), he chooses autoregressive orders with AIC.

Returning to the TAR modeling scope, one can use the MCMC and Bayesian approaches for solving the problems (1) and (2) as an alternative to the use of empirical considerations and the AIC criterion. For the location of the thresholds, the likelihood-function based philosophy is maintained in a similar way to that used by Hausman et al. (1992) in a financial time series context. With the MCMC and Bayesian perspectives, one can set prior distributions on the number of regimes and the autoregressive orders for each regime and then proceed to obtain the posterior ones, in order to compute the desired optimal estimates.

4.1. Identification of the Number of Regimes

Following Carlin and Chib (1995) or Green (1995), the problem of identifying l can be cast as a Bayesian model choice problem via the MCMC approach. With this wisdom, let L be a discrete random variable from which the number of regimes l is a realization and, conditional on l , let K_{1l}, \dots, K_{ll} be the discrete random variables generating the autoregressive orders k_{1l}, \dots, k_{ll} . Let us assume that L takes values on the set $\{2, 3, \dots, l_0\}$, for a known l_0 , with a prior distribution $\pi = \{\pi_2, \dots, \pi_{l_0}\}$ and that the random variables K_{1l}, \dots, K_{ll} take

values on the respective sets $\{0, 1, \dots, \bar{k}_{il}\}$ ($i = 1, \dots, l$), with corresponding prior distributions $\mathbf{q}_i = \{q_{i0}, q_{i1}, \dots, q_{i\bar{k}_i}\}$ ($i = 1, \dots, l$). Then, one can say that there are $l_0 - 1$ candidate models for the data \mathbf{y} , which is denoted as M_j ($j = 2, \dots, l_0$), and that they are parameterized, respectively, by the row vectors

$$\theta_j = (\boldsymbol{\theta}_{x,j}, \mathbf{k}_j, \mathbf{r}_j),$$

where $\mathbf{k}_j = (k_{1j}, \dots, k_{jj})$, with k_{ij} the autoregressive order of regime i ($i = 1, \dots, j$) and $\mathbf{r}_j = (r_{1j}, \dots, r_{j-1,j})$, and r_{ij} the i th threshold ($i = 1, \dots, j - 1$) of model j . $\boldsymbol{\theta}_{x,j}$ denotes the vector of nonstructural parameters for model M_j . Notice that we have not included in this parameter vector the process $\{Z_t\}$ because it is assumed to be known. In a more compact way, one can set $\Theta = (\theta'_2, \dots, \theta'_{l_0})'$ as the matrix of all the possible parameters in a certain numeric matrix space.

In what follows, one needs to be careful in interpreting two-argument functions where, for example, the first argument is a parameter θ with an interval in the real line as its parameter space and the second one is a parameter d that can take discrete values. Then the joint distribution of θ and d is represented by the function $p(\theta, d) = f(\theta|d)g(d)$, where $f(\cdot|d)$ is the conditional density of θ given d (which is assumed to exist) and $g(\cdot)$ is the probability mass function of d .

The interest is on the posterior distribution $p(l|\mathbf{y})$, which can be obtained marginally from the joint posterior distribution $p(\text{vec } \Theta, l|\mathbf{y})$ ($l = 2, \dots, l_0$). Here, “vec” denotes the operation of stacking the columns of a matrix into a column vector. To use the Gibbs sampler for drawing from $p(\text{vec } \Theta, l|\mathbf{y})$, one needs to know the full conditional distributions $p(\text{vec } \Theta|l, \mathbf{y})$ and $p(l|\text{vec } \Theta, \mathbf{y})$ or, equivalently, the set of distributions $p(\theta_j|\theta_i; i \neq j; l, \mathbf{y})$ ($j = 2, \dots, l_0$) and $p(l|\text{vec } \Theta, \mathbf{y})$. Using Carlin and Chib (1995) approach, one obtains for TAR models that

$$p(\theta_j|\theta_i; i \neq j; l, \mathbf{y}) \propto \begin{cases} f(\mathbf{y}|\theta_i, l)p(\theta_i|l) & \text{if } j = l \\ p(\theta_j|l) & \text{if } j \neq l \end{cases}$$

and that

$$p(l|\text{vec } \Theta, \mathbf{y}) = c^{-1} f(\mathbf{y}|\theta_l, l) \left[\prod_{j=2}^{l_0} p(\theta_j|l) \right] \pi_l, \quad (l = 2, \dots, l_0),$$

with c the normalization constant. The distributions $p(\theta_j|l)$, $j \neq l$, are called the link distributions and they should be chosen with some care in order to have a reasonable rate of convergence of the underlying Gibbs sampling (see Carlin and Chib, 1995 paper for details).

Additionally, one can decompose the full conditional distributions

$$p(\theta_j|\theta_i; i \neq j; l, \mathbf{y})$$

in the following equivalent set of full distributions: $p(\boldsymbol{\theta}_{x,j} | (\text{vec } \Theta)_{-(x,j)}, l, \mathbf{y})$, $p(\mathbf{k}_j | (\text{vec } \Theta)_{-(k,j)}, l, \mathbf{y})$, and $p(\mathbf{r}_j | (\text{vec } \Theta)_{-(r,j)}, l, \mathbf{y})$, where $(\text{vec } \Theta)_{-(\dots)}$ denotes the vector $\text{vec } \Theta$ without the subvector indicated in the argument of p and represented by the symbol “(.,.)”.

The identification of the thresholds will be made via minimization of Tong’s (1990) NAIC information criterion by means of a search among quantiles of

the empirical distribution of the data \mathbf{z} in a route similar to that of Hausman et al. (1992). The idea is to postulate a maximum number of regimes l_0 and then to begin a grid search for thresholds in this way: for $l = 2$, we look for r_1 in the specified quantiles set, for $l = 3$, one looks for ordered pairs $r_1 < r_2$, and so on. For computing the NAIC information criterion, which encompasses somehow the conditional likelihood function in expression (6), one needs to know autoregressive orders, as an intrinsic step. To this end, one can set maximum orders for corresponding regimes and thus the grid search will take into account, simultaneously, values of the thresholds and the autoregressive orders. The real data example ahead will illustrate this strategy.

With the above considerations, the set of full conditional distributions for running the Gibbs sampler is given by

$$p(\theta_{x,j} | (\text{vec } \Theta)_{-(x,j)}, l, \mathbf{y}) \propto \begin{cases} f(\mathbf{y} | \theta_l, l) p(\theta_l | l) & \text{if } j = l \\ p(\theta_j | l) & \text{if } j \neq l \end{cases} \quad (13)$$

$$p(\mathbf{k}_j | (\text{vec } \Theta)_{-(k,j)}, l, \mathbf{y}) \propto \begin{cases} f(\mathbf{y} | \theta_l, l) p(\mathbf{k}_l | l) & \text{if } j = l \\ p(\mathbf{k}_j | l) & \text{if } j \neq l, \end{cases} \quad (14)$$

and

$$p(l | \text{vec } \Theta, \mathbf{y}) = c^{-1} f(\mathbf{x} | \mathbf{z}, \theta_l, l) \left[\prod_{j=2}^{l_0} p(\theta_j | l) \right] \pi_l, \quad (l = 2, \dots, l_0), \quad (15)$$

with

$$c = \sum_{l'=2}^{l_0} f(\mathbf{x} | \mathbf{z}, \theta_{l'}, l') \left[\prod_{j=2}^{l_0} p(\theta_j | l') \right] \pi_{l'}.$$

The densities given by expression (15) can be computed using the results about estimation of the non structural parameters of the TAR model presented in Sec 3.

Let

$$d(i, l) = \frac{f(\mathbf{x} | \mathbf{z}, \theta_l, l) p(k_{il} | l)}{\sum_{k'_{il}=0}^{\bar{k}_l} f(\mathbf{x} | \mathbf{z}, \theta_l, l) p(k'_{il} | l)}.$$

Expression (16) can be decomposed and simplified even more as

$$p(k_{ij} | (\text{vec } \Theta)_{-(k_{ij})}, l, \mathbf{y}) = \begin{cases} d(i, l) & \text{if } j = l \\ p(k_{ij} | \theta_l, l) & \text{if } j \neq l, \end{cases} \quad (16)$$

where $i = 1, \dots, j, k_{ij} = 0, 1, \dots, \bar{k}_j, \bar{k}_j$ is a prespecified maximum autoregressive order for all the regimes of model M_j , and $p(k_{ij} | \theta_l, l)$ is the link probability function for the autoregressive order K_{ij} . We recommend taking as the link distributions for the autoregressive orders, the priors \mathbf{q}_i specified above. Another simplification of expression (18), consequently of expression (16), can be accomplished if one notes that $\prod_{l=k+1}^T [h^{(j)}]^{-1}$ does not depend on k_{ij} . Thus, the likelihood function for the

x -data in the numerator in expression (18) is reduced to $\exp(-0.5 \sum_{t=k+1}^T e_t^2)$, with e_t given in Sec. 3, as also happens within each summand in the denominator. Here, one can note that if $\exp(-0.5 \sum_{t=k+1}^T e_t^2)$ is very large, then there is little likelihood about k_{ij} and this in turn implies large uncertainty about this parameter. This fact means that practically the uniform distribution for the posterior of K_{ij} is obtained.

One can also reach a simplification of Eq. (17) noting that $\prod_{j=2}^{l_0} p(\theta_j|l)$ does not depend on l if one sets $p(\theta_j|l) = p(\theta_j|j)$ ($j = 2, \dots, l_0$), i.e., for $j \neq l$ one takes “false” priors while for $j = l$ one puts the “correct” prior. As I wrote before, these priors for the vectors $\theta_{x,j}$ can be chosen as in Sec. 3 and take discrete distributions on the sets $\{0, 1, \dots, \bar{k}_{ij}\}$ ($i = 1, \dots, j; j = 2, \dots, l_0$) for K_{ij} , with \bar{k}_{ij} a specified maximum autoregressive order for regime i of model M_j . In this work, we recommend setting $\bar{k}_{ij} = \bar{k}_j$ ($i = 1, \dots, j$). Essentially, I have proved the following result.

Proposition 4.1. *The full conditional frequency function for L is given by*

$$p(l|\text{vec } \Theta, \mathbf{y}) = c^{-1} f(\mathbf{x}|\mathbf{z}, \theta_l, l) \pi_l, \quad (l = 2, \dots, l_0), \tag{17}$$

with

$$c = \sum_{l'=2}^{l_0} f(\mathbf{x}|\mathbf{z}, \theta_{l'}, l') \pi_{l'}.$$

For each $l = 2, \dots, l_0$, let $\hat{p}_l^{(i)}$ be the value of $p(l|\text{vec } \Theta, \mathbf{y})$ at the i th iterate of the Gibbs sampler. We suggest taking either the mean or the median of the sample $\{\hat{p}_l^{(i)}\}_i$ as the final estimate of $p(l|\text{vec } \Theta, \mathbf{y})$, \hat{p}_l say, and as the identified number of regimes, \hat{l} say, the mode of the distribution $\{\hat{p}_2, \dots, \hat{p}_{l_0}\}$. Another plausible alternative is to take as \hat{l} , the mode of the sample $\{\hat{l}^{(i)}\}_i$, where $\hat{l}^{(i)}$ is the simulated value for L at iteration i . In this stage of the identification procedure, one does not need to track the draws of the remaining parameters. They are reestimated in subsequent steps as indicated below.

4.2. Estimation of the Autoregressive Orders

Conditional on \hat{l} , the estimated number of regimes, we now proceed to estimate $k_1, \dots, k_{\hat{l}}$. For this goal, we take appropriate prior distributions for the random variables $K_1, \dots, K_{\hat{l}}$ on the respective sets $\{0, 1, \dots, \bar{k}_1\}, \dots, \{0, 1, \dots, \bar{k}_{\hat{l}}\}$. These distributions can also be taken as the link distributions $p(k_{\hat{i}}|\theta_{\hat{i}}, \hat{l})$ ($i = 1, \dots, \hat{l}$) of expression (18). For short, we shall write in what follows k_i in place of $k_{\hat{i}}$.

The Gibbs sampler for the estimation of the autoregressive orders is obtained from that for identifying l , i.e., from expressions (15), (18), and (19), just that now I suppress the full conditional density (19). That is to say, I have the following result.

Proposition 4.2. *Let $l := \hat{l}$. The full conditional frequency functions for the autoregressive orders K_1, \dots, K_l are given by*

$$p(k_i | (\text{vec } \Theta)_{-(k_i)}, l, \mathbf{y}) = f(\mathbf{x}|\mathbf{z}, \theta_l, l) p(k_i|l) / \sum_{k'_i=0}^{\bar{k}_i} f(\mathbf{x}|\mathbf{z}, \theta_l, l) p(k'_i|l),$$

where $i = 1, \dots, j$; $k_i = 0, 1, \dots, \bar{k}_i$; \bar{k}_i is a prespecified maximum autoregressive order for all the regimes, and $p(k_i|l)$ is a prior frequency function for K_i .

As a particular point, we can take as initial values for k_1, \dots, k_j in the starting stage of the sampler, those obtained in the last iteration of the Gibbs sampler that gave precisely \hat{l} . Another alternative may be to take random draws from the corresponding prior distributions. The threshold values are those identified for \hat{l} in the threshold identification stage at the beginning.

As in the case of identifying l , consider the sequences $\{\hat{p}_{k_i}^{(g)}\}$ for checking the convergence of the Gibbs sampler, where $\hat{p}_{k_i}^{(g)}$ denotes the computed value for $p(k_i|\hat{l}, \theta_i, \mathbf{y})$ at the g th iteration. After convergence, take either the mean or the median of each sequence as the final estimate of the corresponding posterior probability, say \hat{p}_{k_i} . As \hat{k}_i , take the mode of the corresponding posterior distribution or the most frequent value of K_i in the simulated sample.

In summary, the proposed identification strategy consists of the following steps.

- Step 1.** Select l_0 and then the appropriate thresholds for each $l = 2, \dots, l_0$, via minimization of the NAIC criterion. Here, intermediate draws of the nonstructural parameters are generated for all possible combinations of autoregressive orders.
- Step 2.** Identify l . Intermediate draws of non structural parameters and autoregressive orders are used.
- Step 3.** Conditional on l , identify the orders k_1, \dots, k_j . Intrinsic draws for the non structural parameters are needed.

Overall, the proposed identification procedure is similar to that for identifying, for example, linear ARIMA models via information criteria or to that of George and McCulloch (1993) in a variable selection context.

5. An Empirical Application

Consider now a real application of the proposed methodology in which the dynamical system to be considered consists of the diary rainfall (in mm.), as input variable, and a diary river flow (in m^3/s) as the system output, in a certain Colombian geographical region. The rainfall was measured at the San Rafael Lagoon's meteorological station, with an altitude of 3,420 meters and geographical coordinates 2.23° north (latitude) and 76.23° west (longitude). The flow corresponds to 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 meters 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 data set corresponds to the sample period from January 1, 1992, through November 30, 2000 (3,257 data), and it was assembled by IDEAM, the official Colombian agency for hydrological and meteorological studies. In Fig. 1 one can see the two time series, where the dynamical relationship between the two variables is clear. Additionally, one can see a certain stable path in both variables although there are bursts of large values. This fact is a signal of heteroscedasticity, a major characteristic to be taken into account for explaining the river flow dynamical behavior in terms of precipitation.

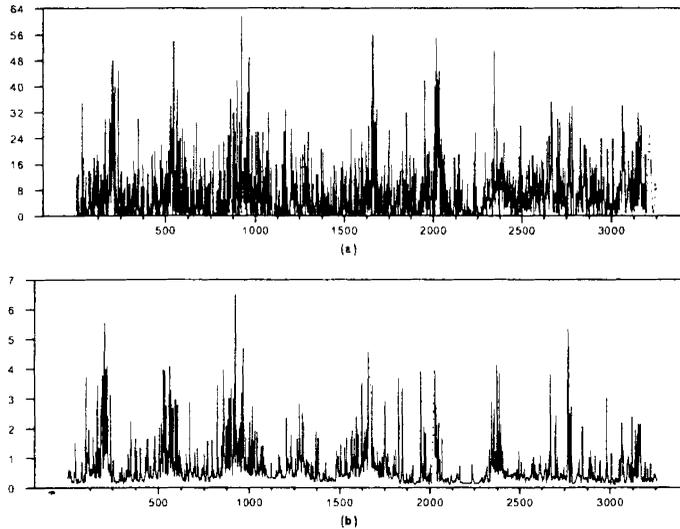


Figure 1. Time series for the real data example: (a) Precipitation; (b) Flow.

Let P_t and X_t be, respectively, the rainfall and the river flow at day t . Because of the universal convention for measuring these two variables, we needed to put $Z_t = P_{t-1}$, that is, translate the precipitation one period back for relating it to the river flow. Empirical evidence and meteorological expertise suggest considering $\{Z_t\}$ as a first-order Markov chain with invariant distribution. However, because of the diary observation frequency of the data, the invariant distribution of $\{Z_t\}$ has no density function in the Lebesgue-measure sense. The problem is that $P(Z_t = 0) > 0$. As a reasonable distribution for the process $\{Z_t\}$, consider the initial distribution function

$$F_1(z) = pF_0(z) + (1 - p)G(z),$$

where $p = \Pr(Z = 0)$, $F_0(z) = I_{[0, \infty)}(z)$, I denotes the indicator function, and $G(z)$ is the distribution function of a normal distribution with mean μ and variance σ^2 truncated at $z = 0$. For the transition kernel, take the distribution function

$$F(z_t|z_{t-1}) = p(z_{t-1})F_0(z_t) + [1 - p(z_{t-1})]G(z_t|z_{t-1}),$$

where $p(z_{t-1}) = P(Z_t = 0|z_{t-1})$ and $G(z_t|z_{t-1})$ is the distribution function of a normal distribution with mean z_{t-1} and variance σ^2 truncated at $z = 0$. Let $g(z)$ and $g(z_t|z_{t-1})$ be the corresponding densities for $G(z)$ and $G(z_t|z_{t-1})$. Also assume that $p(z_{t-1}) = p_j$ if z_{t-1} lies in regime j ($j = 1, \dots, l$). As is noted, the initial and transition kernel distributions are mixture distributions and do not have Lebesgue-measure densities. Consequently, one cannot use directly the proposed Gibbs samplers. To use the results obtained in this article, an approximate distribution for the process $\{Z_t\}$, was considered which is given by the following arguments.

For each positive integer n , let

$$F_{0n}(z) = \begin{cases} 0, & -\infty < z < -1/n; \\ (1/2)[\sin(nz\pi + \pi/2) + 1], & -1/n \leq z \leq 0; \\ 1, & z > 0. \end{cases}$$

Clearly, F_{0n} is a distribution function and the sequence $\{F_{0n}\}$ converges pointwise to F_0 . On the other hand, F_{0n} is differentiable at all real number z and

$$F'_{0n}(z) = \begin{cases} 0, & -\infty < z < -1/n; \\ (n\pi/2)\cos(nz\pi + \pi/2), & -1/n \leq z \leq 0; \\ 0, & z > 0. \end{cases}$$

That is to say, F_{0n} has a density function with respect to the Lebesgue measure on the real line, given by $h_n(z) = F'_{0n}(z)$ for all real number z . In this way, for n large enough take as approximate invariant density

$$f_n(z) = ph_n(z) + (1 - p)g(z)$$

and as approximate transition kernel density

$$f_n(z_t|z_{t-1}) = p(z_{t-1})h_n(z_t) + [1 - p(z_{t-1})]g(z_t|z_{t-1}).$$

Observe here that for drawing from $h_n(z)$, when needed, we will take the value zero for the drawn sample z and that $h_n(0) = n\pi/2$. Finally, it is necessary to remark that some compromise between the size n and computing time should be done because of low rate of convergence detected in the corresponding Markov chains.

After an initial fit of a TAR model to the data, the model residuals suggested the need of taking two transformations to the flow time series: (1) square root of the data and (2) an adjustment for heteroscedasticity. The first one seeks to diminish the influence of extreme values in the model fitting and the second, to delete certain characteristics similar to the financial-time-series stylized facts. This correction was made via the fitting of an ARCH(1) model to the flow data and then correcting them by the estimated standard-deviation time series. Set $X_t = c + \gamma_t$ as the observation equation in the ARCH model, where c is a constant and $\{\gamma_t\}$ is the model noise process. Because the two time series have missing data, 52 in $\{z_t\}$ and 32 in $\{x_t\}$, firstly, they were estimated in a preliminar way and this was done by means of the median of the corresponding time series. From now on, the flow data to be analyzed will be the transformed ones and is denoted as $\{\tilde{x}_t\}$.

With the approach of Sec. 4, the number of regimes l were identified. Observing the regression function between the time series $\{\tilde{x}_t\}$ and $\{z_t\}$, which was estimated using a non parametric kernel approach and is presented in Fig. 2, one notes that approximately the function is conformed by three kinds of behaviors. Here, we are not necessarily looking for piecewise linearity but for change points that determine different geometrical forms in the regression function. Hence, one can postulate that about three regimes would be adequate for the threshold variable Z_t . Then, we proceeded to look for the location of either one threshold r_1 (two regimes) or two thresholds $r_1 < r_2$ (three regimes), as indicated in Sec. 4. The chosen quantiles were

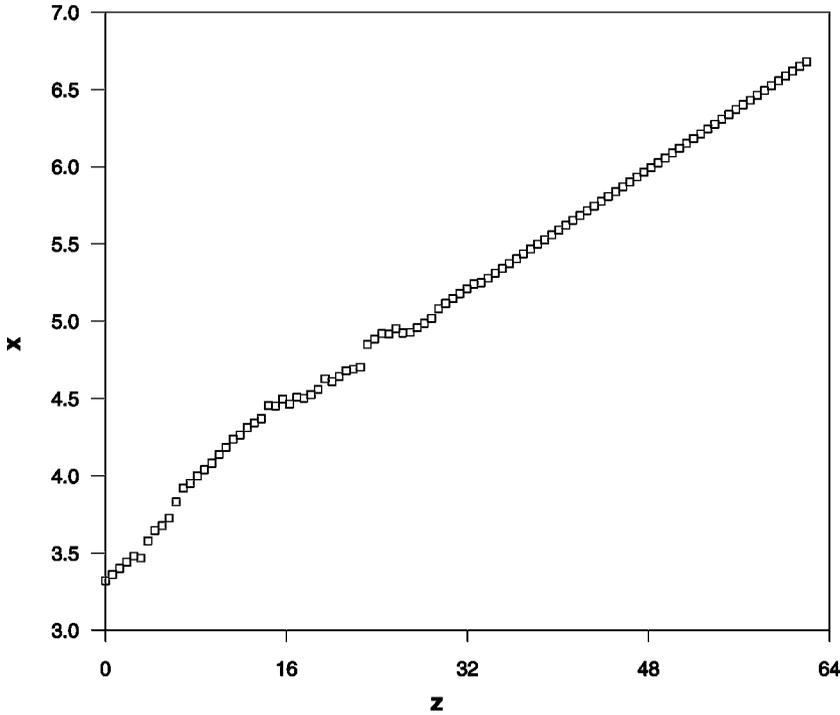


Figure 2. Estimated regression function for precipitation and flow.

the percentiles 1, 5, 10, 25, 50, 75, 90, 95, and 99 with respective values 1.0, 1.0, 1.1, 3.0, 6.0, 10.3, 17.18, 23.0, and 38.0. The results of this search are presented in Table 1. With these possible thresholds sets, the posterior probability function for the number of regimes was computed, finding that $\hat{p}_2 = 0.52$ and $\hat{p}_3 = 0.48$; thus, set $\hat{l} = 2$, although there is not a clear discrimination between the two number of regimes. To check the convergence of the underlying Gibbs sampler, the behavior of the sequences $\{\hat{p}_l^{(i)}\}$ was tracked ($l = 2, 3$) looking for stationarity. I found that for 3000 iterates, the sample autocorrelations functions decay quickly and that the burn-in point is about 10% of the draws.

Conditional on $\hat{l} = 2$, The estimation of the autoregressive orders k_1 and k_2 began finding that $\hat{k}_1 = \hat{k}_2 = 1$, as can be deduced from Table 2. The maximum possible autoregressive order 2 was chosen fitting linear AR models and using the AIC information criterion. As in the case of identifying l , the convergence of the Gibbs samplers was checked via the stationarity approach and found that

Table 1
Set of possible number of regimes for the real data

l	Thresholds	Autoregressive orders	Minimum NAIC
2	6.0	2, 2	4.51
3	6.0, 10.3	2, 2, 1	4.44

Table 2
Posterior probabilities for the AR orders
in the real example

Order	Regime	
	1	2
0	0.31	0.25
1	0.38	0.50
2	0.31	0.25

500 iterates were appropriate, with a very small burn in period of about 1% of the iterates. This concludes the identification stage of the TAR model, taking into account that the identified threshold is $r_1 = 6.0$, the median of the precipitation data.

With the structural parameters identified above, the missing data in the two time series was estimated jointly and the remaining TAR model parameters using the approximate invariant density function $f_n(z)$ with $n = 1000$. To estimate p , was counted the number of zeroes in the sample for the rainfall and computed its relative frequency of occurrence (frequency substitution principle, Bickel and Doksum, 1977) finding $\hat{p} = 0.26$. With the same idea, for estimating the parameters p_j ($j = 1, 2$), the number of pairs $(z_{t-1}, 0)$ with z_{t-1} in the regime j and the number of total pairs $(z_{t-1}, 0)$ in the sample were counted and found that $\hat{p}_1 = 0.87$ and $\hat{p}_2 = 0.13$. The estimated parameters for the truncated normal density g were $\hat{\mu} = 3.24$ and $\hat{\sigma} = 7.76$, where we have made use of the following facts: the truncated distribution attains (1) its mode at the same value as the normal distribution, i.e., at its mean μ , and (2) its inflexion point at $\mu + \sigma$, as the normal density does. Hence, the density of the truncated distribution was estimated using the non parametric kernel approach, and then its mode and inflexion point were computed. As can be noted, these parameter estimates might be accomplished using MCMC methodology, but this route is not considered now.

To generate from the posterior joint density $p(\mathbf{z}|\mathbf{x}, \boldsymbol{\alpha})$, and from these samples to extract draws for the missing data in \mathbf{z} , one needs to generate from $p(z_t|z_{t+1}, \boldsymbol{\alpha}_t, \mathbf{x}_t)$ ($t = T - 1, T - 2, \dots, 1$) as indicated in Sec. 3. This is an intrinsic step of the Gibbs sampler for estimating both the non structural parameters and the missing observations. For this empirical application, the Metropolis-Hastings algorithm was used with a so-called instrumental density given by $ph_n(z_t) + (1 - p)u_{(0, m_z)}(z_t)$, with $n = 1000$, where the density in the second term is that of the uniform distribution on the interval $(0, m_z)$, with m_z the maximum of the time series $\{z_t\}$. For automatically checking the stationarity condition of the inner draws, a non parametric Kolmogorov test was used (Robert and Casella, 1999).

Finally, the initial missing data estimates were replaced with the new ones and repeat the above identifying procedure, finding the same values for the number of regimes, the threshold value, and the autoregressive orders. Consequently, the identified TAR model for these two stochastic processes is given by the structural parameters $\hat{l} = 2$, and $\hat{k}_1 = \hat{k}_2 = 1$.

The final estimates for the non structural parameters are presented in Table 3 and some of the missing data estimates in Table 4, where at time $t = 2012$ a negative value was obtained for the lower extreme of the credible interval. This suggests that

Table 3
Parameter estimates for the real data TAR model

Parameter	Regime j	
	1	2
$a_0^{(j)}$	1.32 [1.23,1.43] (0.06)	1.98 [1.79,2.18] (0.12)
$a_1^{(j)}$	0.59 [0.57,0.62] (0.01)	0.58 [0.53,0.61] (0.03)
$h^{(j)}$	1.35	2.28

Standard errors in parenthesis and 90% credible intervals in brackets.

in an eventual refinement of this real data application, a transformation other than the one used here should be considered for the flow data. The explicit TAR model is then given by

$$\tilde{X}_t = \begin{cases} 1.32 + 0.59\tilde{X}_{t-1} + 1.35\varepsilon_t, & Z_t < 6.0 \\ 1.98 + 0.58\tilde{X}_{t-1} + 2.28\varepsilon_t, & Z_t \geq 6.0, \end{cases}$$

and an interpretation of it is the following: first of all, only two regimes for the rainfall variable are detected, which could be termed as low and high precipitation regimes. Looking at each regime separately, one finds that the stationarity condition for linear AR(1) processes is fulfilled and that the mean values for the transformed flow are 3.22 for the first regime and 4.71 for the second. That is to say, the more the precipitation, the more the transformed flow river (as one would expect). Also, one can see that the noise variance in the first regime is less than the noise variance in the second regime, indicating that the more the

Table 4
Some missing data estimates for the real example

t	\hat{z}_t	90% CI*	t	\hat{z}_t	90% CI
792	7.85	[2.95, 11.36]	1659	4.92	[1.18, 8.74]
3199	3.69	[0.00, 9.79]	1668	3.79	[1.12, 6.64]
3205	8.06	[2.53, 11.88]	2012	4.29	[-0.16, 9.22]
3210	4.78	[0.00, 10.00]	2017	4.47	[0.24, 8.82]
3215	4.82	[0.00, 10.07]	2022	8.02	[3.94, 12.50]
3223	4.02	[0.00, 10.18]	2326	3.88	[0.72, 6.96]
3227	3.24	[0.00, 8.59]			
3237	4.01	[0.00, 9.95]			
3243	4.33	[0.00, 10.09]			
3248	4.44	[0.00, 10.34]			
3255	6.98	[0.00, 11.28]			

*CI: credible interval.

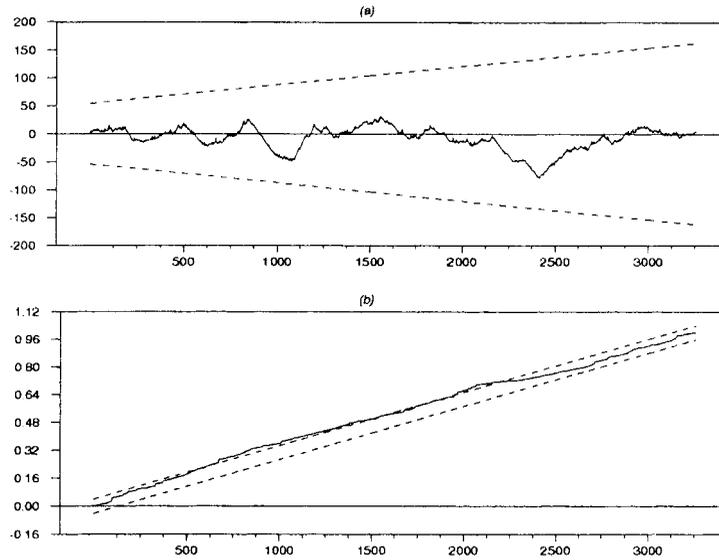


Figure 3. CUSUM and CUSUMSQ charts, (a) and (b) respectively, for the residuals in the empirical example.

rainfall, the more the transformed flow variability. This means that in the flow data transformed initially there is even heteroscedasticity but that it is explained by the rainfall regimes. As can also be observed, the weight of \tilde{X}_{t-1} in the value of \tilde{X}_t is practically the same in both regimes, which means that the dynamic response of the transformed flow data to the influence of its values a day before is not affected by the rainfall regimes.

Concerning the residuals performance, the CUSUMs charts were found to behave well as can be observed at Fig. 3, where the confidence for the bands are 95% for the CUSUM and 99% for the CUSUMSQ. This indicates that there is no evidence for model misspecification or heteroscedasticity in $\{\tilde{X}_t\}$.

6. Conclusions

In this article, a methodology was presented for analyzing bivariate time series with missing data. It was assumed that there is a dynamical nonlinear relationship between the two time series that can be explained by a threshold autoregressive (TAR) model. Some model theoretical properties were found whose implications are useful for its specification in practice.

The time series analysis is based on the identification and estimation stages of the TAR model in the presence of missing data. The model parameters and missing data estimates are addressed jointly and some guidelines are sketched for diagnostic checking. The TAR model analysis is accomplished by means of the MCMC and Bayesian approaches, although some parts of the analysis are developed using empirical Bayes methods. Obviously, the proposed procedure covers the case of complete time series, deleting from the set of full conditional densities those corresponding to the missing data. The forecasting phase of the conventional time series analysis is not considered in this work.

Aside from the use of the MCMC/Bayesian scope, the model and results can be seen as an intermediate approach between the univariate SETAR model and the multivariate threshold model methodologies of Tong (1990) and Tsay (1998).

Acknowledgments

The author gratefully acknowledges Professor Ruey S. Tsay for his valuable advising on the development of this research. He is also grateful to Myriam Muñoz for a useful suggestion on the mathematics for the empirical application, and to IDEAM, the Colombian agency for hydrological and meteorological studies, for the assembling of the real data. Useful comments and suggestions of an anonymous referee lead to improvements in the article, and the author thanks him/her as well.

References

- Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley.
- Bickel, P. J., Doksum, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. San Francisco: Holden-Day, Inc.
- Brockwell, P. J. (1994). On continuous-time threshold ARMA processes. *J. Statist. Plann. Infer.* 39:291–303.
- Brockwell, P. J., Davis, R. A. (1991). *Time Series: Theory and Methods*. New York: Springer-Verlag.
- Carlin, B. P., Chib, S. (1995). Bayesian model choice via Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. Ser. B* 57:473–484.
- Carlin, B. P., Polson, N. G., Stoffer, D. S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *J. Amer. Statist. Assoc.* 87:493–500.
- Carter, C. K., Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika* 81:541–553.
- Carter, C. K., Kohn, R. (1996). Markov chain Monte Carlo in conditionally Gaussian state space models. *Biometrika* 83:589–601.
- Chen, C. W. S., Lee, J. C. (1995). Bayesian inference of threshold autoregressive models. *J. Time Ser. Anal.* 16:483–492.
- Chib, S., Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The Amer. Statistician* 49:327–335.
- George, E. I., McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *J. Amer. Statist. Assoc.* 88:881–889.
- Geweke, J., Terui, N. (1993). Bayesian threshold autoregressive models for nonlinear time series. *J. Time Ser. Anal.* 14:441–454.
- Geweke, J., Tanizaki, H. (1999). On Markov chain Monte Carlo methods for nonlinear and non-Gaussian state-space models. *Commun. Statist. Simul. Computat.* 28:867–894.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82:711–732.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series, and the Kalman Filter*. Cambridge: Cambridge University Press.
- Hausman, J. A., Lo, A. W., MacKinlay, A. C. (1992). An ordered probit analysis of transaction stock prices. *J. Finan. Econ.* 31:319–379.
- Kim, C.-J., Nelson, C. R. (1999). *State Space Models with Regime Switching*. Cambridge: The MIT Press.
- McCulloch, R. E., Tsay, R. S. (1994). Bayesian analysis of threshold autoregressive processes with a random number of regimes. *Comput. Sci. Statist.* 25:253–262.

- Meyn, S. P., Tweedie, R. L. (1993). *Markov Chains and Stochastic Volatility*. London: Springer-Verlag.
- Peña, D., Maravall, A. (1991). Interpolation, outliers, and inverse auto-correlations. *Commun. Statist. Theor. Meth.* 20:3175–3186.
- Robert, C. P., Casella, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer Verlag.
- Shumway, R. H., Stoffer, D. S. (1991). Dynamic linear models with switching. *J. Amer. Statist. Assoc.* 86:411–430.
- Tong, H. (1978). *On a Threshold Model in Pattern Recognition and Signal Processing*. Chen, C. H., ed. Amsterdam: Sijhoff & Noordhoff.
- 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. *J. Roy. Statist. Soc. Ser. B* 42:245–292.
- Tong, H., Yeung, I. (1991a). On tests for self-exciting threshold autoregressive-type non-linearity in partially observed time series. *Appl. Statist.* 40:43–62.
- Tong, H., Yeung, I. (1991b). Threshold autoregressive modeling in continuous time. *Statistica Sinica* 1:411–430.
- Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. *J. Amer. Statist. Assoc.* 84:231–240.
- Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *J. Amer. Statist. Assoc.* 93:1188–1202.

Copyright of Communications in Statistics: Theory & Methods is the property of Marcel Dekker Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.