Bayesian Analysis of Multivariate Threshold Autoregressive Models with Missing Data

Sergio Alejandro Calderón Villanueva
Statistics, Master
Code: 1832255

Universidad Nacional de Colombia
Facultad de Ciencias
Departamento de Estadística
Bogotá, D.C.
Febrero de 2013

Bayesian Analysis of Multivariate Threshold Autoregressive Models with Missing Data

Sergio Alejandro Calderón Villanueva
Statistics, Master
Code: 1832255

Thesis Work to Obtain the Degree of
Doctorade in Statistics

Advisor
Fabio Humerto Nieto Sánchez, Ph.D.
Doctor in Statistics

Research line
Time Series

Research group
Time Series

Universidad Nacional de Colombia
Facultad de Ciencias
Departamento de Estadística
Bogotá, D.C.
Febrero de 2013

Title in English

Bayesian Analysis of Multivariate Threshold Autoregressive Models with Missing Data

Título en español

Analisis Bayesiano de Modelos Multivariados Autoregresivos de Umbrales con datos Faltantes

Abstract: In some fields, we are forced to work with missing data in multivariate time series, unfortunately the analysis in this context can not be done as in the case of complete data. Bayesian analysis of multivariate thresholds autoregressive models(MTAR) with exogenous inputs and missing data is carried out. MCMC methods are used to get samples of the marginal posterior distributions, including threshold values and missing data. In order to identify autoregressive orders, we adapt the Bayesian variable selection method to the MTAR models. Number of regimes is estimated using marginal likelihood and product space strategies. The Forecasting of the output vector is implemented finding its predictive distributions. Simulation experiments and real data examples are presented.

Resumen: En algunos campos, nos vemos forzados a trabajar con datos faltantes en series de tiempo multivaridas, desafortunadamente el análisis en este contexto no puede ser hecho como en caso completo. El análisis de modelos multivaridos autoregresivos de umbrales(MTAR) con entradas exogenas y datos faltantes es llevado a cabo vía el enfoque Bayesiano. Los métodos MCMC son usados para obtener muestras de las distribuciones marginales aposteriori, incluyendo los valores de los umbrales y los datos faltantes. Con el objetivo de identificar los órdenes autoregresivos, el método Bayesiano de selección de varibales es adaptado para modelos MTAR. El número de regímenes es estimado usando la versimilitud marginal y las estrategias de espacio producto. El pronóstico para el vector de salida es implementado encontrando sus densidades predictivas. Experimentos de simulación y ejemplos con datos reales son presentados.

Keywords: Bayesian Analysis; Bayesian variable selection; Monte Carlo Markov Chain; Missing data; Multivariate threshold autoregressive model.

Palabras clave: Análisis Bayesiano; Selección Bayesiana de variables ; Cadenas de Markov Monte Carlo ; Datos Faltantes; Modelos multivariados autoregressivos de umbrales

Acceptation Note

Thesis Work
Aprobado

"Meritoria o Laureada mention"

Jury
——-

Jury
——-

Jury
——-

Advisor
Nieto Sánchez Fabio

Bogotá, D.C., Febrero de 2013

Contents

List of Tables

List of Figures

Introduction

The analysis of time series through statistical non-linear models has become more popular in the last years; in special, because many series exhibit some features that a simple linear model can not reproduce in its sample-paths, for example, asymmetrical cycles, time irreversibility, multimodality, volatility clustering, etc. Another example for the popularity of these models is the computational advance. The numerical solutions for the estimation algorithms in the classical approach have been improved in order to make them more precise and to accelerate the convergence speed. On the other hand, the Markov Chain Monte Carlo (MCMC) methods have facilitated to extract samples of non-standard posterior distributions. However, some drawbacks can arise in the analysis, particularly, in the hydrology and meteorology fields, we have to deal with missing data in the sense of Nieto (2005), the variable(s) of interest had realizations in the sample period considered but were not physically observed. Therefore, the analysis of time series in this context is not straightforward and compromises the following stages: identification of a model, estimation of parameters of the model and missing data, and forecasting of the interest variables.

We can find some advances in the analysis of non-linear time series with missing data in both classical and Bayesian approach. In the classical approach, Tong (1990) pointed out the missing data problem in Markovian models, which specifically can be non-linear time series models. The EM algorithm was suggested for the estimation of the parameters; however, the estimation of the missing data was not contemplated. Another proposal is considered in (Thavaneswaran & Bovas, 1991), they used the idea of estimating functions (optimal estimating) to find a recursive estimation of the parameters vector. The missing data are handled through the recursive algorithm as a non-linear state space model, and the estimation can be obtained following the ideas of Bovas & Thavaneswaran (1991). On the other hand, the analysis of univariate threshold autoregressive (TAR) models with missing data was made in (Nieto, 2005) using the Bayesian methodology of Carlin & Chib (1995), in order to identify the number of regimes and the autoregressive orders in two steps. Estimation of thresholds is made before to start Bayesian procedure via the Non-linear Akaike information criterion (NAIC) (Tong, 1990). Coefficients and variances are estimated as an intrinsic step by Gibbs sampling. Missing data in the output variable are estimated using results in (Carter & Konh, 1994), and a novel smoother was found to estimate missing data in the threshold variable. The convergence of the generated Markov chains is fast. The forecasting was tackled by Nieto (2008), in both output and threshold variables finding the predictive distributions. Now, a new problem lies when we consider to fit vector time series with possibly exogenous inputs to a Multivariate Threshold Autoregressive Model (MTAR) in the context of missing data.

Of course, classical and Bayesian approaches have been proposed to analyse vector time series by means of MTAR models without missing data. In (Tsay, 1998) the parameters were estimated using conditional least squares and Akaike information criterion (AIC); moreover, 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 was studied in (Kwon et al., 2009). A conjugate analysis was used to find posterior distributions of coefficients and covariance matrices, and next, these parameters are integrating out to find posterior distributions of thresholds and lag value. They identified autoregressive orders, that must be the same in each regime, extracting criteria of information. Additionally, they analysed the performance of the proposed criteria, which were becoming better when the sample size is increased. In both approaches, we have to know the number of regimes. At the moment and based on the knowledge of the author, there is not a methodology to analyse MTAR models with missing data in the output, exogenous and threshold time series. Moreover, the methodology available with complete data has some restrictions, namely, in the approaches cited above, the number of regimes must be known. On the other hand, in other approaches where the number of regimes and the autoregressive orders are jointly estimated, no exogenous variables are included in the equation of the model, and in practice the autoregressive orders should be the same in each regime as in (Wu & Lee, 2011) to reduce de number of promising models. Finally, the forecasting for output vector has not be implemented.

We propose a Bayesian analysis of MTAR models with missing data using Markov Chain Monte Carlo methods. The model considered here is a modification with respect to the original Tsay's model, and it is described below. Estimation of matrix parameters and the threshold values is done, also, joint identification of the autoregressive orders and the number of regimes is presented. In addition, we consider a step of forecasting for the output vector. The contents are organized as follows. In chapter 1, the MTAR model is specified with all its assumptions. In chapter 2, we show the results to estimate missing data, coefficients and covariance matrices conditional on the autoregressive orders, the threshold values and the number of regimes. In chapter 3, we use two methods of Bayesian variable selection to identify the autoregressive orders, and we also obtain results to estimate the threshold values without the presence of missing data. Furthermore, the identification of the number of regimes using two methodologies is proposed. In chapter 4, we find the predictive distributions to forecast the output vector. In chapter 5, we carry out a simulation study to assess the performance of the proposal methodology and some suggestions are given to check the proposed model. In chapter 6, an application to the Colombian hydrology is presented; we sketch a practical procedure when there are missing data to analyse the real data. At the end, we make the conclusions of the methodology proposed.

CHAPTER 1

Specifying the multivariate threshold autoregressive model

Let $\{\mathrm{Y}_t\}$ and $\{\mathrm{X}_t\}$ be stochastic processes such that $\mathrm{Y}_t = (\mathrm{Y}_{1t}, \ldots, \mathrm{Y}_{kt})'$, $\mathrm{X}_t = (\mathrm{X}_{1t}, \ldots, \mathrm{X}_{vt})'$ and $\{\mathrm{Z}_t\}$ is a univariate process. $\{\mathrm{Y}_t\}$ follows a MTAR model with threshold variable Z_t if

$$\mathrm{Y}_t = \boldsymbol{\phi}_0^{(j)} + \sum_{i=1}^{p_j} \boldsymbol{\phi}_i^{(j)} \mathrm{Y}_{t-i} + \sum_{i=1}^{q_j} \boldsymbol{\beta}_i^{(j)} \mathrm{X}_{t-i} + \sum_{i=1}^{d_j} \boldsymbol{\delta}_i^{(j)} \mathrm{Z}_{t-i} + \boldsymbol{\Sigma}_{(j)}^{1/2} \varepsilon_t \quad \text{if} \quad r_{j-1} < \mathrm{Z}_t \leq r_j \tag{1.1}$$

where $j = 1, \ldots, l$, $l \in \{2, 3, \ldots\}$ is the number of regimes, $-\infty = r_0 < r_1 < \cdots < r_{l-1} < r_l = \infty$ are the thresholds, which define the regimes. We can see that this model is slight different from the model proposed by Tsay (1998), we have added the threshold variable as a covariate in an autoregressive form; moreover, the delay value does not appear and it must be identified previously. $\{\mathrm{Y}_t\}$, $\{\mathrm{X}_t\}$, $\{\mathrm{Z}_t\}$ are called output, covariates and threshold processes respectively.

Additionally, the innovation process $\{\varepsilon_t\}$ follows a multivariate independent Gaussian zero-mean process with covariance identity matrix I_k, it is mutually independent of $\{\mathrm{X}_t\}$ and $\{\mathrm{Z}_t\}$. This kind of independence is in the sense for any integer n and n time points $t_1, \cdots, t_n$, the random vectors $(\mathrm{X}_{t_1}, \cdots, \mathrm{X}_{t_n})$ and $(\varepsilon_{t_1}, \cdots, \varepsilon_{t_n})$ are independent. 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 structural parameters as $\theta_{\mathrm{yns}} = (\theta_1', \ldots, \theta_l', \mathrm{vec}(\Sigma)')'$, with $\theta_j = \mathrm{vec}(A_j)$, $\eta_j = 1 + k \cdot p_j + v \cdot q_j + d_j$, for $j = 1, \cdots, l$ 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 $\Sigma = (\Sigma_{(1)}, \ldots, \Sigma_{(l)})$. The integer numbers p_j, q_j and d_j with $j = 1, \ldots, l$ are called autoregressive orders for each regime and together with the threshold values $\mathbf{r} = (r_1, \ldots, r_{l-1})'$ and the number of regimes are known as structural parameters and they are denoted as $\theta_{\mathrm{ys}} = (p_1, \ldots, p_l, q_1, \ldots, q_l, d_1, \ldots, d_l, \mathbf{r}', l)'$. Therefore, the full vector of parameters of the model MTAR$(l; p_1, \ldots, p_l, q_1, \ldots, q_l, d_1, \ldots, d_l)$ is $\theta_{\mathrm{y}} = (\theta_{\mathrm{yns}}', \theta_{\mathrm{ys}}')'$. It is important to point out that this model is piecewise linear in the space $\mathrm{Z}_t, \cdots, \mathrm{Z}_{t-d}$,

$\mathrm{X}_{t-1},\cdots,\mathrm{X}_{t-q}$, $\mathrm{Y}_{t-1},\cdots,\mathrm{Y}_{t-p}$ with $p=\max\{p_1,\cdots,p_l\}$, $q=\max\{q_1,\cdots,q_l\}$, $d=\max\{d_1,\cdots,d_l\}$, but it is non-linear in the time. We can find in the literature many applications since the introduction of this model, for instance, in (Tsay, 1998) we can see applications to index futures arbitrage, interest rates and hydrology. In (Hansen, 2011) we can see an extensive review of the applications of threshold models in econometrics and economics.

We assume that there are missing data in all time series, that is, we have unequally-spaced time series, and specifically, the observed data are in the time points $t_1,\cdots,t_{N_1}$ with $1\le t_1\le\cdots\le t_{N_1}\le T$, $s_1,\cdots,s_{N_2}$ with $1\le s_1\le\cdots\le s_{N_2}\le T$, and $h_1,\cdots,h_{N_3}$ with $1\le h_1\le\cdots\le h_{N_3}\le T$ for the time series $\{\mathrm{y}_t\}$,$\{\mathrm{x}_t\}$ and $\{\mathrm{z}_t\}$ respectively. In order to tackle the presence of missing data in the threshold and covariates processes, $\{Z_t\}$ and $\{X_t\}$ respectively, we suppose 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 $f_u(\cdot|\cdot)$ with respect to Lebesgue-measure, where b is an integer number greater than zero. We assume $\{\mathrm{U}_t\}$ is exogenous in the sense that there is not feedback of $\{\mathrm{Y}_t\}$ towards $\{\mathrm{U}_t\}$. We consider this model as a multivariate generalization of the model analysed in (Nieto, 2005). The Markovian assumption on $\{\mathrm{U}_t\}$ ensure that a longer quantity of possible models could be fitted to the new time series $\{\mathrm{u}_t\}$; theory and some examples of Markov chains can be found in (Meyn & Tweedie, 2009).

Let $\theta=(\theta_{\mathrm{y}},\theta_{\mathrm{u}})$ be the full vector of parameters, which is composed by the vector of parameters in the MTAR model θ_y and the vector of parameters of the Markov chain θ_{u}. We suppose the probabilistic mechanism generating $\mathbf{u}=(\mathrm{u}_1,\ldots,\mathrm{u}_T)$ does not depend on θ_y and the joint density of $\mathbf{y}=(\mathrm{y}_1,\ldots,\mathrm{y}_T)$ conditional on $\mathbf{u}$ and $(\theta_{\mathrm{y}},\theta_{\mathrm{u}})$ does not depend on θ_{u}. With the last assumptions, we can see that it is possible to estimate and identify, first, the parameters of the Markov chain $\{\mathrm{u}_t\}$, and next, conditional on the estimated parameters θ_{u}, we can proceed to estimate the parameters of the MTAR model considering the likelihood function

$$\begin{aligned}L(\theta|\mathbf{y},\mathbf{u})=f(\mathbf{y},\mathbf{u}|\ \theta_{\mathrm{y}},\theta_{\mathrm{u}})&=f(\mathbf{u}|\ \theta_{\mathrm{y}},\theta_{\mathrm{u}})f(\mathbf{y}|\mathbf{u},\theta_{\mathrm{y}},\theta_{\mathrm{u}})\\&=f(\mathbf{u}|\ \theta_{\mathrm{u}})f(\mathbf{y}|\mathbf{u},\theta_{\mathrm{y}}),\end{aligned}\tag{1.2}$$

which is exactly the same as proposed in (Nieto, 2005). Hence the parameters of $\{\mathrm{U}_t\}$ can be previously estimated or even known a priori. As a future research, we can consider the joint estimation of the parameters θ_{y} and θ_{u}.

CHAPTER 2

Missing data and non-structural parameters estimation

In this chapter, we give the results to estimate missing data and non-structural parameters conditional on structural parameters. We use a space-state model to handle and estimate the missing data.

2.1 Estimation of missing data

In order to estimate missing data, we write the MTAR model into a state space model with regime-switching where the matrices depend on threshold variable. With this end, we set the following elements, the state vector

$$\alpha_t = \left[\mathrm{Y}_t', \quad \mathrm{Y}_{t-1}', \quad \cdots, \quad \mathrm{Y}_{t-p+1}', \quad \mathrm{X}_t', \quad \mathrm{X}_{t-1}', \quad \cdots, \quad \mathrm{X}_{t-q+1}', \quad \mathrm{Z}_t, \quad \mathrm{Z}_{t-1}, \quad \cdots \quad , \mathrm{Z}_{t-d+1}\right]'$$

which first k components are exactly the same components of the vector Y_t. The vector and matrices $L(\mathrm{Z}_t)$, $H(\mathrm{Z}_t)$, $M(\mathrm{Z}_t)$, $R(\mathrm{Z}_t)$, $K(\mathrm{Z}_t)$, $N(\mathrm{Z}_t)$, with the condition $\phi_i^{(j)} = 0$ for $i > p_j$, $\beta_i^{(j)} = 0$ for $i > q_j$ and $\delta_i^{(j)} = 0$ for $i > d_j$ are

$\mathrm{H}_t(\mathrm{Z}_t) =$

$$\begin{pmatrix}
\phi_1^{(\mathrm{Z}_t)} & \phi_2^{(\mathrm{Z}_t)} & \cdots & \phi_{p-1}^{(\mathrm{Z}_t)} & \phi_p^{(\mathrm{Z}_t)} & \beta_1^{(\mathrm{Z}_t)} & \beta_2^{(\mathrm{Z}_t)} & \cdots & \beta_{q-1}^{(\mathrm{Z}_t)} & \beta_q^{(\mathrm{Z}_t)} & \delta_1^{(\mathrm{Z}_t)} & \delta_2^{(\mathrm{Z}_t)} & \cdots & \delta_{d-1}^{(\mathrm{Z}_t)} & \delta_d^{(\mathrm{Z}_t)} \\
I_k & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & \cdots & I_k & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & I_v & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & I_v & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 & 0
\end{pmatrix}$$

$$L(\mathrm{Z}_t) = \left[\boldsymbol{\phi}_0^{\prime(\mathrm{Z}_t)}, \quad 0, \quad \cdots, \quad 0, \quad 0, \quad 0, \quad \cdots, \quad 0, \quad 0, \quad 0, \cdots, \quad 0\right]'$$

$$M_t(\mathrm{Z}_t) = \begin{bmatrix} 0, & 0, & \cdots, & 0, & 0, & 0, & \cdots, & 0, & 1, & 0, & \cdots, & 0 \\ \mathbf{0}, & \mathbf{0}, & \cdots, & \mathbf{0}, & I_v, & \mathbf{0}, & \cdots, & \mathbf{0}, & \mathbf{0}, & \mathbf{0}, & \cdots, & \mathbf{0} \end{bmatrix}'$$

$$R(\mathrm{Z}_t) = \begin{pmatrix} \mathbf{\Sigma}^{1/2}_{(\mathrm{Z}_t)} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}$$

$$K(\mathrm{Z}_t) = \begin{bmatrix} I_k, & \mathbf{0}, & \cdots, & \mathbf{0}, & \mathbf{0}, & \mathbf{0}, & \cdots, & \mathbf{0}, \mathbf{0}, & \mathbf{0}, \cdots, & \mathbf{0} \end{bmatrix}$$

$N(\mathrm{Z}_t) \equiv \mathbf{0}$ and $a_t = \mathbf{0}$ with probability one for all t, where $\phi_i^{(\mathrm{Z}_t)} = \phi_i^{(j)}$, $\beta_i^{(\mathrm{Z}_t)} = \beta_i^{(j)}$, $\delta_i^{(\mathrm{Z}_t)} = \delta_i^{(j)}$ and $\mathbf{\Sigma}^{1/2}_{(\mathrm{Z}_t)} = \mathbf{\Sigma}^{1/2}_{(j)}$ if $r_{j-1} < \mathrm{Z}_t \le r_j$. Therefore, the state space model considered here is

$$\alpha_t = L(\mathrm{Z}_t) + H(\mathrm{Z}_t)\alpha_{t-1} + M(\mathrm{Z}_t)\begin{pmatrix} \mathrm{Z}_t \\ \mathrm{X}_t \end{pmatrix} + R(\mathrm{Z}_t)\varepsilon_t \tag{2.1}$$

$$\mathrm{Y}_t = K(\mathrm{Z}_t)\alpha_t + N(\mathrm{Z}_t)\begin{pmatrix} \mathrm{Z}_t \\ \mathrm{X}_t \end{pmatrix} + a_t. \tag{2.2}$$

We assume the following general conditions: (i) The processes $\{\varepsilon_t\}$, $\{a_t\}$ are I.I.D. Gaussian with mean vector $\mathbf{0}$, covariance matrix I_k and mutually independent, (ii) $\{\mathrm{u}_t\}$ is mutually independent of $\{\varepsilon_t\}$, $\{a_t\}$, α_0, (iii) α_0 is independent of $\{\varepsilon_t\}$ and $\{a_t\}$. These conditions give us the opportunity to consider more general models that only the MTAR model.

Based on the state space representation, the aim is to find the following posterior distribution

$$p(\alpha_1, \ldots, \alpha_T, \mathrm{z}_1, \ldots, \mathrm{z}_T, \mathrm{x}_1, \ldots, \mathrm{x}_T | \mathrm{y}_1, \ldots, \mathrm{y}_T),$$

which is re-written

$$p(\alpha_1, \ldots, \alpha_T, \mathrm{u}_1, \ldots, \mathrm{u}_T | \mathrm{y}_1, \ldots, \mathrm{y}_T). \tag{2.3}$$

The Gibbs sampling is the MCMC method chosen to find the posterior distribution (2.3). For this purpose, we need to know and extract samples of the following two full conditional distributions

$$p(\alpha_0, \ldots, \alpha_T | \mathrm{y}_1, \ldots, \mathrm{y}_T, \mathrm{u}_1, \ldots, \mathrm{u}_T) \tag{2.4}$$

$$p(\mathrm{u}_1, \ldots, \mathrm{u}_T | \mathrm{y}_1, \ldots, \mathrm{y}_T, \alpha_0, \ldots, \alpha_T). \tag{2.5}$$

Now, in order to extract samples of the full conditional distribution (2.4), the results in (Carter & Konh, 1994) or (Frhüwirth-Schnatter, 1995) can be used with suitable modifications; namely, in time positions where there are missing data, we set $\mathrm{y}_t = (0, \cdots, 0)'$ and $K(\mathrm{z}_t) = \mathbf{0}$. In case the vector y_t is partially observed, that is, not all components of y_t are missing, we complete it with zeros and we put zeros in the rows of the matrix $K(\mathrm{z}_t)$ corresponding to the missing components of the vector y_t. Both last methodologies use the Kalman filter to get the moments of the marginal distributions which are multivariate normal. Finally, we use again the Gibbs sampling to extract samples of distribution (2.5).

It is possible if we can get samples of the following conditional distributions

$$p(\mathrm{u}_1,\dots,\mathrm{u}_b|\mathrm{u}_{-(1:b)},\mathrm{y}_1,\dots,\mathrm{y}_T,\alpha_0,\dots,\alpha_T) \tag{2.6}$$

and for $t=b+1,\dots,T$,

$$p(\mathrm{u}_t|\mathrm{u}_{-t},\mathrm{y}_1,\dots,\mathrm{y}_T,\alpha_0,\dots,\alpha_T), \tag{2.7}$$

where $\mathrm{y}_{s:t}=\mathrm{y}_s,\dots,\mathrm{y}_t$, $\mathrm{u}_{s:t}=\mathrm{u}_1,\dots,\mathrm{u}_t$, $\mathrm{u}_{-t}=(\mathrm{u}_1,\dots,\mathrm{u}_{t-1},\mathrm{u}_{t+1},\dots,\mathrm{u}_T)$ and $\mathrm{u}_{-(s:t)}=(\mathrm{u}_1,\dots,\mathrm{u}_{s-1},\mathrm{u}_{t+1},\dots,\mathrm{u}_T)$ with $s<t$. This methodology is known as single-move Gibbs sampling and was implemented in (Carlin et al., 1992) to analyse non-normal and non-linear state-space models. In the following proposition, we decompose the full conditional distributions (2.6) and (2.7) to obtain the appropriated expressions for extracting samples from them.

Proposition 1. *Assume that the multivariate process $\{Y_t\}$ obeys a state space model with regime-switching defined in (2.1-2.2) with the conditions (i)-(iii) established above. Then, the full conditional distributions 2.6 and 2.7 are given by*

$$\begin{aligned} &p(u_{1:b}|u_{-(1:b)},y_{1:T},\alpha_{0:T}) \\ &\quad\propto p(u_1,\dots,u_b)\prod_{t=1}^{b}p(\alpha_t|\alpha_{t-1},u_t)\prod_{t=1}^{b}\cancel{p(y_t|\alpha_t,u_t)}\prod_{t=b+1}^{2b}p(u_t|u_{t-1:t-b}) \end{aligned} \tag{2.8}$$

and

$$p(u_t|u_{-t},y_{1:T},\alpha_{0:T})\propto p(\alpha_t|\alpha_{t-1},u_t)\cancel{p(y_t|\alpha_t,u_t)}\prod_{s=t}^{t+b}p(u_s|u_{s-b:s-1}). \tag{2.9}$$

Proof. Note that for the distribution (2.6)

$$p(\mathrm{u}_1,\dots,\mathrm{u}_b|\mathrm{u}_{-(1:b)},\mathrm{y}_1,\dots,\mathrm{y}_T,\alpha_0,\dots,\alpha_T) \tag{1}$$

$$=\frac{p(\mathrm{u}_{1:T},\alpha_{1:T},y_{1:T})}{p(\mathrm{u}_{-(1:b)},\mathrm{y}_1,\dots,\mathrm{y}_T,\alpha_0,\dots,\alpha_T)} \tag{2}$$

$$\propto p(\mathrm{u}_{1:T},\alpha_{0:T},y_{1:T}) \tag{3}$$

$$\begin{aligned} =&p(\mathrm{u}_1,\dots,\mathrm{u}_b)p(\alpha_0|\mathrm{u}_1,\dots,\mathrm{u}_b)p(\alpha_1|\alpha_0,\mathrm{u}_1,\dots,\mathrm{u}_b)p(\mathrm{y}_1|\alpha_1,\alpha_0,\mathrm{u}_1,\dots,\mathrm{u}_b)\cdots \\ &p(\alpha_b|\alpha_{0:b-1},\mathrm{y}_{1:b-1},\mathrm{u}_{1:b})p(\mathrm{y}_b|\alpha_{0:b},\mathrm{y}_{1:b-1},\mathrm{u}_{1:b})p(\mathrm{u}_{b+1}|\alpha_{0:b},\mathrm{y}_{1:b},\mathrm{u}_{1:b}) \\ &p(\alpha_{b+1}|\alpha_{0:b},\mathrm{y}_{1:b},\mathrm{u}_{1:b+1})p(\mathrm{y}_{b+1}|\alpha_{0:b+1},\mathrm{y}_{1:b},\mathrm{u}_{1:b+1})\cdots \\ &p(\mathrm{u}_T|\alpha_{0:T-1},\mathrm{y}_{1:T-1},\mathrm{u}_{1:T-1})p(\alpha_T|\alpha_{0:T-1},\mathrm{y}_{1:T-1},\mathrm{u}_{1:T})p(\mathrm{y}_T|\alpha_{0:T},\mathrm{y}_{1:T-1},\mathrm{u}_{1:T}) \end{aligned} \tag{4}$$

$$\begin{aligned} =&p(\mathrm{u}_1,\dots,\mathrm{u}_b)p(\alpha_0)p(\alpha_1|\alpha_0,\mathrm{u}_1)p(\mathrm{y}_1|\alpha_1,\mathrm{u}_1)\prod_{t=2}^{T}p(\alpha_t|\alpha_{t-1},\mathrm{u}_t)\prod_{t=2}^{T}p(\mathrm{y}_t|\alpha_t,\mathrm{u}_t) \\ &\prod_{t=2}^{T}p(\mathrm{u}_t|\mathrm{u}_{t-1:t-b}) \end{aligned} \tag{5}$$

$$\propto p(\mathrm{u}_1,\dots,\mathrm{u}_b)\prod_{t=1}^{b}p(\alpha_t|\alpha_{t-1},\mathrm{u}_t)\prod_{t=1}^{b}p(\mathrm{y}_t|\alpha_t,\mathrm{u}_t)\prod_{t=b+1}^{2b}p(\mathrm{u}_t|\mathrm{u}_{t-1:t-b}) \tag{6}$$

it is possible to go from the line four to line five because

(a) α_0 is independent of the process $\{\mathrm{u}_t\}$.

(b) Conditional on α_{t-1} and u_t, α_t is independent of other terms by the assumptions (i), (ii) and the equation (2.1).

(c) Conditional on α_t and u_t, y_t is independent of other terms by the assumptions (i), (ii) and the equation (2.2).

(d) Conditional on $\mathrm{u}_{1:t-1}$, u_t is independent of $\mathrm{y}_{1:t-1}$ and $\alpha_{0:t-1}$ by assumption (ii) and equations (2.1), (2.2).

Finally, we can go from line five to six dropping out terms independent of $\mathrm{u}_{1:b}$.

Now, the distribution (2.7) can be decomposed in the following way:

$$\begin{aligned}
&p(\mathrm{u}_t|\mathrm{u}_{-t},\mathrm{y}_1,\ldots,\mathrm{y}_T,\alpha_0,\ldots,\alpha_T) && (1)\\
&\quad=\frac{p(\mathrm{u}_{1:T},\mathrm{y}_{1:T},\alpha_{0:T})}{p(\mathrm{u}_{-t},\mathrm{y}_{1:T},\alpha_{0:T})} && (2)\\
&\quad\propto p(\mathrm{u}_{1:T},\alpha_{0:T},y_{1:T}) && (3)\\
&\quad=p(\alpha_0)p(\mathrm{u}_1|\alpha_0)p(\alpha_1|\alpha_0,\mathrm{u}_1)p(\mathrm{y}_1|\alpha_1,\alpha_0,\mathrm{u}_1)\cdots\\
&\qquad p(\mathrm{u}_2|\mathrm{u}_1,\alpha_1,\alpha_0,\mathrm{y}_1)p(\alpha_2|\mathrm{u}_2,\mathrm{u}_1,\alpha_1,\alpha_0,\mathrm{y}_1)p(\mathrm{y}_2|\alpha_2,\alpha_1,\alpha_0,\mathrm{u}_2,\mathrm{u}_1,\mathrm{y}_1)\cdots\\
&\qquad p(\mathrm{u}_T|\mathrm{u}_{1:T-1},\alpha_{0:T-1},\mathrm{y}_{1:T-1})p(\alpha_T|\alpha_{0:T-1},\mathrm{u}_{1:T},\mathrm{y}_{1:T-1})p(\mathrm{y}_T|\mathrm{y}_{1:T-1},\alpha_{0:T},\mathrm{u}_{1:T}) && (4)\\
&\quad=p(\alpha_0)p(\mathrm{u}_1)p(\alpha_1|\alpha_0,\mathrm{u}_1)\prod_{t=2}^{T}p(\mathrm{u}_t|\mathrm{u}_{1:t-1},\alpha_{0:t-1},\mathrm{y}_{1:t-1})\prod_{t=2}^{T}p(\alpha_t|\alpha_{0:t-1},\mathrm{u}_{1:t},\mathrm{y}_{1:t-1})\\
&\qquad\prod_{t=1}^{T}p(\mathrm{y}_t|\mathrm{y}_{1:t-1},\alpha_{0:t},\mathrm{u}_{1:t}) && (5)\\
&\quad=p(\alpha_0)p(\mathrm{u}_1)p(\alpha_1|\alpha_0,\mathrm{u}_1)\prod_{t=2}^{T}p(\mathrm{u}_t|\mathrm{u}_{1:t})\prod_{t=2}^{T}p(\alpha_t|\alpha_{t-1},\mathrm{u}_t)\prod_{t=1}^{T}p(\mathrm{y}_t|\alpha_t,\mathrm{u}_t) && (6)\\
&\quad\propto p(\alpha_t|\alpha_{t-1},\mathrm{u}_t)p(\mathrm{y}_t|\alpha_t,\mathrm{u}_t)\prod_{s=t}^{t+b}p(\mathrm{u}_s|\mathrm{u}_{s-b:s-1}). && (7)
\end{aligned}$$

Using (a) and (b), we can go from line five to six and from six to seven if we drop out terms independent of u_t. Then, the decomposed full conditional distributions are as follows

$$\begin{aligned}
&p(\mathrm{u}_1,\ldots,\mathrm{u}_b|\mathrm{u}_{-(1:b)},\mathrm{y}_1,\ldots,\mathrm{y}_T,\alpha_0,\ldots,\alpha_T)\\
&\qquad\propto p(\mathrm{u}_1,\ldots,\mathrm{u}_b)\prod_{t=1}^{b}p(\alpha_t|\alpha_{t-1},\mathrm{u}_t)\prod_{t=1}^{b}p(\mathrm{y}_t|\alpha_t,\mathrm{u}_t)\prod_{t=b+1}^{2b}p(\mathrm{u}_t|\mathrm{u}_{t-1:t-b})
\end{aligned}$$

and

$$p(\mathrm{u}_t|\mathrm{u}_{-t},\mathrm{y}_1,\ldots,\mathrm{y}_T,\alpha_0,\ldots,\alpha_T)\propto p(\alpha_t|\alpha_{t-1},\mathrm{u}_t)p(\mathrm{y}_t|\alpha_t,\mathrm{u}_t)\prod_{s=t}^{t+b}p(\mathrm{u}_s|\mathrm{u}_{s-b:s-1})$$

□

It is important to point out that the distributions Y_t conditional on α_t, u_t, in (2.8) and (2.9) are degenerate for MTAR models, that is $p(\mathrm{y}_t|\alpha_t, \mathrm{u}_t) = 1$, consequently we can drop out these terms. This fact was mentioned by Nieto (2005) for 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, we can use either the accept-reject algorithm or the Metropolis-Hastings algorithm in order to extract samples of these distributions. If we choose the acept-reject algorithm, the initial distribution or the kernel of the Markov chain $\{\mathrm{u}_t\}$ can be selected as proposal distribution. Now, if we use Metropolis-Hastings algorithm, the random walk version is a solution to extract samples, based on a multivariate normal distribution centred in the origin with a covariance matrix cI, where c is a positive constant, and I is the identity matrix. If there are not missing data in a time t, we do not have to extract samples here, and we skip the procedure at this time. ~~In case we have only some missing components, we get marginal distributions for them, keeping fixed the observed components, then based on these distributions, we can extract samples of the missing components.~~ We have to point out that for MTAR models, the density $p(\alpha_t|\alpha_{t-1}, \mathrm{u}_t)$ is a singular multivariate distribution which collapses to a $k-$variate normal distribution, due to the fact that many components in the state vector are only for completeness, then

$$p(\alpha_t|\alpha_{t-1}, \mathrm{u}_t) = \frac{1}{(2\pi)^{k/2}} |\Sigma_{(Z_t)}|^{-1/2} \exp(-\tfrac{1}{2} e_t' e_t),$$

where

$$e_t = \boldsymbol{\Sigma}_{(\mathrm{z}_t)}^{-1/2} \left(\mathrm{y}_t - \boldsymbol{\phi}_0^{(\mathrm{z}_t)} - \sum_{i=1}^{p_j} \boldsymbol{\phi}_i^{(\mathrm{z}_t)} \mathrm{y}_{t-i} - \sum_{i=1}^{q_j} \boldsymbol{\beta}_i^{(\mathrm{z}_t)} \mathrm{x}_{t-i} - \sum_{i=1}^{d_j} \boldsymbol{\delta}_i^{(\mathrm{z}_t)} \mathrm{z}_{t-i} \right).$$

With the knowledge of the full conditional distributions, we can implement the Gibbs sampling to extract a sample of size G, $(\alpha_t^{(1)}, \mathrm{u}_t^{(1)}), \ldots, (\alpha_t^{(G)}, \mathrm{u}_t^{(G)})$ for some $t = 1, \ldots, T$; with this sample, we can obtain posterior means, variances and credible intervals, which give us estimations of the missing data because the first k components of the state vector are the components of the output vector.

2.2 Estimation of non-structural parameters

In the last section, we showed how to estimate the missing data in the output vector, covariates and threshold variable conditional on the structural and non-structural parameters, although in practice these parameters are unknown and must be estimated. The aim of this section is to estimate jointly non-structural parameters and missing data conditional on structural parameters through Gibbs sampling. To this end, we have to extract samples of the full conditional distributions $p(\theta_{\mathrm{yns}}|\alpha_1, \ldots, \alpha_T, \mathrm{y}_{1:T}, \mathrm{u}_{1:T}; \theta_{\mathrm{ys}})$ and $p(\alpha_1, \ldots, \alpha_T, \mathrm{u}_1, \ldots, \mathrm{u}_T|\mathrm{y}_{1:T}, \theta_{\mathrm{y}})$ in order to implement the Gibbs sampling. We focus in the full conditional distribution $p(\theta_{\mathrm{yns}}|\alpha_1, \ldots, \alpha_T, \mathrm{y}_{1:T}, \mathrm{u}_{1:T}; \theta_{\mathrm{ys}})$ because the other full conditional distribution was analysed in the last section.

In order to extract samples of the distribution $p(\theta_{\mathrm{yns}}|\alpha_1, \ldots, \alpha_T, \mathrm{y}_{1:T}, \mathrm{u}_{1:T}; \theta_{\mathrm{ys}})$, we need to find the following full conditional distributions, for $j = 1, \ldots, l$

$$p(\theta_j|\theta_i, i \neq j, \Sigma, \alpha_{0:T}, \mathrm{u}_{1:T}, \mathrm{y}_{1:T}),$$

$$p(\Sigma_{(j)}|\Sigma_{(i)}, i \neq j, \theta_1, \ldots, \theta_l, \alpha_{0:T}, \mathrm{u}_{1:T}, \mathrm{y}_{1:T}),$$

to use the Gibbs sampling. Hereinafter the vector θ_{ys} is not shown in the distributions, but all are conditional on it. We assume that parameters among regimes are independent; θ_j and $\Sigma_{(j)}$ are also independent. Furthermore, the prior distribution for θ_j is a multinormal distribution with mean θ_{0j} and covariance matrix Σ_{0j}. Prior Normal distribution focuses the knowledge of the coefficients around the mean θ_{0j} with uncertainty quantified in the covariance matrix Σ_{0j}. The prior distribution for $\Sigma_{(j)}$ is an inverse Wishart with covariance matrix S_{0j} and ν_{0j} degrees of freedom. These assumptions were accepted in (Nieto, 2005) following the ideas of Chen & Lee (1995). The following two propositions give us the full conditional distributions that are necessary to estimate the non-structural parameters.

Proposition 2. *For each $j = 1, \ldots, l$ let the matrices*

$$\boldsymbol{W}_j = (w_{t_{1,j}}, \ldots, w_{t_{N_j},j})_{\eta_j \times N_j},$$

with the vectors

$$w_{t,j} = (1, y'_{t-1}, \ldots, y'_{t-p_j}, x'_{t-1}, \ldots, x'_{t-q_j}, z_{t-1}, \ldots, z_{t-d_j})'_{\eta_j \times 1},$$

$$Y_j = (y_{t_{1,j}}, \ldots, y_{t_{N_j},j})_{k \times N_j},$$

and $y_j = \mathrm{vec}(Y_j)$, for $\mathrm{vec}(Y_j - A_j\mathbf{W}_j) = y_j - (\mathbf{W}'_j \otimes I_k)\theta_j$, where $t_{1,j}, \ldots, t_{N_j,j}$ are the points where $r_{j-1} < Z_t \leq r_j$ and N_j is the quantity of these points, called the number of observations in regime j. With the assumptions and the prior distribution specified above, the full conditional distribution $\theta_j|\theta_i, i \neq j, \Sigma, \alpha_{0:T}, u_{1:T}, y_{1:T}$ is multinormal with mean

$$\theta_j^* = V_j([\boldsymbol{W}_j \otimes \Sigma_{(j)}^{-1}]y_j + \Sigma_{0j}^{-1}\theta_{0j}),$$

and covariance matrix

$$V_j = [\,\boldsymbol{W}_j\,\boldsymbol{W}'_j \otimes \Sigma_{(j)}^{-1} + \Sigma_{0j}^{-1}]^{-1}.$$

Proof. Note that

$$\begin{aligned}
p(\theta_j|\theta_i,&i \neq j, \Sigma, \alpha_{0:T}, \mathrm{u}_{1:T}, \mathrm{y}_{1:T}) \\
&\propto p(\mathrm{u}_{1:T}, \alpha_{0:T}, \mathrm{y}_{1:T}|\theta_{\mathrm{yns}})p(\theta_j|\theta_i, i \neq j, \Sigma, \theta_u) \\
&= p(\mathrm{u}_{1:T}, \alpha_{0:T}, \mathrm{y}_{1:T}|\theta_{\mathrm{yns}})p(\theta_j) \quad \text{by independence of the parameters} \\
&= p(\mathrm{u}_{1:T}|\theta_{\mathrm{yns}})p(\alpha_{0:T}|\mathrm{u}_{1:T}, \theta_{\mathrm{yns}})p(\mathrm{y}_{1:T}|\alpha_{0:T}, \mathrm{u}_{1:T}, \theta_{\mathrm{yns}})p(\theta_j) \\
&\propto p(\alpha_{0:T}|\mathrm{u}_{1:T}, \theta_{\mathrm{yns}})p(\theta_j)
\end{aligned}$$

last line is due to the independence of Markov chain between $\{u_t\}$ and θ_y; and $p(\mathrm{y}_{1:T}|\alpha_{0:T}, \mathrm{u}_{1:T}, \theta_{\mathrm{yns}})$ is a product of degenerate densities for MTAR models that do not depend on θ_{yns}. Hence for $j = 1, \ldots, l$

$$p(\theta_j|\theta_i, i \neq j, \Sigma, \alpha_{0:T}, \mathrm{u}_{1:T}, \mathrm{y}_{1:T}) \propto p(\alpha_{0:T}|\mathrm{u}_{1:T}, \theta_{\mathrm{yns}})p(\theta_j),$$

where

$$p(\alpha_{0:T}|\mathrm{u}_{1:T},\theta_{\mathrm{yns}}) = p(\alpha_0)p(\alpha_1|\alpha_0,\mathrm{u}_{1:T},\theta_{\mathrm{yns}})\cdots p(\alpha_T|\alpha_{T-1},\mathrm{u}_{1:T},\theta_{\mathrm{yns}})$$

and every density $p(\alpha_t|\alpha_{t-1},\mathrm{u}_{1:T},\theta_{\mathrm{yns}})$ was found in the last section, therefore

$$\begin{aligned}
p(\alpha_t|\alpha_{t-1},&\mathrm{u}_{1:T},\theta_{\mathrm{yns}}) \propto \\
&|\Sigma_{(j)}|^{-1/2}\exp\left[-\frac{1}{2}\left(\mathrm{y}_t-\boldsymbol{\phi}_0^{(j)}-\sum_{i=1}^{p_j}\boldsymbol{\phi}_i^{(j)}\mathrm{y}_{t-i}+\sum_{i=1}^{q_j}\boldsymbol{\beta}_i^{(j)}\mathrm{x}_{t-i}+\sum_{i=1}^{d_j}\boldsymbol{\delta}_i^{(j)}\mathrm{z}_{t-i}\right)'\right.\\
&\left.\Sigma_{(j)}^{-1}\left(\mathrm{y}_t-\boldsymbol{\phi}_0^{(j)}-\sum_{i=1}^{p_j}\boldsymbol{\phi}_i^{(j)}\mathrm{y}_{t-i}+\sum_{i=1}^{q_j}\boldsymbol{\beta}_i^{(j)}\mathrm{x}_{t-i}+\sum_{i=1}^{d_j}\boldsymbol{\delta}_i^{(j)}\mathrm{z}_{t-i}\right)\right].
\end{aligned}$$

Now, for the independence of the parameters among regimes and the independence between θ_j and $\Sigma_{(j)}$ we have

$$\begin{aligned}
p(\alpha_{0:T}&|\mathrm{u}_{1:T},\theta_{\mathrm{yns}}) \propto \\
&\exp\left[-\frac{1}{2}\sum_{\{t:j_t=j\}}\left(\mathrm{y}_t-\boldsymbol{\phi}_0^{(j)}-\sum_{i=1}^{p_j}\boldsymbol{\phi}_i^{(j)}\mathrm{y}_{t-i}+\sum_{i=1}^{q_j}\boldsymbol{\beta}_i^{(j)}\mathrm{x}_{t-i}+\sum_{i=1}^{d_j}\boldsymbol{\delta}_i^{(j)}\mathrm{z}_{t-i}\right)'\right.\\
&\left.\Sigma_{(j)}^{-1}\left(\mathrm{y}_t-\boldsymbol{\phi}_0^{(j)}-\sum_{i=1}^{p_j}\boldsymbol{\phi}_i^{(j)}\mathrm{y}_{t-i}+\sum_{i=1}^{q_j}\boldsymbol{\beta}_i^{(j)}\mathrm{x}_{t-i}+\sum_{i=1}^{d_j}\boldsymbol{\delta}_i^{(j)}\mathrm{z}_{t-i}\right)\right]
\end{aligned}$$

where $\{t : j_t = j\} = \{t_{1,j},\ldots,t_{N_j,j}\}$, then

$$p(\alpha_{0:T}|\mathrm{u}_{1:T},\theta_{\mathrm{yns}}) \propto \exp\left\{-\frac{1}{2}[y_j-(\mathbf{W}_j'\otimes I_k)\theta_j]'(I_{N_j}\otimes\Sigma_{(j)}^{-1})[y_j-(\mathbf{W}_j'\otimes I_k)\theta_j]\right\}. \quad (2.10)$$

Based on the prior distribution for θ_j, we have that for $j = 1,\ldots,l$ the full conditional distributions are as follows

$$\begin{aligned}
p(\theta_j|\theta_i,&i\neq j,\Sigma,\alpha_{0:T},\mathrm{u}_{1:T},\mathrm{y}_{1:T},\theta_{\mathrm{yns}})\\
&\propto \exp\left\{-\frac{1}{2}[y_j-(\mathbf{W}_j'\otimes I_k)\theta_j]'(I_{N_j}\otimes\Sigma_{(j)}^{-1})[y_j-(\mathbf{W}_j'\otimes I_k)\theta_j]\right\}\\
&\quad\exp\left\{-\frac{1}{2}(\theta_j-\theta_{0j})'(\Sigma_{0j}^{-1})(\theta_j-\theta_{0j})\right\}\\
&\propto \exp\left\{(\theta_j-\theta_j^*)'V_j^{-1}(\theta_j-\theta_j^*)\right\} \quad (2.11)
\end{aligned}$$

which is the kernel of a multivariate normal distribution with covariance matrix

$$V_j = [\mathbf{W}_j \mathbf{W}_j' \otimes \Sigma_{(j)}^{-1} + \Sigma_{0j}^{-1}]^{-1}$$

and mean

$$\theta_j^* = V_j([\mathbf{W}_j \otimes \Sigma_{(j)}^{-1}] y_j + \Sigma_{0j}^{-1} \theta_{0j}).$$

□

Proposition 3. *With the same assumptions of the proposition 2, and the prior distribution for* $\Sigma_{(j)}$ *specified above, the full conditional distribution of* $\Sigma_{(j)} | \Sigma_{(i)}, i \neq j, \theta_1, \ldots, \theta_l, \alpha_{0:T}, u_{1:T}, y_{1:T}$ *for* $j = 1, \cdots, l$ *is an inverse-Wishart with covariance matrix* $(S_j + S_{0j})^{-1}$ *and* $N_j + \nu_j + 1$ *degrees of freedom, where*

$$S_j = \sum_{\{t: j_t = j\}} \left(y_t - \boldsymbol{\phi}_0^{(j)} - \sum_{i=1}^{p_j} \boldsymbol{\phi}_i^{(j)} y_{t-i} + \sum_{i=1}^{q_j} \boldsymbol{\beta}_i^{(j)} x_{t-i} + \sum_{i=1}^{d_j} \boldsymbol{\delta}_i^{(j)} z_{t-i} \right)' \left(y_t - \boldsymbol{\phi}_0^{(j)} - \sum_{i=1}^{p_j} \boldsymbol{\phi}_i^{(j)} y_{t-i} + \sum_{i=1}^{q_j} \boldsymbol{\beta}_i^{(j)} x_{t-i} + \sum_{i=1}^{d_j} \boldsymbol{\delta}_i^{(j)} z_{t-i} \right).$$

Proof. With similar explanations in the proof of the proposition 2 we have

$$p(\Sigma_{(j)} | \Sigma_{(i)}, i \neq j, \theta_1, \ldots, \theta_l, \alpha_{0:T}, \mathrm{u}_{1:T}, \mathrm{y}_{1:T}, \theta_{\mathrm{yns}}) \propto p(\alpha_{0:T} | \mathrm{u}_{1:T}, \theta_{\mathrm{yns}}) p(\Sigma_j).$$

Now, we can write

$$\begin{aligned} p(\alpha_{0:T} | \mathrm{u}_{1:T}, \theta_{\mathrm{yns}}) &\propto |\Sigma_{(j)}|^{-N_j/2} \\ &\times \exp\left[-\frac{1}{2} \operatorname{tr} \left\{ \Sigma_{(j)}^{-1} \sum_{\{t: j_t = j\}} \left(\mathrm{y}_t - \boldsymbol{\phi}_0^{(j)} - \sum_{i=1}^{p_j} \boldsymbol{\phi}_i^{(j)} \mathrm{y}_{t-i} + \sum_{i=1}^{q_j} \boldsymbol{\beta}_i^{(j)} \mathrm{x}_{t-i} + \sum_{i=1}^{d_j} \boldsymbol{\delta}_i^{(j)} \mathrm{z}_{t-i} \right)' \right.\right. \\ &\qquad \left.\left. \left(\mathrm{y}_t - \boldsymbol{\phi}_0^{(j)} - \sum_{i=1}^{p_j} \boldsymbol{\phi}_i^{(j)} \mathrm{y}_{t-i} + \sum_{i=1}^{q_j} \boldsymbol{\beta}_i^{(j)} \mathrm{x}_{t-i} + \sum_{i=1}^{d_j} \boldsymbol{\delta}_i^{(j)} \mathrm{z}_{t-i} \right) \right\} \right] \\ &= |\Sigma_{(j)}|^{-N_j/2} \exp\left[-\frac{1}{2} \operatorname{tr} \left\{ \Sigma_{(j)}^{-1} S_j \right\} \right] \end{aligned}$$

where

$$S_j = \sum_{\{t: j_t = j\}} \left(\mathrm{y}_t - \boldsymbol{\phi}_0^{(j)} - \sum_{i=1}^{p_j} \boldsymbol{\phi}_i^{(j)} \mathrm{y}_{t-i} + \sum_{i=1}^{q_j} \boldsymbol{\beta}_i^{(j)} \mathrm{x}_{t-i} + \sum_{i=1}^{d_j} \boldsymbol{\delta}_i^{(j)} \mathrm{z}_{t-i} \right)' \left(\mathrm{y}_t - \boldsymbol{\phi}_0^{(j)} - \sum_{i=1}^{p_j} \boldsymbol{\phi}_i^{(j)} \mathrm{y}_{t-i} + \sum_{i=1}^{q_j} \boldsymbol{\beta}_i^{(j)} \mathrm{x}_{t-i} + \sum_{i=1}^{d_j} \boldsymbol{\delta}_i^{(j)} \mathrm{z}_{t-i} \right),$$

then for $j = 1, \ldots, l$

$$
\begin{aligned}
&p(\Sigma_{(j)}|\Sigma_{(i)}, i \neq j, \theta_1, \ldots, \theta_l, \alpha_{0:T}, \mathrm{u}_{1:T}, \mathrm{y}_{1:T}, \theta_{\mathrm{yns}}) \\
&\qquad \propto |\Sigma_{(j)}|^{-N_j/2} \exp\left[-\frac{1}{2}\operatorname{tr}\left\{\Sigma_{(j)}^{-1} S_j\right\}\right] |\Sigma_{(j)}|^{-(\nu_{0j}+k+1)/2} \exp\left[-\frac{1}{2}\operatorname{tr}\left\{\Sigma_{(j)}^{-1} S_{0j}\right\}\right] \\
&\qquad = |\Sigma_{(j)}|^{-(N_j+\nu_{0j}+k+1)/2} \exp\left[-\frac{1}{2}\operatorname{tr}\left\{\Sigma_{(j)}^{-1}(S_j + S_{0j})\right\}\right] \qquad (2.12)
\end{aligned}
$$

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

Based on (2.11) and (2.12), we can implement the Gibbs sampling in a straightforward form because the full conditional distributions are standard distributions and it is easy to extract samples of them. In order to estimate both missing data and non-structural parameters, we can use results of current and previous sections together. Specifically, the procedure to estimate the missing data and the non-structural parameters is as follows.

Step 1. Complete the y− and u−time series with reasonable initial values for the missing data.

Step 2. With the completed series, generating random draws from the missing data posterior distributions.

Step 3. Complete again the y− and u−time series with draw values obtained in the Step 2.

Step 4. Generating random draws from the non-structural parameters posterior distributions.

Step 5. With those samples, compute the means and the credible intervals in order to obtain the estimates of the missing data and non-structural parameters.

That implemented procedure can be considered as a multivariate extension of the procedure proposed in (Nieto, 2005).

CHAPTER 3

Estimation of structural parameters of the MTAR model

Based on the model (1.1), we remember that the structural parameters of the MTAR model are: autoregressive orders $p_1, \ldots, p_l, q_1, \ldots, q_l, d_1, \ldots, d_l$, the thresholds $\mathbf{r} = (r_1, \ldots, r_l)'$ and the number of regimes l. We can see an intrinsic nesting of the structural parameters in the sense that the number of thresholds and the number of autoregressive orders depend on the number of regimes, moreover, the dimension of the parameters vector changes when either the number of regimes changes or the autoregressive orders change; this leads us to the Bayesian problem of changing dimension in choice model. Therefore, we have to keep in mind this issue on the estimation of structural parameters in MTAR model. Some advances in the estimation or identification of structural parameters in threshold models have been done, for instance, Campbell (2004) used reversible jump Markov Chain Monte Carlo (RJMCMC) method proposed by Green (1995) to estimate the autoregressive orders in self-exciting autoregressive threshold (SETAR) models, also estimation of the threshold values was proposed in this methodology, however the number of regimes must be known. Another approach based on the idea of changing dimension and the results of Carlin & Chib (1995) was developed by Nieto (2005) for threshold models, in this case, the methodology estimates the autoregressive orders and the number of regimes in two steps, the aim in the first step is the estimation of the number of regimes. In the second step and conditional on the number of the estimated regimes in the first step, the autoregressive orders are estimated. Threshold were estimated using minimization of NAIC suggested in (Tong, 1990) as a preliminary step.

Nieto et al. (2013) proposed a similar approach to the previous one but using the RJMCMC methodology to identify the autoregressive orders and the number of regimes. A different approach to the changing dimension was employed by So & Chen (2003) for univariate SETAR models; they used stochastic search idea to identify autoregressive orders in each regime. The same idea was worked by Chen et al. (2011) for threshold autoregressive moving-average models. In two last works, the thresholds are estimated jointly with the autoregressive orders and the other parameters, by using the Bayesian approach and assuming that the number of regimes is known. A general Bayesian methodology in multivariate threshold autoregressive with heteroscedastic errors is proposed in (Wu & Lee, 2011); the structural and non-structural parameters are jointly estimated and the model choice is done through marginal likelihood. Nevertheless, the equation of the model does not include exogenous variables. Kwon et al. (2009) proposed a procedure to identify the autoregressive orders of the MTAR model using information criteria and Bayes factors, however the number of regimes have to be known. In the following sections, we give some methodologies to identify the structural parameters of MTAR model. In the first part conditional on the number of regimes, a strategy for identifying jointly the autoregressive orders and the threshold values is carried out. In the second part, two conceptually different methodologies are given to estimate the number of regimes.

3.1 Identification of the autoregressive orders and threshold values

We use stochastic search ideas in order to identify autoregressive orders in MTAR models because it permits us that the estimation is done in only one step. To this end, we write the MTAR model (1.1) in the following way

$$\mathrm{Y}_t = (I_k \otimes \mathrm{w}'_{t,j})\theta_j + \Sigma_{(j)}^{1/2}\varepsilon_t \quad \text{if} \quad r_{j-1} < \mathrm{Z}_t \leq r_j, \tag{3.1}$$

where $\theta_j = \mathrm{vec}(A'_j)$ for $j = 1, \ldots, l$. We add 0-1 indicator variables $\gamma_{i,j}$, $i = 1, \ldots, \eta_j$, $j = 1, \ldots, l$; 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. Now, let $\boldsymbol{\gamma}_j = (\gamma_{1,j}, \ldots, \gamma_{\eta_j,j})'$ be the full vector of indicators in the regime j, based on that, we can re-write the MTAR model with the vectors γ_j, $j = 1, \ldots, l$ as follows,

$$\begin{aligned} \mathrm{Y}_t &= (I_k \otimes \mathrm{w}'_{t,j})\Gamma_j\theta_j + \Sigma_{(j)}^{1/2}\varepsilon_t & (3.2) \\ &= (I_k \otimes \mathrm{w}'_{t,j})\vartheta_j + \Sigma_{(j)}^{1/2}\varepsilon_t & (3.3) \end{aligned}$$

if $r_{j-1} < \mathrm{Z}_t \leq r_j$, where $\vartheta_j = (\gamma_{1,j}\theta_{1,j}, \cdots, \gamma_{\eta_j,j}\theta_{\eta_j,j})$ and $\Gamma_j = \mathrm{Diag}\{\gamma_j\}$. The vectors $\boldsymbol{\gamma}_j$, $j = 1, \ldots, l$ with high posterior probability give us information about which lags of the variables should be included, then stochastic search helps us to identify autoregressive orders in indirect way and now the vector of structural parameters is $\theta_{\mathrm{ys}} = (\boldsymbol{\gamma}'_1, \ldots, \boldsymbol{\gamma}'_l, \mathbf{r}, l)$.

Two methods for stochastic search are selected in this work to identify autoregressive order in MTAR models. The first method, called Kuo and Mallick (Kuo), was introduced in (Kuo & Mallick, 1998) for variable selection in regression models. The second one was proposed in (George & McCulloch, 1993) and it is called Stochastic Search Variable Selection (SSVS). Other approaches of Bayesian variable selection can be found in the practical review (O'Hara & Sillanpaa, 2009). In the following two sections we focus on the full conditional distributions of the parameters $\boldsymbol{\gamma}_j$ for $j = 1, \ldots, l$ and $\mathbf{r}$ so as to implement the Gibbs sampling. The full conditional distributions of the non-structural parameters are calculated again to show the dependence on the γ's vectors, and because in this section θ_j $j = 1, \ldots, l$ was defined a slight different to the Chapter 1.

Firstly, we give the marginal likelihood function of a MTAR model given by (3.2)-(3.3) of the decomposition 1.2

$$L(\theta_{\mathrm{y}}|\mathbf{y}, \mathbf{u}) = f(\mathbf{y}|\mathbf{u}, \theta_{\mathrm{y}}).$$

Additionally, we assume that the p first values of $\mathbf{y}$, $\mathbf{y}_p = (y_1, \ldots, y_p)$ are fixed and known, therefore

$$f(\mathbf{y}|\mathbf{u},\theta_{\mathrm{y}}) = f(y_{p+1}|\mathbf{u},\mathbf{y}_p,\theta_{\mathrm{y}})\cdots f(y_T|y_1,\ldots,y_{T-1},\mathbf{u},\theta_{\mathrm{y}}).$$

Now, we have for $t = p+1,\ldots,T$

$$f(y_t|y_1,\ldots,y_{t-1},\mathbf{u},\theta_{\mathrm{y}}) = (2\pi)^{-k/2}|\Sigma_{(j_t)}|^{-1/2}\exp\left\{-\frac{1}{2}e_t'e_t\right\},$$

where

$$e_t = \Sigma_{(j_t)}^{-1/2}\left(\mathrm{y}_t - \boldsymbol{\phi}_0^{(j_t)} - \sum_{i=1}^{p_{j_t}}\boldsymbol{\phi}_i^{(j_t)}\mathrm{y}_{t-i} + \sum_{i=1}^{q_{j_t}}\boldsymbol{\beta}_i^{(j_t)}\mathrm{x}_{t-i} + \sum_{i=1}^{d_{j_t}}\boldsymbol{\delta}_i^{(j_t)}\mathrm{z}_{t-i}\right),$$

then

$$\begin{aligned}
f(\mathbf{y}|\mathbf{u},\theta_y) &= (2\pi)^{-(T-p)/2}\prod_{t=p+1}^{T}|\Sigma_{(j_t)}|^{-1/2}\exp\left\{-\frac{1}{2}\sum_{t=p+1}^{T}e_t'e_t\right\}\\
&= (2\pi)^{-(T-p)/2}\prod_{i=1}^{l}|\Sigma_{(j_t)}|^{-N_i/2}\times\\
&\quad\exp\left\{-\frac{1}{2}\sum_{t=p+1}^{T}\Big[y_t-(I_k\otimes\mathrm{w}_{t,j_t}')\vartheta_{j_t}\Big]'\Sigma_{(i)}^{-1}\Big[y_t-(I_k\otimes\mathrm{w}_{t,j_t}')\vartheta_{j_t}\Big]\right\}\\
&= (2\pi)^{-(T-p)/2}\prod_{i=1}^{l}|\Sigma_{(i)}|^{-N_i/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{l}\Big[y_i-\mathbf{X}_i\vartheta_i\Big]'(I_{N_i}\otimes\Sigma_{(i)}^{-1})\Big[y_i-\mathbf{X}_i\vartheta_i\Big]\right\}
\end{aligned} \tag{3.4}$$

where

$$\mathbf{X}_i = \begin{pmatrix} I_k\otimes\mathrm{w}_{t_1,i}' \\ I_k\otimes\mathrm{w}_{t_2,i}' \\ \vdots \\ I_k\otimes\mathrm{w}_{t_{N_i},i}' \end{pmatrix}.$$

It is important to point out that N_j, $j = 1,\ldots,l$ depend on thresholds $\mathbf{r}$, then the likelihood depends on thresholds through the number of observations in each regime.

3.1.1 Posterior distributions in the Kuo method

In order to implement Kuo and Mallick methodology, we consider the following prior assumptions: the variables $\gamma_{i,j}$ are mutually independent with Bernoulli distribution and probability $p_{i,j} = P[\gamma_{i,j} = 1]$ for $i = 1,\cdots,\eta_j$ with $j = 1,\cdots,l$; θ_j is independent of $\boldsymbol{\gamma}_j$ for $j = 1,\ldots,l$ and $\mathbf{r}$; the threshold vector $\mathbf{r}$ is independent of θ_j, $\Sigma_{(j)}$ and $\boldsymbol{\gamma}_j$ for $j = 1,\ldots,l$ and it follows a uniform distribution over $S = \big\{(r_1,\ldots,r_{l-1})\in\mathbb{R}^{l-1} : z_{(a)} < r_i < z_{(b)}, i = 1,\ldots,l-1; r_1 < r_2 < \cdots < r_{l-1}\big\}$, and its volume is calculated as $V(S) = \int_{z_{(a)}}^{z_{(b)}}\int_{r_1}^{z_{(b)}}\cdots\int_{r_{l-2}}^{z_{(b)}} dr_{l-1}\cdots dr_2 dr_1$, where $z_{(a)}$ and $z_{(b)}$ are the

a-th and b-th percentile of the sample $z_1, \dots, z_T$. Then

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

where $1_S(\mathbf{r})$ is the indicator function of S on $\mathbb{R}^{l-1}$. In the case of one threshold, we consider that $r \sim U(z_{(a)}, z_{(b)})$.

The assumptions in Section 2.2 are also considered to get the full conditional distributions for the structural and non-structural parameters. The full conditional distributions $\theta_j|\theta_{y-\theta_j}, \mathbf{y}, \mathbf{u}$ and $\Sigma_{(j)}|\theta_{y-\Sigma_j}, \mathbf{y}, \mathbf{u}$ for $j = 1, \dots, l$ have to be computed again because the non-observable variables $\gamma_{i,j}$ with $i = 1, \dots, \eta_j$, $j = 1, \dots, l$ were added to the MTAR model. Hence for $j = 1, \dots, l$ we have

$$\begin{aligned}
p(\theta_j|\theta_{y-\theta_j}, \mathbf{y}, \mathbf{u}) &\propto p(\mathbf{y}|\theta_y, \mathrm{u}) p(\theta_j) \\
&\propto \exp\left\{-\frac{1}{2}\Big[\Big(y_j - \mathbf{X}_j\vartheta_j\Big)'\big(I_{N_j} \otimes \Sigma_{(j)}^{-1}\big)\Big(y_j - \mathbf{X}_j\vartheta_j\Big) + (\theta_j - \theta_{0j})'\Sigma_{0j}^{-1}(\theta_j - \theta_{0j})\Big]\right\} \\
&= \exp\Bigg\{ -\frac{1}{2}\Big[\Big(y_j - \mathbf{X}_j\Gamma_j\theta_j\Big)'\big(I_{N_j} \otimes \Sigma_{(j)}^{-1}\big)\Big(y_j - \mathbf{X}_j\Gamma_j\theta_j\Big) + \\
&\qquad (\theta_j - \theta_{0j})'\Sigma_{0j}^{-1}(\theta_j - \theta_{0j})\Big]\Bigg\} \\
&= \exp\left\{-\frac{1}{2}\Big(\theta_j - \theta_j^*\Big)' V_j^{*-1}\Big(\theta_j - \theta_j^*\Big)\right\}
\end{aligned}$$

where $V_j^* = \Big[\Gamma_j'\mathbf{X}_j'\big(I_{N_j} \otimes \Sigma_{(j)}^{-1}\big)\mathbf{X}_j\Gamma_j + \Sigma_{0j}^{-1}\Big]^{-1}$, $\theta_j^* = V_j^*(\Gamma_j'\mathbf{X}_j'\big(I_{N_j} \otimes \Sigma_{(j)}^{-1}\big)y_j + \Sigma_{0j}^{-1}\theta_{0j})$, $\theta_{\mathrm{y}-\theta_j}$ is the parameters vector θ_{y} without the sub-vector θ_j. This is the kernel of a normal multivariate distribution with mean θ_j^* and covariance matrix V_j^*.

Now, for $j = 1, \dots, l$

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

where

$$S_j = \sum_{\{t:j_t=j\}} \big(\mathrm{y}_t - (I_k \otimes \mathrm{w}_{t,j_t}')\vartheta_{j_t}\big)\big(\mathrm{y}_t - (I_k \otimes \mathrm{w}_{t,j_t}')\vartheta_{j_t}\big)'. \tag{3.5}$$

This is the kernel of an inverse-Wishart with covariance matrix $(S_j + S_{0j})^{-1}$ and degrees of freedom $N_j + \nu_{0j}$. The full conditional distributions described above have similar expressions to the found in 2.11 and 2.12 with the difference that in the matrix Γ_j appears the influence of the vector γ_j. The following two propositions give us the full conditional distributions for the threshold values and the vector γ_j for $j = 1, \dots, l$.

Proposition 4. *With the assumptions on the threshold vector $\boldsymbol{r}$ at the beginning of the section, the full conditional distribution $\boldsymbol{r}|\theta_{y-\boldsymbol{r}}, \boldsymbol{y}, \boldsymbol{u}$ is proportional to*

$$\frac{1}{V(\boldsymbol{S})} \mathit{1}_{\boldsymbol{s}}(\boldsymbol{r}) \left\{\prod_{i=1}^{l} |\Sigma_{(i)}|^{-N_i/2}\right\} \exp\left\{-\frac{1}{2}\sum_{i=1}^{l}\Big[y_i - \boldsymbol{X}_i\vartheta_i\Big]'\big(I_{N_i} \otimes \Sigma_{(i)}^{-1}\big)\Big[y_i - \boldsymbol{X}_i\vartheta_i\Big]\right\}. \tag{3.6}$$

Proof. Note that

$$\begin{aligned}
&p(\mathbf{r}|\theta_{y-\mathbf{r}}, \mathbf{y}, \mathbf{u}) \propto p(\mathbf{y}|\theta_y, \mathrm{u})p(\mathbf{r}) \\
&\quad \propto \frac{1}{V(\mathbf{S})} 1_{\mathbf{S}}(\mathbf{r}) \left\{ \prod_{i=1}^{l} |\Sigma_{(i)}|^{-N_i/2} \right\} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{l} \Big[y_i - \mathbf{X}_i \vartheta_i \Big]' \big(I_{N_i} \otimes \Sigma_{(i)}^{-1} \big) \Big[y_i - \mathbf{X}_i \vartheta_i \Big] \right\}
\end{aligned}$$

where N_j for $j = 1, \ldots, h$ depends on $\mathbf{r}$. □

This full conditional distribution has similar expression to that found by Safadi & Morettin (2000) in univariate threshold autoregressive moving average model.

Now, we can extract samples of the posterior distribution of $\boldsymbol{\gamma}_j|\theta_{y-\boldsymbol{\gamma}_j}, \mathbf{y}, \mathbf{u}$ for $j = 1, \ldots, l$ if we extract samples of every individual distribution $\gamma_{i,j}|\theta_{y-\boldsymbol{\gamma}_{i,j}}, \mathbf{y}, \mathbf{u}$ for $i = 1, \ldots, \eta_j$ and $j = 1, \ldots, l$.

Proposition 5. *The full conditional distribution of $\gamma_{i,j}|\theta_{y-\boldsymbol{\gamma}_{i,j}}, \boldsymbol{y}, \boldsymbol{u}$ for $i = 1, \ldots, \eta_j$ and $j = 1, \ldots, l$ is a Bernoulli distribution with probability*

$$P(\gamma_{i,j} = 1|\theta_{y-\gamma_{i,j}}, \boldsymbol{y}, \boldsymbol{u}) = \frac{a_{i,j}}{a_{i,j} + b_{i,j}}, \tag{3.7}$$

where

$$a_{i,j} = p(\boldsymbol{y}|\boldsymbol{u}, \theta_{y-\gamma_{i,j}}, \gamma_{i,j} = 1, l)p_{i,j}$$

and

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

Proof. Note that

$$\begin{aligned}
&P(\gamma_{i,j} = 1|\theta_{y-\gamma_{i,j}}, \mathbf{y}, \mathbf{u}) \\
&\qquad = \frac{P(\gamma_{i,j} = 1)p(\theta_{y-\gamma_{i,j}}, \mathbf{y}, \mathbf{u}|\gamma_{i,j} = 1)}{P(\gamma_{i,j} = 1)p(\theta_{y-\gamma_{i,j}}, \mathbf{y}, \mathbf{u}|\gamma_{i,j} = 1) + P(\gamma_{i,j} = 0)p(\theta_{y-\gamma_{i,j}}, \mathbf{y}, \mathbf{u}|\gamma_{i,j} = 0)} \\
&\qquad = \frac{p(\theta_{y-\gamma_{i,j}}, \mathbf{y}, \mathbf{u}|\gamma_{i,j} = 1)p_{i,j}}{p(\theta_{y-\gamma_{i,j}}, \mathbf{y}, \mathbf{u}|\gamma_{i,j} = 1)p_{i,j} + p(\Theta_{y-\gamma_{i,j}}, \mathbf{y}, \mathbf{u}|\gamma_{i,j} = 0)(1 - p_{i,j})}.
\end{aligned}$$

Now, we can see that

$$\begin{aligned}
&p(\theta_{y-\gamma_{i,j}}, \mathbf{y}, \mathbf{u}|\gamma_{i,j} = 1) \\
&\qquad = p(\theta_{y-\gamma_{i,j}}|\gamma_{i,j} = 1)p(\mathbf{u}|\theta_{y-\gamma_{i,j}}, \gamma_{i,j} = 1)p(\mathbf{y}|\mathbf{u}, \theta_{y-\gamma_{i,j}}, \gamma_{i,j} = 1) \\
&\qquad = p(\theta_{y-\gamma_{i,j}})p(\mathbf{u})p(\mathbf{y}|\mathbf{u}, \theta_{y-\gamma_{i,j}}, \gamma_{i,j} = 1) \quad \text{by prior independence,}
\end{aligned}$$

$p(\theta_{y-\gamma_{i,j}}, \mathbf{y}, \mathbf{u}|\gamma_{i,j} = 0)$ is similar to above; the factors that do not depend on $\gamma_{i,j}$ can be factorized and they do not appear in the posterior distribution, then

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

where

$$a_{i,j} = p(\mathbf{y}|\mathbf{u}, \theta_{y-\gamma_{i,j}}, \gamma_{i,j} = 1, l)p_{i,j}$$

and

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

therefore $\gamma_{i,j}|\theta_{y-\gamma_{i,j}}, \mathbf{y}, \mathbf{u}$ has the Bernoulli distribution with probability given by (3.7).

□

All distributions above are standard distributions except the distribution for thresholds; in that situation, we can use the random walk Metropolis-Hastings algorithm to extract samples of that distribution. In this algorithm, the instrumental distribution depends on the current value of the chain, then the acceptance rates must be carefully analysed, see (Robert & Casella, 2004, p.295)

3.1.2 Posterior distribution in the SSVS method

The posterior distributions for the parameters $\mathbf{r}$ and $\Sigma_{(j)}$ for $j = 1, \ldots, l$ are exactly the same as above, the other ones change a little because the SSVS method suggest that each individual parameter $\theta_{i,j}$ has distribution conditional on the variable $\gamma_{i,j}$ for $i = 1, \ldots, \eta_j$, $j = 1, \ldots, l$, that is,

$$\theta_{i,j}|\gamma_{i,j} \sim (1 - \gamma_{i,j})N(0, \tau_{i,j}^2) + \gamma_{i,j}N(0, c_{i,j}^2\tau_{i,j}^2) \tag{3.8}$$

for specified values $c_{i,j} > 1$, $\tau_{i,j} > 0$. This mixture gives the following desirable characteristics: when $\gamma_{i,j} = 0$ we set $\tau_{i,j}$ close to zero, then $\theta_{i,j} \sim N(0, \tau_{i,j})$ which permits estimate $\theta_{i,j}$ by zero. On the other hand, when $\gamma_{i,j} = 1$ we set $c_{i,j}$ large, then $\theta_{i,j} \sim N(0, c_{i,j}^2\tau_{i,j}^2)$ which permits estimate $\theta_{i,j}$ as non-zero. Now, to obtain 3.8 as prior distribution of $\theta_{i,j}|\gamma_{i,j}$, we use the multivariate normal prior distribution

$$\theta_j|\boldsymbol{\gamma}_j \sim N(\mathbf{0}, D_{\gamma_j}RD_{\gamma_j})$$

where $D_{\gamma_j} = Diag[a_{1,j}\tau_{1,j}, \ldots, a_{g_j,j}\tau_{g_j,j}]$ with $a_{i,j} = 1$ if $\gamma_{i,j} = 0$ and $a_{i,j} = c_{i,j}$ if $\gamma_{i,j} = 1$. R is a known matrix where we can assign dependence among parameters. Then

$$\begin{aligned} p(\theta_j|\theta_{y-\theta_j}, \mathbf{y}, \mathbf{u}) &\propto p(y|\theta_y, \mathrm{u})p(\theta_j|\gamma_j) \\ &\propto \exp\left\{ -\frac{1}{2}\Big[\Big(y_j - \mathbf{X}_j\Gamma_j\theta_j\Big)'\big(I_{N_j} \otimes \Sigma_{(j)}^{-1}\big)\Big(y_j - \mathbf{X}_j\Gamma_j\theta_j\Big) + \right. \\ &\qquad \left. \theta_j'(D_{\gamma_j}RD_{\gamma_j})^{-1}\theta_j\Big]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\Big(\theta_j - \theta_j^*\Big)'V_j^{*-1}\Big(\theta_j - \theta_j^*\Big)\right\} \end{aligned}$$

where $V_j^* = \Big[\Gamma_j'\mathbf{X}_j'\big(I_{N_j} \otimes \Sigma_{(j)}^{-1}\big)\mathbf{X}_j\Gamma_j + (D_{\gamma_j}RD_{\gamma_j})\Big]^{-1}$, $\theta_j^* = V_j^*(\Gamma_j'\mathbf{X}_j'\big(I_{N_j} \otimes \Sigma_{(j)}^{-1}\big)y_j)$

that is the kernel of a normal multivariate distribution with mean θ_j^* and covariance matrix V_j^*.

Proposition 6. *With the assumptions in this section, the full conditional distribution* $\gamma_{i,j}|\theta_{y-\boldsymbol{\gamma}_{i,j}}, \boldsymbol{y}, \boldsymbol{u}$ *for* $i = 1, \ldots, \eta_j$ *and* $j = 1, \ldots, l$ *is a Bernoulli distribution with probability*

$$p(\gamma_{i,j} = 1|\theta_{y-\gamma_{i,j}}, \boldsymbol{y}, \boldsymbol{u}) = \frac{a_{i,j}}{a_{i,j} + b_{i,j}}, \tag{3.9}$$

where

$$a_{i,j} = p(\boldsymbol{y}|\boldsymbol{u}, \theta_{y-\gamma_{i,j}}, \gamma_{i,j} = 1)p(\theta_j|\gamma_{j-\gamma_{i,j}}, \gamma_{i,j} = 1, l)p_{i,j}$$

and

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

Proof. Note that

$$\begin{aligned}
&P(\gamma_{i,j} = 1|\theta_{y-\gamma_{i,j}}, \mathbf{y}, \mathbf{u}) \\
&= \frac{P(\gamma_{i,j} = 1)p(\theta_{y-\gamma_{i,j}}, \mathbf{y}, \mathbf{u}|\gamma_{i,j} = 1)}{P(\gamma_{i,j} = 1)p(\theta_{y-\gamma_{i,j}}, \mathbf{y}, \mathbf{u}|\gamma_{i,j} = 1) + P(\gamma_{i,j} = 0)p(\theta_{y-\gamma_{i,j}}, \mathbf{y}, \mathbf{u}|\gamma_{i,j} = 0)} \\
&= \frac{p(\theta_{y-\gamma_{i,j}}, \mathbf{y}, \mathbf{u}|\gamma_{i,j} = 1)p_{i,j}}{p(\theta_{y-\gamma_{i,j}}, \mathbf{y}, \mathbf{u}|\gamma_{i,j} = 1)p_{i,j} + p(\theta_{-\gamma_{i,j}}, \mathbf{y}, \mathbf{u}|\gamma_{i,j} = 0)(1 - p_{i,j})}
\end{aligned}$$

now, we can see that

$$\begin{aligned}
&p(\theta_{y-\gamma_{i,j}}, \mathbf{y}, \mathbf{u}|\gamma_{i,j} = 1) \\
&= p(\theta_{y-\gamma_{i,j}}|\gamma_{i,j} = 1)p(\mathbf{u}|\theta_{y-\gamma_{i,j}}, \gamma_{i,j} = 1)p(\mathbf{y}|\mathbf{u}, \theta_{y-\gamma_{i,j}}, \gamma_{i,j} = 1) \\
&= p(\gamma_{j-\gamma_{i,j}})p(\theta_j|\gamma_{j-\gamma_{i,j}}, \gamma_{i,1} = 1)p(\theta_{y-(\theta_j,\gamma_j)})p(\mathbf{u})p(\mathbf{y}|\mathbf{u}, \theta_{y-\gamma_{i,j}}, \gamma_{i,j} = 1) \quad \text{By prior independence}
\end{aligned}$$

$p(\theta_{y-\gamma_{i,j}}, \mathbf{y}, \mathbf{u}|\gamma_{i,j} = 0)$ is similar to above, the factors that do not depend on $\gamma_{i,j}$ can be factorized and not appear in the posterior distribution, then

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

where

$$a_{i,j} = p(\mathbf{y}|\mathbf{u}, \theta_{y-\gamma_{i,j}}, \gamma_{i,j} = 1)p(\theta_j|\gamma_{j-\gamma_{i,j}}, \gamma_{i,j} = 1, l)p_{i,j}$$

and

$$b_{i,j} = p(\mathbf{y}|\mathbf{u}, \theta_{y-\gamma_{i,j}}, \gamma_{i,j} = 0)p(\theta_j|\gamma_{j-\gamma_{i,j}}, \gamma_{i,j} = 0)(1 - p_{i,j})$$

then the posterior distribution is a Bernoulli distribution with probability $\frac{a_{i,j}}{a_{i,j}+b_{i,j}}$. □

With these distributions, we can implement Gibbs sampling to get the estimations of the $\gamma's$ variables, threshold vector $\mathbf{r}$ and non-structural parameters. We can see that these distributions have similar expressions to those found in (So & Chen, 2003) for univariate threshold models.

The threshold values can be estimated in a similar way to the proposed by Nieto (2005) as a previous step to the estimation of the other parameters of the MTAR model. To this end, the ideas in (Tong, 1990) and (Tsay, 1998) are followed to calculate the NAIC in MTAR model. The NAIC for a MTAR model with l regimes is

$$NAIC = \left\{ \sum_{j=1}^{l} AIC_j(\mathbf{r}) \right\} / \left\{ \sum_{j=1}^{l} N_j \right\} \tag{3.10}$$

where

$$AIC_j(\mathbf{r}) = N_j \ln\left(|S_j/N_j|\right) + 2k\eta_j.$$

3.2 Estimation of the number of regimes

Two methodologies to estimate the number of regimes in the MTAR model are proposed in this work. The first one is based on the methodology of Dellaportas et al. (2002)(Metropolised Carlin and Chib) that takes into the changing dimension in the vector of parameters when the number of regimes changes; the second one uses the marginal likelihood concept and the form to calculate it when we have outputs of the Gibbs sampling (Chib, 1995). We use the results of the last section to obtain next results. The following sections give us a brief description of the methods and implementation in MTAR models. It is important to keep in mind that the number of parameters change if the number of regimes is changed.

3.2.1 Metropolised Carlin and Chib for MTAR models

The Metropolised Carlin and Chib is a hybrid modification of the methodology proposed by Carlin & Chib (1995), where a Metropolis step is added to propose and accept or reject a new model. In this case, if the current state of the model is m, that means a model with m regimes, and a m' model is proposed with probability $j(m; m')$, then the acceptance probability for the MTAR model with m' regimes is

$$\alpha = \min\left(1, \frac{f(\mathbf{y}|\theta_{\mathrm{y,m'}}, m')p(\theta_{\mathrm{y,m'}}|m')p(\theta_{\mathrm{y,m}}|m')p(m')j(m', m)}{f(\mathbf{y}|\theta_{\mathrm{y,m}}, m)p(\theta_{\mathrm{y,m}}|m)p(\theta_{\mathrm{y,m'}}|m)p(m)j(m, m')}\right) \tag{3.11}$$

where $f(\mathbf{y}|\theta_{\mathrm{y,m}}, m)$ is the likelihood function in (3.4), $p(\theta_{\mathrm{y,m}}|m)$ is the prior distribution of the parameters, $p(\theta_{\mathrm{y,m'}}|m)$ is the pseudo prior distribution and $p(m)$ is the prior probability for a MTAR model with m regimes; $\theta_{\mathrm{y,m}} = (\theta_{\mathrm{yns}}, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_m, \mathbf{r})$. An advantage of this approach arises when there are many models to consider, because here we only have to generate samples of the posterior distribution of the actual model and of the pseudo prior distribution of the proposal model instead all of them.

In order to implement the metropolised Carlin and Chib, we consider that the assumptions about the parameters a priori are the same. The pseudo prior distributions for the vector of parameters are: (i) Multivariate normal distribution for threshold values and $\theta's$ vectors; (ii) Wishart distribution for the covariance matrix with large number of degree freedom and (iii) Bernoulli distributions for the individual $\gamma's$ variables, with independence between the parameters. In order to estimate the parameters of the pseudo prior distributions we run the Gibbs sampling for each one of the proposed models, and based on the runs, we estimate the parameters for each of these pseudo prior distributions; this is because the pseudo prior distributions have to be close to the posterior distributions for every model in order to improve the convergence. This recommendation was given in (Dellaportas et al., 2002) which it is easy to implement. We used traditional sample estimators to estimate each one of the parameters of the pseudo prior distributions. The probability distribution for the jumps $j(m, m')$ holds $j(m, m) = 0$ and it is uniform in sense of the probability for each jump is the same and symmetric. The prior distribution for the model indicator is a discrete uniform distribution and this complete the implementation of the metropolised Carlin and Chib algorithm.We summarize the implementation of the Metropolised Carlin and Chib in the following steps:

Step 1. Choose l_0 and then for each model with $l = 2, \ldots, l_0$ regimes extract previous samples to estimate the parameters of the pseudo prior distributions.

Step 2. Propose a new MTAR model m' with probability $j(m, m')$, where m is the current model.

Step 3. Extract $\theta_{\mathrm{y,m}}$ of the posterior distribution $p(\theta_{\mathrm{y,m}}|\mathrm{y}, m)$ following the results of the Chapter 3.

Step 4. Extract $\theta_{\mathrm{y},m'}$ of the pseudo prior $p(\theta_{\mathrm{y},m'}|m \neq m')$.

Step 5. Compute the probability 3.11 based on the samples of the steps 3 and 4 and accept or reject the model m'.

We can identify the number of regimes as the mode of $\{m^{(i)}\}_i$, where $m^{(i)}$ is the accepted value of m in the iteration i. The convergence of the chain for the model is checked monitoring the behaviour of the probabilities $p_m^{(i)}$ for $m = 1, \ldots, l_0$, where

$$p_m^{(i)} = \frac{\#\text{ of } m \text{ until iteration } i}{i}. \tag{3.12}$$

3.2.2 Marginal likelihood

The marginal likelihood can be used as a criterion in the problem of model choice. In order to implement this methodology in MTAR models, we have to estimate the threshold values as Nieto (2005) suggested using NAIC, because it is necessary to increase the speed of convergence of the chains, this does not happen if the thresholds are estimated jointly with the other parameters as we could observe a slow mixing for the estimation of the threshold values when the model is no the correct. The methodology developed by Chib (1995) is used here to estimate the marginal likelihood in a logarithm scale, for a partition in B parts of a parameter vector $\theta^\star_{\mathrm{y},l}$ of the model l when the outputs of the Gibbs sampling are available, that is

$$\begin{aligned}\ln \widehat{m}(\mathrm{y}) &= \ln f\left(\mathrm{y}|\theta^\star_{\mathrm{y},l}\right) + \ln \pi\left(\theta^\star_{\mathrm{y},l}\right) - \ln \widehat{\pi}\left(\theta^\star_{\mathrm{y},l}|\mathrm{y}\right)\\ &= \ln f\left(\mathrm{y}|\theta^\star_{\mathrm{y},l}\right) + \ln \pi\left(\theta^\star_{\mathrm{y},l}\right) - \sum_{f=1}^{B} \ln \widehat{\pi}\left(\theta^\star_{\mathrm{y},f,l}|\mathrm{y}, \theta^\star_{\mathrm{y},s,l}(s<f)\right)\end{aligned}$$

where

$$\widehat{\pi}\left(\theta^{\star}_{\mathrm{y},f,l}|\mathrm{y},\theta^{\star}_{\mathrm{y},s,l}(s<f)\right)=G^{-1}\sum_{j=1}^{G}\pi\left(\theta^{\star}_{\mathrm{y},f,l}|\mathrm{y},\theta^{\star}_{\mathrm{y},1,l},\cdots,\theta^{\star}_{\mathrm{y},f-1,l},\theta_h^{(j)}(h>r)\right)$$

and $\theta_h^{(j)}(h>r)$ are draws $\left\{\theta_r^{(j)},\theta_{r+1}^{(j)},\ldots,\theta_B^{(j)}:j=1\ldots,G\right\}$ of the reduced completed conditional Gibbs sampling.

For MTAR models we choose a posterior ordinate $\theta^{\star}_{\mathrm{y},l} = (\theta_1^{\star\prime},\cdots,\theta_l^{\star\prime},\gamma_1^{\star\prime},\cdots,\gamma_l^{\star\prime},\mathrm{vec}(\Sigma^{\star}_{(1)})',\cdots,\mathrm{vec}(\Sigma^{\star}_{(l)})')'$ as the posterior mode for γ's variables and the mean posterior for the other parameters based on a previous run of the Gibbs sampling for each model. Now, when we use the Kuo method to identify the autoregressive orders, we can see that likelihood in logarithmic scale is

$$\begin{aligned}\ln f(\mathrm{y}|\theta^{\star}_{\mathrm{y},l}) =& -\frac{1}{2}(T-p)\ln 2\pi\\ & -\frac{1}{2}\sum_{i=1}^{l}\left[N_i\ln\left|\Sigma^{\star}_{(i)}\right| + \left(y_j-\mathbf{X}_j\Gamma_j^{\star}\theta_j^{\star}\right)'\left(I_{N_j}\otimes\Sigma^{\star-1}_{(j)}\right)\left(y_j-\mathbf{X}_j\Gamma_j^{\star}\theta_j^{\star}\right)\right],\end{aligned}$$

the prior distribution in a logarithmic scale is

$$\begin{aligned}\ln\pi\left(\theta^{\star}_{\mathrm{y},l}\right) =& \sum_{i=1}^{l}\ln\pi(\gamma_i^{\star})+\ln\pi\left(\Sigma^{\star}_{(i)}\right)+\ln\pi(\theta_i^{\star})\\ =& \sum_{i=1}^{l}\left\{\sum_{j=1}^{\eta_i}\left[\gamma^{\star}_{i,j}\ln p_{i,j}+\left(1-\gamma^{\star}_{i,j}\right)\ln(1-p_{i,j})\right]\right.\\ & -\left[\frac{k\nu_{0,i}}{2}\ln 2+\frac{k(k-1)}{4}\ln\pi+\sum_{j=1}^{k}\ln\Gamma\left(\frac{\nu_{0,i}+1-j}{2}\right)+\frac{\nu_{0,i}}{2}\ln|S_{0,i}|\right.\\ & \left.+\frac{\nu_{0,i}+k+1}{2}\ln\left|\Sigma^{\star}_{(i)}\right|+\frac{1}{2}\mathrm{tr}\left(\Sigma^{\star-1}_{(i)}S_{0,i}\right)\right]\\ & \left.-\frac{1}{2}\left[\eta_i\ln 2\pi+\ln|\Sigma_{0,i}|+(\theta_i^{\star}-\theta_{0,i})'\,\Sigma_{0,i}^{-1}\,(\theta_i^{\star}-\theta_{0,i})\right]\right\}\end{aligned}$$

in the SSVS method, we only have to change the prior to $\theta's$ parameters. Finally, the posterior distribution can be split in the following form

$$\pi\left(\theta^{\star}_{\mathrm{y},l}|\mathrm{y}\right)=\pi(\gamma_1^{\star}|\mathrm{y})\pi(\gamma_2^{\star}|\gamma_1^{\star}\mathrm{y})\cdots\pi(\gamma_l^{\star}|\gamma_1^{\star},\cdots,\gamma_{l-1}^{\star},\mathrm{y}) \tag{3.13}$$

$$\pi(\theta_1^{\star}|\gamma_1^{\star},\cdots,\gamma_l^{\star},\mathrm{y})\cdots\pi(\theta_l^{\star}|\gamma_1^{\star},\cdots,\gamma_l^{\star},\theta_1^{\star},\cdots,\theta_{l-1}^{\star},\mathrm{y}) \tag{3.14}$$

$$\pi(\Sigma^{\star}_{(1)}|\gamma_1^{\star},\cdots,\gamma_l^{\star},\theta_1^{\star},\cdots,\theta_l^{\star},\mathrm{y})\cdots\pi(\Sigma^{\star}_{(l)}|\gamma_1^{\star},\cdots,\gamma_l^{\star},\theta_1^{\star},\cdots,\theta_l^{\star},,\Sigma_1^{\star},\cdots,\Sigma_{l-1}^{\star},\mathrm{y}) \tag{3.15}$$

the line (3.13) must be split for each individual $\gamma^\star_{i,j}$, then the distribution for every individual variable has the following form

$$\pi(\gamma^\star_{i,j} \mid \gamma^\star_1, \cdots, \gamma^\star_{j-1}, \gamma^\star_{1,j}, \cdots, \gamma^\star_{i-1,j}, \mathrm{y})$$

for $i = 1, \ldots, \eta_j$ with $j = 1, \ldots, l$ and it can be estimated based on G iterations of the reduced Gibbs sampling to the parameters $\gamma_{i,j}, \cdots, \gamma_{\eta_j,j}, \gamma_{i+1}, \gamma_l, \theta_1, \cdots, \theta_l, \Sigma_{(1)}, \cdots, \Sigma_{(l)}$, conditional on $\gamma^\star_1, \cdots, \gamma^\star_{j-1}, \gamma^\star_{1,j}, \cdots, \gamma^\star_{i-1,j}$, except to $\gamma^\star_{1,1}$ that uses a run of the full Gibbs Sampling, then

$$\widehat{\pi}(\gamma^\star_{i,j} \mid \gamma^\star_1, \cdots, \gamma^\star_{j-1}, \gamma^\star_{1,j}, \cdots, \gamma^\star_{i-1,j}, \mathrm{y}) = \frac{1}{G}\sum_{g=1}^{G} Be\left(\gamma^\star_{i,j}; p^{(g)}\right)$$

where $Be(\cdot;\cdot)$ means the Bernoulli probability function and

$$p^{(g)} =$$
$$p\left(\gamma_{i,j} = 1 | \gamma^\star_1, \cdots, \gamma^\star_{j-1}, \gamma^\star_{1,j}, \cdots, \gamma^\star_{i-1,j}, \gamma^{(g)}_{i+1,j}, \cdots, \gamma^{(g)}_{\eta_i,j}, \cdots, \gamma^{(g)}_l, \theta^{(g)}_1, \cdots, \theta^{(g)}_l, \Sigma^{(g)}_{(1)}, \cdots, \Sigma^{(g)}_{(l)}, \mathbf{y}\right)$$

is obtained with 3.7 or 3.9 in the KUO and SSVS method respectably. Now, every $\pi\left(\theta^\star_j \mid \gamma^\star_1, \cdots, \gamma^\star_l, \theta^\star_1, \cdots, \theta^\star_{j-1}\right)$ for $j = 1, \cdots, l$ in the line (3.14) can be estimated based on G iterations of the reduced Gibbs sampling for the parameters $\theta_j, \cdots, \theta_l$ conditional on $\gamma^\star_1, \cdots, \gamma^\star_l, \theta^\star_1, \cdots, \theta^\star_{j-1}$, then

$$\pi\left(\theta^\star_j \mid \gamma^\star_1, \cdots, \gamma^\star_l, \theta^\star_1, \cdots, \theta^\star_{j-1}\right) = \frac{1}{G}\sum_{g=1}^{G} N\left(\theta^\star_j; \theta^{(g)}_j, V^{(g)}_j\right)$$

where

$$V^{(g)}_j = \left[\Gamma^{\star\prime}_j \mathbf{X}'_j \big(I_{N_j} \otimes \Sigma^{(g)-1}_{(j)}\big) \mathbf{X}_j \Gamma^\star_j + \Sigma^{-1}_{0j}\right]^{-1}$$
$$\theta^{(g)}_j = V^{(g)}_j (\Gamma^{\star\prime}_j \mathbf{X}'_j \big(I_{N_j} \otimes \Sigma^{(g)-1}_{(j)}\big) y_j + \Sigma^{-1}_{0j}\theta_{0j}).$$

Finally, we have to estimate the distributions

$$\pi\left(\Sigma^\star_{(j)} \mid \gamma^\star_1, \cdots, \gamma^\star_l, \theta^\star_1, \cdots, \theta^\star_l, \Sigma^\star_{(1)}, \cdots, \Sigma^\star_{(j-1)}\right)$$

for $j = 1, \cdots, l$. These distributions correspond to a inverse Wishart with $N_j + \nu_{0j}$ degree freedom and covariance matrix $(S_j + S_{0j})^{-1}$ which only depends on $\theta^\star_j$ and the data; the matrix S_j can be calculated using (3.5), then we only have to evaluate directly the inverse Wishart distribution in $\Sigma^\star_{(j)}$, that is

$$\pi\left(\Sigma^\star_{(j)} \mid \gamma^\star_1, \cdots, \gamma^\star_l, \theta^\star_1, \cdots, \theta^\star_l, \Sigma^\star_{(1)}, \cdots, \Sigma^\star_{(j-1)}\right) = IG(\Sigma^\star_{(j)}; N_j + \nu_{0j}, (S_j + S_{0j})^{-1}).$$

Last two distributions are similar to the showed in (Nieto et al., 2013).

It is important to point out that we can implement some modifications to the last approaches. In the case of the metropolised Carlin and Chib methodology, the NAIC can be used to estimate previously the thresholds, then this methodology is employed exactly the same that above but without including the component of the threshold values. On the other hand, when the marginal likelihood needs to be computed, and the threshold values are estimated jointly with other parameters, then it is necessary to use the methodology proposed in (Chib & Jeliazkov, 2001) due to use of the Metropolis-Hastings algorithm to extract samples of the full conditional of the thresholds.

CHAPTER 4

Forecasting

This chapter is devoted to the forecasting phase with MTAR models. For this end, we need to find $E[\mathrm{Y}_{T+h}|\mathrm{y}_{1:T}, \mathrm{u}_{1:T}, m]$, which is the best prediction in the sense of MMSE(minimum mean square error) for a model with m regimes and $h \geq 1$. Nevertheless, an exact analytical expression of that conditional expectation is no easy to obtain in this context of non-linear models, this fact was pointed out in the 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 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. However we focus on the joint predictive distribution $p(\mathrm{y}_{T+1:T+h}, \mathrm{u}_{T+1:T+h}|\mathrm{y}_{1:T}, \mathrm{u}_{1:T}, m)$ from which the marginal distributions of interest can be obtained. The joint predictive distribution can be obtained in the following way:

$$p(\mathrm{y}_{T+1:T+h}, \mathrm{u}_{T+1:T+h}|\mathrm{y}_{1:T}, \mathrm{u}_{1:T}, m) = \int p(\mathrm{y}_{T+1:T+h}, \mathrm{u}_{T+1:T+h}|\mathrm{y}_{1:T}, \mathrm{u}_{1:T}, \theta_{\mathrm{y},m}, m)p(\theta_{\mathrm{y},m}|\mathrm{y}_{1:T}, \mathrm{u}_{1:T}, m)d\theta_{\mathrm{y},m}, \tag{4.1}$$

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

$$p(\mathrm{y}_{T+1}, \cdots, \mathrm{y}_{T+h}, \mathrm{u}_{T+1}, \cdots, \mathrm{u}_{T+h}|\mathrm{y}_{1:T}, \mathrm{u}_{1:T}, \theta_{\mathrm{y},m}, m) = \prod_{i=1}^{h} p(\mathrm{u}_{T+i}|\mathrm{u}_{1:T+i-1})p(\mathrm{y}_{T+i}|\mathrm{u}_{T+i}, \mathrm{y}_{1:T+i-1}, \mathrm{u}_{1:T+i-1}, \theta_{\mathrm{y},m}, m). \tag{4.2}$$

It is worth noticing that the densities in (4.2) for $i = 1, \ldots, h$ satisfy that:

(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 $\mathrm{w}'_{t,j}, \Gamma_j, \theta_j$ for $j = 1, \cdots, m$ described in Chapter 3.

The distribution (4.1) can be accessed by simulation as follows for $i-th$ 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 in the Chapter 3.

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 $p(\mathrm{y}_{T+1}|\mathrm{u}^{(i)}_{T+1}, \mathrm{y}_{1:T}, \mathrm{u}_{1:T}, \theta^{(i)}_{\mathrm{y},m}, m)$.

Step 4. Extract a random draw $\mathrm{u}^{(i)}_{T+2}$ of the kernel density $p(\mathrm{u}_{T+2}|\mathrm{u}^{(i)}_{T+1}, \mathrm{u}_{1:T})$.

Step 5. Extract a random draw $\mathrm{y}^{(i)}_{T+2}$ of the density $p(\mathrm{y}_{T+2}|\mathrm{u}^{(i)}_{T+2}, \mathrm{u}^{(i)}_{T+1}, \mathrm{y}_{1:T}, \mathrm{u}_{1:T}, \theta^{(i)}_{\mathrm{y},m}, m)$.

Continue extracting random draws recursively until $\mathrm{u}^{(i)}_{T+h}$ and $\mathrm{y}^{(i)}_{T+h}$. With the set $\{\mathrm{u}^{(i)}_{t+h}, \mathrm{y}^{(i)}_{t+h}\}_{i,h}$, it is possible to calculate: mean of the predictive distribution (point forecast), covariance matrix of the predictive distribution (measure of uncertainty of the forecast) and credible intervals for the point forecast. This procedure allows us to include the uncertainty of the parameters of the MTAR model to the forecast, which generalizes the forecasting procedure proposed by Nieto (2008) and Vargas (2012). In the following section we give ex post simulation examples for checking the performance of the forecasting procedure.

CHAPTER 5

Simulation results

In this chapter we illustrate the proposed methodology. The first three sections show us the performance of the estimation of the missing data and both structural and non-structural parameters of the MTAR model. In the fourth section, we can find some exploratory diagnostic for checking model adequacy. The last section is devoted to forecasting of output vector and exogenous variables. The results of the simulations are based on the models that we define below. Let $\mathrm{U}_t = (\mathrm{Z}_t, \mathrm{X}_t)'$ be a stable $VAR(1)$ process defined as:

Model 1.

$$\mathrm{U}_t = \mathrm{A}\mathrm{U}_{t-1} + \mathrm{r}_t \tag{5.1}$$

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 set the model with two regimes as follows:

Model 2.

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

where

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

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

$$\boldsymbol{\phi}_1^{(2)} = \begin{pmatrix} 0.3 & 0.5 \\ 0.2 & 0.7 \end{pmatrix}$$

$$\boldsymbol{\beta}_1^{(1)} = \begin{pmatrix} 0.3 \\ -0.4 \end{pmatrix}$$

$$\boldsymbol{\delta}_1^{(1)} = \begin{pmatrix} 0.6 \\ 1.0 \end{pmatrix}$$

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

with $r = \widehat{z}_{0.4}$ which is the $40-th$ percentile.

We can see in the Figures 5.1 and 5.2 a realization of the MTAR model with two regimes proposed above and a realization of the VAR model specified in 5.1, respectively.

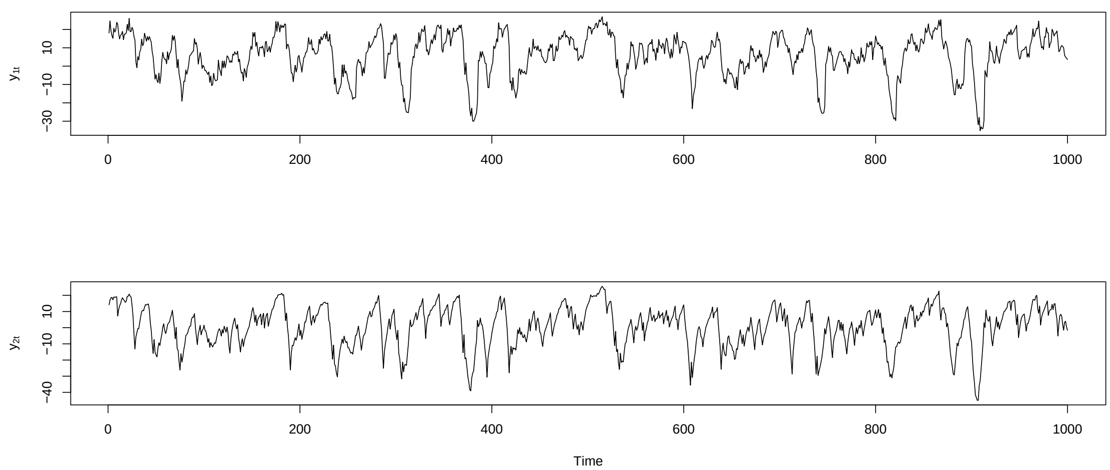

FIGURE 5.1. Simulated output vector of MTAR model 2.

Now, we set the model with three regimes as:

Model 3.

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

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

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

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

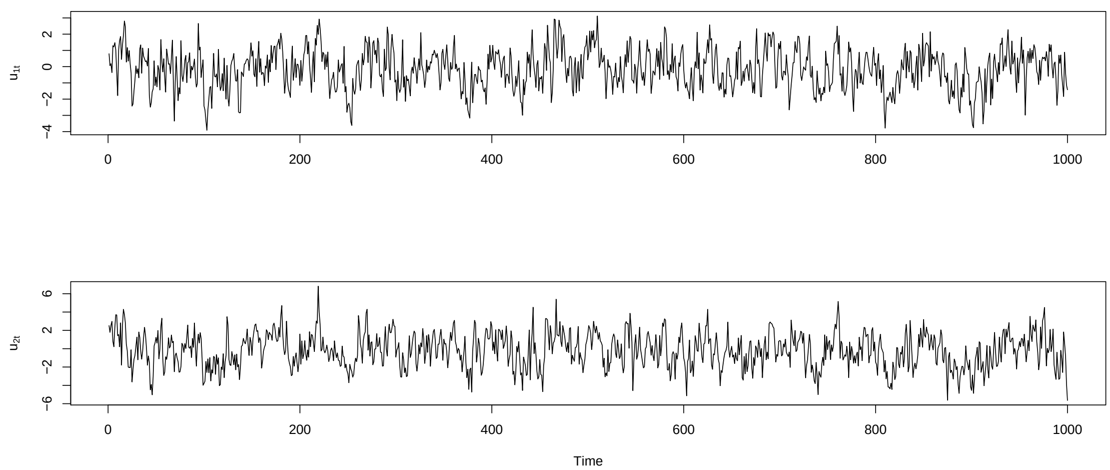

FIGURE 5.2. Simulated exogenous variables for model 1.

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

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

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

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

with $r_1 = \widehat{z}_{0.25}$ and $r_2 = \widehat{z}_{0.75}$.

The Figure 5.3 is a realization of the MTAR with three regimes defined above. We can see a stable behaviour of the series, however there are some bursts of large values that indicate us the presence of marginal heteroscedasticity.

5.1 Estimation of missing data and non-structural parameters

In this section we show the performance of the estimation of the missing data and the non-structural parameters assuming that the structural parameters are known. The first results are related only with the estimation of the missing data. To this end, we simulate a realization of length 1000 of the processes $\{\mathrm{Y}_t\}$ and $\{\mathrm{U}_t\}$ and randomly we eliminate the observations of five positions in both multivariate time series $\{\mathrm{y}_t\}$ and $\{\mathrm{u}_t\}$. Next, we proceed to run the Gibbs sampling as was explained in the section two. In the Tables 5.1 and 5.2 we report the point estimation and the credible intervals for the missing data, it was based on last 5000 of 10000 iterations of the Gibbs sampler. The random

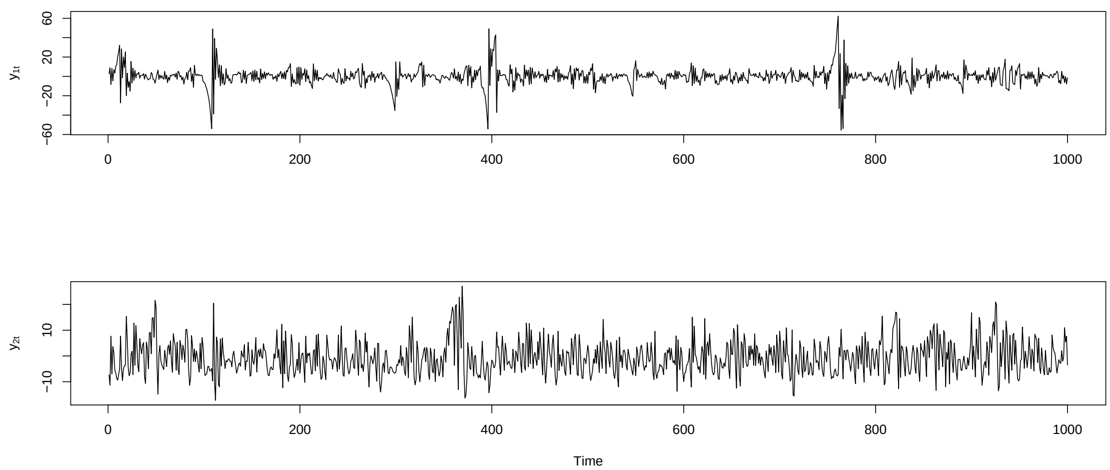

FIGURE 5.3. Simulated output vector of MTAR model 3.

t	y_t	$\widehat{y}_t$	C.I. 95%	t	u_t	$\widehat{u}_t$	C.I. 95%
136	11.83	9.80	$[7.70; 11.88]$	44	1.51	0.94	$[-0.15; 2.19]$
	5.79	5.22	$[2.92; 7.49]$		-0.25	1.00	$[-0.81; 2.85]$
2.98	0.34	1.16	$[-1.86; 4.23]$	327	-2.06	-0.97	$[-2.12; -0.33]$
	-9.68	-8.44	$[-10.45; -6.45]$		-0.15	-0.68	$[-2.77; 1.48]$
493	11.03	10.43	$[6.60; 14.42]$	505	0.89	1.19	$[-0.06; 2.62]$
	9.62	7.11	$[5.53; 8.68]$		2.23	2.07	$[0.16; 4.20]$
695	14.15	14.07	$[12.34; 15.79]$	707	1.08	0.44	$[-0.26; 1.51]$
	5.20	5.25	$[3.26; 7.30]$		0.10	0.28	$[-1.85; 2.42]$
914	—	—	$---$	940	-0.46	-0.55	$[-1.23; -0.31]$
	8.03	8.91	$[7.63; 10.19]$		2.41	1.91	$[0.13; 3.66]$

TABLE 5.1. Missing data estimation for output vector of MTAR model 2 and exogenous variables of model 1 with with $r = -0.30$. $-$ means that component is not missing.

walk Metropolis-Hastings algorithm was employed to extract samples of the posterior distributions for the missing data in $\{u_t\}$ with $c = 0.1$. Recall that the distributions for the estimation of the missing data in $\{y_t\}$ are multinormal which moments are found throughout Kalman filter. The point estimation and the credible intervals are calculated using the average and quantiles $q_{\alpha/2}$ and $q_{1-\alpha/2}$ of the values obtained of the sample for $\alpha = 0.05$. We can see that the estimation and the true data are close in many cases, and almost all 95% credible intervals contain the true data (all the true data are included in the 99% credible intervals). The point estimation of the missing data in the threshold variable belongs to the same regime of the true data, it indicates us that the missing data estimation of exogenous variables is working reasonably well. We checked the convergence of the distributions for the missing data vector monitoring first and second moments and some quantiles of the marginal chains throughout the iterations; all these quantities have a numerical stabilization, and it is fast in the case of the output vector y_t, but it takes many more iterations (around 5000) for the vector u_t. Only in one case $t = 353$ in Table 5.2, it was necessary more than 5000 iterations to attain the stabilization.

t	y_t	$\widehat{y}_t$	C.I. 95%	t	u_t	$\widehat{u}_t$	C.I. 95%
172	-2.80	-3.09	$[-4.69; -1.47]$	41	-1.44	-1.19	$[-1.89; -0.87]$
	11.24	10.97	$[7.97; 14.02]$		3.03	-0.05	$[-1.59; 1.74]$
283	3.78	3.55	$[2.05; 5.06]$	353	-1.55	-1.08	$[-1.60; -0.86]$
	4.48	6.01	$[-1.34; 12.79]$		-3.11	-0.76	$[-2.50; -1.07]$
403	4.99	3.75	$[2.08; 5.34]$	492	0.17	-0.21	$[-0.83; 0.69]$
	2.01	2.23	$[0.54; 3.94]$		0.46	0.04	$[-1.98; 2.22]$
627	-7.85	-8.97	$[-13.06; -4.94]$	621	2.97	2.31	$[1.11; 3.68]$
	8.11	3.48	$[1.38; 5.58]$		3.90	2.53	$[0.73; 4.38]$
865	11.03	10.52	$[8.80; 12.09]$	923	2.88	2.49	$[1.19; 3.92]$
	-8.34	-9.51	$[-11.20; -7.83]$		4.91	4.71	$[2.86; 6.53]$

TABLE 5.2. Missing data estimation for output vector of MTAR model 3 and exogenous variables of model 1 with $r_1 = -0.85 \quad r_2 = 0.90$.

parameter		Regime 1				Regime 2	
$\phi_{0,1}$		1.04				4.88	
$\phi_{0,2}$		-0.96				1.97	
		$y_{t-1,1}$	$y_{t-1,2}$	$y_{t-2,1}$	$y_{t-2,2}$	$y_{t-1,1}$	$y_{t-1,2}$
$\phi_{\cdot,km}$	$y_{t,1}$	0.50	-0.18	0.10	0.57	0.28	0.50
	$y_{t,2}$	-0.20	0.80	-0.39	0.48	0.20	0.69
			x_{t-1}				
$\beta_{\cdot,1}$			0.35				
$\beta_{\cdot,2}$			-0.35				
			z_{t-1}				
$\delta_{\cdot,1}$			0.62				
$\delta_{\cdot,2}$			1.02				
			1	2		1	2
$\Sigma^{(1/2)}_{\cdot,ik}$	1		1.01			2.64	
	2		0.62	1.55		0.53	1.0

TABLE 5.3. Non-structural parameters estimation for MTAR model 2. All the true parameters lie in the 95% credible intervals

In the second part of the results, we estimated the missing data and the non-structural parameters jointly. We ran 15000 iterations of the Gibbs sampling with a burning period of 5000 iterations, and we took the mean of the sample for last iterations as estimation of the parameters and missing data. In the Tables 5.3 and 5.4, we report the point estimation for the non-structural parameters in MTAR models with two and three regimes, they show us that the point estimation of the non-structural parameters are close to the true parameters, even the true values lie in the not reported 95% credible intervals. In the Tables 5.5 and 5.6 we report the point estimations and credible intervals of the missing data of the proposed models, the results of the estimation are similar to the obtained when only the missing data are estimated. The convergence to the stationary distribution is fast because the Gibbs sampling converged in the first 100 iterations for the non-structural parameters and the first 1000 for the missing data. It is important to point out that we started the Gibbs sampling with different values and always the convergence was attained.

Parameter		Regime 1		Regime 2				Regime 3	
$\phi_{0,1}$		2.03		0.41				-2.74	
$\phi_{0,2}$		0.63		-3.93				1.81	
		$y_{t-1,1}$	$y_{t-1,2}$	$y_{t-1,1}$	$y_{t-1,2}$	$y_{t-2,1}$	$y_{t-2,2}$	$y_{t-3,1}$	$y_{t-3,2}$
$\phi_{\cdot,km}$	$y_{t,1}$	-0.89	-0.00	0.70	0.005	0.79	0.20	-0.78	0.01
	$y_{t,2}$	0.22	-0.49	-0.002	0.58	-0.00	-0.41	0.20	0.78
				x_{t-1}				x_{t-2}	
$\beta_{\cdot,1}$				1.17				-0.71	
$\beta_{\cdot,2}$				-0.80				0.70	
								z_{t-1}	
$\delta_{\cdot,1}$								0.34	
$\delta_{\cdot,2}$								2.07	
		1	2		1	2		1	2
$\Sigma^{(1/2)}_{\cdot,ik}$	1	0.97			1.00			1.92	
	2	0.31	3.92		0.03	0.98		-0.43	1.02

TABLE 5.4. Non-structural parameters estimation for MTAR model 3. All the true parameters lie in the 95% credible intervals.

t	y_t	$\widehat{y}_t$	C.I. 95%	t	u_t	$\widehat{u}_t$	C.I. 95%
33	14.48	17.84	$[13.94; 21.87]$	107	-0.46	-0.91	$[-2.20; -0.25]$
	16.06	16.60	$[15.01; 18.24]$		-1.05	-0.87	$[-2.99; 1.28]$
368	-40.44	-40.49	$[-42.19; -38.79]$	387	0.84	0.42	$[-0.80; 1.64]$
	-45.98	-45.44	$[-47.43; -43.45]$		-0.03	0.05	$[-1.92; 2.00]$
597	-0.53	-1.23	$[-4.27; 1.85]$	597	-1.23	-0.87	$[-2.06; -0.25]$
	-3.92	-4.67	$[-6.69; -2.60]$		-0.19	-0.53	$[-3.02; 1.68]$
664	6.15	8.99	$[5.01; 13.11]$	651	-1.10	-1.46	$[-3.12; -0.33]$
	4.24	5.39	$[3.82; 6.94]$		-0.75	-0.96	$[-3.43; 1.57]$
887	14.43	14.41	$[11.30; 17.50]$	887	-0.17	1.30	$[-0.16; 3.60]$
	15.43	14.42	$[12.38; 16.48]$		-0.59	-0.10	$[-6.18; 5.62]$

TABLE 5.5. Missing data estimation for output vector of MTAR model 2 and exogenous variables of model 1 with $r = -0.22$.

t	y_t	$\widehat{y}_t$	C.I. 95%	t	u_t	$\widehat{u}_t$	C.I. 95%
105	1.89	2.60	$[0.98; 4.21]$	168	1.26	1.04	$[-2.50; 2.62]$
	-5.03	-4.71	$[-6.35; -3.05]$		1.74	0.75	$[-2.95; 3.55]$
215	—	—	—	375	-1.15	-2.21	$[-4.54; -0.73]$
	-4.95	-2.16	$[-8.67; 4.45]$		-1.42	-1.59	$[-6.90; 3.25]$
432	1.67	2.81	$[1.31; 4.27]$	424	0.75	0.33	$[-0.54; 0.93]$
	2.23	4.97	$[-2.08; 11.94]$		-0.19	1.43	$[-0.97; 3.84]$
759	0.64	-0.33	$[-1.75; 1.10]$	777	0.13	0.16	$[-0.61; 0.91]$
	-0.59	0.01	$[-1.96; 1.95]$		1.20	1.20	$[-1.20; 3.54]$
927	-6.41	-7.26	$[-8.77; -5.73]$	839	1.38	2.02	$[0.98; 2.82]$
	6.25	1.97	$[-5.01; 8.93]$		2.79	1.62	$[-0.52; 4.54]$

TABLE 5.6. Missing data estimation for output vector MTAR of model 3 and exogenous variables of model 1 with $r_1 = -0.67 \quad r_2 = 0.96$. $-$ means that component is not missing.

5.2 Estimation of the autoregressive orders and the thresholds values

In this section we check the performance of the joint estimation of the non-structural parameters, autoregressive orders and threshold values of the MTAR model. The identification of the autoregressive orders is carried out in indirect form and depends on $\gamma's$ variables as follows. For the regime j we have the vector $\boldsymbol{\gamma}_j$, it is possible to extract samples of its posterior distribution following Chapter 3, then the vector $\boldsymbol{\gamma}_j$ with the highest posterior probability is called "the best associated autoregressive vector (model)"for this regime, and the autoregressive orders are based on this vector. The autoregressive orders p_j, q_j and d_j in the regime j are identified looking for the individual variables $\gamma_{i,j} = 1$ associated with the parameters $\theta_{i,j}$ of highest lag. The following example give us an explanation about how the identification of the autoregressive orders is based on the γ's variables. Assume that in the state of the nature the bivariate process $\{\mathrm{Y}_t\}$ follows a MTAR(2;1;1) process

$$\mathrm{Y}_t = \begin{cases} \begin{pmatrix} \phi_{0,1}^{(1)} \\ \phi_{0,2}^{(1)} \end{pmatrix} + \begin{pmatrix} \phi_{1,11}^{(1)} & \phi_{1,12}^{(1)} \\ \phi_{1,21}^{(1)} & \phi_{1,22}^{(1)} \end{pmatrix} \mathrm{Y}_{t-1} + \boldsymbol{\Sigma}_{(1)}^{1/2}\varepsilon_t, & \text{if } \mathrm{Z}_t \leq r \\ \begin{pmatrix} \phi_{0,1}^{(2)} \\ \phi_{0,2}^{(2)} \end{pmatrix} + \begin{pmatrix} \phi_{1,11}^{(2)} & \phi_{1,12}^{(2)} \\ \phi_{1,21}^{(2)} & \phi_{1,22}^{(2)} \end{pmatrix} \mathrm{Y}_{t-1} + \boldsymbol{\Sigma}_{(2)}^{1/2}\varepsilon_t, & \text{if } \mathrm{Z}_t > r, \end{cases}$$

where all entries of matrices are different of zero except for $\boldsymbol{\Sigma}_{(j)}^{1/2}$ for $j = 1, 2$. Then, it is necessary to propose initial autoregressive orders to the identification; for the example $p_1 = 2$ and $p_2 = 2$ are the proposed initial autoregressive orders. With these values we have that

$$\theta_j = (\theta_{1,j}, \theta_{2,j}, \cdots, \theta_{10,j})' = \left(\theta_{0,1}^{(j)}, \theta_{1,11}^{(j)}, \theta_{1,12}^{(j)}, \theta_{2,11}^{(j)}, \theta_{2,12}^{(j)}, \theta_{0,2}^{(j)}, \theta_{1,21}^{(j)}, \theta_{1,22}^{(j)}, \theta_{2,21}^{(j)}, \theta_{2,22}^{(j)}\right)'$$

and $\gamma_j = (\gamma_{1,j}, \gamma_{2,j}, \cdots, \gamma_{10,j})'$ for $j = 1, 2$. The parameters $\theta_{2,11}^{(j)}, \theta_{2,12}^{(j)}, \theta_{2,21}^{(j)}, \theta_{2,22}^{(j)}$ for $j = 1, 2$ are the components of the matrices for the second lag of the vector Y_t in each regime. The best associated autoregressive vector for each regime must be $(1, 1, 1, 0, 0, 1, 1, 1, 0, 0)$ because in the state of the nature the parameters $\theta_{2,11}^{(j)}, \theta_{2,12}^{(j)}, \theta_{2,21}^{(j)}, \theta_{2,22}^{(j)}$ for $j = 1, 2$ must be zero, which is obtained if the associated variables $\gamma_{4,j}, \gamma_{5,j}, \gamma_{9,j}, \gamma_{10,j}$ are zero for $j = 1, 2$. Therefore the autoregressive orders for each regimes are identified as $\hat{p}_1 = 1$ and $\hat{p}_2 = 1$.

To check the performance in the estimations of the structural parameters, we simulate 100 replications of the MTAR model and count how many times the correct associated autoregressive vector is the best associated vector or the second best in each regime. Every row in the matrices 5.2 and 5.3 indicate us the correct associated autoregressive vector in each regime. In order to implement the methodology, it is necessary to propose initial autoregressive orders which be $p_j = q_j = d_j = 3$ for the model with two regimes and $p = q = d = 4$ for the model with three regimes. The prior probabilities $p_{i,j}$ for $i = 1, \ldots, \eta_j$, $j = 1, \ldots, l$ were set in 0.5 because there is not prior knowledge to include or not the parameter $\theta_{i,j}$ in the model.

$$\text{Two Regimes} \begin{pmatrix} 1,1,1,1,1,0,0,1,0,0,1,0,0,1,1,1,1,1,0,0,1,0,0,1,0,0 \\ 1,1,1,0,0,0,0,0,0,0,0,0,0,1,1,1,0,0,0,0,0,0,0,0,0,0 \end{pmatrix} \tag{5.2}$$

		Regime 1	Regime 2
Percent of correct model	Kuo	89[96]	76[97]
	SSVS	90[100]	92[95]
Mean probability	Kuo	0.52[0.49]	0.30[0.26]
	SSVS	0.55[0.52]	0.37[0.36]
Percent Threshold	Kuo	99	
	SSVS	95	

Table 5.7. Efficiency of Kuo and SSVS methods for model 2.

		Regime 1	Regime 2	Regime 3
Percent of correct model	Kuo	61[69]	85[94]	72[86]
	SSVS	64[77]	89[97]	83[93]
Probability mean	Kuo	0.052[0.049]	0.34[0.32]	0.17[0.16]
	SSVS	0.11[0.10]	0.46[0.44]	0.28[0.27]
Percent Threshold	Kuo		95	
	SSVS		89	

Table 5.8. Efficiency of Kuo and SSVS methods for model 3.

$$\text{Three Regimes}\begin{pmatrix}1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0\\ 1,1,0,1,1,0,0,0,0,1,0,0,0,0,0,0,0,1,0,1,0,1,0,0,0,0,1,0,0,0,0,0,0,0\\ 1,0,0,0,0,1,0,0,0,0,1,0,0,1,0,0,0,1,0,0,0,0,1,1,0,0,0,1,0,0,1,0,0,0\end{pmatrix} \tag{5.3}$$

Tables 5.7 and 5.8 summarize the results of the simulation based on the Kuo and SSVS methods where in every replication, we ran the Gibbs sampling for 15000 iterations and we took the last 10000 to get the results. The percentage of the correct model is the percentage of times that the correct model is the best model, and the percentage in brackets is the percentage that the correct model is either the best model or the second best model in each regime. The mean probability is the mean of the probabilities that the correct model is the best model, and in brackets is the mean of the probabilities that the correct model is either the best model or second best model. Furthermore the percent threshold is the percentage of times that the true thresholds lie in the individual credible interval of 95%. We used the following prior values $c_{i,j} = 25$, $\tau_i = 1.25$ and $\tau_{i,j} = 1.5$ for all i, j in the models with two and three regimes respectively for the SSVS method.

In terms of the percentage of the correct model, we can see that Kuo and SSVS methods work similar with a little difference in favour to SSVS method. However, the SSVS method depends strongly on the prior values $c_{i,j}$ and $\tau_{i,j}$ which must be chosen carefully following the recommendations in (George & McCulloch, 1993). On the other hand, for the MTAR(2;3,3,3;3,3,3) model considered above, we have 134217728 possible subset models, but the stochastic search is done only for about 2000 subset models (in the worst of the cases); it gives us a considerable gain in terms of time. If we consider the complexity of the models and the many subset promising models, we can see that the identification of the autoregressive orders is successful through the Bayesian variable selection. It is important to point out that the methodologies SSVS and KUO choose models whose average probabilities are ergodic, see Figures 5.4 and 5.5. To extract samples of the posterior distribution of the threshold values, we employed a random walk Metropolis-Hastings based on the uniform distribution on the unitary circle with small squared radius 0.005 and 0.00375 for the models with two regimes and three regimes respectively, of course the samples are correlated but the autocorrelation decrease fast.

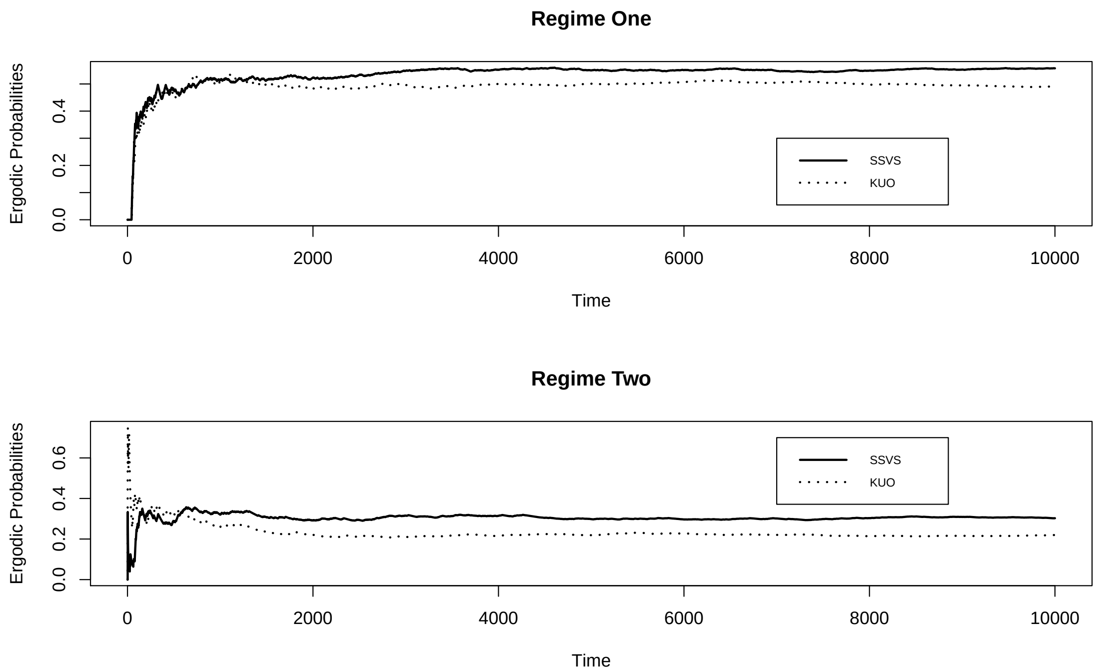

FIGURE 5.4. Ergodic posterior probabilities for the best vector $\boldsymbol{\gamma}$ in the model 2.

The estimation of the threshold values was effective based on threshold percentage. When we searched the thresholds based on NAIC as a previous step, we found that these values are similar to the obtained with Bayesian estimation.

The Tables 5.9-5.12 gives us the point estimation and the credible intervals of the non-structural parameters and threshold values for a specific example. We can see that all true values lie in the 95% credible intervals, and many of the estimations are close to the true values. The credible intervals obtained with the Kuo method are smaller than the obtained with the SSVS method, and it is the only appreciated difference that we found for the estimation of the non-structural parameters and threshold values using Kuo and SSVS methods.

5.3 Identification of the number of regimes

Metropolised Carlin and Chib approach and marginal likelihood were the two methods implemented to identify the number of regimes in a MTAR model. In this section, we check the relevance of both proposed methods based on the models with two and three regimes described at the beginning of the chapter. For the marginal likelihood method, as a previous step, we search the thresholds values using NAIC with the autoregressive orders fixed for the proposed models, this is due to the slow mixing in the estimation of the thresholds when the number of the regimes is not the correct. Conditional on the found thresholds and for KUO and SSVS methods, we calculate the marginal likelihood in a logarithmic scale for each proposed model based on the last 2000 of 3000 iterations of the full and reduced Gibbs sampling. On the other hand, we run the Metropolised Carlin

	Regime 1		Regime 2	
Parameter	Estimation	Credible Interval 95%	Estimation	Credible Interval 95%
$\phi_{0,1}$	0.900	(0.70;1.11)	4.860	(4.45;5.26)
$\phi_{1,11}$	0.512	(0.47;0.55)	0.280	(0.22;0.33)
$\phi_{1,12}$	-0.200	(-0.23;-0.17)	0.510	(0.47;0.55)
$\phi_{2,11}$	0.093	(0.05;0.12)	-0.002	(-2.00;1.92)
$\phi_{2,12}$	0.600	(0.56;0.64)	0.006	(-1.90;1.92)
$\phi_{3,11}$	0.007	(-1.92;1.98)	0.000	(-1.89;1.93)
$\phi_{3,12}$	-0.019	(-1.98;1.92)	-0.001	(-1.93;1.88)
$\beta_{1,1}$	0.330	(0.24;0.43)	0.015	(-1.90;1.90)
$\beta_{2,1}$	-0.004	(-1.94;1.92)	0.006	(-1.88,1.94)
$\beta_{3,1}$	0.003	(-1.94;2.02)	-0.002	(-1.93;1.82)
$\delta_{1,1}$	0.550	(0.37;0.72)	-0.002	(-1.91;1.76)
$\delta_{2,1}$	-0.007	(-1.96;1.91)	-0.033	(-1.93;1.86)
$\delta_{3,1}$	-0.009	(-1.98;1.92)	0.003	(-1.92;1.93)
$\phi_{0,2}$	-1.170	(-1.50;-0.86)	1.970	(1.80;2.13)
$\phi_{1,21}$	-0.170	(-0.23;-0.10)	0.190	(0.17;0.21)
$\phi_{1,22}$	0.800	(0.75;0.85)	0.690	(0.68;0.71)
$\phi_{2,21}$	-0.460	(-0.46;-0.36)	-0.012	(-1.91;1.93)
$\phi_{2,22}$	0.470	(0.41;0.53)	-0.002	(-2.03;2.00)
$\phi_{3,21}$	0.010	(-1.92;1.90)	-0.004	(-1.93;1.97)
$\phi_{3,22}$	-0.003	(-1.93;1.98)	0.010	(-1.97;1.91)
$\beta_{1,2}$	-0.330	(-0.47;-0.18)	-0.008	(-1.93;1.96)
$\beta_{2,2}$	-0.008	(-1.89;1.97)	-0.015	(-1.92;1.93)
$\beta_{3,2}$	0.012	(-1.96;1.90)	0.014	(-1.90;1.97)
$\delta_{1,2}$	0.880	(0.61;1.15)	0.013	(-1.91;1.93)
$\delta_{2,2}$	0.009	(-1.90;1.80)	0.031	(-1.93;1.98)
$\delta_{3,2}$	0.003	(-1.87;1.87)	0.002	(-1.93;1.99)
Σ_{11}	0.910	(0.83;1.0001)	2.620	(2.41;2.84)
Σ_{12}	0.620	(0.53;0.71)	0.510	(0.43;0.60)
Σ_{22}	1.580	(1.44;1.74)	0.920	(0.85;1.00)
r	-0.19	$(-0.20;-0.18)$		

TABLE 5.9. Point estimation and credible intervals with Kuo method for the model 2 with $r = -0.1889$.

	Regime 1		Regime 2		Regime 3	
Parameter	Estimation	C.I. 95%	Estimation	C.I. 95%	Estimation	C.I. 95%
$\phi_{0,1}$	1.69	(1.48;1.89)	0.46	(0.33;0.59)	-2.702	(-3.19;-2.22)
$\phi_{1,11}$	-0.90	(-0.92;-0.88)	0.70	(0.69;0.72)	0.00	(-1.92;1.98)
$\phi_{1,12}$	-0.02	(-1.94;-1.95)	0.01	(-1.86;1.95)	-0.00	(-1.83;1.91)
$\phi_{2,11}$	-0.00	(-1.99;1.97)	0.79	(0.77;0.80)	0.02	(-1.94;1.98)
$\phi_{2,12}$	0.01	(-1.88;1.98)	0.19	(0.16;0.21)	-0.00	(-1.92;1.87)
$\phi_{3,11}$	0.02	(-1.91;2.02)	0.01	(-1.86;1.93)	-0.78	(-0.81;-0.76)
$\phi_{3,12}$	-0.01	(-1.92;1.89)	-0.00	(-1.92;1.94)	0.02	(-1.87;1.92)
$\phi_{4,11}$	-0.00	(-1.98:1.98)	-0.01	(-1.97;1.91)	-0.03	(-2.05;1.95)
$\phi_{4,12}$	0.00	(-1.87:1.92)	-0.02	(-1.99;1.93)	0.00	(-1.96;1.92)
$\beta_{1,1}$	-0.01	(-1.91:1.84)	1.12	(1.04;1.21)	-0.02	(-1.86;1.88)
$\beta_{2,1}$	0.00	(-1.85;1.88)	0.00	(-1.88;1.89)	-0.53	(-0.76;-0.28)
$\beta_{3,1}$	0.01	(-1.87;1.95)	0.00	(-1.97;1.91)	-0.00	(-1.87;.86)
$\beta_{4,1}$	0.02	(-1.91;1.90)	-0.01	(-1.96;1.92)	-0.01	(-1.91;1.89)
$\delta_{1,1}$	-0.02	(-1.97;1.89)	-0.00	(-1.99;1.96)	0.30	(-1.40;1.38)
$\delta_{2,1}$	-0.02	(-1.95;1.93)	-0.02	(-1.95;1.87)	-0.00	(-1.90;1.83)
$\delta_{3,1}$	-0.01	(-1.95;1.90)	-0.01	(-1.84;1.89)	0.02	(-1.86;1.90)
$\delta_{4,1}$	0.00	(1.94;1.92)	0.01	(-1.88;1.93)	-0.00	(-1.83;1.90)
$\phi_{0,2}$	0.75	(-1.03;1.78)	-3.94	(-4.06;-3.82)	1.76	(1.50;2.00)
$\phi_{1,21}$	0.21	(0.13;0.29)	-0.00	(-1.95;1.97)	-0.00	(-1.93;1.91)
$\phi_{1,22}$	-0.36	(-0.52;-0.21)	0.60	(0.58;0.62)	-0.00	(-1.91;1.87)
$\phi_{2,21}$	-0.01	(-1.88;1.90)	0.00	(-1.95;2.02)	0.01	(-1.87;1.94)
$\phi_{2,22}$	-0.00	(-1.93;1.93)	-0.41	(-0.43;-0.39)	-0.00	(-1.93;2.02)
$\phi_{3,21}$	-0.03	(-1.93;1.90)	-0.00	(-1.95;1.97)	0.19	(0.18;0.21)
$\phi_{3,22}$	0.01	(-1.88;1.94)	0.02	(-1.94;2.02)	0.79	(0.76;0.83)
$\phi_{4,21}$	-0.00	(-1.87;1.94)	-0.00	(-2.00;1.98)	-0.01	(-1.99;1.94)
$\phi_{4,22}$	-0.00	(-1.92;1.88)	-0.00	(-1.92;1.94)	0.00	(-1.96;1.94)
$\beta_{1,2}$	-0.04	(-1.94;1.86)	-0.77	(-0.86;-0.69)	-0.01	(-1.94;1.94)
$\beta_{2,2}$	0.09	(-1.83;1.72)	0.03	(-1.86;1.84)	0.76	(0.64;0.88)
$\beta_{3,2}$	-0.01	(-1.87;1.87)	-0.02	(-1.88;1.94)	-0.01	(-1.90;1.96)
$\beta_{4,2}$	-0.01	(-1.73;1.84)	-0.00	(-1.85;1.94)	0.01	(-1.90;1.94)
$\delta_{1,2}$	0.02	(-1.83;1.80)	0.00	(-1.88;1.88)	2.12	(1.88;2.36)
$\delta_{2,2}$	0.01	(-1.79;1.79)	0.00	(-1.92;1.97)	-0.01	(-1.95;1.89)
$\delta_{3,2}$	-0.02	(-1.84;1.80)	0.00	(-1.94;1.93)	-0.02	(-1.99;1.91)
$\delta_{4,2}$	-0.06	(-1.79;1.80)	0.00	(-1.86;1.90)	-0.00	(-1.93;1.93)
Σ_{11}	0.85	(0.71; 1.01)	1.03	(0.92;1.16)	2.19	(1.89;2.56)
Σ_{12}	0.33	(0.15;0.53)	0.06	(-0.01;0.14)	-0.51	(-0.69;-0.35)
Σ_{22}	4.05	(3.39;4.86)	0.91	(0.81;1.02)	1.08	(0.94;1.26)
r_1	-0.78	$(-0.79; -0.73)$		r_2	0.763	(0.7633; 0.7645)

TABLE 5.10. Point estimation and credible intervals with Kuo method for model 3 with $r_1 = -0.7965$ $r_2 = 0.7636$.

	Regime 1		Regime 2	
Parameter	Estimation	Credible Interval 95%	Estimation	Credible Interval 95%
$\phi_{0,1}$	0.799	(0.53;1.05)	4.59	(4.16;5.03)
$\phi_{1,11}$	0.567	(0.50;0.65)	0.338	(0.28;0.39)
$\phi_{1,12}$	-0.208	(-0.25;-0.16)	0.461	(0.41;0.51)
$\phi_{2,11}$	0.077	(-0.81;0.91)	0.004	(-2.46;2.44)
$\phi_{2,12}$	0.584	(0.53;0.63)	0.040	(-2.29;2.48)
$\phi_{3,11}$	0.031	(-2.43;2.47)	-0.005	(-2.41;2.38)
$\phi_{3,12}$	-0.004	(-2.46;2.48)	0.004	(-2.41;2.47)
$\beta_{1,1}$	0.233	(0.13;0.33)	0.011	(-2.44;2.39)
$\beta_{2,1}$	0.035	(-2.36;2.46)	0.042	(-2.33;2.55)
$\beta_{3,1}$	0.023	(-2.40;2.43)	0.003	(-2.42;2.44)
$\delta_{1,1}$	0.477	(0.28;0.67)	-0.005	(-2.30;2.34)
$\delta_{2,1}$	0.006	(-2.38;2.27)	0.018	(-2.38;2.45)
$\delta_{3,1}$	-0.002	(-2.39;2.36)	0.003	(-2.3;2.38)
$\phi_{0,2}$	-1.386	(-1.76;-1.01)	1.905	(1.70;2.09)
$\phi_{1,21}$	-0.161	(-0.26;-0.04)	0.215	(0.18;0.24)
$\phi_{1,22}$	0.798	(0.73;0.86)	0.684	(0.66;0.70)
$\phi_{2,21}$	-0.380	(-0.49;-0.30)	0.003	(-2.45;2.46)
$\phi_{2,22}$	0.465	(0.39;0.53)	-0.002	(-2.38;2.48)
$\phi_{3,21}$	-0.019	(-2.44;2.43)	-0.012	(-2.51;2.44)
$\phi_{3,22}$	-0.011	(-2.43;2.47)	0.009	(-2.48;2.44)
$\beta_{1,2}$	-0.443	(-0.59;-0.29)	0.009	(-2.36;2.42)
$\beta_{2,2}$	0.009	(-2.37;2.43)	0.015	(-2.40;2.38)
$\beta_{3,2}$	0.025	(-2.36;2.36)	-0.000	(-2.45;2.48)
$\delta_{1,2}$	0.771	(0.48;1.07)	0.018	(-2.37;2.37)
$\delta_{2,2}$	0.009	(-1.90;1.80)	-0.002	(-2.42;2.31)
$\delta_{3,2}$	0.003	(-1.87;1.87)	0.030	(-2.38;2.44)
Σ_{11}	0.939	(0.85;1.03)	2.521	(2.32;2.73)
Σ_{12}	0.545	(0.46;0.64)	0.565	(0.48;0.65)
Σ_{22}	1.482	(1.34;1.64)	0.991	(0.92;1.06)
r	-0.261	$(-0.264;-0.257)$		

TABLE 5.11. Point estimation and credible intervals with SSVS method for model 2 with $r = -0.2610$.

	Regime 1		Regime 2		Regime 3	
Parameter	Estimation	C.I. 95%	Estimation	C.I. 95%	Estimation	C.I. 95%
$\phi_{0,1}$	1.88	(1.59;2.11)	0.29	(0.16;0.42)	-2.88	(-3.48;-2.30)
$\phi_{1,11}$	-0.89	(-0.91;-0.87)	0.69	(0.67;0.71)	0.01	(-2.85;2.90)
$\phi_{1,12}$	-0.02	(-2.98;2.85)	0.00	(-2.95;2.90)	-0.05	(-2.93;2.82)
$\phi_{2,11}$	0.00	(-3.00;2.97)	0.79	(0.78;0.81)	0.02	(-2.83;3.03)
$\phi_{2,12}$	0.00	(-2.91;2.86)	0.19	(0.17;0.21)	-0.00	(-2.91;2.98)
$\phi_{3,11}$	0.00	(-2.87;2.95)	0.00	(-3.04;2.75)	-0.82	(-0.87;-0.76)
$\phi_{3,12}$	-0.01	(-2.90;2.92)	-0.03	(-1.92;1.94)	-0.00	(-2.99;3.01)
$\phi_{4,11}$	0.00	(-2.88;2.93)	-0.02	(-2.98;2.86)	-0.00	(-2.95;2.99)
$\phi_{4,12}$	-0.00	(-2.99;2.86)	0.04	(-2.92;3.04)	-0.00	(-2.87;2.91)
$\beta_{1,1}$	-0.01	(-2.94;2.95)	1.20	(1.12;1.28)	0.05	(-2.76;2.81)
$\beta_{2,1}$	-0.00	(-2.88;2.81)	-0.00	(-2.90;2.94)	-0.71	(-0.94;-0.49)
$\beta_{3,1}$	-0.01	(-2.90;2.88)	-0.03	(-2.95;2.93)	0.00	(-2.84;2.85)
$\beta_{4,1}$	-0.00	(-2.92;2.85)	-0.00	(-2.93;2.99)	0.11	(-2.49;2.44)
$\delta_{1,1}$	-0.04	(-2.72;2.64)	0.00	(-2.80;2.72)	0.32	(-2.06;2.00)
$\delta_{2,1}$	-0.01	(-2.89;2.91)	-0.02	(-2.84;2.84)	-0.04	(-2.84;2.68)
$\delta_{3,1}$	0.02	(-2.86;0.95)	-0.01	(-2.89;2.88)	-0.06	(-2.93;2.79)
$\delta_{4,1}$	-0.00	(-2.92;2.92)	0.02	(-2.81;2.91)	0.00	(-2.77;2.85)
$\phi_{0,2}$	0.25	(-2.57;2.42)	-4.08	(-4.20;-3.95)	1.71	(1.40;2.03)
$\phi_{1,21}$	0.16	(0.06;0.25)	0.00	(-2.92;2.93)	-0.00	(-2.83;2.84)
$\phi_{1,22}$	-0.56	(-0.67;-0.44)	0.60	(0.58;0.62)	-0.00	(-2.82;2.83)
$\phi_{2,21}$	-0.00	(-2.97;2.85)	-0.04	(-3.12;2.84)	-0.00	(-2.94;2.98)
$\phi_{2,22}$	0.00	(-2.95;2.89)	-0.39	(-0.41;-0.37)	0.01	(-2.99;2.95)
$\phi_{3,21}$	0.01	(-2.87;2.76)	-0.03	(-2.99;2.97)	0.21	(0.18;0.24)
$\phi_{3,22}$	0.03	(-2.95;3.06)	-0.00	(-2.87;2.95)	0.80	(0.77;0.83)
$\phi_{4,21}$	0.00	(-2.88;2.92)	-0.01	(-2.98;2.88)	0.00	(-2.89;2.88)
$\phi_{4,22}$	-0.01	(-2.95;2.98)	-0.03	(-2.95;2.82)	-0.03	(-2.98;2.87)
$\beta_{1,2}$	-0.03	(-1.94;1.86)	-0.85	(-0.94;-0.77)	0.00	(-2.73;2.80)
$\beta_{2,2}$	0.06	(-2.76;2.76)	0.03	(-2.84;3.03)	0.73	(0.60;0.86)
$\beta_{3,2}$	-0.00	(-2.78;2.78)	-0.01	(-2.91;2.82)	-0.00	(-2.98;2.85)
$\beta_{4,2}$	-0.01	(-2.82;2.81)	-0.00	(-2.95;2.82)	-0.01	(-2.99;2.86)
$\delta_{1,2}$	-0.17	(-2.68;2.60)	0.03	(-2.80;2.84)	2.23	(1.96;2.52)
$\delta_{2,2}$	-0.04	(-2.69;2.83)	0.06	(-2.73;2.86)	-0.07	(-2.58;2.66)
$\delta_{3,2}$	0.00	(-2.68;2.80)	0.02	(-2.86;2.90)	-0.02	(-2.89;2.73)
$\delta_{4,2}$	0.03	(-2.82;2.84)	0.01	(-2.91;2.82)	0.00	(-2.78;2.95)
Σ_{11}	1.02	(0.91;1.16)	0.95	(0.87;1.04)	1.94	(1.85;2.20)
Σ_{12}	0.35	(0.21;0.51)	-0.03	(-0.09;0.023)	-0.39	(-0.44;-0.28)
Σ_{22}	4.06	(3.61;4.61)	0.95	(0.87;1.04)	0.97	(0.93;1.10)
r_1	-0.854	$(-0.85; -0.851)$		r_2	0.927	$(0.924; 0.929)$

TABLE 5.12. Point estimation and credible intervals with SSVS method for model 3 with $r_1 = -0.8557$ $r_2 = 0.9257$.

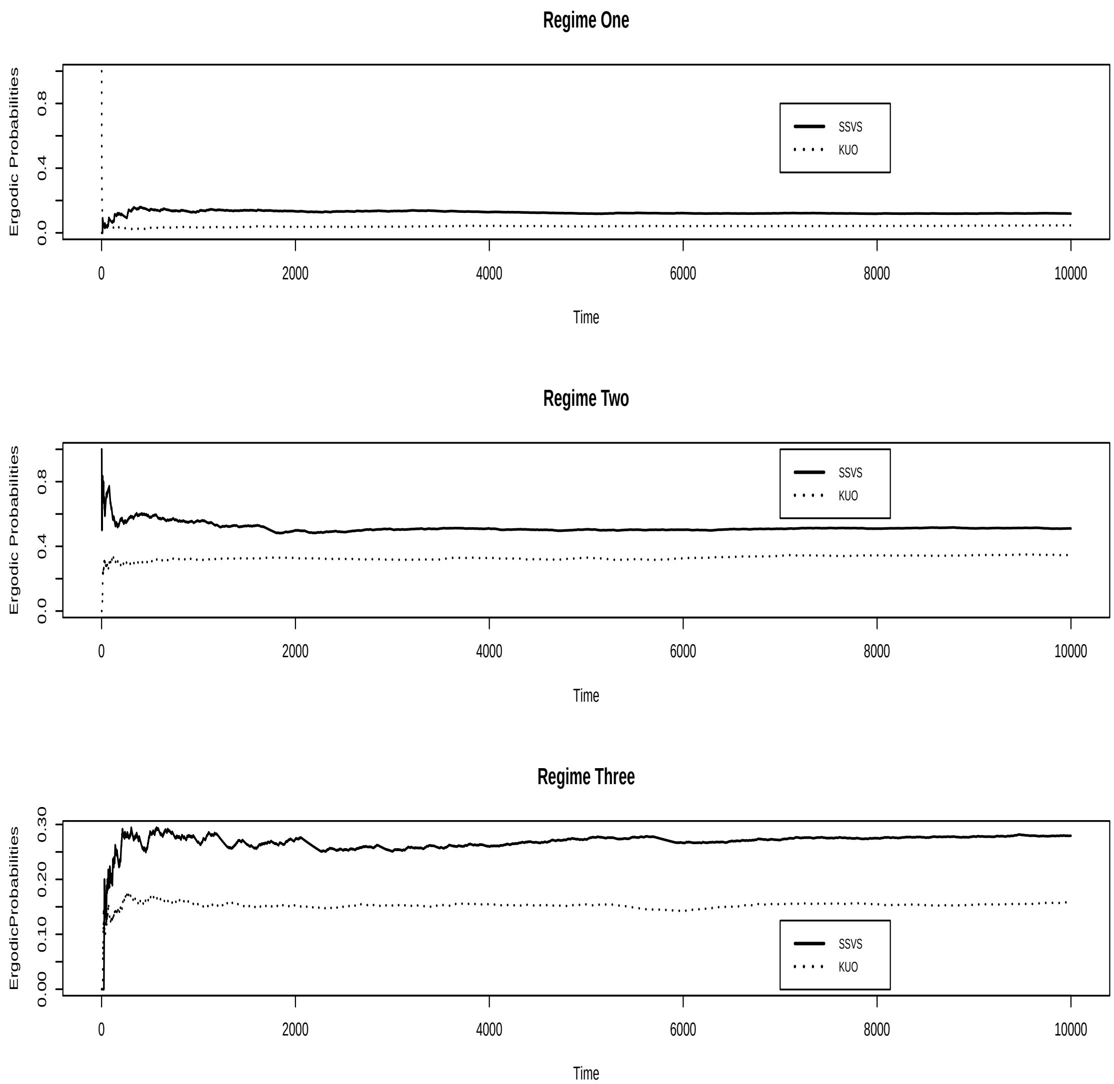

FIGURE 5.5. Ergodic posterior probabilities for the best vector $\boldsymbol{\gamma}$ in the model 3.

and Chib sampler in order to get the posterior probabilities for each model based on 10000 iterations with a burn-in period of 5000 iterations.

In the Tables 5.13-5.16 are shown the results of the marginal likelihood and Carlin and Chib methods in order to identify the number of regimes in MTAR models with two and three regimes. Both methods identify correctly the number of regimes of the simulated model. The convergence of the chain for the model was checked monitoring the probabilities in 3.12 and it was always attained. The convergence is fast when the pseudo prior distributions are close to the posterior distributions. We can see in the Figures 5.6 and 5.7 the behaviour of the last 10000 values of the generated chains for some parameters in MTAR model with two and three regimes using the SSVS method. The behaviour is stable and confirms the convergence of the chains; similar results were found when we used the KUO method. In spite of the accuracy of the two methods to identify the number of regimes of the MTAR model, both methods require moderate computation time for obtaining the identification. The following suggestions are given here in order

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

Table 5.13. Marginal likelihood when model 2 is the true

	Model		
	2 regimes	3 Regimes	4 Regimes
SSVS	1	0	0
KUO	1	0	0

Table 5.14. Posterior probabilities for the Metropolised Carlin and Chib when model 2 is the true

to decrease the computation time for each method, however they have to be investigated as a future research. When the marginal likelihood is being calculated, we have to run the reduced Gibbs sampling many times in order to extract samples of all variables γ's which involves a large expenditure of time. We can decrease the computation time if we are able to extract samples of the full block of the variables γ's, which is possible if we use the methodology developed in (Paroli & Spezia, 2008). In the Metropolised Carlin and Chib method, we have to make a previous run of the Gibbs sampling to extract samples that permits to obtain estimations of the parameters of the pseudo prior distributions, for which it is necessary to spend much time in this previous stage. We can see that it is possible to reduce the computation time if we can set the prior distributions as the pseudo prior distributions. On the other hand, it is important to be careful with the Metropolised Carlind and Chib algorithm for the identification of the number of regimes in small samples, because the identification was not successful in many cases possibly due to pseudo prior distributions chosen in this methodology. However, marginal likelihood methodology was successful in small samples.

5.4 Diagnostic checking

In this section we give some suggestions to check the model adequacy that was chosen with the methodology proposed above. We take some ideas in the works of Hong & Lee (2003), Ling & Li (1997) and Nieto (2005) to propose some diagnostics which confirm the goodness of the fit of the MTAR model. To this end, we define the residuals of the model as follows for $t = \max\{p, q, d\} + 1, \ldots, T$, where p, d, q were defined in Section 2.1 :

$$e_t = \Sigma_{(j)}^{(-1/2)} \left(\mathrm{Y}_t - \left(I_k \otimes \mathrm{w}'_{t,j} \right) \vartheta_j \right) \quad \text{if} \quad r_{j-1} < \mathrm{Z}_t \leq r_j. \tag{5.4}$$

Model	SSVS log(marginal)	KUO log(marginal)
2 Regimes	-5305.064	-5087.263
3 Regimes	-2710.262	-2645.785
4 Regimes	-2787.504	-2703.042

Table 5.15. Marginal likelihood when model 3 is the true

	Model		
	2 regimes	3 Regimes	4 Regimes
SSVS	0	1	0
KUO	0	0.6561	0.3439

Table 5.16. Posterior probabilities for the Metropolised Carlin and Chib when model 3 is the true

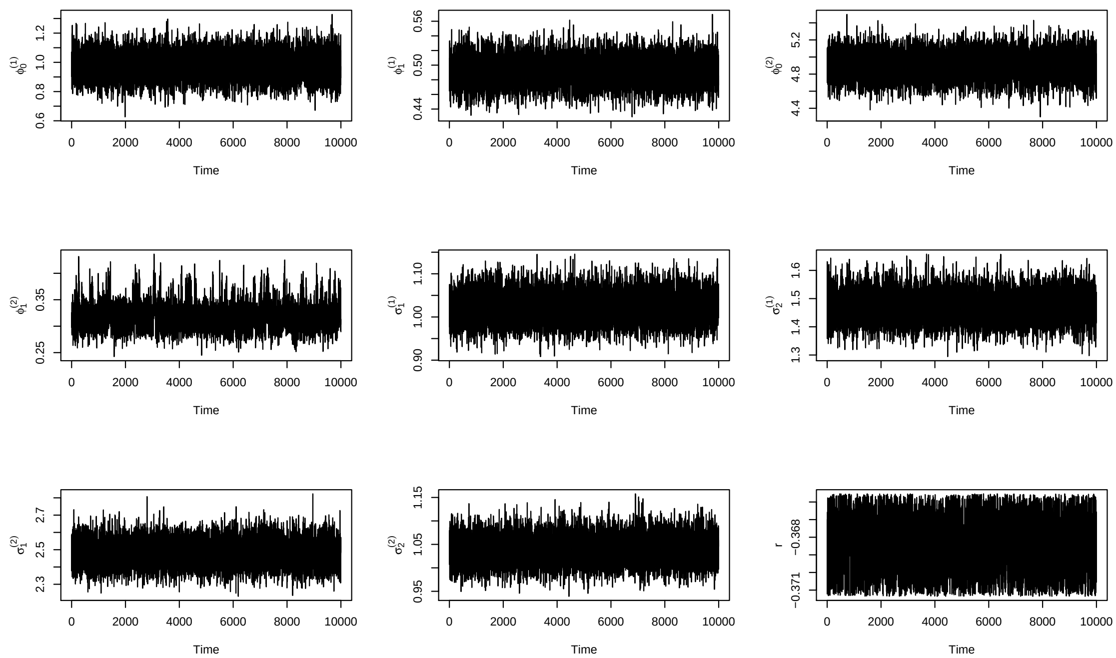

Figure 5.6. Values of the chains for some parameters of the model 2

If there are missing data, they must be replaced for their estimations which makes that the residuals in (5.4) undergo a change, then now the process $\{e_t\}$ is called pseudo residuals. This phenomenon, which was pointed out by Nieto (2005), can affect the distributions of diagnostic tests, then these diagnostics must be considered with caution. It also is a topic for a future research in MTAR models. We use the CUSUM and CUSUM of Squares as diagnostics of the adequacy of the model. We show in the Figures 5.8 and 5.9 the results of the diagnostics for the pseudo residuals in the models with two and three regimes identified above. We can see that the behaviour of the CUSUM and CUSUMSQ graphics is reasonably well which indicate us that the data have been fitted correctly.

5.5 Forecasting

In this section we make a simulation study in order to illustrate the results of the Chapter 4 and verify the behaviour of the forecasts. To achieve this goal we add the following new MTAR model with two regimes:

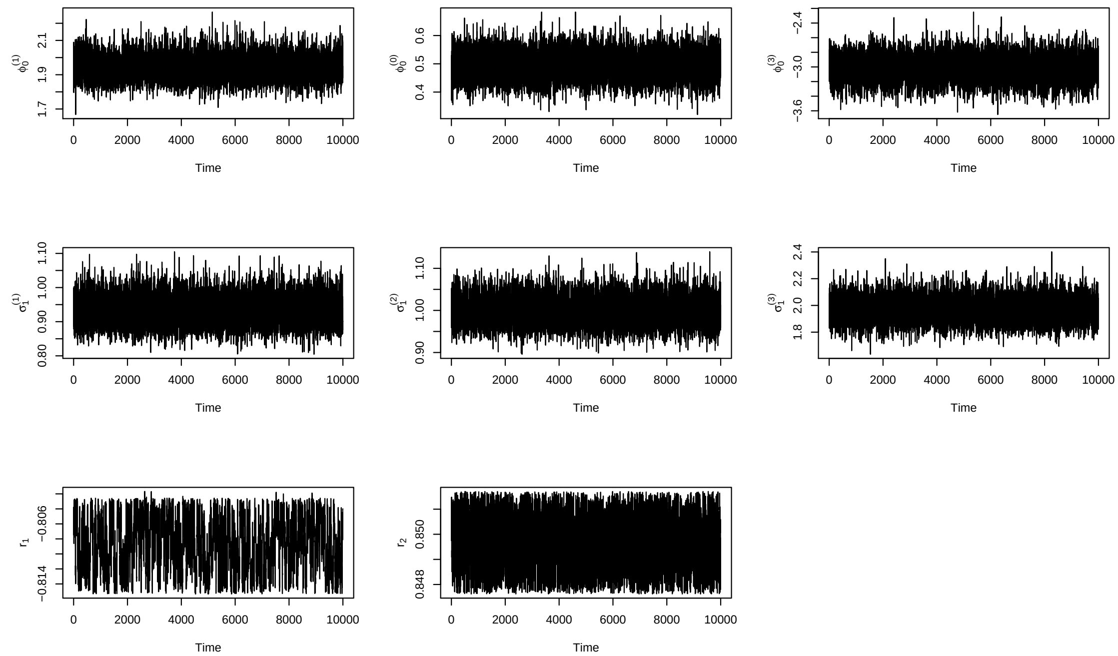

FIGURE 5.7. Values of the chains for some parameters of model 3.

Model 4.

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

where $\{\mathrm{U}_t\}$ follows the model 1. That model has less influence of the exogenous variables and its variability is less than the exhibited for the model 2. We simulate a realization of size $1000 + h$ of the models 2, 3 and 4, 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 root square of trace of the variance of the predictive distribution (RVPD), that is the covariance matrix of prediction error, for checking the performance of the forecast. The Table 5.17 gives us the results of the forecasting for the proposed model 4, while Tables 5.18 and 5.19 give us the summary of the forecasting results for the models 1 and 3 proposed at the beginning of the chapter with $h = 10$ and last 10000 iterations. We can see that all the true values lie in the 95% individual credible intervals and many of the forecasts are close to them. It is important to point out that the forecasts for the model 4 proposed in this section have RVPD less than the forecasts for the model 2, this may be due to less influence of the exogenous variables and less heteroscedasticity of the new proposed model. Another feature of the forecasting in MTAR models is that RVPD increases as the forecast horizon increases. We can establish that the method proposed to forecast MTAR models behaves appropriately.

It is important to point out that in the current Chapter we used different values to initialize the chains and the parameters of the prior distributions. The results in each case were

h	y_{1000+h}	$\widehat{y}_{1000+h}$	RVPD	95% C.I.	u_{1000+h}	$\widehat{u}_{1000+h}$	RVPD	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 5.17. Forecasting output and exogenous vectors for model 4.

h	y_{1000+h}	$\widehat{y}_{1000+h}$	RVPD	95% C.I.	u_{1000+h}	$\widehat{u}_{1000+h}$	RVPD	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 5.18. Forecasting output and exogenous vectors for model 2.

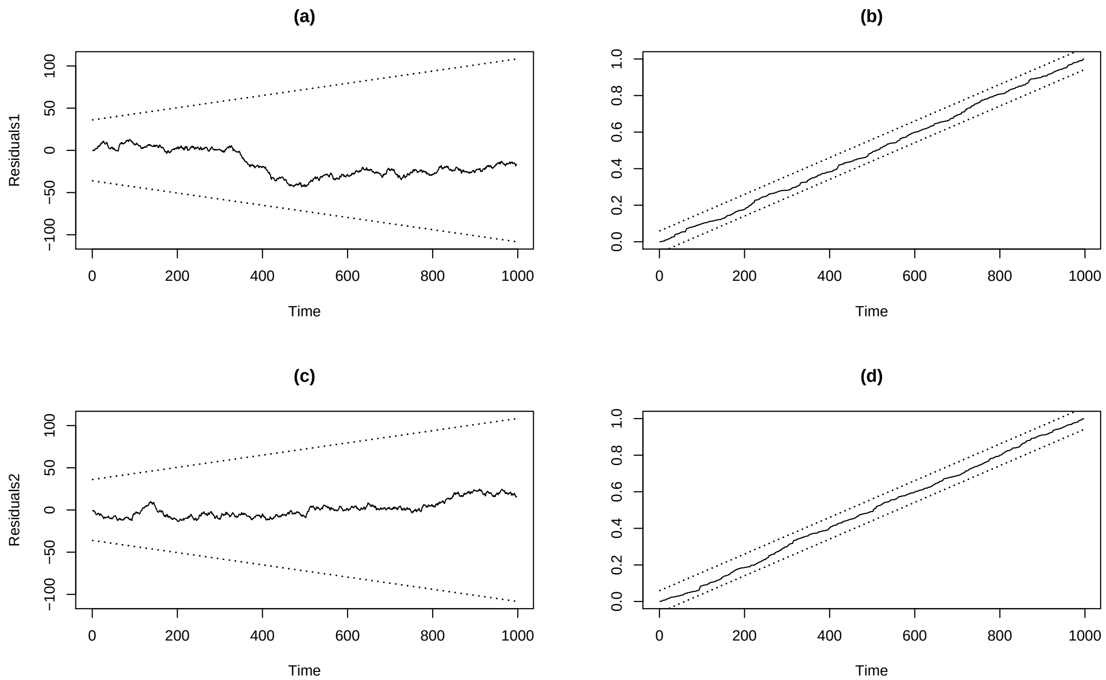

FIGURE 5.8. (a),(c) CUSUM statistic for residuals of model 2. (b),(d) Corresponding CUSUMSQ statistic.

similar to the obtained in the current chapter and are not presented here. All procedures were implemented in R package (R Core Team, 2013).

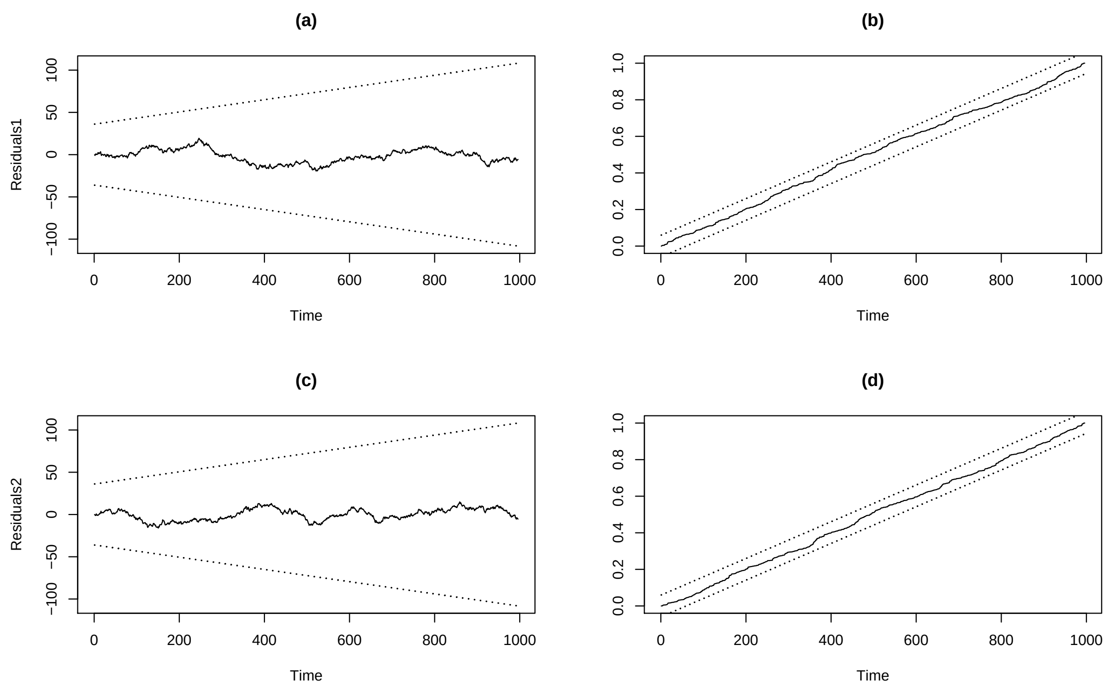

FIGURE 5.9. (a),(c) CUSUM statistic of residuals of model 3. (b),(d) Corresponding CUSUMSQ statistic.

h	y_{1000+h}	$\widehat{y}_{1000+h}$	RVPD	95% C.I.	u_{1000+h}	$\widehat{u}_{1000+h}$	RJMSE	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 5.19. Forecasting output and exogenous vectors for model 3.

CHAPTER 6

Application

In this section we employ the developed methodology to hydrological data. We try to find the relationship between diary rainfall (in mm) and the diary river flow (in m^3/s) of two rivers where a river empties into the other one in a region of department of Cauca in Colombia. The rainfall was measured in the San Juan's meteorological station with an altitude of 2400 meters and geographical coordinates 2° 2' 7.1"north and 76° 29' 47.1"west. The first flow river was measured in the El Trebol's hydrological station of the Bedon river with altitude of 1720 meters and geographical coordinates 2° 15' 0.1"north and 76° 7' 42.6"west; the second flow river was measured on La Plata river in the Villalosada's hydrological station with an altitude of 1300 meters 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 to control for hydrological/meteorological factors, which may distort the kind of dynamical relationship explained by the MTAR model, this fact was mentioned in (Nieto, 2005) 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 Bedon river and 213 in the series of the river flow of La Plata river. These data were provided by IDEAM, the official Colombian agency for hydrological and meteorological studies. We can see in the Figure 6.1 the time series of the variables proposed for the analysis, these figures show us a strong relationship of the rainfall with the river flows, and a stable behaviour in mean with some bursts of large values which should be 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. Now, we need to check the non-linearity of $\{\widetilde{\mathrm{Y}}_t\}$ caused by Z_t using the non-linearity test proposed by Tsay (1998). This test is briefly described. Consider the null hypothesis that $\{\mathrm{Y}_t\}$ is linear versus the alternative hypothesis that it follows the multivariate threshold model given in 1.1. Based on the observations $\mathrm{y}_t, \mathrm{x}_t$ and z_t for $t = 1, \cdots, T$, and for p, q, d known, the arranged regression is employed to implement the test. Let $h = \max\{p, d, q\}$, $\mathbf{W}_t = (1, \mathrm{y}'_1, \ldots, \mathrm{y}'_p, \mathrm{x}'_1, \ldots, \mathrm{x}'_q, \mathrm{z}_1, \ldots, \mathrm{z}_d)$ a

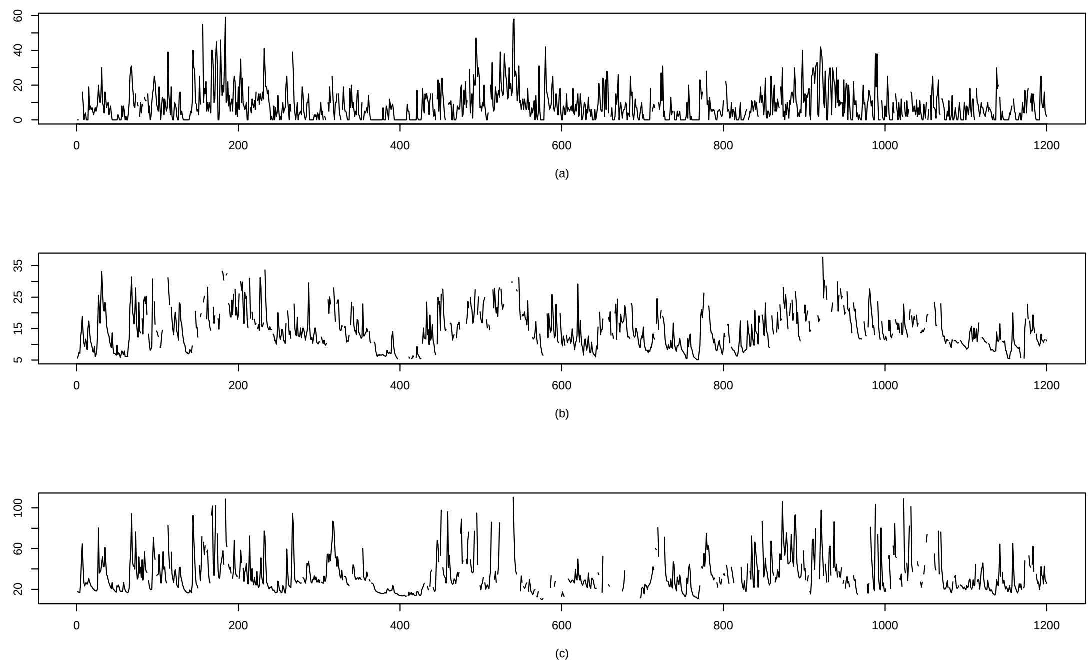

FIGURE 6.1. Time series of real application: (a) Rainfall; (b) River flow of Bedon's river; (c) River flow of La Plata's river

vector of regressors variables, and $\mathbf{\Phi}$ an unknown matrix. If the null hypothesis holds, the MTAR model collapses to

$$\mathrm{Y}'_t = \mathbf{W}'_t\mathbf{\Phi} + \epsilon'_t \tag{6.1}$$

for $t = h, \ldots, T$. Let $S = \{\mathrm{z}_h, \ldots, \mathrm{z}_T\}$ be the set of values of Z_t. Consider the order statistics of S and denote its ith smallest element by $\mathrm{z}_{(i)}$. Then, the arranged regression based on the increasing order of the threshold variable Z_t is

$$\mathrm{Y}'_{t(i)} = \mathbf{W}'_{t(i)}\mathbf{\Phi} + \epsilon'_{t(i)} \tag{6.2}$$

for $i = 1, \ldots, T-h$. Let $\hat{\mathbf{\Phi}}_m$ be the least square estimator of $\mathbf{\Phi}$ of the regression in 6.2 based on the first m observations, in other words, the regression associated with the m smallest values of S. Let

$$\hat{\mathbf{e}}_{t(m+1)} = \mathrm{Y}'_{t(m+1)} - \hat{\mathbf{\Phi}_m}\mathbf{W}_{t(m+1)}$$

and

$$\hat{\boldsymbol{\eta}}_{t(m+1)} = \hat{\mathbf{e}}_{t(m+1)}/[1 + \mathbf{W}'_{t(m+1)}V_m\mathbf{W}_{t(m+1)}]^{1/2}$$

where $V_m = \left[\sum_{i=1}^{m}\mathbf{W}_{t(i)}\mathbf{W}'_{t(i)}\right]^{-1}$, be the predictive residual and the standardized predictive residual of the arranged regression 6.2. These quantities can be obtained by the recursive least square algorithm. Next, consider the regression

$$\hat{\boldsymbol{\eta}}'_{t(l)} = \mathbf{W}'_{t(l)}\mathbf{\Psi} + \varepsilon'_{t(l)} \tag{6.3}$$

for $l = m_0+1, \ldots, T-h$, where m_0 denotes the starting point of the recursive least squares estimation. Here it is necessary to test the null hypothesis $\mathbf{\Psi} = \mathbf{0}$ versus the alternative $\mathbf{\Psi} \neq \mathbf{0}$. It is important to point out that in the original Tsay's (1998) test the delay parameter was took into account, and here it is considered as zero. Then, the proposed test statistic is

$$C(0) = [T - h - m_0 - (kp + vq + d + 1)][\ln |S_0| - \ln |S_1|] \tag{6.4}$$

where $|M|$ denotes the determinant of the matrix A,

$$S_0 = \frac{1}{T - h - m_0} \sum_{l=m_0+1}^{T-h} \hat{\boldsymbol{\eta}}_{t(l)} \hat{\boldsymbol{\eta}}'_{t(l)}$$

and

$$S_0 = \frac{1}{T - h - m_0} \sum_{l=m_0+1}^{T-h} \hat{\boldsymbol{\varepsilon}}_{t(l)} \hat{\boldsymbol{\varepsilon}}'_{t(l)}$$

where $\hat{\boldsymbol{\varepsilon}}_{t(l)}$ is the least squares residual of regression 6.3. Under of null hypothesis that $\{Y_t\}$ is linear, Nieto & Hoyos (2011) showed by simulation, that $C(0)$ has a distribution chi-square with with $k(kp + vq + d + 1)$ degrees freedom based on model of Tsay (1998). Nevertheless, the time series of river flow and rainfall that are analysed here have missing data, and then the non-linearity test can not be applied directly. To tackle this problem, Nieto & Hoyos's (2011) ideas are again used to implement the non-linearity test in presence of missing data. To this end, it is important to note that the output process is linear under null hypothesis; it also is assumed that the threshold process is linear, therefore the fixed-point smoother algorithm can be applied to estimate and complete the missing data. If threshold process is non-linear but it is a Markov chain, it is possible to estimate the missing data in the time series $\{\mathrm{u}_t\}$ following the ideas in (Nieto & Hoyos, 2011). Hence the equations 2.8 and 2.9 are reduced to

$$p(\mathrm{u}_1, \cdots, \mathrm{u}_b | \mathrm{u}_{-1:b}, \mathrm{y}_{1:T}, \alpha_{0:T}) \propto p(\mathrm{u}_1, \cdots, \mathrm{u}_b) \prod_{t=b+1}^{2b} p(\mathrm{u}_t | \mathrm{u}_{t-1}) \tag{6.5}$$

and

$$p(\mathrm{u}_t | \mathrm{u}_{-t}, \mathrm{y}_{1:T}, \alpha_{0:T}) \propto \prod_{s=t}^{t+b} p(\mathrm{u}_s | \mathrm{u}_{s-b:s-1}) \tag{6.6}$$

for $t = b + 1, \ldots, T$. With the completed series, the non-linearity test is applied as in the case of complete data if the percentage of missing data in less than 20%. For this application, the percentage of missing data is less than 20%, then we proceed as in the case of complete data. The missing data were estimated using the fixed-point smoother algorithm to estimate missing data in $\{\widetilde{\mathbf{y}}_t\}$ time series, and the reduced equations 6.5 and 6.6 to estimate missing data in $\{\mathrm{z}_t\}$ time series. The Table 6.1 give us the statistic $\widehat{C}(0)$, quantile $\chi^2_{(\cdot,\alpha)}$ and $p-$value for non-linearity test with two possible $VAR(p)$ process $p = 1$ and $p = 5$ as null hypothesis and $\alpha = 0.05$ We can see that null hypothesis is rejected for both $VAR(p)$ processes. Therefore, the results are a signal of the strong threshold non-linearity of $\{\widetilde{\mathbf{Y}}_t\}$, which is explained by $\{\mathrm{Z}_t\}$. Now it is possible to start the Bayesian analysis of the MTAR model.

p	$\widehat{C}(0)$	$\chi^2_{(\cdot,\alpha)}$	$p-$value
1	149.686	12.591	$9.014e^{-30}$
5	91.249	33.924	$2.098e^{-10}$

TABLE 6.1. Non-linearity test results.

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

TABLE 6.2. Thresholds for the models considered

In the first part of the analysis, we completed the series with the medians of each variable and apply the methodology developed to estimate the parameters of the MTAR model. The estimation of the threshold for the proposed models with two, three and four regimes are given in the Table 6.2, they were found using NAIC. Based on the thresholds, we calculate the marginal likelihood as was implemented in the Chapter 3 in order to identify the number of regimes. SSVS method was implemented to identify the autoregressive orders with prior values $c_{i,j} = 2.0$ and $\tau_{i,j} = 0.01$ and maximum autoregressive orders $p = d = 5$. The results for marginal likelihood are shown in the Table 6.3. We can see that the procedure suggest a model with two regimes for the data. Table 6.4 give us the autoregressive orders for model with two regimes. We have identified the structural parameters by means of the proposed methodology, aditionally we have also obtained the non-structural parameters. With the estimation of the parameters of the model, we proceed to estimate the missing data following the first part of the Chapter 2, then we fill the missing data with their estimation and run again the last procedure. To implement the estimation of the missing data we use the ideas in (Nieto, 2005) due to 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 depens on z_{t-1} with Lebesgue-measure density $g(z_t|z_{t-1})$ that the support on $(0, \infty)$. The approxi-

Model	Log(marginal)
2 regimes	3895.979
3 regimes	1476.012
4 regimes	1186.459

TABLE 6.3. Identification of number of regimes for real data

Regime 1	Regime 2
$p_1 = 5$	$p_2 = 5$
$q_1 = 4$	$q_2 = 5$

TABLE 6.4. Identification of autoregressive orders for real data

mation 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 large enough 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 estimation which is $\hat{p} = 0.23$; we consider that $g(z)$ is a normal density truncated at $z = 0$. A numerical procedure is used to estimate the mean

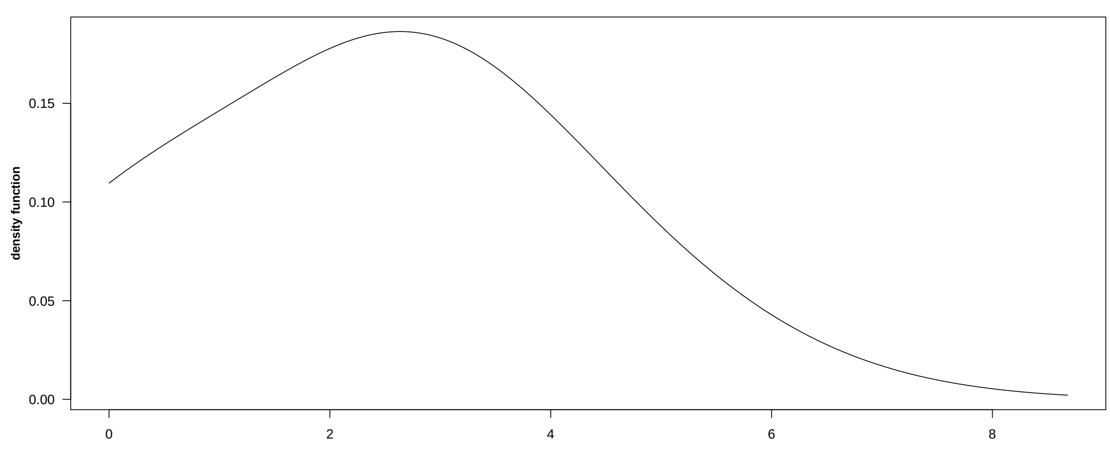

FIGURE 6.2. Density estimate throughout kernel Gaussian.

and standard deviation of density $g(z)$, it is based on a non-parametric approximation with Gaussian Kernel see Figure 6.2. The maximum of the estimated kernel is $\hat{\mu} = 2.64$ and the inflexion point must have the form $\hat{\mu} + \hat{\sigma}$ that can be found using the second derivative of the estimated kernel, Figure 6.3. We can observe a inflexion point located in 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 $j = 1, \cdots, l$ 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 $l = 2$ 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 $\hat{\sigma} = 1.885$. We use the independent Metropolis-Hastings algorithm to extract samples of the posterior densities for the estimation of the missing data in the threshold variable; the proposal density considered here is given by $ph_n(z_t) + (1-p)u_{(0,m_z)}(z_t)$ 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 varies between 12 and 34 percent and the convergence was checked using the non-parametric Kolmogorov test with different batches between 10 and 20 and last 10000 iterations with a burn-in period of 5000 iterations, we could see that all chains converge to the stationary distribution. In this context the estimator used

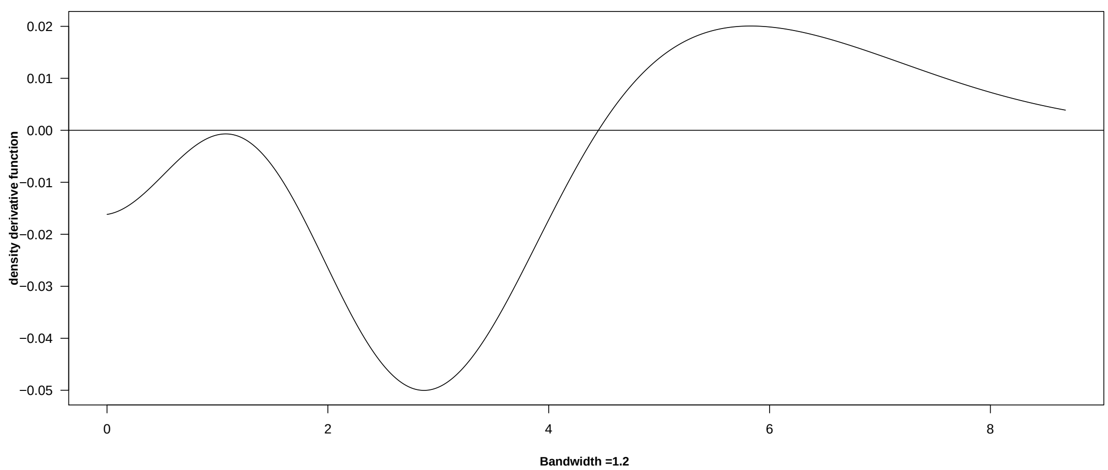

FIGURE 6.3. Second derivative of estimated density for threshold variable.

for the estimation of the missing data in the threshold variable is the median, because the posterior distribution is skewed and takes only positive values, therefore if the real missing data was zero, we would obtain zero as estimation using the median instead of the average. Now, we can replace the previous estimation of missing data with current estimates and repeat the procedure for estimating the structural and non-structural parameters of the MTAR model. The estimation of the thresholds for each proposed model and the results for marginal likelihood are in the Tables 6.5 and 6.6. In this case, the number of regimes obtained upon completion of the procedure is the same as in the previous analysis $\hat{l} = 2$, although the thresholds found differ somewhat from the above. The autoregressive orders based on $\gamma's$ variables are shown in the Table 6.7 and with this, the identification of the structural parameters of the MTAR model is completed. Finally, the estimation of the non-structural parameters of the NMTAR model are shown in Table 6.8. Now, in order to verify the adequacy of the model, we checked the residuals of the model based on 99% Cusum

Model	Thresholds	NAIC
2 regimes	2.6	8.371
3 regimes	2.0 3.2	7.553
4 regimes	0.0 2.4 3.2	6.9810

TABLE 6.5. Thresholds for the models considered

Model	Log(marginal)
2 regimes	4279.243
3 regimes	1837.574
4 regimes	1491.498

TABLE 6.6. Identification of number of regimes for real data

and CusumSQ plots, see Figure 6.4. The diagnostic checking plots for the model with 2 regimes have a satisfactory behaviour with a moderate heteroscedasticity exhibited in the CusumSQ plots for both components of the residuals. Therefore, it possible to conclude that the model was fitted adequately. It is important to point out that distribution of the residuals has heavy tails and a better fit can be done using innovations with that characteristic, therefore an interesting research topic can be carried out if we consider innovations with heavy-tailed and biased distributions in MTAR models. Based on the

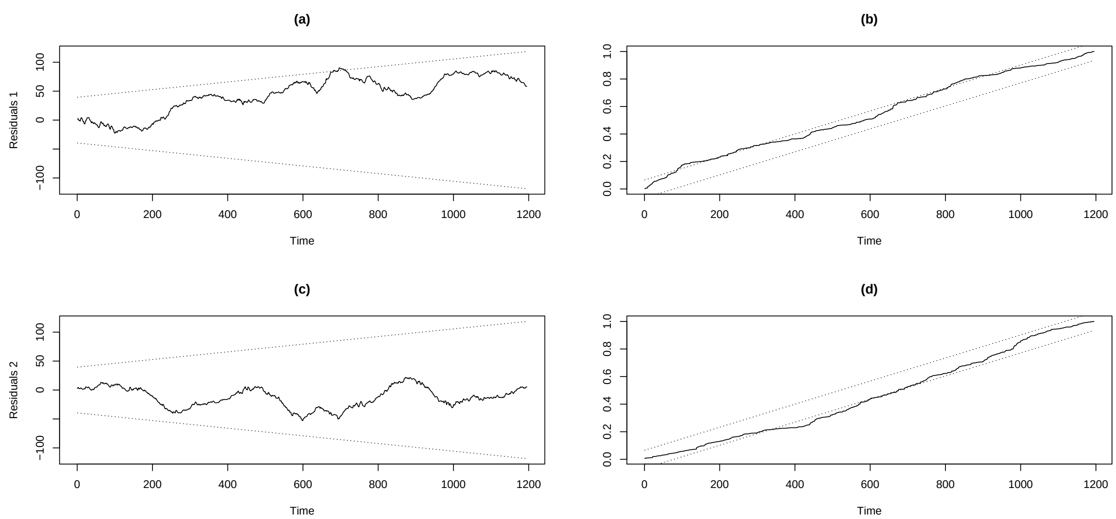

FIGURE 6.4. (a),(c) Cusum of residuals; (b),(d) CusumSQ of residuals for the model with 2 regimes.

estimation of the non-structural parameters we can see interesting facts, for instance there is a feedback relation between two rivers, this relation can be possible from the Bedon flow river to La Plata flow river but no in the other direction. This feedback feature tells us that it is possible to add any other variables to explain better the river flows. Another fact is that in the second regime the variances are larger than the first regime, it indicates us that the transformed rainfall produces more variability in the transformed flow river. The last feature that we can see is that the impact of the precipitation in the second regime is bigger than the first regime in many cases (see parameters $\delta_{i,s}^{(j)}$, for $j = 1, 2; i = 1, \cdots, 5; s = 1, 2$ in Table 6.8) or in estimated Model 5.

Regime 1
$p_1 = 5, q_1 = 4$
$\boldsymbol{\gamma}_1 = (11111111111110001111111111111101)$ with probability 0.0599
Regime 2
$p_2 = 5, q_2 = 5$
$\boldsymbol{\gamma}_2 = (11111111111101011111111111111101)$ with probability 0.0872

TABLE 6.7. Identification of autoregressive orders for the real data example

Model 5.

$$
\mathrm{Y}_t = \begin{cases}
\begin{bmatrix}0.060\\0.112\end{bmatrix} + \begin{bmatrix}0.164 & 0.073\\0.073 & 0.224\end{bmatrix}\mathrm{Y}_{t-1} + \begin{bmatrix}0.112 & 0.061\\0.046 & 0.141\end{bmatrix}\mathrm{Y}_{t-2} + \begin{bmatrix}0.112 & 0.061\\0.046 & 0.141\end{bmatrix}\mathrm{Y}_{t-3} + \\
\begin{bmatrix}0.084 & 0.024\\0.036 & 0.121\end{bmatrix}\mathrm{Y}_{t-4} + \begin{bmatrix}0.094 & 0.039\\0.039 & 0.112\end{bmatrix}\mathrm{Y}_{t-5} + \begin{bmatrix}0.022\\0.013\end{bmatrix}\mathrm{Z}_{t-1} + \begin{bmatrix}0.005\\0\end{bmatrix}\mathrm{Z}_{t-2} + \\
\begin{bmatrix}0\\-0.010\end{bmatrix}\mathrm{Z}_{t-3} + \begin{bmatrix}-0.002\\-0.019\end{bmatrix}\mathrm{Z}_{t-4}\begin{bmatrix}0.097 & 0.040\\0.049 & 0.130\end{bmatrix}\varepsilon_t, & \text{if } \mathrm{Z}_t \leq 2.6 \\
\\
\begin{bmatrix}0.083\\0.101\end{bmatrix} + \begin{bmatrix}0.127 & 0.076\\0..083 & 0.187\end{bmatrix}\mathrm{Y}_{t-1} + \begin{bmatrix}0.105 & 0.056\\0.060 & 0.133\end{bmatrix}\mathrm{Y}_{t-2} + \begin{bmatrix}0.079 & 0.059\\0.067 & 0.125\end{bmatrix}\mathrm{Y}_{t-3} + \\
\begin{bmatrix}0.081 & 0.079\\0.069 & 0.117\end{bmatrix}\mathrm{Y}_{t-4} + \begin{bmatrix}0.083 & 0.076\\0.069 & 0.136\end{bmatrix}\mathrm{Y}_{t-5} + \begin{bmatrix}0.032\\0.036\end{bmatrix}\mathrm{Z}_{t-1} + \begin{bmatrix}0\\-0.006\end{bmatrix}\mathrm{Z}_{t-2} + \\
\begin{bmatrix}0\\-0.009\end{bmatrix}\mathrm{Z}_{t-3} + \begin{bmatrix}0\\0\end{bmatrix}\mathrm{Z}_{t-4} + \begin{bmatrix}-0.011\\-0.023\end{bmatrix}\mathrm{Z}_{t-5} + \begin{bmatrix}0.120 & 0.040\\0.040 & 0.186\end{bmatrix}\varepsilon_t, & \text{if } \mathrm{Z}_t > 2.6
\end{cases}
$$

The final estimates of some missing data are shown in Table 6.9, we can see that none of credible intervals in special the related with output vector take negative values in spite of the Gaussian distribution of the innovations. The forecasting for both threshold variable and output vector was carried out for 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 RVPD are shown in the Table 6.10 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 hydrology application.

	Regime 1		Regime 2	
Parameter	Estimation	Credible Interval 95%	Estimation	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)
$\phi_{4,22}$	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)
$\phi_{5,22}$	0.112	(0.081;0.144)	0.136	(0.102;0.171)
$\delta_{1,2}$	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)
Σ_{11}	0.093	(0.087;0.098)	0.120	(0.114;0.128)
Σ_{12}	0.015	(0.010;0.019)	0.040	(0.033;0.047)
Σ_{22}	0.104	(0.098;0.111)	0.186	(0.175;0.198)

TABLE 6.8. Estimation of the non-structural parameters for the real data with $\hat{r} = 2.6$.

t	$\widehat{y}_t$	C.I. 95%	t	$\widehat{z}_t$	C.I. 95%
	1.322	[1.132; 1.516]			
80	− − −−	− − −−	3	0	[0; 1.887]
	1.365	[1.163; 1.570]			
112	1.737	[1.537; 1.944]	74	2.695	[0; 5.186]
	1.677	[1.417; 1.941]			
183	2.023	[1.654; 2.401]	155	5.961	[3.876; 7.780]
	0.907	[0.714; 1.097]			
398	− − −−	− − −−	309	0	[0; 2.910]
	0.960	[0.700; 1.212]			
409	1.293	[0.970; 1.636]	328	0	[0; 2.340]
	1.510	[1.261; 1.755]			
657	1.781	[1.460; 2.125]	458	0.308	[0; 4.378]
	1.621	[1.362; 1.887]			
724	2.049	[1.673; 2.440]	511	1.674	[0; 4.895]
	− − −−	− − −−			
845	1.821	[1.509; 2.156]	718	4.282	[2.256; 6.762]
	1.368	[1.166; 1.570]			
1042	1.801	[1.589; 2.018	844	0	[0; 4.102]
	1.260	[1.070; 1.459]			
1198	− − −−	− − −−	1179	3.138	[0; 5.731]

TABLE 6.9. Missing Data Estimation for real Data. − means that component is not missing.

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

TABLE 6.10. Forecasting of output and threshold variables for the real data.

Conclusions

In this thesis, a methodology was developed to analyse multivariate time series with missing data that follow a multivariate threshold autoregressive(MTAR) model with exogenous variables. A modification of the original Tsay's (1998) model was considered here, where lags of threshold variable appear as regressors. The following steps are considered here to analyse time series via MTAR models: identification of threshold values, autoregressive orders and number of regimes of the model (structural parameters); estimation of the autoregressive and covariance matrices (non-structural parameters); exploratory diagnostic checking with the residuals of model; missing data estimation and forecasting procedure.

The Bayesian approach and the MCMC techniques were employed to the identification and estimation of the parameters. Specifically, the Metropolised Carlin and Chib and marginal likelihood methodologies were used to identify the number of regimes. The SSVS and KUO variable selection strategies were adapted to identify the autoregressive orders indirectly. The threshold values were identified following one of two ways: the first way was joint estimation with other parameters of the model extracting samples of its posterior distribution; the second way was to search the threshold values via NAIC as a previous step of the Bayesian estimation. The autoregressive and covariance matrices were estimated intrinsically into the global Bayesian procedure. It was necessary only one step for the identification and estimation of the parameters of the MTAR model. The properties of the state space models were exploited to estimate the missing data in the output time series. However, it was necessary to propose a smoother to estimate missing data in the exogenous and threshold time series using the idea of single-move Gibbs sampling. The estimation of the missing data and non-structural parameters was carried out jointly conditional on the structural parameters. A forecasting procedure for the output vector was developed incorporating the uncertainty of the non-structural parameters, autoregressive orders and threshold values conditional on the number of regimes. Additionally, the forecasting for the covariates is obtained as an intrinsic step of forecasting procedure.

Some research problems for the future were detected during the development of this thesis: (1) The parameters of the exogenous variables could be estimated jointly with the parameters of the MTAR model using the Bayesian approach. (2) Investigate the distribution of the statistics for the diagnostic checking in presence of missing data. (3) Develop a similar methodology when the distribution of the innovations is not Gaussian. (4) The delay can be considered as parameter to be identified with other parameters of the MTAR model. (5) Cointegration relationship in MTAR models might be identified and estimated.

Bibliography

Bovas, A. & Thavaneswaran, A. (1991). A nonlinear time series model and estimation of missing data., *Annals of the Institute of Statistical Mathematics* **43**(3): 493–504.

Campbell, E. (2004). Bayesian selection of threshold autoregressive models., *Journal of Time Series Analysis* **25**(4): 467–482.

Carlin, B. P. & Chib, S. (1995). Bayesian Model Choice via Markov Chain Monte Carlo Methods., *Journal of the Royal Statistical Society. Series B (Methodological)* **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., *Journal of the American Statistical Association* **87**(418): 473–484.

Carter, C. & Konh, R. (1994). On Gibbs Sampling for State Space Models., *Biometrika* **81**(3): 541–553.

Chen, C. & Lee, J. (1995). Bayesian inference of threshold autoregressive models., *Journal of Time Series Analysis* **16**(5): 483–492.

Chen, C. W. S., Liu, F. & Gerlach, R. (2011). Bayesian subset selection for threshold autoregressive moving-average models., *Computational Statistics* **26**(1): 1–30.

Chib, S. (1995). Marginal Likelihood from the Gibbs Output., *Journal of the American Statistical Association* **90**(432): 1313–1321.

Chib, S. & Jeliazkov, I. (2001). Marginal Likelihood from the Metropolis-Hastings Output., *Journal of the American Statistical Association* **96**(453): 270–281.

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

Frhüwirth-Schnatter, S. (1995). Data augmentation and dynamic linear models., *Journal of Time Series Analysis* **15**(2): 183–202.

George, E. & McCulloch, R. (1993). Variable Selection Via Gibbs Sampling., *Journal of the American Statistical Association* **8**(423): 881–889.

Green, P. (1995). Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination., *Biometrika* **82**(4): 711–732.

Hansen, B. (2011). Threshold autoregression in economics., *Statistics and Its Interface* **4**(2): 123–127.

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

Hong, T. & Lee, T. (2003). Diagnostic Checking for the Adequacy of Nonlinear Time Series Models., *Econometric Theory* **19**(6): 1065–1121.

Kuo, L. & Mallick, B. (1998). Variable Selection for Regression Models., *Sankhyā: The Indian Journal of Statistics, Series B* **60**(1): 65–81.

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

Ling, S. & Li, W. (1997). Diagnostic checking of nonlinear multivariate time series with multivariate arch errors., *Journal of Time Series Analysis.* **18**(5): 447–464.

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.

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., *Communications in Statistics - Simulation and Computation* **42**(4): 814–840.

O'Hara, R. & Sillanpaa, M. (2009). A Review of Bayesian Variable Selection Methods: What, How and Which., *Bayesian Analysis* **4**(1): 85–118.

Paroli, R. & Spezia, L. (2008). Bayesian Variable Selection in Markov Mixture Models., *Communications in Statistics-Simulation and Computation* **37**(1): 25–47.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *http://www.R-project.org/*

Robert, C. & Casella, G. (2004). *Monte Carlo Statistical Methods.*, Springer texts in Statistics, Springer Science+Business Media Inc., 233 Spring Street, New York, NY 10013, USA.

Safadi, T. & Morettin, P. (2000). Bayesian Analysis of Threshold Autoregressive Moving Average Models., *Sankhyā: The Indian Journal of Statistics, Series B.* **62**(3): 353–371.

So, M. & Chen, C. W. S. (2003). Subset Threshold Autoregression., *Journal of Forecasting* **22**(1): 49–66.

Thavaneswaran, A. & Bovas, A. (1991). Estimation of multivariate non-linear time series models, *Journal of Statistical Planning and Inference* **29**(3): 351–363.

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

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

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.