Bayesian analysis of multivariate threshold autoregressive models with missing data

Sergio A. Calderón V. and Fabio H. Nieto
Departamento de Estadística, Universidad Nacional de Colombia, Bogotá D.C., Colombia

ABSTRACT
In some fields, we are forced to work with missing data in multivariate time series. Unfortunately, the data analysis in this context cannot be carried out in the same way as in the case of complete data. To deal with this problem, a Bayesian analysis of multivariate threshold autoregressive models with exogenous inputs and missing data is carried out. In this paper, Markov chain Monte Carlo methods are used to obtain samples from the involved posterior distributions, including threshold values and missing data. In order to identify autoregressive orders, we adapt the Bayesian variable selection method in this class of multivariate process. The number of regimes is estimated using marginal likelihood or product parameter-space strategies.

1. Introduction
Sometimes, in the real life, we are faced with the issue of having missing data in the time series which we are going to analyze. Then, we have the problem of fitting a statistical model to the data in the presence of the missing observations and, if needed, estimate them. In the case that the time series are realizations of univariate linear stochastic processes and the interest variable had realizations in the sample period under study but those values were either not recorded or recorded with numerical mistakes, the methods for fitting models in this situation are abundant and well known in the literature (e.g., Gómez and Maravall, 1994). It is worth noting here that there is not a random mechanism behind the occurrence of the missing observations. In case the time series are realizations of univariate non-linear processes, as for example threshold autoregressive (TAR) models, some results have been presented in the literature, as is the case, among others, of Nieto’s (2005) paper and the references therein. In that paper, the Bayesian paradigm was used for fitting a TAR model in the presence of the missing data and, then, for designing a smoothing algorithm that allows us to estimate the missing observations. As noted by Nieto (2005) and Nieto et al. (2013), the Bayesian approach is favored over the frequentist one because of the difficult numerical maximization of the likelihood function. This fact says that extracting information from the data about the state of nature (TAR) with only this function can be an unsuccessful task. The problems quoted above get more complex if we have several variables under study. We then are faced with analyzing multivariate time series with absent observations. More even, the multivariate stochastic process that originates the time series can be non-linear and, in particular, of TAR type. Thus, a multivariate threshold autoregressive (MTAR) model is called for addressing the described problem.
Frequentist and Bayesian approaches have been proposed to analyze vector time series by means of MTAR models without missing data. In Tsay’s (1998) paper, the parameters of the model were estimated using conditional least squares and the Akaike information criterion (AIC). In that paper, the MTAR model was applied in finance, specifically to study index future arbitrage; in economy, where the U.S. interest rates were analyzed; and in hydrology, where the river flows data of Iceland were considered. A Bayesian analysis of MTAR models was considered by Kwon et al. (2009), where conjugate analysis was used to find posterior distributions of coefficients and covariance matrices, and then, these parameters were integrated out to find the posterior distributions of thresholds and the so-called lag value. These authors identified autoregressive orders, which were assumed to be the same in each regime, using information criteria. In the approaches mentioned above, the number of regimes must be known. In the approach proposed by Wu and Lee (2011), where the number of regimes and the autoregressive orders are jointly estimated, no exogenous variables are included in the equation of the model, as, for example, Tsay (1998) did.

In this paper we present a Bayesian analysis of an MTAR model, when there are missing data in some or all of the endogenous, exogenous, and threshold variables. The main idea is to develop procedures to (i) identify the model, (ii) estimate the unknown model parameters, and (iii) design a smoothing algorithm for estimating the missing observations. The paper is organized as follows. In Section 2, we specify the MTAR model. In Section 3, the smoothing algorithm and the method for estimating the so-called non structural parameters are presented assuming that the structural parameters are known. The procedures for the estimation of the number of regimes, the thresholds, and the autoregressive orders in each regime are included in Section 4. In order to illustrate the use of the proposed methodology, some simulated models are included in Section 5 and a real data application in the hydrology field is presented in Section 6; finally, Section 7 concludes.

2. Specifying the multivariate threshold autoregressive model

Let \( \{Y_t\} \) and \( \{X_t\} \) be multivariate stochastic processes such that \( Y_t = (Y_{1t}, \ldots, Y_{kt})' \), and \( X_t = (X_{1t}, \ldots, X_{lt})' \) and let \( \{Z_t\} \) be a univariate process. \( \{Y_t\} \) follows an MTAR model with threshold variable \( Z_t \) if

\[
Y_t = \phi^{(j)}_0 + \sum_{i=1}^{p_j} \phi_i^{(j)} Y_{t-i} + \sum_{i=1}^{q_j} \beta_i^{(j)} X_{t-i} + \sum_{i=1}^{d_j} \delta_i^{(j)} Z_{t-i} + \Sigma^{1/2}(j) \epsilon_t \quad \text{when} \quad r_{j-1} < Z_t \leq r_j
\]

for some \( j = 1, \ldots, l \), where \( l \geq 2 \) is the number of regimes, and \( r_0, r_1, \ldots, r_{l-1}, r_l \) are real numbers such that \( -\infty = r_0 < r_1 < \cdots < r_{l-1} < r_l = \infty \). They are called the thresholds and define the model regimes.

Additionally, \( \{\epsilon_t\} \) is a \( k \)-dimensional Gaussian zero-mean white process with covariance matrix \( I_k \), the identity matrix of order \( k \), and it is mutually independent of \( \{X_t\} \) and \( \{Z_t\} \). For \( j = 1, \ldots, l \), the coefficients \( \phi_i^{(j)} \) for \( i = 0, 1, \ldots, p_j \), \( \beta_i^{(j)} \) for \( i = 1, \ldots, q_j \), \( \delta_i^{(j)} \) for \( i = 1, \ldots, d_j \), and \( \Sigma^{1/2}(j) \) are real matrices of suitable dimensions and we call them non structural parameters. The non negative integer numbers \( p_j \) with \( j = 1, \ldots, l \) are called the autoregressive orders for each regime; \( q_j \) and \( d_j \) denote the maximum lags for the vector \( X_t \) and the threshold variable \( Z_t \), respectively, in each regime; all of them together with the threshold values \( r = (r_1, \ldots, r_{l-1})' \) and the number of regimes \( l \) are known as structural parameters. We define the vector of non structural parameters as \( \theta_{\text{ns}} = (\theta_1', \ldots, \theta_l', \text{vec}(\Sigma)') \), with \( \theta_j = \text{vec}(A_j) \) and \( A_j = (\phi_0^{(j)}, \phi_1^{(j)}, \ldots, \phi_{p_j}^{(j)}, \beta_1^{(j)}, \ldots, \beta_{q_j}^{(j)}, \delta_1^{(j)}, \ldots, \delta_{d_j}^{(j)})_{k \times n_j} \), where \( n_j =
1 + k \cdot p_j + v \cdot q_j + d_j \text{ for } j = 1, \ldots, l, \text{ and } \Sigma = (\Sigma(1), \ldots, \Sigma(l)). \text{ We set the vector of structural parameter as } \theta_{ys} = (p_1, \ldots, p_l, q_1, \ldots, q_l, d_1, \ldots, d_l, r', l'). \text{ Therefore, the full parameter vector of the model, which is denoted MTAR}(l; p_1, \ldots, p_l, q_1, \ldots, q_l, d_1, \ldots, d_l) \text{, is } \theta_y = (\theta'_{ys}, \theta'_y)'.

Now, we assume that there are missing data in all of the time series, that is, we have unequally spaced time series, and, specifically, the observed data are at the time points \(t_1, \ldots, t_{n_1}\) with \(1 \leq t_1 \leq \cdots \leq t_{n_1} \leq T, s_1, \ldots, s_{n_2}\) with \(1 \leq s_1 \leq \cdots \leq s_{n_2} \leq T, \) and \(h_1, \ldots, h_{n_3}\) with \(1 \leq h_1 \leq \cdots \leq h_{n_3} \leq T\) for the time series \{\(y_i\), \{\(x_i\), and \{\(z_i\), respectively.

In order to tackle the presence of missing data in the threshold and covariate process \{\(Z_t\) and \{\(X_t\), respectively, we assume that \(\{U_t = (Z_t, X_t)'\) is a \((v + 1)\)-dimensional homogeneous \(b\)th order Markov chain with initial density \(f_u(\cdot)\) and transition kernel density \(f_u(\cdot|\cdot)\), with respect to the Lebesgue-measure, where \(b\) is an integer number greater than zero. We assume that \(\{U_t\}\) is exogenous in the sense that there is no feedback of \(\{Y_t\}\) toward \(\{U_t\}\).

Let \(\theta = (\theta_y, \theta_u)\) be the full vector of parameters, which is composed of the vector of parameters \(\theta_y\) in the MTAR model and the vector of parameters \(\theta_u\) of the Markov chain. We assume that the probabilistic mechanism generating the data \(u = (u_1, \ldots, u_T)\) does not depend on \(\theta_y\) and the joint density of \(Y = (Y_1, \ldots, Y_T)\) conditional on \(u\) and \((\theta_y, \theta_u)\) does not depend on \(\theta_u\). With the last assumptions, we can see that it is possible to estimate and identify, first of all, the parameters of the Markov chain \(\{u_t\}\), and, then, conditional on the estimated parameters \(\theta_u\), we can proceed to estimate the parameters of the MTAR model. Following Nieto (2005), we found that the model likelihood function is given by

\[
L(\theta|y, u) = f(y, u|\theta_y, \theta_u) = f(u|\theta_y, \theta_u) f(y|u, \theta_y, \theta_u)
= f(u|\theta_u) f(y|u, \theta_y)
\] (2)

for each \(\theta\) in a convenient parameter space. \(\{y_t\}, \{x_t\}, \) and \(\{z_t\}\) are called respectively output, covariates, and threshold processes.

**Remarks.** This MTAR model is an open-loop multivariate system as was mentioned in Tong (1990). We can also observe that the MTAR model is slightly different from the model proposed by Tsay (1998), because we have added lags of the threshold variable in the regression for each regime.

### 3. Estimation of missing data and non structural parameters

We consider here the estimation of the missing data and the non structural parameters conditional on structural parameters.

#### 3.1. Estimation of missing data

We shall write the MTAR model in a state-space form with regime-switching as in Nieto’s (2005) paper. To do that, let

\[
\alpha_t = [Y_{t}', Y_{t-1}', \ldots, Y_{t-p+1}', X_{t}', X_{t-1}', \ldots, X_{t-q+1}', Z_t, Z_{t-1}, \ldots, Z_{t-d+1}]
\]

be the state vector, and let

\[
H(Z_t) = \]

\[
1 + k \cdot p_j + v \cdot q_j + d_j \text{ for } j = 1, \ldots, l, \text{ and } \Sigma = (\Sigma(1), \ldots, \Sigma(l)). \text{ We set the vector of structural parameter as } \theta_{ys} = (p_1, \ldots, p_l, q_1, \ldots, q_l, d_1, \ldots, d_l, r', l'). \text{ Therefore, the full parameter vector of the model, which is denoted MTAR}(l; p_1, \ldots, p_l, q_1, \ldots, q_l, d_1, \ldots, d_l) \text{, is } \theta_y = (\theta'_{ys}, \theta'_y)'.

Now, we assume that there are missing data in all of the time series, that is, we have unequally spaced time series, and, specifically, the observed data are at the time points \(t_1, \ldots, t_{n_1}\) with \(1 \leq t_1 \leq \cdots \leq t_{n_1} \leq T, s_1, \ldots, s_{n_2}\) with \(1 \leq s_1 \leq \cdots \leq s_{n_2} \leq T, \) and \(h_1, \ldots, h_{n_3}\) with \(1 \leq h_1 \leq \cdots \leq h_{n_3} \leq T\) for the time series \{\(y_i\), \{\(x_i\), and \{\(z_i\), respectively.

In order to tackle the presence of missing data in the threshold and covariate process \{\(Z_t\) and \{\(X_t\), respectively, we assume that \(\{U_t = (Z_t, X_t)'\) is a \((v + 1)\)-dimensional homogeneous \(b\)th order Markov chain with initial density \(f_u(\cdot)\) and transition kernel density \(f_u(\cdot|\cdot)\), with respect to the Lebesgue-measure, where \(b\) is an integer number greater than zero. We assume that \(\{U_t\}\) is exogenous in the sense that there is no feedback of \(\{Y_t\}\) toward \(\{U_t\}\).

Let \(\theta = (\theta_y, \theta_u)\) be the full vector of parameters, which is composed of the vector of parameters \(\theta_y\) in the MTAR model and the vector of parameters \(\theta_u\) of the Markov chain. We assume that the probabilistic mechanism generating the data \(u = (u_1, \ldots, u_T)\) does not depend on \(\theta_y\) and the joint density of \(Y = (Y_1, \ldots, Y_T)\) conditional on \(u\) and \((\theta_y, \theta_u)\) does not depend on \(\theta_u\). With the last assumptions, we can see that it is possible to estimate and identify, first of all, the parameters of the Markov chain \(\{u_t\}\), and, then, conditional on the estimated parameters \(\theta_u\), we can proceed to estimate the parameters of the MTAR model. Following Nieto (2005), we found that the model likelihood function is given by

\[
L(\theta|y, u) = f(y, u|\theta_y, \theta_u) = f(u|\theta_y, \theta_u) f(y|u, \theta_y, \theta_u)
= f(u|\theta_u) f(y|u, \theta_y)
\] (2)

for each \(\theta\) in a convenient parameter space. \(\{y_t\}, \{x_t\}, \) and \(\{z_t\}\) are called respectively output, covariates, and threshold processes.

**Remarks.** This MTAR model is an open-loop multivariate system as was mentioned in Tong (1990). We can also observe that the MTAR model is slightly different from the model proposed by Tsay (1998), because we have added lags of the threshold variable in the regression for each regime.

### 3. Estimation of missing data and non structural parameters

We consider here the estimation of the missing data and the non structural parameters conditional on structural parameters.

#### 3.1. Estimation of missing data

We shall write the MTAR model in a state-space form with regime-switching as in Nieto’s (2005) paper. To do that, let

\[
\alpha_t = [Y_{t}', Y_{t-1}', \ldots, Y_{t-p+1}', X_{t}', X_{t-1}', \ldots, X_{t-q+1}', Z_t, Z_{t-1}, \ldots, Z_{t-d+1}]
\]

be the state vector, and let

\[
H(Z_t) = \]
obtained. Hereinafter, we do not write the vector \( \theta \) and \( p \) or Frühwirth-Schnatter (1995), with suitable modifications due to the missing data.

Here, \( \phi(\cdot) \) is to find the posterior distribution of the vectors of corresponding observed data. Based on the state-space representation, the goal is addressed via the conditional distribution \( p(y_t, x_m, z_m | y_o, x_o, z_o, \theta) \) where \( y_t, x_m, z_m \) denote the vectors of missing data in the time series \( \{y_t\}, \{x_t\}, \{z_t\} \), respectively, and \( y_o, x_o, z_o \) denote the vectors of corresponding observed data. Based on the state-space representation, the goal is to find the posterior distribution \( p(\alpha_1, \ldots, \alpha_T, z_1, \ldots, z_T, x_1, \ldots, x_T | y_1, \ldots, y_T, \theta) \), because from this the required distribution \( p(y_m, x_m, z_m | y_o, x_o, z_o) \) can be obtained. Hereinafter, we do not write the vector \( \theta \) in the conditioning set. \( p(x_1, \ldots, x_T, z_1, \ldots, z_T, x_1, \ldots, x_T | y_1, \ldots, y_T) \) can be rewritten as \( p(x_1, \ldots, x_T, u_1, \ldots, u_T | y_1, \ldots, y_T) \) and drawing samples from that distribution is equivalent to drawing samples from the full conditional distributions \( p(\alpha_0, \ldots, \alpha_T | y_1, \ldots, y_T, u_1, \ldots, u_T) \) and \( p(u_1, \ldots, u_T | y_1, \ldots, y_T, \alpha_0, \ldots, \alpha_T) \). It is possible to simulate samples from \( p(\alpha_0, \ldots, \alpha_T | y_1, \ldots, y_T, u_1, \ldots, u_T) \) using the results of Carter and Kohn (1994) or Frühwirth-Schnatter (1995), with suitable modifications due to the missing data.

\[
L(Z_t) = \begin{bmatrix} \phi^{(Z_t)}_1, 0, \ldots, 0, 0, 0, \ldots, 0, 0, \ldots, 0 \end{bmatrix}
\]

\[
M(Z_t) = \begin{bmatrix} 0, 0, \ldots, 0, 0, 0, \ldots, 0, 1, 0, \ldots, 0 \\ 0, 0, \ldots, 0, I_o, 0, \ldots, 0, 0, 0, \ldots, 0 \end{bmatrix}
\]

\[
R(Z_t) = (\Sigma_{(Z_t)}^{1/2}, 0, \ldots, 0), \quad K(Z_t) = [I_k, 0, \ldots, 0, 0, 0, \ldots, 0, 0, 0, \ldots, 0] \quad \text{and} \quad N(Z_t) = 0.
\]

Here, \( \phi^{(Z_t)}_i = \phi^{(j)}_i, \beta^{(Z_t)}_i = \beta^{(j)}_i, \delta^{(Z_t)}_i = \delta^{(j)}_i \) and \( \Sigma_{(Z_t)}^{1/2} = \Sigma_{(j)}^{1/2} \) if \( r_{j-1} < Z_t \leq r_j \) for some \( j = 1, \ldots, l \). Therefore, the state-space model becomes

\[
\alpha_t = L(Z_t) + H(Z_t)\alpha_{t-1} + M(Z_t)\begin{bmatrix} Z_t \\ X_t \end{bmatrix} + R(Z_t)\epsilon_t \tag{3}
\]
as the state equation, and

\[
Y_t = K(Z_t)\alpha_t + N(Z_t)\begin{bmatrix} Z_t \\ X_t \end{bmatrix} + a_t \tag{4}
\]
as the observation equation. We need to assume additionally that: (i) The process \( \{\epsilon_t\} \) is I.I.D. Gaussian with mean vector 0, covariance matrix \( I_k \); (ii) \( \{U_t\} \) is mutually independent of \( \{\epsilon_t\} \) and \( \alpha_0 \); and (iii) \( \alpha_0 \) is independent of \( \{\epsilon_t\} \). Trivially, the processes \( \{\epsilon_t\}, \{U_t\} \), and the vector \( \alpha_0 \) are mutually independent of \( \{\alpha_t\} \). The statistical problem of estimating missing data can be addressed via the conditional distribution \( p(y_m, x_m, z_m | y_o, x_o, z_o, \theta) \) where \( y_m, x_m, z_m \) denote the vectors of missing data in the time series \( \{y_t\}, \{x_t\}, \{z_t\} \), respectively, and \( y_o, x_o, z_o \) denote the vectors of corresponding observed data.
These methodologies use the Kalman filter to obtain the moments of the distributions involved here, which are multivariate normal distributions. Finally, it is possible to draw samples from \( p(u_1, \ldots, u_T | y_1, \ldots, y_T, \alpha_0, \ldots, \alpha_T) \), if we can obtain samples from \( p(u_1, \ldots, u_b | u_{-(1:b)}, y_1, \ldots, y_T, \alpha_0, \ldots, \alpha_T) \) and \( p(u_{-(t:b)} | y_t, \alpha_0, \ldots, \alpha_T) \), for \( t = b + 1, \ldots, T \), where we define \( u_{-(t)} = (u_t, \ldots, u_b) \), \( u_{-(t:b)} = (u_t, \ldots, u_{-(1:b)}, u_{-(1:b)+1}, \ldots, u_T) \), and \( u_{-(t:b)} = (u_t, \ldots, u_{-(1:b)}, u_{-(1:b)+1}, \ldots, u_T) \) with \( s < t \). This methodology is known as single-move Gibbs sampling and was implemented by Carlin et al. (1992) to analyze non normal and non linear state-space models. Now we decompose the full conditional distribution \( p(u_1, \ldots, u_b | u_{-(1:b)}, y_1, \ldots, y_T, \alpha_0, \ldots, \alpha_T) \) and the distribution \( p(u_t | u_{-(t:b)}, y_1, \ldots, y_T, \alpha_0, \ldots, \alpha_T) \) for \( t = b + 1, \ldots, T \), in order to obtain the appropriated expressions for drawing samples from them.

Note that under assumptions (i)–(iii) and Equations (3) and (4), we can deduce that: (a) \( \alpha_0 \) is independent of the process \( \{U_t\} \); (b) \( \alpha_t \) is independent of \( \alpha_{0:b-2}, Y_{1:t} \) and \( U_{1:t-1} \) conditional on \( \alpha_{t-1} \) and \( U_t \); (c) \( Y_t \) is independent of \( \alpha_{0:b-1}, Y_{1:t-1} \), and \( U_{1:t-1} \) conditional on \( \alpha_t \) and \( U_t \); (d) \( U_t \) is independent of \( Y_{1:t-1} \) and \( \alpha_{0:b-2} \) conditional on \( U_{1:t-1} \). Finally, dropping out terms that are independent of \( U_{1:b} \) we can write the distribution \( p(u_1, \ldots, u_b | u_{-(1:b)}, y_1, \ldots, y_T, \alpha_0, \ldots, \alpha_T) \) as

\[
p(u_1, \ldots, u_b | u_{-(1:b)}, y_1, \ldots, y_T, \alpha_0, \ldots, \alpha_T) \propto p(u_1, \ldots, u_b) \\
\times \prod_{t=1}^{b} p(\alpha_t | \alpha_{t-1}, u_t) \prod_{t=1}^{b} p(y_t | \alpha_t, u_t) \prod_{t=b+1}^{2b} p(u_t | u_{t-1:b-1})
\]

(5)

In the same manner, we can write the distributions \( p(u_t | u_{-(t:b)}, y_1, \ldots, y_T, \alpha_0, \ldots, \alpha_T) \) for \( t = b + 1, \ldots, T \) as

\[
p(u_t | u_{-(t:b)}, y_1, \ldots, y_T, \alpha_0, \ldots, \alpha_T) \propto p(\alpha_t | \alpha_{t-1}, u_t) p(y_t | \alpha_t, u_t) \prod_{s=t}^{t+b} p(u_s | u_{s-b-1})
\]

(6)

for \( t = b + 1, \ldots, T \). It is important to point out that the distribution of \( Y_t \) conditional on \( \alpha_t, u_t \) for all \( t \), in (5) and (6) is degenerated for MTAR models, that is, \( p(y_t | \alpha_t, u_t) = 1 \) for all \( y_t \in \mathbb{R}^k \), consequently we can drop out these terms. This fact was mentioned by Nieto (2005) regarding univariate threshold models.

We can see that the posterior distributions found to estimate missing data in the covariates and threshold variable are not in general standard distributions; therefore, the random-walk Metropolis-Hastings algorithm is used to extract samples based on a multivariate normal distribution centered in the origin with a covariance matrix \( cI \), where \( c \) is a positive constant and \( I \) is the identity matrix. For details we can see the works of Tierney (1994) or Robert and Casella (2004). If there are not missing data at time \( t \), we do not need to extract samples herein, and we skip the procedure at this time point. We note that for MTAR models, the density \( p(\alpha_t | \alpha_{t-1}, u_t) \) corresponds to a singular multivariate normal distribution which collapses to a \( k \)-variate normal distribution, due to the fact that many components are in the state vector only for the sake of completeness. Then \( p(\alpha_t | \alpha_{t-1}, u_t) = \frac{1}{(2\pi)^{k/2} |\Sigma_{(\alpha)}|^{1/2}} \exp(-\frac{1}{2} e^T \Sigma_{(\alpha)}^{-1} e) \), where

\[
e_t = \Sigma_{(\alpha)}^{-1/2} (y_t - \phi_{0}(\alpha) - \sum_{i=1}^{p} \phi_{i}(\alpha) y_{t-i} - \sum_{i=1}^{d} \beta_{i}(\alpha) x_{t-i} - \sum_{i=1}^{d} \delta_{i}(\alpha) z_{t-i}).
\]

In summary and assuming that the parameters are known, draws for \( p(y_n, x_m, z_m | y_n, x_n, z_n, \theta) \) are obtained via marginalization of the draws set \( \{\alpha_{i}, \ldots, \alpha_{T}, u_{1}, \ldots, u_{T} \} | i = 1, \ldots, G \)\), which are obtained through expressions (5), (6), and the results of Carter and Kohn (1994) or Frühwirth-Schnatter (1995). The estimate
of a missing data is accomplished by means of the ergodic average of the draws from the corresponding elements of the draws set.

### 3.2. Estimation of non structural parameters

Now, we are going to estimate the non structural parameters assuming that there are not missing data. We focus on the full conditional distribution \( p(\theta_{\text{yns}}|\alpha_1, \ldots, \alpha_T, y_{1:T}, u_{1:T}; \theta_{\text{ys}}) \). In order to draw samples for this distribution, we consider the conditional distributions \( p(\theta_j|\theta_i, i \neq j, \Sigma, \alpha_{0:T}, u_{1:T}, y_{1:T}) \) and \( p(\Sigma(j)|\Sigma(i), i \neq j, \theta_1, \ldots, \theta_l, \alpha_{0:T}, u_{1:T}, y_{1:T}) \), for \( j = 1, \ldots, l \). Hereinafter, the vector \( \theta_{\text{ys}} \) is omitted in the conditioning set. We assume that the set \( \{\theta_1,\Sigma(1),\ldots,\theta_l,\Sigma(l)\} \) is mutually independent. Furthermore, the prior distribution for \( \theta_j \) is a multi-normal distribution with mean \( \theta_{0j} \) and covariance matrix \( \Sigma_{0j} \). Prior normal distribution focuses the knowledge of the coefficients around the mean \( \theta_{0j} \) with uncertainty quantified by the covariance matrix \( \Sigma_{0j} \). The prior distribution for \( \Sigma(j) \) is an inverse Wishart with covariance matrix \( S_{0j} \) and \( \nu_{0j} \) degrees of freedom. Now, to obtain the full conditional distributions for the non structural parameters we define the matrices, for \( j = 1 \ldots l \), \( W_j = (w_{t_1,j}, \ldots, w_{t_{N_j},j})_{N_j \times N_j} \), with \( w_{t,j} = (1, y'_{t-1}, \ldots, y'_{t-p}, x'_{t-1}, \ldots, x'_{t-q}, z_{t-1}, \ldots, z_{t-d})_{t \times 1} \). \( Y_j = (y_{t_1,j}, \ldots, y_{t_{N_j},j})_{k \times N_j} \), and \( y_j = \text{vec}(Y_j) \), where \( t_1,j, \ldots, t_{N_j},j \) are the time points where \( r_{j-1} < Z_t \leq r_j \). Note that \( N_j \) is the number of these points. Note that due to the independence of the parameters, the independence between the Markov chain \( \{U_i\} \) and \( \theta_i \) and that of \( p(y_{1:T}|\alpha_{0:T}, u_{1:T}, \theta_{\text{yyns}}) \) is a product of degenerated densities for MTAR models that do not depend on \( \theta_{\text{yyns}} \), we have for \( j = 1, \ldots, l \),

\[
p(\theta_j|\theta_i, i \neq j, \Sigma, \alpha_{0:T}, u_{1:T}, y_{1:T}) \propto p(\alpha_{0:T}|u_{1:T}, \theta_{\text{yyns}}) p(\theta_j)
\]

where

\[
p(\alpha_{0:T}|u_{1:T}, \theta_{\text{yyns}}) = p(\alpha_0) p(\alpha_1|\alpha_0, u_{1:T}, \theta_{\text{yyns}}) \cdots p(\alpha_T|\alpha_{T-1}, u_{1:T}, \theta_{\text{yyns}})
\]

when it is considered that \( \text{vec}(Y_j - A_j W_j) = y_j - (W'_j \otimes I_k) \theta_j \) and each density \( p(\alpha_l|\alpha_{l-1}, u_{1:T}, \theta_{\text{yyns}}) \) was found in the last section. Because of the mutual independence of the parameters, we obtain

\[
p(\alpha_{0:T}|u_{1:T}, \theta_{\text{yyns}}) \propto \exp \left\{ -\frac{1}{2} [y_j - (W'_j \otimes I_k) \theta_j]' (I_{N_j} \otimes \Sigma^{-1}_{(j)}) [y_j - (W'_j \otimes I_k) \theta_j] \right\}
\]

Hence, for \( j = 1, \ldots, l \),

\[
p(\theta_j|\theta_i, i \neq j, \Sigma, \alpha_{0:T}, u_{1:T}, y_{1:T}, \theta_{\text{yyns}}) \propto \exp \left\{ (\theta_j - \theta_j^*)' V_j^{-1}(\theta_j - \theta_j^*) \right\}
\]

which is the kernel of a multivariate normal distribution with covariance matrix \( V_j = [W_j W'_j \otimes \Sigma^{-1}_{(j)} + \Sigma_{0j}]^{-1} \) and mean \( \theta_j^* = V_j ([W_j \otimes \Sigma^{-1}_{(j)}] y_j + \Sigma_{0j}^{-1} \theta_{0j}) \). With similar arguments as above, we obtain

\[
p(\Sigma(j)|\Sigma(i), i \neq j, \theta_1, \ldots, \theta_l, \alpha_{0:T}, u_{1:T}, y_{1:T}, \theta_{\text{yyns}}) \propto p(\alpha_{0:T}|u_{1:T}, \theta_{\text{yyns}}) p(\Sigma(j))
\]
Now,

\[
p(\alpha_{0:T}|u_{1:T}, \theta_{yns}) \propto |\Sigma(j)|^{-N_j/2} \times \exp \left[ -\frac{1}{2} \operatorname{tr} \left\{ \Sigma_1^{-1}(j) \sum_{t:t'=j} \left( y_t - \phi_0^{(j)} - \sum_{i=1}^{p_j} \phi_i^{(j)} y_{t-i} + \sum_{i=1}^{q_j} \beta_i^{(j)} x_{t-i} + \sum_{i=1}^{d_j} \delta_i^{(j)} z_{t-i} \right) \right\} \right]
\]

\[
\times \left( y_t - \phi_0^{(j)} - \sum_{i=1}^{p_j} \phi_i^{(j)} y_{t-i} + \sum_{i=1}^{q_j} \beta_i^{(j)} x_{t-i} + \sum_{i=1}^{d_j} \delta_i^{(j)} z_{t-i} \right) \right]\]

\[
= |\Sigma(j)|^{-N_j/2} \exp \left[ -\frac{1}{2} \operatorname{tr} \left\{ \Sigma_1^{-1} S_j \right\} \right]
\]

where, in general, \( \operatorname{tr}(A) \) denotes the trace of a matrix \( A \) and

\[
S_j = \sum_{t:t'=j} \left( y_t - \phi_0^{(j)} - \sum_{i=1}^{p_j} \phi_i^{(j)} y_{t-i} + \sum_{i=1}^{q_j} \beta_i^{(j)} x_{t-i} + \sum_{i=1}^{d_j} \delta_i^{(j)} z_{t-i} \right)
\]

Consequently, for \( j = 1, \ldots, l, \)

\[
p(\Sigma(j) | \Sigma(i), i \neq j, \theta_1, \ldots, \theta_1, \alpha_{0:T}, u_{1:T}, y_{1:T}, \theta_{yns}) \propto |\Sigma(j)|^{-N_j/2} \exp \left[ -\frac{1}{2} \operatorname{tr} \left\{ \Sigma_1^{-1} S_j \right\} \right] |\Sigma(j)|^{-(\nu_{0j} + k_1 + 1)/2} \exp \left[ -\frac{1}{2} \operatorname{tr} \left\{ \Sigma_2^{-1} S_{ij} \right\} \right]
\]

\[
= |\Sigma(j)|^{-(N_j + \nu_{0j} + k_1 + 1)/2} \exp \left[ -\frac{1}{2} \operatorname{tr} \left\{ \Sigma_1^{-1} (S_j + S_{0j}) \right\} \right] \tag{8}
\]

which is the kernel of an inverse Wishart distribution with covariance matrix \((S_j + S_{0j})^{-1}\) and \(N_j + \nu_{0j} + 1\) degrees of freedom.

**Summary:** Conditional on structural parameters, we can estimate non structural parameters with the samples simulated from the full conditional distributions (7) and (8).

### 4. Estimation of the structural parameters of the MTAR model

In this section we propose the methodologies that enable us to identify the autoregressive orders, the thresholds, and the number of regimes of the MTAR model.

#### 4.1. Identification of the autoregressive orders and the threshold values

We use stochastic search ideas for identifying the autoregressive orders in the MTAR models. To this end, we write the MTAR model (1) in the following way:

\[
Y_t = (I_k \otimes w_{t,j}) \theta_j + \Sigma_1^{1/2} \xi_t \text{ if } r_{j-1} < Z_t \leq r_j \tag{9}
\]

where \( \theta_j = \text{vec}(A_i^{j}) \) for \( j = 1, \ldots, l \). We add 0–1 indicator variables \( \gamma_{i,j}, i = 1, \ldots, \eta_j, j = 1, \ldots, l \); such that if \( \gamma_{i,j} = 1 \), the associated parameter \( \theta_{i,j} \) should be included in the model; and if \( \gamma_{i,j} = 0 \), the associated parameter should not be. Now, let \( \nu_j = (\gamma_{1,j}, \ldots, \gamma_{\eta_j,j})' \) be the
vector of indicators in the regime \( j \). Then, using the vectors \( \gamma_j \), \( j = 1, \ldots, l \), we can rewrite the MTAR model in Equation (9) as

\[
Y_t = (I_k \otimes w'_{t,j}) \Gamma_j \theta_j + \Sigma^{1/2}_{(j)} \varepsilon_t
\]

\[
= (I_k \otimes w'_{t,j}) \theta_j + \Sigma^{1/2}_{(j)} \varepsilon_t
\]

(10)

(11)

if \( r_{j-1} < Z_t \leq r_j \), where \( \theta_j = (\gamma_1, \theta_{t,j}, \ldots, \gamma_{c(j)}, \theta_{y_j}) \) and \( \Gamma_j = \text{Diag}(\gamma_j) \). Now, the structural parameter vector becomes \( \theta_{ys} = (\gamma_1', \ldots, \gamma_l', r, l) \). The Kuo and Mallick (KUO) and Gibbs Variable Selection (GVS) methods developed by Kuo and Mallick (1998) and Dellaportas et al. (2002), respectively, are implemented here for MTAR models.

We assume that the \( p \) first values of \( y_t, y_p = (y_1, \ldots, y_p) \) say, are fixed. Then, from (2), (10), and (11), we obtain

\[
f(y|u, \theta_j) = (2\pi)^{-(T-p)/2} \prod_{i=1}^{l} |\Sigma(i)|^{-N_i/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^{l} [y_i - X_i(\theta_j)]' (I_{N_i} \otimes \Sigma^{-1}_{(i)}) [y_i - X_i(\theta_j)] \right\}
\]

(12)

where

\[
X_i = \begin{pmatrix}
I_k \otimes w'_{1,i} \\
I_k \otimes w'_{2,i} \\
\vdots \\
I_k \otimes w'_{N_i,i}
\end{pmatrix}
\]

It is important to point out that \( N_j, j = 1, \ldots, l \), depends on the threshold vector \( r \); then, the likelihood depends on thresholds through the number of observations in each regime.

4.1.1. Posterior distributions using the KUO method

In order to implement KUO methodology, we consider the following prior assumptions: the variables \( \gamma_{t,j} \) are mutually independent with Bernoulli distribution each one and \( P[\gamma_{t,j} = 1] = p_{t,j} \) for \( i = 1, \ldots, n_j \) and \( j = 1, \ldots, l \); \( \theta_j \) is independent of \( \gamma_j \) and \( r \) for \( j = 1, \ldots, l \); and the threshold vector \( r \) has, a priori, a uniform density over the region \( S = \{(r_1, \ldots, r_{l-1}) \in \mathbb{R}^{l-1} : z(a) < r_i < z(b), i = 1, \ldots, l-1; r_1 < r_2 < \cdots < r_{l-1}\} \), where \( z(a) \) and \( z(b) \) denote the \( a \)th and \( b \)th quantiles of the sample \( z_1, \ldots, z_T \), respectively. Thus, the a priori density distribution is

\[
p(r) = \frac{1}{V(S)} 1_S(r), \quad r \in \mathbb{R}^{l-1}
\]

where \( 1_S(r) \) is the indicator function of \( S \) on \( \mathbb{R}^{l-1} \) and \( V(S) = \int_{z(a)}^{z(b)} \int_{r_1}^{r_{l-2}} \cdots \int_{r_{l-2}}^{r_1} dr_{l-1} \cdots dr_2 dr_1 \). Additionally, we assume that \( r \) is independent of \( \theta_j, \Sigma_j \), and \( \gamma_j \) for \( j = 1, 2, \ldots, l \). In the case of one threshold, a priori, \( r \sim U(z(a), z(b)) \). The assumptions in Section 3.2 are considered to obtain again the full conditional distributions for the structural and non structural parameters in this context. Its expressions are similar to those found in (7) and (8), noting only that \( \gamma_j \) appears in the matrix \( \Gamma_j \). Hence, for \( j = 1, \ldots, l \), we have

\[
p(\theta_j|\theta_{j-1}, y, u) \propto p(y|\theta_j, u) p(\theta_j)
\]

\[
\propto \exp \left\{ -\frac{1}{2} (\theta_j - \theta_j^*)' V_j^{-1} (\theta_j - \theta_j^*) \right\}
\]
where \( V_j = [\Gamma_j X'_j (I_{N_j} \otimes \Sigma^{-1}_{(j)}) X_j \Gamma_j + \Sigma^{-1}_{0j}]^{-1} \), \( \theta^*_j = V_j (\Gamma_j X'_j (I_{N_j} \otimes \Sigma^{-1}_{(j)}) y_j + \Sigma^{-1}_{0j} \theta_{0j}) \), and \( \theta_{y-\theta_j} \) is the parameter vector \( \theta_y \) without the sub-vector \( \theta_j \). This is the kernel of a normal multivariate distribution with mean \( \theta^*_j \) and covariance matrix \( V^*_j \).

Now, for \( j = 1, \ldots, l \),

\[
p(\Sigma_{(j)} | \theta_{y-\Sigma_j}, y, u) \propto |\Sigma_{(j)}|^{-(N_j + \nu_{0j})-k/2} \exp \left[ -\frac{1}{2} \text{tr} \left\{ \Sigma_{(j)}^{-1} (S_j + S_{0j}) \right\} \right]
\]

where

\[
S_j = \sum_{[i:j]} (y_i - (I_k \otimes w_{t,j}) \theta_{jt}) (y_i - (I_k \otimes w_{t,j}) \theta_{jt})'
\]

This is the kernel of an inverse Wishart distribution with covariance matrix \( (S_j + S_{0j})^{-1} \) and \( N_j + \nu_{0j} \) degrees of freedom. We now consider the full conditional distribution for the threshold vector. It is no difficult to see that

\[
p(r|\theta_{y-r}, y, u) \propto \frac{1}{V(S)} 1_S(r) \left\{ \prod_{i=1}^l |\Sigma_{(i)}|^{-N_i/2} \right\} \times \exp \left\{ -\frac{1}{2} \sum_{i=1}^l [y_i - X_i \theta_i]' (I_{N_i} \otimes \Sigma_{(i)}^{-1}) [y_i - X_i \theta_i] \right\}
\]

where \( N_j \) depends on \( r; j = 1, \ldots, l \).

Random-walk Metropolis-Hastings algorithm can be implemented for extracting samples from this full conditional distribution. Notice that this distribution has a similar expression to that found by Safadi and Morettin (2000) in the univariate threshold autoregressive moving average model. Now, the full conditional distribution of \( \gamma_{i,j} \) is Bernoulli with

\[
P(\gamma_{i,j} = 1|\theta_{y-\gamma_{i,j}}, y, u) = \frac{a_{i,j}}{a_{i,j} + b_{i,j}}
\]

where \( a_{i,j} = p(y|u, \theta_{y-\gamma_{i,j}}, \gamma_{i,j} = 1, l) p_{i,j} \) and \( b_{i,j} = p(y|u, \theta_{y-\gamma_{i,j}}, \gamma_{i,j} = 0) (1 - p_{i,j}) \) for \( i = 1, \ldots, \eta_j \) and \( j = 1, \ldots, l \).

### 4.1.2. Posterior distribution using the GVS method

If we use the GVS approach, it is necessary to assume that

\[
\theta_{i,j} | \gamma_{i,j} \sim (1 - \gamma_{i,j}) N(0, \tau_{i,j}^2) + \gamma_{i,j} N(0, c_{i,j}^2 \tau_{i,j}^2)
\]

for specified values \( c_{i,j} > 1, \tau_{i,j} > 0 \), and for \( i = 1, \ldots, \eta_j, j = 1, \ldots, l \). This mixture gives desirable characteristics that are explained in George and McCulloch’s (1993) paper. The assumed expression in (16) can be obtained from a general framework using the multivariate normal prior distribution \( \theta_j | y_j \sim N(0, D_{y_j} D_{y_j}') \) where \( D_{y_j} = \text{Diag} [a_{1,j} \tau_{1,j}, \ldots, a_{\eta_j,j} \tau_{\eta_j,j}] \) with \( a_{i,j} = 1 \) if \( \gamma_{i,j} = 0 \) and \( a_{i,j} = c_{i,j} \) if \( \gamma_{i,j} = 1 \). Then

\[
p(\theta_j | \theta_{y-\theta_j}, y, u) \propto p(y|\theta_{y}, u) p(\theta_j | y_j)
\]

\[
\propto \exp \left\{ -\frac{1}{2} (\theta_j - \theta_j^*) V_j^{-1} (\theta_j - \theta_j^*) \right\}
\]

where \( V_j = [\Gamma_j X'_j (I_{N_j} \otimes \Sigma^{-1}_{(j)}) X_j \Gamma_j + (D_{y_j} RD_{y_j})]^{-1} \) and \( \theta_j^* = V_j (\Gamma_j X'_j (I_{N_j} \otimes \Sigma^{-1}_{(j)}) y_j) \). This is the kernel of a normal multivariate distribution with mean \( \theta_j^* \) and covariance matrix \( V_j^* \).
Now, we proceed in a similar way to the KUO method for obtaining that the full conditional distributions of $\gamma_{i,j}|\theta_{j-\gamma_{i,j}}, y, u; i = 1, \ldots, \eta_j$ and $j = 1, \ldots, l$; are Bernoulli distributions with probability

$$P(\gamma_{i,j} = 1|\theta_{j-\gamma_{i,j}}, y, u) = \frac{a_{i,j}}{a_{i,j} + b_{i,j}}$$

(17)

where $a_{i,j} = p(y|u, \theta_{j-\gamma_{i,j}}, \gamma_{i,j} = 1)p(\theta_{j-\gamma_{i,j}}|\gamma_{i,j})$ and $b_{i,j} = p(y|u, \theta_{j-\gamma_{i,j}}, \gamma_{i,j} = 0)(1 - p_{i,j})$.

To see this, note that

$$P(\gamma_{i,j} = 1|\theta_{j-\gamma_{i,j}}, y, u) = \frac{P(\gamma_{i,j} = 1)p(\theta_{j-\gamma_{i,j}}|y_{i,j} = 1)}{P(\gamma_{i,j} = 1)p(\theta_{j-\gamma_{i,j}}|y_{i,j} = 1) + P(\gamma_{i,j} = 0)p(\theta_{j-\gamma_{i,j}}|y_{i,j} = 0)}$$

and that

$$p(\theta_{j-\gamma_{i,j}}|y_{i,j} = 1) = p(\gamma_{j-\gamma_{i,j}})$$

and

$$p(\theta_{j-\gamma_{i,j}}|y_{i,j} = 0)$$

(17) has a similar expression to the above one, putting $\gamma_{i,j} = 0$ instead of $\gamma_{i,j} = 1$.

We can see that these distributions have similar expressions to those found in So and Chen’s (2003) paper for univariate threshold models. The posterior distributions for the parameters $\gamma$ and $\Sigma_{ij}$ for $j = 1, \ldots, l$ are exactly the same as in the KUO method. It is important to point out that in order to implement Kuo and GVS methods, it is necessary to propose initial autoregressive orders and maximum lags $p, q, d$ for each regime.

In summary, conditional on the number of regimes, we can estimate the threshold values and identify the autoregressive order via the posterior distributions of the $\gamma$’s vectors. It is worth noting that intrinsic values of the non structural parameters are necessary to implement the methodology.

4.2. Estimation of the number of regimes

Two methodologies to estimate the number of regimes in the MTAR model are set forth in this section. The first one is based on the Metropolized Carlin-and-Chib procedure of Dellaportas et al. (2002), which takes into account the changing dimension in the parameter vector when the number of regimes changes; the second uses the marginal likelihood approach of Chib (1995).

4.2.1. Metropolized Carlin-and-Chib for MTAR models

In the context of Carlin and Chib (1995), the problem of identifying the number of regimes $l$ consists, basically, in drawing samples from the joint distribution $p(\Theta, l|y)$, and then marginalizing with respect to $\Theta$, where $\Theta = (\theta_{y,2}, \theta_{y,3}, \ldots, \theta_{y,b})$, and $\theta_{y,m} = (\theta_{ym}, \gamma_1, \ldots, \gamma_m, \gamma_n)$, for a known $b_0$. The Metropolized Carlin-and-Chib procedure is a hybrid modification of the methodology proposed by Carlin and Chib (1995), where a Metropolis step is added to propose a new model. In this case, if the current state model is $m$, i.e., a model
with $m$ regimes, and a model $m'$ is proposed with probability $j(m; m')$, then the acceptance probability of the MTAR model with $m'$ regimes is

$$
\alpha = \min \left( 1, \frac{f(y|\theta_{ym'}, m') p(\theta_{ym'}|m') p(m') j(m', m)}{f(y|\theta_{ym}, m) p(\theta_{ym}|m) p(m) j(m, m')} \right)
$$

where $f(y|\theta_{ym}, m)$ is the likelihood function in (12), $p(\theta_{ym}|m)$ is the prior distribution of all parameters under model $m$, $p(\theta_{ym'}|m)$ is so-called the pseudo prior distribution, and $p(m)$ is the prior probability for a model with $m$ regimes.

In order to implement the Metropolized Carlin-and-Chib method, we use the same prior distributions of Section 4.1. The pseudo prior distributions are the following: (i) Multivariate normal distribution for threshold values and $\theta$’s vectors, (ii) Wishart distribution for the covariance matrix in each regime with a large number of degrees of freedom, and (iii) Bernoulli distributions for the individual $\gamma$’s variables. We also assume that the parameters are independent. In order to estimate the parameters of the pseudo prior distributions, we run a separate sampling for each of the proposed models, and based on the draws, we estimate the parameters for each of these pseudo prior distributions. This is because the pseudo prior distributions have to be approximated to the posterior distributions for every model in order to improve the convergence. This recommendation was given by Dellaportas et al. (2002). In the probability distribution for the jumps $j(m, m')$, we set $j(m, m) = 0$ and $j(m, m') = j(m', m)$; $m \neq m'$. The prior distribution for the model indicator is a discrete uniform distribution and this completes the implementation of the Metropolized Carlin-and-Chib algorithm. We can identify the number of regimes as the mode of $\{m^{(i)}\}_i$, where $m^{(i)}$ is the accepted value of $m$ at iteration $i$.

### 4.2.2. Marginal likelihood procedure

Marginal likelihood can be used as a criterion in the problem of model choice. The idea is as follows: once some MTAR models with $l = 2, 3, \ldots, l_0$ regimes have been proposed, for some $l_0 \geq 2$, we proceed to identify the respective threshold values as in Nieto’s (2005) work using the following non linear AIC:

$$
NAIC = \frac{\left\{ \sum_{j=1}^{l} AIC_j(\mathbf{r}) \right\}}{\left\{ \sum_{j=1}^{l} N_j \right\}}
$$

where $AIC_j(\mathbf{r}) = N_j \ln(|S_j/N_j|) + 2k\eta_j$, $j = 1, \ldots, l$, and $\mathbf{r} \in S$. In doing so, the convergence speed of the chains increases. The methodology developed by Chib (1995) is used here to estimate the marginal likelihood on a logarithmic scale for each of the proposed models. For this case, we chose a posterior ordinate $\theta_{y,l}^* = (\theta_1^*, \ldots, \theta_l^*, \gamma_1^*, \ldots, \gamma_l^*, \text{vec}(\Sigma_{(1)}^*)', \ldots, \text{vec}(\Sigma_{(l)}^*)')'$ as the posterior mode for all involved parameters, based on a previous run of the Gibbs sampling for each model. Now, based on the Kuo method for identifying the autoregressive orders, we can see that the log-likelihood is

$$
\ln f(y|\theta_{y,l}^*) = -\frac{1}{2} (T - p) \ln 2\pi
- \frac{1}{2} \sum_{i=1}^{l} \left[ N_i \ln |\Sigma_{(i)}^*| + (y_j - X_j \Gamma_j^* \theta_j^*)'(I_{N_j} \otimes \Sigma_{(i)}^{*-1})(y_j - X_j \Gamma_j^* \theta_j^*) \right]
$$
and the prior distribution on a logarithmic scale is given by
\[
\ln \pi (\theta^*_i, y) = \sum_{i=1}^{l} \ln \pi (\gamma^*_i) + \ln \pi (\Sigma^*_i) + \ln \pi (\theta^*_i)
\]
\[
= \sum_{i=1}^{l} \left\{ \sum_{j=1}^{n_i} \left[ \gamma^*_i \ln p_{i,j} + (1 - \gamma^*_i,j) \ln(1 - p_{i,j}) \right] \right\}
\]
\[
- \left[ \frac{k v_{0,i}}{2} \ln 2 + \frac{k(k - 1)}{4} \ln \pi + \sum_{j=1}^{k} \ln \Gamma \left( \frac{v_{0,i} + 1 - j}{2} \right) + \frac{v_{0,i}}{2} \ln |S_{0,i}| \right. 
\]
\[
+ \frac{v_{0,i} + k + 1}{2} \ln |\Sigma^*_{(i)}| + \frac{1}{2} \text{tr} \left( \Sigma^{-1}_{(i)} S_{0,i} \right) 
\]
\[
- \frac{1}{2} \left[ \eta_i \ln 2\pi + \ln |\Sigma_{0,i}| + (\theta^*_i - \theta_{0,i}) \Sigma^{-1}_{0,i} (\theta^*_i - \theta_{0,i}) \right] \right\}
\]

If we use the GVS method, we only have to change the prior distribution for the parameters \( \theta \). Finally, the posterior distribution can be split up in the following form:
\[
\pi (\theta^*_i | y) = \pi (\gamma^*_i | y) \pi (\gamma^*_i | y) \cdots \pi (\gamma^*_i | y) \pi (\Sigma^*_i) \pi (\theta^*_i)
\]
\[
\pi (\theta^*_1, \gamma^*_1, \ldots, \gamma^*_l) \cdots \pi (\theta^*_1, \gamma^*_1, \ldots, \gamma^*_l)
\]
\[
\pi (\Sigma^*_1) \cdots \pi (\Sigma^*_l)
\]

Expression (20) can be split up for each individual \( \gamma^*_i \); then, the value of the probability distribution function for every individual variable is \( \pi (\gamma^*_i | \gamma^*_i, \ldots, \gamma^*_i, \ldots, \gamma^*_i, y) \) for \( i = 1, \ldots, n_i \) with \( j = 1, \ldots, l \) and it can be estimated based on \( G \) iterations of the reduced Gibbs sampling for the parameters \( y_i, \ldots, y_{n_i}, y_{i+1}, \ldots, y_l, \Sigma(1), \ldots, \Sigma(i) \), conditional on \( \gamma^*_1, \ldots, \gamma^*_i, \ldots, \gamma^*_i \), except for \( \gamma^*_i \), which uses a run of the full Gibbs Sampling. Then,
\[
\hat{\pi} (\gamma^*_i | \gamma^*_1, \ldots, \gamma^*_i, \ldots, \gamma^*_i, y) = \frac{1}{G} \sum_{g=1}^{G} \text{Be} (\gamma^*_i, p^{(g)})
\]

where \( \text{Be} (\cdot, \cdot) \) means the value of the Bernoulli probability function and
\[
p^{(g)} = p \left( y_{i,j} = 1 | y_i^*, \ldots, y_{i,j}^*, y_{i+1}^*, \ldots, y_{i,j}^*, y_{i+1}^*, \ldots, y_{n_i}^*, \ldots, y_i^* \right)
\]
is obtained using either (15) or (17). Now, every density value \( \pi (\theta^*_j | \gamma^*_1, \ldots, \gamma^*_i, \ldots, \theta^*_j, \ldots, \theta^*_j) \), \( j = 1, \ldots, l \), in Expression (21) can be estimated using \( G \) iterations of the reduced Gibbs sampling for parameters \( \theta_j, \ldots, \theta_i \) conditional on \( \gamma^*_1, \ldots, \gamma^*_i, \ldots, \theta^*_j \). Thus,
\[
\hat{\pi} (\theta^*_j | \gamma^*_1, \ldots, \gamma^*_i, \ldots, \theta^*_j) = \frac{1}{G} \sum_{g=1}^{G} \text{N} (\theta^*_j, \theta^{(g)}_j, V^{(g)}_j)
\]
where $V_j^{(q)} = [\Gamma_j^* X_j'(I_{N_j} \otimes \Sigma_j^{(1)} - 1) X_j \Gamma_j^* + \Sigma_j^{-1}]$ and $\theta_j^{(q)} = V_j^{(q)} (\Gamma_j^* X_j'(I_{N_j} \otimes \Sigma_j^{(0)} - 1) y_j + \Sigma_0^{-1} \theta_0)$. Finally, we have to estimate $\pi(\Sigma_j^{(1)}, \gamma_{j1}^{*}, \ldots, \gamma_{j}^{*}, \theta_{j1}^{*}, \ldots, \theta_{j}^{*}, \Sigma_{1j}^{*}, \ldots, \Sigma_{(j-1)}^{*})$ in (22) for $j = 1, \ldots, l$. These density values correspond to an inverse Wishart distribution with $N_j + v_{0j}$ degrees of freedom and covariance matrix $(S_j + S_{0j})^{-1}$, which only depends on $\theta_j^{*}$ and the data and matrix $S_j$ can be calculated using (13); hence, we only have to evaluate the inverse Wishart density in $\Sigma_j^{(1)}, \pi(\Sigma_j^{(1)}, \gamma_{j1}^{*}, \ldots, \gamma_{j}^{*}, \theta_{j1}^{*}, \ldots, \theta_{j}^{*}, \Sigma_{1j}^{*}, \ldots, \Sigma_{(j-1)}^{*}) = IW(\Sigma_j^{(1)}, N_j + v_{0j}, (S_j + S_{0j})^{-1})$. These last two distributions are similar to those shown in Nieto et al. (2013).

Globally, the methodology to fit an MTAR model in the presence of missing data is given by the following steps:

Step 1. Fill the missing data in the time series with initial values.

Step 2. Identify the number of regimes using the Metropolized Carlin-and-Chib approach or marginal likelihood procedure. If marginal likelihood procedure is used, then the threshold values are identified in this step using the NAIC in (19).

Step 3. Identify the autoregressive orders using the KUO or GVS variable selection methods.

Step 4. Estimate the non structural parameters and the missing data.

Step 5. Go to Step 1 putting now as initial missing-data estimates those obtained in Step 4 and continue the loop until the estimation of the parameters and missing data are practically the same (numerically stable).

5. Simulation results

In this section we illustrate the proposed methodology using an MTAR model with two regimes. The sample size for the simulated series is $T = 1000$ where the first initial 200 values were discarded. Let $U_t = (Z_t, X_t)'$ be a stable VAR(1) process defined as

$$U_t = AU_{t-1} + \nu_t$$

where $A = \begin{pmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{pmatrix}$ and $\{\nu_t\} \sim IIDN(0, \Sigma_\nu)$, where $\Sigma_\nu = \begin{pmatrix} 1.0 & 0.4 \\ 0.4 & 2.0 \end{pmatrix}$. It is well known that this process is a homogeneous Markov chain of order 1.

Now, we set

$$Y_t = \begin{cases} \phi_0^{(1)}(1)Y_{t-1} + \phi_1^{(1)}Y_{t-2} + \beta_0^{(1)}X_{t-1} + \delta_0^{(1)}Z_{t-1} + \Sigma_{(1)}^{1/2} \epsilon_t, & \text{if } Z_t \leq r \\ \phi_0^{(2)}(1)Y_{t-1} + \Sigma_{(2)}^{1/2} \epsilon_t, & \text{if } Z_t > r \end{cases}$$

where

$$\phi_0^{(1)} = \begin{pmatrix} 1.0 \\ -1.0 \end{pmatrix}, \quad \phi_0^{(2)} = \begin{pmatrix} 5.0 \\ 2.0 \end{pmatrix}, \quad \phi_1^{(1)} = \begin{pmatrix} 0.5 & -0.2 \\ -0.2 & 0.8 \end{pmatrix}, \quad \phi_2^{(1)} = \begin{pmatrix} 0.1 & 0.6 \\ -0.4 & 0.5 \end{pmatrix},$$

$$\phi_1^{(2)} = \begin{pmatrix} 0.3 & 0.5 \\ 0.2 & 0.7 \end{pmatrix}, \quad \beta_0^{(1)} = \begin{pmatrix} 0.3 \\ -0.4 \end{pmatrix}, \quad \delta_0^{(1)} = \begin{pmatrix} 0.6 \\ 1.0 \end{pmatrix}, \quad \Sigma_{(1)}^{1/2} = \begin{pmatrix} 1.0 & 0.6 \\ 0.6 & 1.5 \end{pmatrix},$$

$$\Sigma_{(2)} = \begin{pmatrix} 2.5 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}$$

with $r$ the 40th percentile of the variable $Z$. Once the time series was simulated, we found that $r = -0.22$. We set initial values $p = q = d = 3$ in each regime, and with them we obtain the
Table 1. Thresholds for the proposed models in the simulated example.

| Model            | Thresholds | NAIC  |
|------------------|------------|-------|
| 2 Regimes        | -0.2196    | 13.46 |
| 3 Regimes        | -0.2196    | 12.79 |
| 4 Regimes        | -1.0196    | 12.18 |

Additionally, we set $c_{i,j} = 25$, and $\tau_{i,j} = 1.5$ for all $i, j$ for the GVS method. Prior probabilities $p_{i,j}$ are set equal to 0.5. Finally, in five time points, we removed the simulated data of each time series in order to obtain missing observations.

Following the steps proposed at the end of the last section, first, we proceed to fill in the missing data with the medians as initial values. Second, we identify the number of regimes using the Metropolized Carlin-and-Chib marginal likelihood procedures using $l_0 = 4$. The thresholds that were identified, using the NAIC for the proposed models, are presented in Table 1. These thresholds are necessary in order to calculate the marginal likelihood.

With that information, we proceeded to run 12,000 and 3000 iterations with a burn-in period of 2000 and 1000 for the Metropolized Carlin-and-Chib and marginal likelihood procedures, respectively. The results of Metropolized Carlin-and-Chib and marginal likelihood procedures for the identification of the number of regimes are summarized in Tables 2 and 3. We can see that the two procedures identified the true value of the number of regimes.

Now we go to Step 3. Here, we need to identify the autoregressive orders based on the $\gamma$'s vectors with highest posterior probability. These vectors for the GVS and KUO methods are shown in Table 4.

We can see that the vectors with highest posterior probabilities are the same as the vectors in (23). The estimation of the threshold value was $\hat{r} = -0.22$ with 95% credible interval $(-0.22; -0.21)$. This result was obtained using simulated samples from the distribution (14). We can observe that this point estimation is similar to that obtained via NAIC in Table 1. We present in Table 5 the point estimation and the 95% credible intervals of the non structural parameters when GVS method is used. This was Step 4. We can see that all the true values lie within the credible intervals.

Table 2. Marginal likelihood values for the simulated example.

| Model         | GVS log(marginal) | KUO log(marginal) |
|---------------|-------------------|-------------------|
| 2 Regimes     | -2660.008         | -2655.670         |
| 3 Regimes     | -2703.529         | -2700.695         |
| 4 Regimes     | -2800.543         | -2768.929         |

Table 3. Posterior probabilities for the Metropolized Carlin-and-Chib procedure.

|                | 2 regimes | 3 Regimes | 4 Regimes |
|----------------|-----------|-----------|-----------|
| GVS            | 1         | 0         | 0         |
| KUO            | 1         | 0         | 0         |
distribution with mean present their sample autocorrelation functions. Now, we illustrate the estimation of the missing data. The random-walk Metropolis-Hastings algorithm based on a multivariate normal distribution with mean \( \theta \) and covariance matrix \( 0.1I \) was employed for extracting samples from distributions (5) and (6). The results of the estimation of the missing data are presented in Table 6.

**Table 4.** \( \gamma \)'s vectors with the highest posterior probabilities.

|         | \( \gamma \) | Highest probability |
|---------|--------------|---------------------|
| GVS     | (1,1,1,1,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0) | 0.796 |
| KUO     | (1,1,1,1,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0) | 0.670 |

**Table 5.** Point estimation and credible intervals with GVS method for the simulated example.

| Parameter | Regime 1 | Regime 2 |
|-----------|----------|----------|
|           | Estimation | Credible interval 95% | Estimation | Credible interval 95% |
| \( \phi_{0,1} \) | 1.100 | (0.96; 1.24) | 4.923 | (4.66; 5.18) |
| \( \phi_{1,11} \) | 0.467 | (0.44; 0.49) | 0.309 | (0.28; 0.34) |
| \( \phi_{1,12} \) | 0.191 | (−0.21; −0.17) | 0.485 | (0.46; 0.51) |
| \( \phi_{2,11} \) | 0.122 | (0.10; 0.14) | −0.004 | (−2.43; 2.44) |
| \( \phi_{2,12} \) | 0.599 | (0.57; 0.63) | −0.029 | (−0.27; 2.41) |
| \( \phi_{3,11} \) | 0.005 | (−0.23; 2.49) | 0.002 | (−0.45; 2.39) |
| \( \phi_{3,12} \) | 0.014 | (−0.21; 2.45) | −0.022 | (−0.36; 2.43) |
| \( \beta_{1,1} \) | 0.352 | (0.29; 0.42) | −0.037 | (−0.46; 2.39) |
| \( \beta_{2,1} \) | 0.007 | (−0.24; 2.47) | 0.033 | (−0.44; 2.47) |
| \( \beta_{3,1} \) | 0.005 | (−0.25; 2.43) | −0.014 | (−0.48; 2.34) |
| \( \delta_{1,1} \) | 0.592 | (0.47; 0.71) | −0.004 | (−0.39; 2.39) |
| \( \delta_{2,1} \) | 0.025 | (−0.20; 2.44) | 0.019 | (−0.32; 2.37) |
| \( \delta_{3,1} \) | 0.028 | (−0.26; 2.36) | 0.017 | (−0.37; 2.39) |
| \( \phi_{0,2} \) | 0.870 | (−1.07; −0.67) | 1.965 | (1.85; 2.08) |
| \( \phi_{1,21} \) | 0.220 | (−0.26; −0.18) | 0.204 | (0.19; 0.22) |
| \( \phi_{1,22} \) | 0.794 | (0.76; 0.83) | 0.695 | (0.68; 0.71) |
| \( \phi_{2,21} \) | 0.388 | (−0.42; −0.36) | 0.011 | (−0.45; 2.45) |
| \( \phi_{2,22} \) | 0.513 | (0.47; 0.55) | 0.002 | (−0.44; 2.37) |
| \( \phi_{3,21} \) | 0.026 | (−0.25; 2.48) | −0.007 | (−0.46; 2.43) |
| \( \phi_{3,22} \) | −0.000 | (−0.20; 2.41) | 0.004 | (−0.41; 2.46) |
| \( \gamma_{1,2} \) | −0.294 | (−0.39; −0.20) | 0.004 | (−0.40; 2.37) |
| \( \gamma_{2,2} \) | 0.016 | (−0.23; 2.41) | −0.026 | (−0.47; 2.43) |
| \( \gamma_{3,2} \) | 0.000 | (−2.45; 2.46) | 0.010 | (−0.42; 2.44) |
| \( \delta_{2,2} \) | 1.014 | (0.84; 1.18) | 0.022 | (−0.40; 2.46) |
| \( \delta_{3,2} \) | −0.007 | (−0.24; 2.44) | −0.029 | (−0.46; 2.40) |
| \( \Sigma_{11} \) | 0.925 | (0.87; 1.00) | 2.43 | (2.28; 2.58) |
| \( \Sigma_{12} \) | 0.551 | (0.49; 0.61) | 0.461 | (0.39; 0.62) |
| \( \Sigma_{22} \) | 1.427 | (1.34; 1.52) | 0.994 | (0.93; 1.05) |
We can observe that all true values lie within 95% credible intervals except in the first component of $u_t$ for $t = 887$ that lies within the 98% credible interval. With this step we finish the estimation procedure.

| $t$  | $y_t$ | $\hat{y}_t$ | C.I. 95%          | $t$  | $u_t$ | $\hat{u}_t$ | C.I. 95%          |
|------|-------|-------------|-------------------|------|-------|-------------|-------------------|
| 33   | 14.48 | 17.84       | [13.94; 21.87]    | 107  | -0.46 | -0.91       | [-2.20; -0.25]    |
| 368  | -40.44| -40.49      | [-42.19; -38.79]  | 387  | 0.84  | 0.42        | [-0.80; 1.64]     |
| 597  | -0.53 | -1.23       | [-4.27; 1.85]     | 597  | -1.23 | -0.87       | [-2.06; -0.25]    |
| 664  | 6.15  | 8.99        | [5.01; 13.11]     | 651  | -1.10 | -1.46       | [-3.12; -0.33]    |
| 887  | 14.43 | 14.41       | [11.30; 17.50]    | 887  | -0.17 | 1.30        | [-0.16; 3.60]*    |
| 15.43| 14.34 | 14.42       | [12.38; 16.48]    | -0.59| -0.10 | [-6.18; 5.62]|

* True value lies within the 98% credible interval.
We give some suggestions to check the adequacy of the fitted model. To this end, we define the residuals of the model for $t = \max\{p, q, d\} + 1, \ldots, T$, as

$$e_t = \Sigma_{(j)}^{(-1/2)}(Y_t - (I_k \otimes w_k)\theta_j) \text{ if } r_{j-1} < Z_t \leq r_j$$  \hspace{1cm} (24)$$

If there are missing data, they must be replaced by their estimations and then the process $\{e_t\}$ in (24) is called a pseudo-residuals process as in Nieto’s (2005) paper. In Figure 3, we see that the behavior of the CUSUM and CUSUMSQ charts indicates an adequate model specification.

As an interesting fact, we check the performance of the proposed procedure for estimating the autoregressive orders and threshold values. To do that, we simulated 100 replications of the MTAR model and count how many times the correct $\gamma$’s vectors match with the estimated $\gamma$’s vectors with highest or second highest posterior probabilities in each regime; we call that the percentage of correct model. Moreover, the mean probability is calculated, which is the mean value of the probabilities of the $\gamma$’s vectors with the highest or the second highest posterior probabilities. Additionally, we calculate the threshold coverage using the 95% credible intervals.

We show in Table 7 the summary of the above results. In terms of the percentage of the correct model, we can see that the KUO and GVS methods work in a similar manner with little difference in favor of the GVS method. However, the GVS method depends strongly on the prior values $c_{i,j}$ and $\tau_{i,j}$ which must be chosen carefully following the recommendations

Table 7. Efficiency of Kuo and GVS methods and threshold estimation. * Values in brackets are for models with highest or second highest probabilities.

|                      | Regime 1 | Regime 2 |
|----------------------|----------|----------|
| Percentage of correct model | Kuo: 89 [96]* | 76 [97] |
|                      | GVS: 90 [100] | 92 [95] |
| Mean probability | Kuo: 0.52 [0.49] | 0.30 [0.26] |
|                      | GVS: 0.55 [0.52] | 0.37 [0.36] |
| Threshold coverage | Kuo: 99 | 95 |
|                      | GVS: 95 | |
set forth by George and McCulloch (1993). It is important to point out that the GVS and KUO methodologies choose models whose average probabilities are stable (see Figure 4). Additionally, we observe that the estimation of the threshold value was very close to its true value, in the sense that threshold coverage is high.

6. A real data application

In this section we apply the developed methodology set hydrological/meteorological data. The time series are the daily rainfall (in mm) and two river flows (in m$^3$/s) in southern Colombia. The rainfall was measured at the San Juan meteorological station with an altitude of 2400 m and geographical coordinates $2^\circ 2' 7.1''$ north and $76^\circ 29' 47.1''$ west. The first river flow was measured at the El Trebol hydrological station of the Bedon river with altitude of 1720 m and geographical coordinates $2^\circ 15' 0.1''$ north and $76^\circ 7' 42.6''$ west; the second river flow was measured at the La Plata river at the Villalosada hydrological station with an altitude of 1300 m and geographical coordinates $2^\circ 18' 43.9''$ north and $75^\circ 58' 12.5''$ west. The stations are located close to the Earth’s equator in a very dry geographical zone. This last characteristic permits the control for hydrological/meteorological factors, which may distort the kind of dynamic relationship that will be explained by the MTAR model below as was mentioned by Nieto (2005). The sample period is from January 1, 2006, to April 14, 2009 (1200 time points). There are 57 missing data in the rainfall series, 214 in the river flow of the Bedon river and 213 in the flow of the La Plata river. These data were provided by IDEAM, the official Colombian agency for hydrological and meteorological studies. In Figure 5 we present the plots of the time series noting a clear relationship between rainfall and river flows. Let $P_t$ and $Y_t = (Y_{1,t}, Y_{2,t})'$ be the rainfall and the bivariate river flows at day $t$. We consider that the threshold variable is $Z_t = \sqrt{P_{t-1}}$ because of the universal convention for the measurement of this variable and taking in account the marginal heteroscedasticity in $\{P_t\}$. The original bivariate time series $\{Y_t\}$ was transformed due to the marginal heteroscedasticity detected in both variables; thus we set $\tilde{Y}_t = \log(\sqrt{Y_t})$, which means that the transformations are made.
componentwise. Now, we need to check the non-linearity of \( \{\tilde{Y}_t\} \) caused by \( Z_t \) using the non-linearity test proposed by Tsay (1998). Because the time series have less than 20% of missing data, we use the results of Nieto and Hoyos (2011) that allow us to apply Tsay’s (1998) test to the incomplete data. Using Tsay’s notation, Table 8 gives us the statistic \( \hat{C}(0) \), the 0.05-th quantile of the corresponding distribution, \( \chi^2_{0.05} \) say, and the \( p \)-value for the statistical test with null hypothesis a \( \text{VAR}(5) \) process.

We can see that the null hypothesis is rejected at the 5% level. Therefore, the results are a signal of the strong threshold non-linearity of \( \{\tilde{Y}_t\} \), which is explained by \( \{Z_t\} \). Thus, we shall fit an MTAR model to these data. In the first part of the analysis, we completed the time series with the medians of each time series and applied the proposed methodology to fit the model. We propose to use the marginal likelihood procedure to identify the number of regimes, and the results for calculating the thresholds are presented in Table 9 with \( l_0 = 4 \).

Based on these thresholds, the results for marginal likelihood are presented in Table 10 and they show that a model with two regimes is adequate for the data.

Table 8. Non-linearity test result for the empirical data.

| \( p \) | \( \hat{C}(0) \) | \( \chi^2 \) | \( p \)-Value |
|-------|--------|--------|--------|
| 5     | 91.249 | 33.924 | 2.098e−10 |

Table 9. Thresholds for the considered models in the real data example.

| Model     | Thresholds | NAIC |
|-----------|------------|------|
| 2 Regimes | 2.4        | 8.739|
| 3 Regimes | 1.8 3.2    | 7.950|
| 4 Regimes | 1.0 2.4 3.0| 7.424|
Nieto (2005) suggested an approximation to the initial and kernel distributions of the truncated normal distributions at $p = d = 5$. We obtained that $p_1 = 5, d_1 = 4$, and $p_2 = 5, d_2 = 5$. In order to estimate the missing data, we follow the results in Section 3. To do that, we used Nieto’s (2005) Markov chain specification for the threshold variable. Essentially, Nieto (2005) suggested an approximation to the initial and kernel distributions of $[z_t]$, which are given by $f_{in}(z) = p h_n(z) + (1 - p)g(z)$ and $f_n(z_t|z_{t-1}) = p(z_{t-1}) h_n(z_t) + [1 - p(z_{t-1})] g(z_t|z_{t-1})$, respectively, where $p = P[Z = 0] > 0$; $g(z)$ and $g(z_t|z_{t-1})$ are the densities of the truncated normal distributions at $z = 0$ with means $\mu$ and $z_{t-1}$, respectively.

### Table 10. Identification of the number of regimes for the real data example.

| Model      | Log(marginal) |
|------------|---------------|
| 2 Regimes  | 3895.979      |
| 3 Regimes  | 1476.012      |
| 4 Regimes  | 1186.459      |

The GVS method was implemented to identify the autoregressive orders with prior values $c_{i,j} = 2.0$ and $\tau_{i,j} = 0.01$ and initial values $p = d = 5$. We obtained that $p_1 = 5, d_1 = 4$, and $p_2 = 5, d_2 = 5$. In order to estimate the missing data, we follow the results in Section 3. To do that, we used Nieto’s (2005) Markov chain specification for the threshold variable. Essentially, Nieto (2005) suggested an approximation to the initial and kernel distributions of $[z_t]$, which are given by $f_{in}(z) = p h_n(z) + (1 - p)g(z)$ and $f_n(z_t|z_{t-1}) = p(z_{t-1}) h_n(z_t) + [1 - p(z_{t-1})] g(z_t|z_{t-1})$, respectively, where $p = P[Z = 0] > 0$; $g(z)$ and $g(z_t|z_{t-1})$ are the densities of the truncated normal distributions at $z = 0$ with means $\mu$ and $z_{t-1}$, respectively.

### Table 11. Estimation of the non structural parameters for the real data application.

| Parameter | Regime 1       | Credible interval 95% | Regime 2       | Credible interval 95% |
|-----------|----------------|-----------------------|----------------|-----------------------|
| $\phi_{0,1}$ | 0.060 | (0.026; 0.094) | 0.083 | (0.047; 0.121) |
| $\phi_{1,11}$ | 0.164 | (0.129; 0.199) | 0.127 | (0.093; 0.163) |
| $\phi_{1,12}$ | 0.073 | (0.040; 0.106) | 0.076 | (0.044; 0.108) |
| $\phi_{2,11}$ | 0.112 | (0.077; 0.147) | 0.105 | (0.070; 0.140) |
| $\phi_{2,12}$ | 0.061 | (0.028; 0.095) | 0.056 | (0.024; 0.088) |
| $\phi_{3,11}$ | 0.097 | (0.063; 0.131) | 0.079 | (0.045; 0.115) |
| $\phi_{3,12}$ | 0.040 | (0.003; 0.0074) | 0.059 | (0.025; 0.091) |
| $\phi_{4,11}$ | 0.084 | (0.050; 0.119) | 0.081 | (0.046; 0.115) |
| $\phi_{4,12}$ | 0.024 | (–0.010; 0.060) | 0.079 | (0.046; 0.111) |
| $\phi_{5,11}$ | 0.094 | (0.061; 0.127) | 0.083 | (0.049; 0.116) |
| $\phi_{5,12}$ | 0.039 | (0.002; 0.071) | 0.076 | (0.043; 0.108) |
| $\delta_{1,1}$ | 0.022 | (0.016; 0.029) | 0.032 | (0.024; 0.039) |
| $\delta_{2,1}$ | 0.005 | (–0.012; 0.014) | 0.000 | (–0.017; 0.018) |
| $\delta_{3,1}$ | 0.000 | (–0.017; 0.018) | 0.000 | (–0.017; 0.017) |
| $\delta_{4,1}$ | 0.000 | (–0.017; 0.016) | 0.000 | (–0.018; -0.017) |
| $\delta_{5,1}$ | 0.000 | (–0.017; 0.017) | 0.011 | (–0.018; -0.003) |
| $\phi_{0,2}$ | 0.112 | (0.077; 0.149) | 0.101 | (0.064; 0.138) |
| $\phi_{1,21}$ | 0.073 | (0.037; 0.108) | 0.083 | (0.046; 0.120) |
| $\phi_{1,22}$ | 0.224 | (0.190; 0.258) | 0.187 | (0.152; 0.222) |
| $\phi_{2,21}$ | 0.046 | (0.007; 0.082) | 0.060 | (0.021; 0.097) |
| $\phi_{2,22}$ | 0.141 | (0.108; 0.174) | 0.133 | (0.098; 0.168) |
| $\phi_{3,21}$ | 0.049 | (0.006; 0.085) | 0.067 | (0.031; 0.103) |
| $\phi_{3,22}$ | 0.130 | (0.097; 0.163) | 0.125 | (0.090; 0.160) |
| $\phi_{4,21}$ | 0.036 | (–0.004; 0.072) | 0.069 | (0.031; 0.105) |
| $\phi_{4,22}$ | 0.121 | (0.088; 0.155) | 0.117 | (0.082; 0.151) |
| $\phi_{5,22}$ | 0.039 | (–0.001; 0.075) | 0.069 | (0.033; 0.105) |
| $\delta_{1,2}$ | 0.013 | (0.006; 0.019) | 0.036 | (0.025; 0.047) |
| $\delta_{2,2}$ | 0.000 | (–0.018; 0.017) | 0.006 | (–0.019; 0.013) |
| $\delta_{3,2}$ | 0.010 | (–0.001; 0.000) | 0.009 | (–0.022; 0.011) |
| $\delta_{4,2}$ | 0.019 | (–0.026; -0.011) | 0.000 | (–0.018; 0.016) |
| $\delta_{5,2}$ | 0.002 | (–0.016; 0.016) | 0.023 | (–0.033; -0.012) |
| $\Sigma_{11}$ | 0.093 | (0.087; 0.098) | 0.120 | (0.114; 0.128) |
| $\Sigma_{12}$ | 0.015 | (0.010; 0.019) | 0.040 | (0.033; 0.047) |
| $\Sigma_{22}$ | 0.104 | (0.098; 0.111) | 0.186 | (0.175; 0.198) |
and standard deviation \( \sigma; p(z_{t-1}) = P(Z_t = 0|z_{t-1} \in B_j) \) where \( B_j = \{z|r_{j-1} < z \leq r_j\} \) for \( j = 1, \ldots, \hat{l} = 2; \)

\[
h_n(z) = \begin{cases} 
0, & -\infty < z < -1/n \\
(n\pi/2)[\cos(nz\pi + \pi/2) + 1], & -1/n \leq z \leq 0 \\
0, & z > 0 
\end{cases}
\]

with \( n = 100 \). The estimates of those parameters for the approximation in this application are: \( \hat{\rho} = 0.23 \), and \( \hat{P}[Z_t = 0|z_{t-1} \in B1] = 0.8532, \hat{P}[Z_t = 0|z_{t-1} \in B2] = 0.1468, \hat{\mu} = 2.64, \) and \( \hat{\sigma} = 1.885 \). We used the random-walk Metropolis-Hastings algorithm to draw samples of the posterior densities for the estimation of the missing data in the threshold variable where the proposal density considered here is \( p_{h_{n}}(z_{t}) + (1 - p)u(0,m_{z})(z_{t}), \) where \( u(0,m_{z})(z_{t}) \) is the uniform density on the interval \((0, m_{z})\) and \( m_{z} \) is the maximum of the series \( \{z_{t}\} \). The acceptance rates vary between 12\% and 34\% and the convergence was checked using the non parametric Kolmogorov test with different batches between 10 and 20 and with 10,000 iterations, using a burn-in period of 5000 iterations. Thus, we accept that all the chains converged to the stationary distribution. In this context, the estimation of the missing data in the threshold variable was the median of the drawn samples, because the posterior distribution is skewed and takes only positive values or zeroes. Only two iterates were needed to find the final missing-data and parameter estimate. The final estimation of the structural parameters are: \( \hat{l} = 2 \) with \( \log(\text{marginal}) = 4279.243; \hat{r} = 2.6 \) with \( \text{NAIC} = 8.371 \), and the autoregressive orders are \( p_1 = 5, d_1 = 4, \) and \( p_2 = 5, d_2 = 5 \). The estimation of the non structural parameters of the MTAR model is shown in Table 11.

The estimates of some missing data are shown in Table 12, where we can see that none of the credible intervals, particularly the ones related with the output vector, take on negative values in spite of the Gaussian distribution of the innovations (as did happen in Nieto’s, 2005 paper).

Now, in order to verify the adequacy of the model, we checked the pseudo-residuals of the model based on CUSUM and CUSUMSQ charts at the 99\% confidence level, indicating an

**Table 12.** Missing data estimates for real data example. The symbol “—” means that component is not missing.

| \( t \) | \( \hat{y}_t \) | C.I. 95\% | \( t \) | \( \hat{x}_t \) | C.I. 95\% |
|---|---|---|---|---|---|
| 80 | 1.322 | [1.132; 1.516] | 3 | 0 | [0; 1.887] |
| 112 | 1.365 | [1.163; 1.570] | 74 | 2.695 | [0; 5.186] |
| 183 | 1.737 | [1.537; 1.944] | 155 | 5.961 | [3.876; 7.780] |
| 398 | 2.023 | [1.654; 2.401] | 309 | 0 | [0; 2.910] |
| 409 | 0.907 | [0.714; 1.097] | 1.293 | 0 | [0; 2.340] |
| 657 | 1.571 | [1.261; 1.755] | 458 | 0.308 | [0; 4.378] |
| 724 | 1.621 | [1.362; 1.887] | 2.049 | 1.674 | [0; 4.895] |
| 845 | 1.821 | [1.509; 2.156] | 718 | 4.282 | [2.256; 6.762] |
| 1042 | 1.368 | [1.166; 1.570] | 844 | 0 | [0; 4.102] |
| 1198 | 1.801 | [1.589; 2.018] | 1179 | 3.138 | [0; 5.731] |
adequate model fit (see Figure 6). Based on the fitted model we can note interesting empirical facts. First, there is a feedback relationship from the Bedon river flow to the La Plata river flow but not at the contrary, it also happened in Tsay’s (1998) paper. Second, in the second regime, the diagonal entries in the noise weighting matrix are larger than those in the first regime, indicating that the more rainfall the more the variability in the transformed flow rivers. Third, the impact of the precipitation in the second regime is greater than that in the first regime in some cases (see parameters \( \hat{\delta}_{i,s}^{(j)} \), for \( j = 1, 2; i = 1, \ldots, 5; s = 1, 2 \) in Table 11).

7. Conclusions

In this paper, we have proposed a methodology for fitting an MTAR model to time series with missing data. The methodology is based on the Bayesian approach and consists in the phases of (i) model identification via the estimation of the structural parameters and (ii) joint estimation of non-structural parameters and missing data. Obviously, the procedures can be adapted and then applied if there are not missing data in the time series involved. As main facts of the proposed approach, we have (i) included the threshold variable and eventual exogenous variables as covariates in the model equations, (ii) obtained the posterior distributions of certain indicator variables instead of directly obtaining the posterior distributions for the autoregressive orders, via stochastic simulation or Bayesian variable selection approaches, (iii) obtained the posterior distribution of the thresholds conditional on the number of regimes, and (iv) used the Metropolized Carlin-and-Chib procedure or the marginal likelihood approach for estimating the number of regimes.

Acknowledgments

We gratefully acknowledge the very useful comments and suggestions of an anonymous referee, which helped to improve substantially the article.

References

Carlin, B.P., Chib, S. (1995). Bayesian model choice via Markov Chain Monte Carlo methods. J. Royal Stat. Soc. Ser. B. (Methodol.) 57(3):473–484.
Carlin, B.P., Polson, N.G., Stoffer, D.S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *J. Am. Stat. Assoc.* 87(418):473–484.

Carter, C., Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika* 81(3):541–553.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *J. Am. Stat. Assoc.* 90(432):1313–1321.

Dellaportas, P., Forster, J., Ntzoufras, I. (2002). On Bayesian model and variable selection using MCMC. *Stat. Comput.* 12(1):27–36.

Frühwirth-Schnatter, S. (1995). Data augmentation and dynamic linear models. *J. Time Ser. Anal.* 15(2):183–202.

George, E., McCulloch, R. (1993). Variable selection via Gibbs sampling. *J. Am. Stat. Assoc.* 8(423):881–889.

Gómez, V., Maravall, A. (1994). Estimation prediction and interpolation for nonstationary series with the Kalman filter. *J. Am. Stat. Assoc.* 89(426):611–624.

Kuo, L., Mallick, B. (1998). Variable selection for regression models. *Sankhyä: Indian J. Stat. Ser. B* 60(1):65–81.

Kwon, Y., Bozdogan, H., Bensmail, H. (2009). Performance of model selection criteria in Bayesian threshold VAR (TVAR) models. *Economet. Rev.* 28(1-3):83–101.

Nieto, F. (2005). Modeling bivariate threshold autoregressive processes in the presence of missing data. *Commun. Stat. - Theory Methods* 34(4):905–930.

Nieto, F., Hoyos, M. (2011). Testing linearity against a univariate TAR specification in time series with missing data. *Revista Colombiana de Estadística* 34(1):73–94.

Nieto, F., Zhang, H., Li, W. (2013). Using the reversible jump MCMC procedure for identifying and estimating univariate TAR models. *Commun. Stat. - Simul. Comput.* 42(4):814–840.

Robert, C., Casella, G. (2004). *Monte Carlo Statistical Methods.* New York, NY: Springer texts in Statistics. Springer Science+Business Media Inc.

Safadi, T., Morettin, P. (2000). Bayesian analysis of threshold autoregressive moving average models. *Sankhyä: Indian J. Stat. Ser. B* 62(3):353–371.

So, M., Chen, C.W.S. (2003). Subset threshold autoregression. *J. Forecasting* 22(1):49–66.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Stat.* 22(4):1701–1728.

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

Tsay, R.S. (1998). Testing and modeling multivariate threshold models. *J. Am. Stat. Assoc.* 93(443):467–482.

Wu, C., Lee, J. (2011). Forecasting time-varying covariance with a robust Bayesian threshold model. *J. Forecasting* 30(5):451–468.