Revista Colombiana de Estadística - Theoretical Statistics
July 2021, volume 44, issue 2, pp. 369–383
http://doi.org/10.15446/rce.v44n2.91356

Forecasting with Multivariate Threshold Autoregressive Models

Pronósticos con modelos multivariados autorregresivos de umbrales

Sergio A. Calderón-Villanueva[1,a], Fabio H. Nieto[1,b]

[1]Statistics Department, Faculty of Science, Universidad Nacional de Colombia city, Colombia

Abstract

An important stage in the analysis of time series is forecasting of the interest variables. However, the forecasting in non-linear time series models is not straightforward as in linear time series models because an exact analytical expression for the conditional expectation it is not easy to obtain. In this paper, a procedure for forecasting with multivariate threshold autoregressive(MTAR) models is proposed via the so-called predictive distributions in the Bayesian approach. This strategy gives us the forecasts for the response and exogenous variable vectors. The coverage percentages of the forecast intervals and the variability of the predictive distributions are analyzed in this work. An application in the Hydrology field is presented.

***Key words*:** Bayesian approach; Forecasting; Predictive distributions; Coverage percentages; Multivariate threshold autoregressive Model.

Resumen

Una etapa importante en el análisis de series de tiempo es el pronóstico de las variables de interés. Sin embargo, el pronóstico en modelos de series de tiempo no lineales no es directo como en el caso de modelos lineales de series de tiempo porque obtener la forma analítica exacta de la esperanza condicional no es fácil. En este artículo, un procedimiento de pronóstico con modelos multivariados autorregresivos de umbrales(MTAR) es propuesta vía las las llamadas distribuciones predictivas en el enfoque Bayesiano. Esta estrategia nos entrega tanto los pronósticos del vector de respuesta, como el de las variables exógenas. Los porcentajes de cobertura de los intervalos de pronósticos y la variabilidad de las distribuciones predictivas son analizadas en este trabajo. Una aplicación al campo de la hidrología es presentada.

***Palabras clave*:** Enfoque bayesiano; Pronóstico; Distribuciones predictivas; Porcentajes de cobertura; Modelo multivariado Autoregresivo de umbrales.

[a]Ph.D. E-mail: sacalderonv@unal.edu.co
[b]Ph.D. E-mail: fhnietos@unal.edu.co

1. Introduction

Time series analysis is referred to the comprehension of dependence in real problems when the observations are taken in time. Linear and nonlinear models are used to understand those problems; however, sometimes, we can get better results using nonlinear models instead of linear ones. Many nonlinear time series models have been studied in the literature. We can find in Tsay & Chen (2018) an extensive compendium of statistical methodologies for analyzing time series through nonlinear models. For instance, a family of that nonlinear time series models that have been useful in econometrics is the threshold autoregressive models(TAR), see Tong (2015) for a specific review of those models.

Some developments in the analysis of multivariate threshold autoregressive (MTAR) models have been carried out in the last years. Tsay (1998) introduced a methodology to estimate the parameters of de model using conditional least squares and Akaike Information Criterion (AIC) for identifying the so-called structural parameters. Additionally, recursive least squares and predictive residuals in the arranged regression were used to construct a non-linearity test. On the other hand, Bayesian analysis of MTAR models has been studied by some authors. Kwon et al. (2009) proposed a conjugate analysis to find the posterior distributions of the coefficients and covariance matrices, and next, these parameters were integrated out of the posterior distributions to obtain the threshold and the lag value. These authors identified the autoregressive orders via some information criteria. Wu & Lee (2011) also developed a Bayesian methodology to analyze MTAR models with conditional heteroscedasticity without exogenous variables as covariates. Calderon & Nieto (2017) carried out the analysis of MTAR models with missing data in the output and exogenous vectors. Variable selection methodology was employed to identify the autoregressive orders; marginal likelihood and the Metropolised Carlin-and- Chib gave in (Dellaportas et al., 2002) were used to determinate the number of regimes. Informative prior distributions were established in order to get a Bayesian conjugate analysis.

In spite of the latest developments in the MTAR models, the forecasting with these models has not been taken into account in the phase of the model fitting in these studies. A specific problem that is studied by time series analysis is the forecasting of interest variables. We can find an interesting survey of forecasting with nonlinear time series models in Kock & Teräsvirta (2011), including the neural networks model. In Geweke & Amisano (2010), we can also find a forecasting strategy through predictive distribution for asset returns. However, we can find some advances in forecasting with multivariate threshold autoregressive models. De Gooijer & Vidiella-i Anguera (2004) proposed two schemes of simulation(Monte Carlo and Bootstrapping) for obtaining the multi-step ahead forecasts and evaluate the forecasts using three criteria. The schemes do not include the uncertainty of the parameters. Another forecasting procedure that uses Monte Carlo(MC) simulation and the ability to extract samples from the posterior distribution of the parameters was proposed by Kwon (2003). This procedure includes the uncertainty of the some of parameters of the model. Nevertheless, these forecasting procedures do not incorporate, neither, the forecasting of the exogenous variables nor the uncertainty of autoregressive orders. Therefore, in this paper, a methodology is proposed to compute the point forecasts of variables-of-interest that it is based on Calderon & Nieto's (2017) model. Additionally, the uncertainty of the point forecasts is obtained using the predictive distributions. The key issue of the methodology is the use of the uncertainty of the parameters of the model, including the autoregressive orders. This work also allows to deal with missing data and the forecasting of the exogenous variables in the model. The methodology is illustrated with simulation examples and a real-data hydrological/meteorological application. The article is organized as follows. In Section 2, the model with its assumptions and the forecasting procedure are presented. In Section 3, a simulation study is carried out to illustrate the proposed methodology. A real-data application is presented in Section 4 and Section 5 concludes.

2. The model and Forecasting Procedure

2.1. Specifying the MTAR Model

This section is a summary of Calderon & Nieto's (2017) paper, where a Bayesian methodology was developed for fitting the so-called MTAR model. In that paper, we found, particular details about the chosen prior distributions and the MCMC methods that were implemented. Let $\{\mathrm{Y}_t\}$, $\{\mathrm{X}_t\}$ $\{\mathrm{Z}_t\}$ be stochastic processes such that $\mathrm{Y}_t = (\mathrm{Y}_{1t}, \ldots, \mathrm{Y}_{kt})'$, $k \geq 1$, $\mathrm{X}_t = (\mathrm{X}_{1t}, \ldots, \mathrm{X}_{vt})'$, $\nu \geq 1$, and $\{\mathrm{Z}_t\}$ is a univariate process. We say that $\{\mathrm{Y}_t\}$ follows a MTAR model with threshold variable Z_t if, for all t,

$$\mathrm{Y}_t = \boldsymbol{\phi}_0^{(j)} + \sum_{i=1}^{p_j} \boldsymbol{\phi}_i^{(j)} \mathrm{Y}_{t-i} + \sum_{i=1}^{q_j} \boldsymbol{\beta}_i^{(j)} \mathrm{X}_{t-i} + \sum_{i=1}^{d_j} \boldsymbol{\delta}_i^{(j)} \mathrm{Z}_{t-i} + \boldsymbol{\Sigma}_{(j)}^{1/2} \varepsilon_t \quad \text{when} \quad r_{j-1} < \mathrm{Z}_t \leq r_j, \tag{1}$$

where $j = 1, \ldots, l$, $l \in \{2, 3, \ldots\}$ is the number of regimes and $-\infty = r_0 < r_1 < \cdots < r_{l-1} < r_l = \infty$ are the thresholds, which define the regimes. The non-negative integer numbers p_j, q_j and d_j, with $j = 1, \ldots, l$ are called autoregressive orders for each regime. Notice that we have added the threshold variable as a covariate in an autoregressive form. $\{\mathrm{Y}_t\}$, $\{\mathrm{X}_t\}$ and $\{\mathrm{Z}_t\}$ are called respectively the output, covariates and threshold processes. Additionally, the innovation process $\{\varepsilon_t\}$ follows a multivariate independent Gaussian zero-mean process with covariance matrix I_k, the identity matrix of order k, and it is mutually independent of $\{\mathrm{X}_t\}$ and $\{\mathrm{Z}_t\}$.

For $j = 1, \ldots, l$, the coefficients $\boldsymbol{\phi}_i^{(j)}$ for $i = 0, 1, \ldots, p_j$, $\boldsymbol{\beta}_i^{(j)}$ for $i = 1, \ldots, q_j$, $\boldsymbol{\delta}_i^{(j)}$ for $i = 1, \ldots, d_j$ and $\boldsymbol{\Sigma}_{(j)}^{1/2}$ are real matrices of suitable dimensions and we call them non-structural parameters. We also define the vector of non-structural parameters as $\theta_{\text{yns}} = (\theta_1', \ldots, \theta_l', \text{vec}(\Sigma)')'$, with $\theta_j = \text{vec}(A_j)$, $\eta_j = 1 + k \cdot p_j + v \cdot q_j + d_j$, for $j = 1, \ldots, l$ where

$$\mathrm{A}_j = (\phi_0^{(j)}, \phi_1^{(j)}, \ldots, \phi_{p_j}^{(j)}, \beta_1^{(j)}, \ldots, \beta_{q_j}^{(j)}, \delta_1^{(j)}, \ldots, \delta_{d_j}^{(j)})_{k \times \eta_j},$$

and $\Sigma = (\Sigma_{(1)}, \dots, \Sigma_{(l)})$. The autoregressive orders, the threshold values $\mathbf{r} = (r_1, \dots, r_{l-1})'$ and the number of regimes l are known as structural parameters and they are denoted as $\theta_{\text{ys}} = (p_1, \dots, p_l, q_1, \dots, q_l, d_1, \dots, d_l, \mathbf{r}', l)'$. Therefore, the full vector of parameters of the model, denoted MTAR$(l; p_1, \dots, p_l, q_1, \dots, q_l, d_1, \dots, d_l)$, is $\theta_{\text{y}} = (\theta'_{\text{yns}}, \theta'_{\text{ys}})'$.

The MTAR model (1) can be written in the following way:

$$\mathrm{Y}_t = (I_k \otimes \mathrm{w}'_{t,j})\theta_j + \Sigma_{(j)}^{1/2}\varepsilon_t \quad \text{if} \quad r_{j-1} < \mathrm{Z}_t \leq r_j, \tag{2}$$

where $\theta_j = \text{vec}(A'_j)$ for $j = 1, \dots, l$, and $\otimes$ denotes the Kronecker product. We add 0-1 indicator variables $\gamma_{i,j}$, $i = 1, \dots, \eta_j$, $j = 1, \dots, l$; such that if $\gamma_{i,j} = 1$, the associated parameter $\theta_{i,j}$ should be included; if $\gamma_{i,j} = 0$ the associated parameter should not be included. These indicator variables are added to the MTAR model in order to identify the autoregressive orders. Now, let $\boldsymbol{\gamma}_j = (\gamma_{1,j}, \dots, \gamma_{\eta_j,j})'$ be the full vector of indicators in the regime j. Based on that, we can re-write the MTAR model in (2) with the vectors γ_j, $j = 1, \dots, l$ as follows,

$$\mathrm{Y}_t = (I_k \otimes \mathrm{w}'_{t,j})\Gamma_j\theta_j + \Sigma_{(j)}^{1/2}\varepsilon_t \tag{3}$$

$$= (I_k \otimes \mathrm{w}'_{t,j})\vartheta_j + \Sigma_{(j)}^{1/2}\varepsilon_t \tag{4}$$

if $r_{j-1} < \mathrm{Z}_t \leq r_j$, where $\vartheta_j = (\gamma_{1,j}\theta_{1,j}, \dots, \gamma_{\eta_j,j}\theta_{\eta_j,j})$ and $\Gamma_j = \text{Diag}\{\gamma_j\}$ and now the vector of structural parameters is $\theta_{\text{ys}} = (\boldsymbol{\gamma}'_1, \dots, \boldsymbol{\gamma}'_l, \mathbf{r}, l)$. Additionally, we assume that $\{\mathrm{U}_t = (\mathrm{Z}_t, \mathrm{X}'_t)'\}$ is a $(v+1)$-dimensional homogeneous bth order Markov chain with stationary density $f_u(\cdot)$ and transition kernel density $f_u(\cdot|\cdot)$, both with respect to the Lebesgue-measure, where b is an integer number greater than zero. We assume $\{\mathrm{U}_t\}$ is exogenous in the sense that there is no feedback of $\{\mathrm{Y}_t\}$ towards $\{\mathrm{U}_t\}$. The Markovian assumption on $\{\mathrm{U}_t\}$ ensures that a greater quantity of possible models could be used to describe the time series $\{\mathrm{u}_t\}$. Theory and some examples of Markov chains can be found in (Meyn & Tweedie, 2009). In order to identify the models and estimate its parameters, we recommend following these steps in the presence of missing data: 1) fill the missing data in the time series with initial values, 2) identify the number of regimes using the Metropolised Carlin-and-Chib approach or the marginal likelihood procedure. If the marginal likelihood procedure is used, then the threshold values are identified in this step using the non-linear AIC(NAIC). 3) Identify the autoregressive orders using the KUO or GVS variable selection methods. If the Metropolised Carlin-and-Chib approach was employed to identify the number of regimes, use the appropriated full conditional distribution for estimating the threshold values. 4) Estimate the non-structural parameters and the missing data. 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).

2.2. The Forecasting Procedure

In this section, a procedure for forecasting with an MTAR model is proposed. This is the core of this work, and the in-sample fit of the model is only an intermediate step, following Calderon & Nieto's (2017) methodology. The goal is to find $E[\mathrm{Y}_{T+h}|\mathrm{y}_{1:T},\mathrm{u}_{1:T},m]$, where T is the sample size, $h\geq 1$, $\mathrm{y}_{1:T}=(\mathrm{y}_1,\dots,\mathrm{y}_T)$, $\mathrm{x}_{1:T}=(\mathrm{x}_1,\dots,\mathrm{x}_T)$ and $\mathrm{u}_{1:T}=(\mathrm{u}_1,\dots,\mathrm{u}_T)$. This is the best prediction in the sense of the MMSE (minimum mean square error). Nevertheless, an exact analytical expression of that conditional expectation is not easy to obtain in this context of non-linear models. This fact was pointed out in Nieto's (2008) article for univariate TAR models. Therefore, using Bayesian analysis and the quadratic loss function as the optimality criterion, we proceed to find the joint predictive distributions $p(\mathrm{y}_{T+h}|\mathrm{y}_{1:T},\mathrm{u}_{1:T},m)$ for $h\geq 1$, with which we can obtain the target conditional expectations. This distribution can be obtained in the following way:

$$
\begin{aligned}
p(\mathrm{y}_{T+1:T+h},&\mathrm{u}_{T+1:T+h}|\mathrm{y}_{1:T},\mathrm{u}_{1:T},m)= \\
&\int p(\mathrm{y}_{T+1:T+h},\mathrm{u}_{T+1:T+h}|\mathrm{y}_{1:T},\mathrm{u}_{1:T},\theta_{\mathrm{y},m},m)p(\theta_{\mathrm{y},m}|\mathrm{y}_{1:T},\mathrm{u}_{1:T},m)d\theta_{\mathrm{y},m},
\end{aligned}
\tag{5}
$$

where $p(\theta_{\mathrm{y},m}|\mathrm{y}_{1:T},\mathrm{u}_{1:T},m)$ is the posterior distribution of the parameters of an MTAR model with m regimes and $p(\mathrm{y}_{T+1:T+h},\mathrm{u}_{T+1:T+h}|\mathrm{y}_{1:T},\mathrm{u}_{1:T},\theta_{\mathrm{y},m},m)$ is a distribution that must be specified with the assumptions of the MTAR model. In order to specify that distribution, all the assumptions about the MTAR model presented in last section are accepted and also for all t, $\mathrm{Y}_{1:t}$ does not Granger-cause U_t (see Harvey, 1989), then

$$
\begin{aligned}
p(\mathrm{y}_{T+1},\dots,\mathrm{y}_{T+h},&\mathrm{u}_{T+1},\dots,\mathrm{u}_{T+h}|\mathrm{y}_{1:T},\mathrm{u}_{1:T},\theta_{\mathrm{y},m},m)= \\
&\prod_{i=1}^{h}p(\mathrm{u}_{T+i}|\mathrm{u}_{1:T+i-1})p(\mathrm{y}_{T+i}|\mathrm{u}_{T+i},\mathrm{y}_{1:T+i-1},\mathrm{u}_{1:T+i-1},\theta_{\mathrm{y},m},m).
\end{aligned}
\tag{6}
$$

It is worth noting that the densities in (6) for, $i=1,\dots,h$, satisfy the following:

(i) $p(\mathrm{u}_{T+i}|\mathrm{u}_{1:T+i-1})$ is the kernel density of the Markov chain $\{\mathrm{U}_t\}$.

(ii) $p(\mathrm{y}_{T+i}|\mathrm{u}_{T+i},\mathrm{y}_{1:T+i-1},\mathrm{u}_{1:T+i-1},\theta_{\mathrm{y},m},m)$ is a multinormal distribution with mean

$$(I_k\otimes \mathrm{w}'_{T+i,j})\Gamma_j\theta_j$$

and covariance matrix $\mathbf{\Sigma}_{(j)}$ if $r_{j-1}<\mathrm{z}_{T+i}\leq r_j$, with $\mathrm{w}'_{T+i,j},\Gamma_j,\theta_j$ for $j=1,\dots,m$ described in Calderon & Nieto (2017).

To draw samples from (5), we can use **MC** simulation in the following way: for the $i-th$ iteration:

Step 1. Extract a random draw $\theta^{(i)}_{\mathrm{y},m}$ of the posterior distribution $p(\theta_{\mathrm{y},m}|\mathrm{y}_{1:T},\mathrm{u}_{1:T},m)$ following the results shown in Calderon & Nieto (2017).

Step 2. Extract a random draw $\mathrm{u}^{(i)}_{T+1}$ of the kernel density $p(\mathrm{u}_{T+1}|\mathrm{u}_{1:T})$.

Step 3. Extract a random draw $\mathrm{y}^{(i)}_{T+1}$ of the density $p(\mathrm{y}_{T+1}|\mathrm{u}^{(i)}_{T+1},\mathrm{y}_{1:T},\mathrm{u}_{1:T},\theta^{(i)}_{\mathrm{y},m},m)$.

Continue extracting random draws recursively for $\mathrm{u}_{T+j}^{(i)}$ and $\mathrm{y}_{T+j}^{(i)}$ from $p(\mathrm{u}_{T+j}|\mathrm{u}_{T+1:T+j-1}^{(i)}, \mathrm{u}_{1:T})$ and $p(\mathrm{y}_{T+j}|\mathrm{u}_{T+1:T+j}^{(i)}, \mathrm{y}_{T+1:T+j-1}^{(i)}, \mathrm{y}_{1:T}, \mathrm{u}_{1:T}, \theta_{\mathrm{y},m}^{(i)}, m)$ respectively, for $j = 2, \ldots, h$. With the set $\{\mathrm{u}_{t+h}^{(i)}, \mathrm{y}_{t+h}^{(i)}\}_{i,h}$, it is possible to calculate: the mean of the predictive distribution (point forecast), the covariance matrix of the predictive distribution (measure of uncertainty of the forecast) and credible intervals or region intervals for future value of the vector $(U_{T+1:T+h}, Y_{T+1:T+h})$. This procedure allows us to include the uncertainty of the parameters of the MTAR model in the forecast, which generalizes the forecasting procedure proposed by Nieto (2008) and Vargas (2012).

3. Simulation Examples

Let $\mathrm{U}_t = (\mathrm{Z}_t, \mathrm{X}_t)'$ be a stable $VAR(1)$ process defined as

$$\mathrm{U}_t = \mathrm{A}\mathrm{U}_{t-1} + \mathrm{r}_t,$$

with

$$\mathrm{A} = \begin{pmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{pmatrix}$$

and

$$\{\mathrm{r}_t\} \sim IIDN(\mathbf{0}, \Sigma_\mathrm{r}),$$

where

$$\Sigma_\mathrm{r} = \begin{pmatrix} 1.0 & 0.4 \\ 0.4 & 2.0 \end{pmatrix}.$$

It is well known that this process is a homogeneous Markov chain of order 1.

We consider the following models:

Model 1.

$$\mathrm{Y}_t = \begin{cases} \boldsymbol{\phi}_0^{(1)} + \boldsymbol{\phi}_1^{(1)}\mathrm{Y}_{t-1} + \boldsymbol{\phi}_2^{(1)}\mathrm{Y}_{t-2} + \boldsymbol{\beta}_1^{(1)}\mathrm{X}_{t-1} + \boldsymbol{\delta}_1^{(1)}\mathrm{Z}_{t-1} + \boldsymbol{\Sigma}_{(1)}^{1/2}\varepsilon_t, & \text{if } \mathrm{Z}_t \le r \\ \boldsymbol{\phi}_0^{(2)} + \boldsymbol{\phi}_1^{(2)}\mathrm{Y}_{t-1} + \boldsymbol{\Sigma}_{(2)}^{1/2}\varepsilon_t, & \text{if } \mathrm{Z}_t > r, \end{cases}$$

where

$$\boldsymbol{\phi}_0^{(1)} = \begin{pmatrix} 1.0 \\ -1.0 \end{pmatrix}, \quad \boldsymbol{\phi}_0^{(2)} = \begin{pmatrix} 5.0 \\ 2.0 \end{pmatrix},$$

$$\boldsymbol{\phi}_1^{(1)} = \begin{pmatrix} 0.5 & -0.2 \\ -0.2 & 0.8 \end{pmatrix}, \quad \boldsymbol{\phi}_2^{(1)} = \begin{pmatrix} 0.1 & 0.6 \\ -0.4 & 0.5 \end{pmatrix}$$

$$\boldsymbol{\phi}_1^{(2)} = \begin{pmatrix} 0.3 & 0.5 \\ 0.2 & 0.7 \end{pmatrix}, \quad \boldsymbol{\beta}_1^{(1)} = \begin{pmatrix} 0.3 \\ -0.4 \end{pmatrix}, \quad \boldsymbol{\delta}_1^{(1)} = \begin{pmatrix} 0.6 \\ 1.0 \end{pmatrix}$$

$$\mathbf{\Sigma}_{(1)}^{1/2} = \begin{pmatrix} 1.0 & 0.6 \\ 0.6 & 1.5 \end{pmatrix}, \quad \mathbf{\Sigma}_{(2)}^{1/2} = \begin{pmatrix} 2.5 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}$$

with $r = \widehat{z}_{0.4}$, the $40-th$ percentile of the series $\{z_t\}$.

Now, we set the model with three regimes as:

Model 2.

$$\mathrm{Y}_t = \begin{cases} \boldsymbol{\phi}_0^{(1)} + \boldsymbol{\phi}_1^{(1)}\mathrm{Y}_{t-1} + \mathbf{\Sigma}_{(1)}^{1/2}\varepsilon_t, & \text{if } \mathrm{Z}_t \leq r_1 \\ \boldsymbol{\phi}_0^{(2)} + \boldsymbol{\phi}_1^{(2)}\mathrm{Y}_{t-1} + \boldsymbol{\phi}_2^{(2)}\mathrm{Y}_{t-2} + \boldsymbol{\beta}_1^{(2)}\mathrm{X}_{t-1} + \mathbf{\Sigma}_{(2)}^{1/2}\varepsilon_t, & \text{if } r_1 < \mathrm{Z}_t \leq r_2 \\ \boldsymbol{\phi}_0^{(3)} + \boldsymbol{\phi}_3^{(3)}\mathrm{Y}_{t-3} + \boldsymbol{\beta}_2^{(3)}\mathrm{X}_{t-2} + \boldsymbol{\delta}_1^{(3)}\mathrm{Z}_{t-1} + \mathbf{\Sigma}_{(3)}^{1/2)}\varepsilon_t, & \text{if } \mathrm{Z}_t > r_2 \end{cases}$$

where

$$\boldsymbol{\phi}_0^{(1)} = \begin{pmatrix} 2.0 \\ 1.0 \end{pmatrix}, \quad \boldsymbol{\phi}_0^{(2)} = \begin{pmatrix} 0.4 \\ -4.0 \end{pmatrix}, \quad \boldsymbol{\phi}_0^{(3)} = \begin{pmatrix} -3.0 \\ 2.0 \end{pmatrix}$$

$$\boldsymbol{\phi}_1^{(1)} = \begin{pmatrix} -0.9 & 0.0 \\ 0.2 & -0.5 \end{pmatrix}$$

$$\boldsymbol{\phi}_1^{(2)} = \begin{pmatrix} 0.7 & 0.0 \\ 0.0 & 0.6 \end{pmatrix}, \quad \boldsymbol{\phi}_2^{(2)} = \begin{pmatrix} 0.8 & 0.2 \\ 0.0 & -0.4 \end{pmatrix}$$

$$\boldsymbol{\phi}_3^{(3)} = \begin{pmatrix} -0.8 & 0.0 \\ 0.2 & 0.8 \end{pmatrix}$$

$$\boldsymbol{\beta}_1^{(2)} = \begin{pmatrix} 1.2 \\ -0.8 \end{pmatrix}, \quad \boldsymbol{\beta}_2^{(3)} = \begin{pmatrix} -0.6 \\ 0.7 \end{pmatrix}$$

$$\boldsymbol{\delta}_1^{(3)} = \begin{pmatrix} 0.6 \\ 2.0 \end{pmatrix}$$

$$\mathbf{\Sigma}_1^{1/2} = \begin{pmatrix} 1.0 & 0.3 \\ 0.3 & 4.0 \end{pmatrix}, \quad \mathbf{\Sigma}_2^{1/2} = \begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{pmatrix}, \quad \mathbf{\Sigma}_3^{1/2} = \begin{pmatrix} 2.0 & -0.4 \\ -0.4 & 1.0 \end{pmatrix}$$

with $r_1 = \widehat{z}_{0.25}$ and $r_2 = \widehat{z}_{0.75}$, the $25-th$ and $75-th$ percentiles of the time series $\{z_t\}$.

Finally, we consider the following model:

Model 3.

$$\mathrm{Y}_t = \begin{cases} \begin{pmatrix} 1.0 \\ -1.0 \end{pmatrix} + \begin{pmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{pmatrix}\mathrm{Y}_{t-1} + \begin{pmatrix} 0.0 & 0.0 \\ 0.25 & 0.0 \end{pmatrix}\mathrm{Y}_{t-2} + \begin{pmatrix} 0.3 \\ -0.4 \end{pmatrix}\mathrm{X}_{t-1} + \begin{pmatrix} 1.0 & 0.6 \\ 0.6 & 1.5 \end{pmatrix}\varepsilon_t, & \text{if } \mathrm{Z}_t \leq r \\ \begin{pmatrix} 5.0 \\ 2.0 \end{pmatrix} + \begin{pmatrix} 0.3 & 0.5 \\ 0.2 & 0.7 \end{pmatrix}\mathrm{Y}_{t-1} + \begin{pmatrix} 2.5 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}\varepsilon_t, & \text{if } \mathrm{Z}_t > r. \end{cases}$$

This model has less influence of the exogenous variables and its marginal variability is less than that exhibited by Model 1. This model characteristic can be seen in the simulated realizations of the stochastic processes.

We simulate a realization of size $1000 + h$ of the models 1, 2, and 3; next, we take as the effective sample size $T = 1000$, and we proceed to forecast the vectors for $T + h$ with $h \geq 1$. We use the Frobenius norm of the covariance matrix of the predictive distribution (FNPD) to quantify the joint variability of the forecast. Table 1 gives us the forecasting results for the proposed Model 3, while Tables 2 and 3 give us the summary of the forecasting results for models 1 and 2 with $h = 10$ and the last 10000 iterations. We can see that all the true values lie within the 95% individual credible intervals and many of the forecasts are close to them. It is important to point out that the forecasts for Model 3 proposed in this section have FNPD less than the forecasts for Model 1, this may be due to the lesser influence of the exogenous variables and less heteroscedasticity of the new proposed model. Another feature of the forecasting with MTAR models is that FNPD increases as the forecast horizon increases. Finally, we carried out a simulation study in order to check the performance of the proposed forecasting procedure. We simulated 100 realizations of Model 1, Model 2, and Model 3, next it was calculated the percentage of times that true future values lie within individual 95% prediction credible intervals. We can observe in Table 4 that the coverage percentages are high for all models; however, some percentages are smaller than the expected value 95%. The reason for that fact is that we did not calculate coverage percentages using joint prediction intervals. We can establish that the method proposed to forecast MTAR models behaves appropriately.

Table 1: Forecasting output and exogenous vectors for Model 3.

h	y_{1000+h}	$\widehat{\mathrm{y}}_{1000+h}$	FNPD	95% C.I.	u_{1000+h}	$\widehat{\mathrm{u}}_{1000+h}$	FNPD	95% C.I.
1	13.99	18.05	3.77	(11.05;23.99)	-0.54	0.72	1.74	(-1.24;2.67)
	23.16	18.00		(15.54;21.41)	-1.67	0.85		(-1.93;3.62)
2	9.31	17.25	4.98	(8.10;24.69)	-0.58	0.44	2.03	(-1.80;2.66)
	21.04	18.60		(14.92;22.94)	-0.88	0.71		(-2.60;3.99)
3	7.73	16.64	5.90	(6.34;25.24)	-0.94	0.27	2.15	(-2.06;2.64)
	16.54	18.76		(13.37;24.17)	-1.67	0.50		(-2.98;3.98)
4	6.17	16.21	6.65	(5.11;25.48)	-0.94	0.19	2.23	(-2.21;2.61)
	14.23	18.60		(11.03;24.97)	-0.69	0.35		(-3.34;4.07)
5	6.34	15.77	7.30	(4.11;25.72)	-0.53	0.12	2.25	(-2.23;2.55)
	11.85	18.34		(9.14;25.30)	-2.05	0.25		(-3.50;3.89)
6	10.61	15.37	7.84	(3.50;26.02)	0.69	0.08	2.28	(-2.35;2.51)
	11.65	17.95		(7.34;25.84)	-1.91	0.14		(-3.60;3.89)
7	11.06	15.09	8.27	(3.11;26.14)	1.68	0.04	2.29	(-2.39;2.40)
	17.63	17.63		(6.21;26.23)	1.47	0.09		(-3.56;3.82)
8	18.13	14.75	8.57	(2.70;26.23)	0.00	0.01	2.26	(-2.39;2.43)
	12.38	17.28		(5.32;26.22)	3.19	0.04		(-3.69;3.75)
9	11.80	14.46	8.83	(2.42;26.26)	1.77	0.01	2.27	(-2.43;2.47)
	14.10	16.93		(4.58;26.30)	4.66	0.04		(-3.70;3.71)
10	15.80	14.46	9.05	(2.11;26.19)	3.49	0.01	2.29	(-2.37;2.46)
	12.96	16.62		(4.10;26.30)	4.60	0.03		(-3.77;3.75)

Table 2: Forecasting output and exogenous vectors for Model 1.

h	y_{1000+h}	$\widehat{y}_{1000+h}$	FNPD	95% C.I.	u_{1000+h}	$\widehat{u}_{1000+h}$	FNPD	95% C.I.
1	14.66	15.70	5.42	(11.60;21.03)	1.13	0.05	1.73	(-1.91;2.02)
	13.50	11.43		(2.52;16.73)	-0.11	0.38		(-2.36;3.18)
2	16.03	16.16	6.81	(9.06;22.03)	-0.42	0.06	2.02	(-2.21;2.31)
	10.06	10.52		(-1.02;18.24)	1.33	0.22		(-2.99;3.52)
3	12.74	15.51	9.11	(7.47;22.68)	-0.13	0.05	2.13	(-2.23;2.40)
	11.04	8.99		(-9.77;19.38)	-1.72	0.13		(-3.28;3.65)
4	13.97	14.83	11.34	(4.40;23.25)	0.03	0.03	2.23	(-2.40;2.45)
	13.07	7.42		(-18.02;20.20)	-2.77	0.11		(-3.49;3.76)
5	11.11	13.86	13.51	(0.18;23.70)	-0.83	0.03	2.26	(-2.41;2.49)
	6.52	5.87		(-25.83;20.65)	-2.42	0.08		(-3.63;3.72)
6	14.97	12.73	15.62	(-4.66;23.93)	-0.51	0.02	2.29	(-2.47;2.42)
	4.24	4.50		(-33.00;21.09)	-3.82	0.09		(-3.73;3.86)
7	9.32	11.60	17.54	(-9.95;24.25)	-0.91	0.02	2.28	(-2.41;2.42)
	-3.11	3.29		(-38.46;21.42)	-4.08	0.01		(-3.70;3.64)
8	9.73	10.53	19.08	(-15.56;24.40)	-2.26	0.00	2.26	(-2.47;2.44)
	-8.00	2.20		(-41.94;21.54)	-2.08	0.02		(-3.63;3.73)
9	3.84	9.51	20.40	(-20.96;24.60)	-1.71	0.00	2.27	(-2.39;2.40)
	-17.00	1.28		(-45.67;21.71)	-1.78	0.08		(-3.74;3.76)
10	0.23	8.58	21.35	(-24.53;24.49)	-4.32	0.00	2.28	(-2.45;2.42)
	-25.12	0.54		(-45.41;21.82)	-0.95	0.01		(-3.77;3.77)

Table 3: Forecasting output and exogenous vectors for Model 2.

h	y_{1000+h}	$\widehat{y}_{1000+h}$	FNPD	95% C.I.	u_{1000+h}	$\widehat{u}_{1000+h}$	FNPD	95% C.I.
1	-2.32	-1.34	4.07	(-3.94;1.03)	-1.75	-1.91	1.74	(-3.87;0.11)
	2.09	0.92		(-6.60;8.89)	-4.53	-3.02		(-5.80;-0.20)
2	5.24	1.74	4.86	(-3.80;5.94)	-1.76	-1.25	2.04	(-3.48;1.02)
	-3.17	-0.42		(-7.99;8.37)	-1.36	-2.28		(-5.62;1.03)
3	1.02	-1.18	5.53	(-7.48;4.46)	-0.14	-0.85	2.19	(-3.18;1.49)
	-5.08	-0.42		(-8.49;9.29)	-0.20	-1.65		(-5.28;1.96)
4	5.05	0.54	6.02	(-7.88;7.09)	-0.62	-0.58	2.23	(-2.96;1.85)
	-4.92	-0.95		(-8.92;8.83)	-0.01	-1.17		(-4.82;2.46)
5	-0.77	-0.99	6.43	(-10.02;6.93)	-0.87	-0.41	2.24	(-2.82;2.02)
	7.05	-0.91		(-8.91;9.82)	-0.19	-0.81		(-4.45;2.79)
6	0.81	0.02	6.92	(-10.95;8.32)	-0.24	-0.28	2.29	(-2.74;2.17)
	2.56	-1.17		(-9.13;9.39)	-3.78	-0.55		(-4.28;3.18)
7	0.78	-0.81	7.48	(-11.81;8.82)	-1.46	-0.21	2.29	(-2.64;2.25)
	1.29	-1.17		(-9.52;10.37)	-5.42	-0.39		(-4.17;3.38)
8	0.78	-0.81	7.98	(-12.09;10.13)	-1.47	-0.14	2.28	(-2.53;2.30)
	1.29	-1.17		(-9.93;10.11)	-3.25	-0.26		(-3.98;3.55)
9	-0.50	-0.60	8.35	(-13.11;10.64)	-1.60	-0.09	2.28	(-2.51;2.31)
	-0.97	-1.15		(-9.95;10.73)	-2.79	-0.18		(-3.86;3.62)
10	4.32	-0.24	8.74	(-13.32;12.51)	-1.61	-0.06	2.28	(-2.52;2.40)
	4.91	-1.12		(-9.98;10.95)	-3.57	-0.15		(-3.86;3.61)

Table 4: Coverage percentage of the individual 95% prediction credible intervals.

h	Model 2	Model 3	Model 4
1	90	100	92
2	93	96	90
3	88	95	93
4	100	98	95
5	94	93	96
6	97	92	91
7	100	91	95
8	99	96	96
9	96	95	96
10	97	94	95

4. An Empirical Application

In this section, we apply the forecasting procedure to hydrological data. In Calderon & Nieto (2017) was found the relationship between daily rainfall (in mm) and the daily river flow (in m^3/s) of two rivers where one river empties into the other in a region of the department of Cauca in Colombia. The rainfall was measured at the San Juan's meteorological station with an altitude of 2400 meters above the sea level and geographical coordinates 2° 2' 7.1" north and 76° 29' 47.1" west. The first river flow was measured at the El Trebol's hydrological station of the Bedon river with an altitude of 1720 meters above the sea level and geographical coordinates 2° 15' 0.1"north and 76° 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 meters above the sea level and geographical coordinates 2° 18' 43.9" north and 75° 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 of hydrological/meteorological factors, which may distort the kind of dynamical relationship explained by the MTAR model, this fact was mentioned in Nieto (2005) with regard to univariate TAR models. The period of time that we considered is from January 1st, 2006 to April 14th, 2009 (1200 time points) which has 57 time points with missing data in the rainfall series, 214 in the series of the river flow of the Bedon river and 213 in the series of the river flow of the La Plata river. These data were provided by IDEAM, the official Colombian agency for hydrological and meteorological studies. We can note in Figure 1 the time series of the variables proposed for the analysis, these figures show us a strong relationship between the rainfall and the river flows, and a stable behaviour in mean with some bursts of large values which were taken into account in the modelling process.

Let P_t and $\mathrm{Y}_t = (\mathrm{Y}_{1,t}, \mathrm{Y}_{2,t})'$ be the rainfall and the bivariate river flow of the Bedon and La Plata rivers at day t. We consider that the threshold variable is $\mathrm{Z}_t = \sqrt{\mathrm{P}_{t-1}}$ because of universal convention due to measurement of the variables, and for decreasing the heteroscedasticity in $\{\mathrm{P}_t\}$. The original bivariate time series $\{\mathrm{y}_t\}$ was transformed due to the marginal heteroscedasticity in both variables to $\widetilde{\mathrm{y}}_t = \log(\sqrt{\mathrm{y}_t})$, which means that the transformations are made componentwise.

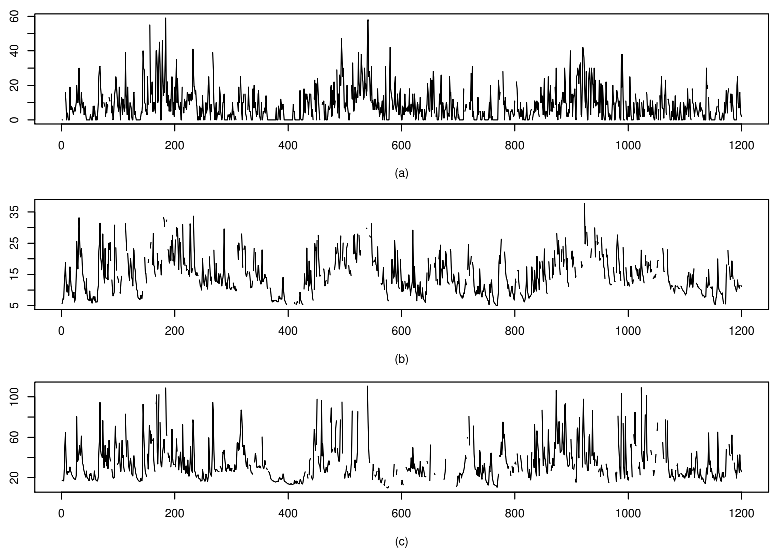

Figure 1: Time series of the real application: (a) Rainfall; (b) River flow of the Bedon River; (c) River flow of the La Plata River.

On the other hand, we can see that the threshold variable has a mixed type distribution, which does not have density with respect to the Lebesgue-measure. Nieto's (2005) ideas suggest an approximation for the distributions that defines the Markov chain $\{\mathrm{Z}_t\}$, the initial distribution

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

where $p = Pr(\mathrm{Z} = 0) > 0$, $F_0(z) = I_{[0,\infty)}(z)$, I denotes the indicator function, and $G(z)$ is a distribution function with Lebesgue-measure density $g(z)$ with support on $(0, \infty)$. The transition kernel is defined as the distribution

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

where $p(z_{t-1}) = Pr(\mathrm{Z}_t = 0|z_{t-1})$ and $G(z_t|z_{t-1})$ is a distribution function that depends on z_{t-1} with Lebesgue-measure density $g(z_t|z_{t-1})$ with support on $(0, \infty)$. The approximation is considered as follows: for each positive integer n, let

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

$\{F_{0n}\}$ is a sequence of distribution functions and converges pointwise to F_0. Additionally, F_{0n} is differentiable at all real number z with first derivative

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

The sequence h_n converges pointwise to δ_0, the Dirac delta function at $z = 0$ which is not a Lebesgue density function. The approximations of the initial and Kernel densities for n sufficiently large are respectively

$$f_{1n} = ph_n(z) + (1 - p)g(z)$$

and

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

we use $n = 100$ for the approximation. We proceed with estimation of the parameters of the initial density. To estimate p, we counted the number of zeroes in the sample of the rainfall and used the relative frequency as an estimation which is $\hat{p} = 0.23$; we consider that $g(z)$ is a normal density truncated at $z = 0$.

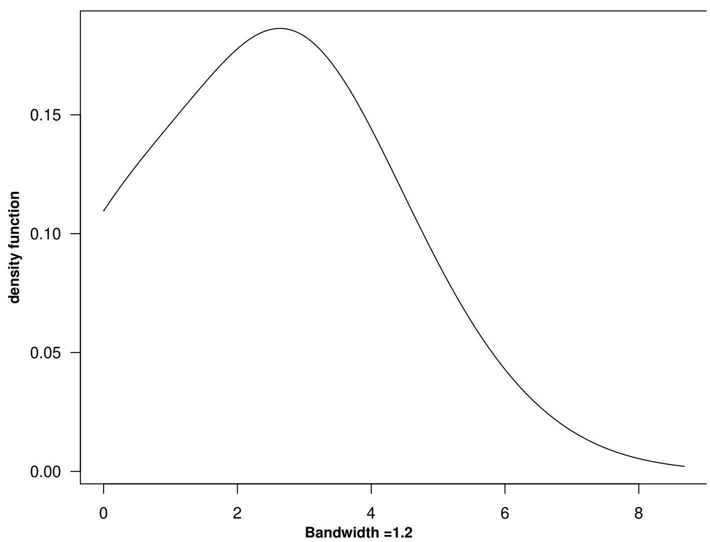

FIGURE 2: Density estimate throughout kernel Gaussian.

A numerical procedure is used to estimate the mean and standard deviation of density $g(z)$, it is based on a non-parametric approximation with Gaussian Kernel, see Figure 2. The maximum of the estimated kernel is $\hat{\mu} = 2.64$ and the inflexion point must have the form $\hat{\mu} + \hat{\sigma}$ which can be found using the second derivative of the estimated kernel, see Figure 3. We can observe an inflexion point located at the interval $(4.44; 4.45)$ and it is approximately 4.445, therefore the estimation of standard deviation is $\hat{\sigma} = 1.885$ and with this we complete the estimation of the parameters for the initial density. For the kernel density, the estimation of $p_j = Pr(\mathrm{Z}_t = 0|z_{t-1} \in B_j)$ where $B_j = \{z|r_{j-1} < z \leq r_j\}$ for $j = 1, \ldots, l$ is done counting the pairs $(z_{t-1} \in B_j, z_t = 0)$ and dividing by the total pairs $(z_{t-1}, z_t = 0)$ which for $m = 2$ they are $\hat{p}_1 = 0.8532$, $\hat{p}_2 = 0.1468$; the kernel is a normal density truncated at $z = 0$ with mean z_{t-1} and standard deviation $\hat{\sigma} = 1.885$.

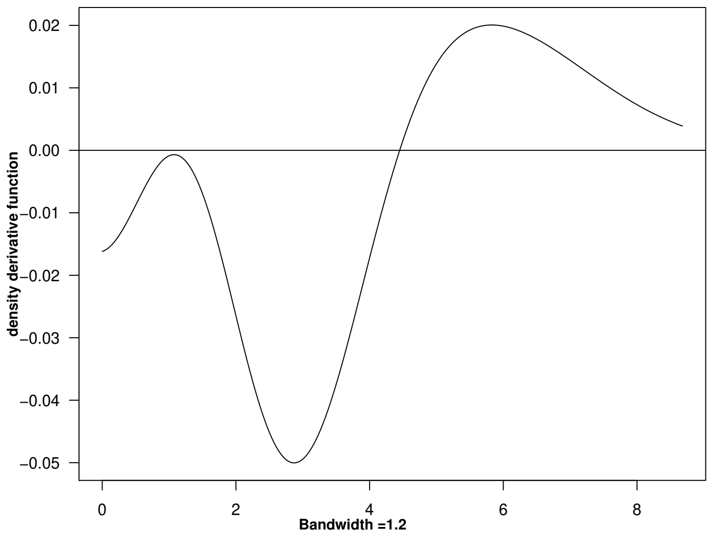

FIGURE 3: Second derivative of estimated density for threshold variable.

TABLE 5: Forecasting of output and threshold variables for the real data.

h	y_{T+h}	$\widehat{y}_{1200+h}$	FNPD	95% C.I.	z_{T+h}	$\widehat{z}_{1200+h}$	FNPD	95% C.I.
1	N.A. N.A.	1.166 1.561	0.160	(0.973;1.394) (1.339;1.820)	3.00	0.601	1.309	[0;4.522)
2	N.A. N.A.	1.153 1.564	0.182	(0.944;1.447) (1.317;1.870)	2.828	0.738	1.566	[0;5.627)
3	1.097 1.458	1.135 1.549	0.197	(0.913;1.466) (1.300;1.893)	2.645	0.701	1.582	[0;5.623)
4	1.042 1.445	1.120 1.540	0.207	(0.888;1.475) (1.278;1.881)	2.449	0.781	1.645	[0;5.827)
5	1.166 1.504	1.112 1.538	0.216	(0.872;1.487) (1.266;1.907)	1.414	0.789	1.668	[0;5.801)
6	1.313 1.739	1.100 1.530	0.225	(0.850;1.483) (1.251;1.921)	4.24	0.801	1.695	[0;5.983)
7	1.216 1.556	1.091 1.520	0.231	(0.835;1.477) (1.222;1.908)	3.605	0.785	1.671	[0;5.937)
8	1.179 1.435	1.080 1.508	0.236	(0.821;1.462) (1.210;1.901)	2.236	0.779	1.665	[0;5.933)
9	1.147 1.394	1.073 1.500	0.239	(0.810;1.473) (1.189;1.885)	1.732	0.785	1.668	[0;5.960)
10	1.142 1.393	1.065 1.491	0.242	(0.793;1.466) (1.179;1.895)	1.414	0.779	1.663	[0;5.887)

N.A. means that it is a missing value.

It was also found that the estimation of the number of regimes and the threshold was $\hat{l} = m = 2$ and $\hat{r} = 2.6$. With this information and replacing the missing data with their estimation, we proceed with the forecasting procedure. The forecasting for both threshold variable and output vector was carried out regarding a forecast horizon of $h = 10$, where the final estimates of the missing data were used to complete the series. The forecasts, credible intervals and the FNPD are shown on Table 5 based on a run of 15000 iterations with a burning period of 5000. All credible intervals take zeroes or positive values only, this is an important fact because the forecasting of the univariate TAR model in Nieto (2008) exhibited negative values in the credible intervals for the output vector in the hydrological application.

5. Conclusions

In this paper, we have derived the Bayesian predictive distributions of the MTAR model, which involve the uncertainty of the model parameters. The forecasts are obtained using simulation methods to draw samples of the joint predictive distribution. Additionally, the procedure includes how to deal with missing data. We note that the forecasting procedure behaves well based on the results of the simulated and real data examples. As future work, we will consider the influence of prior distributions for the parameters models in the forecasting performance.

[Received: November 2020 — Accepted: May 2021]

References

Calderon, S. & Nieto, F. (2017), 'Bayesian analysis of Multivariate Threshold Autoregressive Models with Missing Data', *Communications in Statistics: Theory and Methods* **46**(1), 296–318.

De Gooijer, J. G. & Vidiella-i Anguera, A. (2004), 'Forecasting threshold cointegrated systems', *International Journal of Forecasting* **20**(2), 237–253.

Dellaportas, P., Forster, J. & Ntzoufras, I. (2002), 'On Bayesian model and variable selection using MCMC', *Statistics and Computing* **12**(1), 27–36.

Geweke, J. & Amisano, G. (2010), 'Comparing and evaluating Bayesian predictive distributions of asset returns', *International Journal of Forecasting* **26**(2), 216–230.

Harvey, A. (1989), *Forecasting, structural time series model and the Kalman filter*, Cambridge University Press, Cambridge.

Kock, A. B. & Teräsvirta, T. (2011), Forecasting With Nonlinear Time Series Models, *in* M. P. Clements & D. F. Hendry, eds, 'The Oxford Handbook of Economic Forecasting', Oxford University Press, Oxford, chapter 2, pp. 61–87.

Kwon, Y. (2003), Bayesian Analysis of Threshold Autoregressive Models, PhD thesis, University of Tennessee - Knoxville.

Kwon, Y., Bozdogan, H. & Bensmail, H. (2009), 'Performance of Model Selection Criteria in Bayesian Threshold VAR (TVAR) Models', *Econometric Reviews* **28**(1-3), 83–101.

Meyn, S. & Tweedie, R. (2009), *Markov Chains and Stochastic Stability*, Oxford Statistical Science, Cambridge University Press, Cambridge.

Nieto, F. (2005), 'Modeling Bivariate Threshold Autoregressive Processes in the Presence of Missing Data', *Communications in Statistics - Theory and Methods.* **34**(4), 905–930.

Nieto, F. (2008), 'Forecasting with univariate TAR models', *Statistical Methodology* **5**(3), 263–273.

Tong, H. (2015), 'Threshold models in time series analysis-Some reflections', *Journal of Econometrics* **189**(3), 485–491.

Tsay, R. S. (1998), 'Testing and Modeling Multivariate Threshold Models', *Journal of the American Statistical Association* **93**(443), 467–482.

Tsay, R, S. & Chen, R. (2018), *Nonlinear Time Series Analysis*, Wiley Series in Probability and Statistics, John Wiley and Sons.

Vargas, L. (2012), Cálculo de la distribución predictiva en un modelo TAR, Master's thesis, Universidad Nacional de Colombia, Colombia.

Wu, C. & Lee, J. (2011), 'Forecasting Time-Varying Covariance with a Robust Bayesian Threshold Model', *Journal of Forecasting* **30**(5), 451–468.