Revista Colombiana de Estadística
January 2015, Volume 38, Issue 1, pp. 239 to 266
DOI: http://dx.doi.org/10.15446/rce.v38n1.48813

TAR Modeling with Missing Data when the White Noise Process Follows a Student's t-Distribution

Modelamiento TAR con datos faltantes cuando el proceso del ruido blanco tiene una distribución t de Student

Hanwen Zhang[1,a], Fabio H. Nieto[2,b]

[1]Facultad de Estadística, Universidad Santo Tomás, Bogotá, Colombia
[2]Departamento de Estadística, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia

Abstract

This paper considers the modeling of the threshold autoregressive (TAR) process, which is driven by a noise process that follows a Student's t-distribution. The analysis is done in the presence of missing data in both the threshold process $\{Z_t\}$ and the interest process $\{X_t\}$. We develop a three-stage procedure based on the Gibbs sampler in order to identify and estimate the model. Additionally, the estimation of the missing data and the forecasting procedure are provided. The proposed methodology is illustrated with simulated and real-life data.

***Key words*:** Bayesian Statistics, Gibbs Sampler, Missing Data, Forecasting, Time Series, Threshold Autoregressive Model.

Resumen

En este trabajo consideramos el modelamiento de los modelos autoregresivos de umbrales (TAR) con datos faltantes tanto en la serie de umbrales como la serie de interés cuando el proceso del ruido blanco sigue una distribución t de student. Desarrollamos un procedimiento de tres etapas basado en el muestreador de Gibbs para identificar y estimar el modelo, además de la estimación de los datos faltantes y el procedimiento para el pronóstico. La metodología propuesta fue aplicada a datos simulados y datos reales.

***Palabras clave*:** datos faltantes, estadística Bayesiana, modelo autoregresivo de umbrales, muestreador de Gibbs, pronóstico, series de tiempo.

[a]Professor. E-mail: hanwenzhang@usantotomas.edu.co
[b]Professor. E-mail: fhnietos@unal.edu.co

1. Introduction

TAR models proposed by Tong (1978) assume that the values of a process $\{Z_t\}$ (the threshold process) determine not only the values of the process of interest $\{X_t\}$, but also its dynamics. When the threshold process is the same process of interest but is lagged, the model is known as SETAR (Self-exciting TAR, Briñez & Nieto, 2005). Nieto (2005) developed a Bayesian methodology for the identification and estimation of TAR models allowing missing data in the both threshold process and the process of interest. On the other side, Nieto (2008, 2011) characterized in univariate TAR models in terms of their mean, conditional mean, variance, conditional variance and also found the expressions for the best predictor. Vargas (2012) improved the prediction with TAR models, taking into account the variability in the parameters and Nieto & Moreno (2013) explored three kinds of conditional variance in order to try to compare this type of nonlinear models with GARCH models. Also the TAR models can be easily extended to threshold autoregressive moving average (TARMA) models, and its Bayesian modeling with two regimes has been investigated in Sáfadi & Morettin (2000) and Xia, Liu, Pan & Liang (2012). Another extension of the TAR model is when there are two threshold variables instead of one and this is done by Chen, Chong & Bai (2012) within the particular case of two regimes.

Despite TAR models usefulness, they are not easily identified due to the large number of parameters and the nesting structure between the parameters. For SETAR models, Tsay (1989) provided a simple and widely applicable model-building procedure. However, generally speaking, most of the parameters, as the thresholds and the number of regimes, are assumed to be known; otherwise they can be identified using Tong (1990)'s NAIC criterion together with some graphical techniques. Assuming that the noise process is Gaussian, Nieto (2005) developed a Bayesian procedure in order to identify the number of regimes and estimate the other parameters, once the thresholds are identified, using NAIC criterion for each possible number of regimes. This work is accomplished in the presence of missing data in both the process of interest and the threshold process.

In many cases, the data cannot be appropriately described by the Gaussian distribution; for example, it is well-known that financial time series often have heavy tails, and the t-distribution could be more appropriate for the noise process than the Gaussian distribution. Zhang (2012) estimated the parameters of the TAR model with t distributed error process when the model is completely identified. However, in practice, the identification problem may be difficult to be carry out, because this requires some additional knowledge about the phenomenon and this fact is particularly unrealistic due to the complex model structure. The focus of this work is to propose a Bayesian methodology which includes model identification in the TAR modelling. Specifically, a three-stage methodology based on the Gibbs sampler is proposed: in the first stage, the number of regimes together with the thresholds are estimated; in the second, the autoregressive orders in the regimes are estimated, and finally, in the last stage, the autoregressive orders, the variance weights and other parameters of the noise process are estimated. In this way, the identification of the model takes place in the first two stages, and the

estimation of the model in the last stage. Additionally, a Bayesian methodology for the estimation of missing data and the forecasting issue is developed. The methodologies developed are illustrated with simulated and real-life data in the finance field.

The work is organized as follows: in Section 2, we introduce the TAR model with t distributed noise process; in Section 3, we present the estimation procedure for the non structural parameters when the structural parameters are known; in Section 4, we present the identification for the structural parameters; in Sections 5 and 6, we present forecasting and missing-data estimation procedures. Finally, in Section 7, we illustrate the developed methodology in simulated and real-life data.

2. TAR Model with T-Distributed Noise

In order to introduce the TAR model, first suppose that the set of real numbers is divided in l disjoint intervals as $\mathbb{R} = \bigcup_{j=1}^{l} R_j$ where $R_j = (r_{j-1}, r_j]$, where $r_1 < \cdots < r_{l-1}$ and $r_0 = -\infty$, $r_l = \infty$. The $R_1, \ldots, R_l$ are denomited the regimes and the values $r_1, \ldots, r_{l-1}$ are denominated the thresholds. Let $\{Z_t\}$ be a stochastic process with stochastic behaviour described by a Markov chain of order p called the threshold process and let $\{X_t\}$ be the process of interest.

The dynamic of the process $\{X_t\}$ is determined by the process $\{Z_t\}$ in the way that when $Z_t \in R_j = (r_{j-1}, r_j]$ for some $j = 1, \ldots, l$ the model for $\{X_t\}$ is

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

The values $k_1, \ldots, k_l$ are nonnegative integer numbers representing the autoregressive orders in the l regimes, that is, different autoregressive orders are allowed in different regimes. $h^{(j)} > 0$ for $j = 1, \ldots, l$, Nieto & Moreno (2013) found that the parameters $(h^{(j)})^2$ correspond to the variance of X_t conditional on the regime and the past values of X, the so-called type II conditional variance. With respect to the noise process $\{e_t\}$, Nieto (2005) uses a Gaussian distribution, in this paper a Student's t-distribution is used. However, in order to mantain the interpretation of $(h^{(j)})^2$ mentioned before, we use a t-distribution with degrees of freedom n divided by its standard deviation $\sqrt{n/(n-2)}$, that is, $e_t \sim_{iid} \frac{t_n}{\sqrt{n/(n-2)}}$ with $n > 2$, which is mutually independent from the process $\{Z_t\}$; in this way, $Var(e_t) = 1$ for all t and hence, following Nieto & Moreno (2013), $(h^{(j)})^2 = Var(X_t | R_j, x_{t-1}, \ldots, x_1)$. Additionally, we assume that $\{Z_t\}$ is exogenous in the sense that there is no feedback of $\{X_t\}$ towards it.

Nieto & Moreno (2013) found conditions on the coefficients $a_i^{(j)}$ to achieve stationarity when the distribution for the noise process is Gaussian; however, as pointed out by Nieto (2005), the stationarity is not required for the correct implementation of the proposed Bayesian methodology; we have the same situation in

this paper and the corresponding conditions for stationarity deserve future investigations.

The parameters of the model can be divided into two groups:

- *Structural parameters*: the number of regimes l, the $l-1$ thresholds r_1, ..., r_{l-1} and the autoregressive orders of the l regimes k_1, ..., k_l.
- *Non-structural parameters*: the autoregressive coefficients $a_i^{(j)}$ with $i = 0, \ldots, k_j$ and $j = 1, \ldots, l$, the variance weights $h^{(1)}$, ..., $h^{(l)}$ and the degrees of freedom of the noise process, n.

In this research, we use the following notation: $\boldsymbol{\theta}_j' = (a_0^{(j)}, a_1^{(j)}, \ldots, a_{k_j}^{(j)})'$ for $j = 1, \ldots, l$, $\boldsymbol{\theta}' = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_l')'$. and $\mathbf{h}' = (h^{(1)}, \ldots, h^{(l)})'$.

2.1. Conditional Likelihood Function of the Model

According to Nieto (2005) and conditioned upon the values of the structural parameters, the initial values $\mathbf{x}_k = (x_1, \ldots, x_k)'$, where $k = \max\{k_1, \ldots, k_l\}$ and the observed data of the threshold process $\mathbf{z} = (z_1, \ldots, z_T)'$, the conditional likelihood function is given by:

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

where $\boldsymbol{\theta}_z$ denotes the vector of parameters of the threshold process $\{Z_t\}$ and $\boldsymbol{\theta}_x$ denotes the vector of all the non structural parameters, that is $\boldsymbol{\theta}_x' = (\boldsymbol{\theta}', \mathbf{h}', n)$. As $e_t \sim \frac{t_n}{\sqrt{n/(n-2)}}$, for $t = k+1, \ldots, T$, the variable $x_t|x_{t-1}, \ldots, x_1, \mathbf{z}$ is distributed as a t_n variable multiplied by $\frac{h^{(j_t)}}{\sqrt{n/(n-2)}}$ and adding $a_0^{(j_t)} + \sum_{i=1}^{k_{j_t}} a_i^{(j_t)} x_{t-i}$, where $j_t = j$ if $Z_t \in R_j = (r_{j-1}, r_j]$ for some $j = 1, \ldots, l$. That is, the distribution of x_t, conditioned upon the past values of x and $\mathbf{z}$, is the non-standardized Student's t-distribution with n degrees of freedom, location parameter $a_0^{(j_t)} + \sum_{i=1}^{k_{j_t}} a_i^{(j_t)} x_{t-i}$ and scale parameter $\frac{h^{(j_t)}}{\sqrt{n/(n-2)}}$. Thus,

$$f(x_t|x_{t-1}, \ldots, x_1, \mathbf{z}, \boldsymbol{\theta}_x, \boldsymbol{\theta}_z) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi(n-2)}\Gamma(\frac{n}{2})} \frac{1}{h^{(j_t)}} \left(1 + \frac{\left[x_t - a_0^{(j_t)} - \sum_{i=1}^{k_{j_t}} a_i^{(j_t)} x_{t-i}\right]^2}{(h^{(j_t)})^2 (n-2)}\right)^{-\frac{n+1}{2}}.$$

Consequently, the conditional likelihood function is given by

$$f(\mathbf{x}|\mathbf{z}, \boldsymbol{\theta}_x, \boldsymbol{\theta}_z) = \left[\frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi(n-2)}\Gamma(\frac{n}{2})}\right]^{T-k} \prod_{t=k+1}^{T} \left[h^{(j_t)}\right]^{-1} \prod_{t=k+1}^{T} \left(1 + \frac{e_t^2}{n-2}\right)^{-\frac{n+1}{2}}, \quad (2)$$

with $e_t = \frac{1}{h^{(j_t)}} \left(x_t - a_0^{(j_t)} - \sum_{i=1}^{k_{j_t}} a_i^{(j_t)} x_{t-i}\right)$ and $j_t = j$ when $Z_t \in R_j$.

3. Estimation of Non Structural Parameters

In this part of the research, the structural parameters are assumed to be known, and we focus on finding the posterior conditional distributions of the autoregressive coefficients $\boldsymbol{\theta}_j$, the variance weights $h^{(j)}$ with $j = 1, \ldots, l$ and the degrees of freedom of the noise process n. Additionally we assume prior independence between the parameters $\boldsymbol{\theta}$, $\mathbf{h}$ and n, as well as prior independence among the non structural parameters in each one of the l regimes.

The prior distribution for the vector $\boldsymbol{\theta}_j$ is a multivariate normal distribution with mean vector $\boldsymbol{\theta}_{0,j}$ and covariance matrix $\mathbf{V}_{0,j}^{-1}$, denoted as $\boldsymbol{\theta}_j \sim N(\boldsymbol{\theta}_{0,j}, \mathbf{V}_{0,j}^{-1})$, and the posterior conditional distribution of $\boldsymbol{\theta}_j$ is given by the following result:

Proposition 1. *For each $j = 1, \ldots, l$, the conditional distribution of $\boldsymbol{\theta}_j$ given the structural parameters $\boldsymbol{\theta}_i$, with $i \neq j$, $\mathbf{h}$, and n is given by*

$$p(\boldsymbol{\theta}_j|\boldsymbol{\theta}_i, i \neq j, \mathbf{h}, \mathbf{x}, \mathbf{z}, n) \propto \prod_{\{t:j_t=j\}} \left(1 + \frac{\left[x_t - a_0^{(j)} - \sum_{i=1}^{k_j} a_i^{(j)} x_{t-i}\right]^2}{(h^{(j)})^2(n-2)} \right)^{-\frac{n+1}{2}} \times \exp\left\{ -\frac{1}{2}(\boldsymbol{\theta}_j - \boldsymbol{\theta}_{0,j})' \mathbf{V}_{0,j} (\boldsymbol{\theta}_j - \boldsymbol{\theta}_{0,j}) \right\}. \tag{3}$$

Note that the posterior conditional distribution of $\boldsymbol{\theta}_j$ is affected only by $h^{(j)}$, but not the other components of $\mathbf{h}$ in regimes different from j, so we have some class of posterior independence between regimes.

Now, with respect to the variance weights $h^{(j)}$, we follow the standard Bayesian methodology assigning an inverse Gamma distribution with shape parameter α and scale parameter β, $(IG(\alpha, \beta))$, as the prior distribution of $(h^{(j)})^2$, that is,

$$p((h^{(j)})^2) \propto (h^{(j)})^{-2\alpha-2} \exp\{-\beta/(h^{(j)})^2\} I_{(0,\infty)}((h^j)^2).$$

Combining this prior distribution of $(h^{(j)})^2$ and the conditional likelihood function, we have the following posterior conditional distribution $(h^{(j)})^2$:

Proposition 2. *For each $j = 1, \ldots, l$, the posterior distribution of $(h^{(j)})^2$ given the structural parameters, $\boldsymbol{\theta}_j$, $j = 1, \ldots, l$, $h^{(i)}$, with $i \neq j$ and n, is given by*

$$\begin{aligned} p((h^{(j)})^2|\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_l, h^{(i)}, i \neq j, \mathbf{x}, \mathbf{z}, n) \\ \propto \prod_{\{t:j_t=j\}} \left(1 + \frac{\left[x_t - a_0^{(j)} - \sum_{i=1}^{k_j} a_i^{(j)} x_{t-i}\right]^2}{(h^{(j)})^2(n-2)} \right)^{-\frac{n+1}{2}} \\ \times (h^{(j)})^{-2\alpha-2-n_j} \exp\{-\beta/(h^{(j)})^2\}. \end{aligned} \tag{4}$$

Note that, the posterior conditional distribution of $(h^{(j)})^2$ is affected only by $\boldsymbol{\theta}_j$, but not by $\boldsymbol{\theta}_i$ with $i \neq j$, so again we have the posterior independence between $\boldsymbol{\theta}_j$, $h^{(j)}$ in different regimes.

Finally, we found the posterior conditional distribution of the degrees of freedom of the noise process n. The prior distribution of n is a Gamma distribution following the suggestion of Watanabe (2001), since in the distribution $Gamma(\alpha', \beta')$, the expectation and the variance are given by $\alpha'\beta'$ and $\alpha'\beta'^2$, respectively. Actually, α' and β' can be chosen according to prior knowledge about n, and in case that there is no prior information about n, we can choose a quite large prior variance to represent the high degree of uncertainty in the prior information of n. The prior distribution of n is given by

$$p(n) \propto n^{\alpha'-1} \exp\{-n/\beta'\}.$$

Using this prior distribution, we find the posterior conditional distribution of n.

Proposition 3. *The posterior conditional distribution of the degrees of freedom of the noise process $\{e_t\}$ is given by*

$$\begin{aligned} &p(n|\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_l, \mathbf{h}, \mathbf{x}, \mathbf{z}) \\ &\quad \propto \left[\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n-2}\,\Gamma(\frac{n}{2})}\right]^{T-k} \prod_{t=k+1}^{T} \left(1 + \frac{\left[x_t - a_0^{(j_t)} - \sum_{i=1}^{k_{j_t}} a_i^{(j_t)} x_{t-i}\right]^2}{(h^{(j_t)})^2 (n-2)}\right)^{-\frac{n+1}{2}} \\ &\quad \times n^{\alpha'-1} \exp\{-n/\beta'\}. \end{aligned} \tag{5}$$

In conclusion, the estimation of non structural parameters can be carried out by means of a Gibbs sampler, using the conditional densities (3), (4) and (5). We use the grid method to simulate values from these distributions; in order to define the parameter space, we fit an autoregressive $AR(p)$ model (where $p = \max\{k_1, \ldots, k_l\}$) and use the estimation plus and minus two times the standard error as the parameter space. For example, suppose that $l = 2$, $k_1 = 1$ and $k_2 = 2$, and the estimated coefficients of the $AR(2)$ model are 0.8 and -0.4 with standard error 0.1 and 0.2, then the parameter space for $a_1^{(1)}$ and $a_1^{(2)}$ would be (0.6,1) and the parameter space for $a_2^{(2)}$ would be (-0.8,0). Also for the coefficients $h^{(j)}$ we take into account the estimation of the error variance in the $AR(2)$ fitting. Finally, for the degree of freedom n, we choose the parameter space to be (2,30), since the t distribution has no finite variance when $n \leq 2$ and would be too similar to a normal distribution for $n > 30$.

4. Estimation of Structural Parameters

In this section, we develop the results concerning the estimation of the structural parameters, i.e. the identification of a TAR model. Firstly, we assume that the number of regimes l and the $l-1$ thresholds are known, and we estimate the autoregressive orders in these regimes; then we consider the case where the thresholds are known, and finally, we have the general case, where all the structural parameters are unknown.

4.1. Estimation of the Autoregressive Orders $k_1, \ldots, k_l$

As we assume that the number of the regimes and the thresholds are known, remaining parameters to be estimated are the autoregressive orders and the non-structural parameters.

We assume that the autoregressive orders $k_1, \ldots, k_l$ are realizations of discrete random variables $K_1, \ldots, K_l$, and each of theses variables takes value in the set $\{0, 1, \ldots, k_{\max}\}$. It is important to note that when the values of some autoregressive orders change, the specification of the TAR model changes and the dimension of the vector of the autoregressive coefficients $\boldsymbol{\Theta}$ also changes. Carlin & Chib (1995) developed a Bayesian methodology for the selection of models, and Nieto (2005) adapted this methodology in order to identify the TAR model with Gaussian noise. In this research, we adapt the same methodology to identify the TAR model with t distributed noise. Suppose that M is a discrete random variable indexing the model which takes values $1, \ldots, (k_{max}+1)^l$. For each possible model $M = m$, we define the vector of parameters $\boldsymbol{\Theta}_m$ as $\boldsymbol{\Theta}'_m = (\boldsymbol{\theta}'_1, \ldots, \boldsymbol{\theta}'_l, \mathbf{h}')$ for the model m with $m = 1, \ldots, (k_{max}+1)^l$. The degrees of freedom n can be considered as a nuisance parameter, since its dimension is the same for all models as well as its interpretation.

Carlin & Chib (1995) found the following conditional densities

$$p(M = m|\boldsymbol{\Theta}, \mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\Theta}_m,\ M = m)P(M = m)}{\sum_{m'} p(\mathbf{y}|\boldsymbol{\Theta}'_m,\ M = m')P(M = m')}. \tag{6}$$

where $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_{(k_{max}+1)^l}\}$, $\mathbf{y} = (\mathbf{x}, \mathbf{z})$ is the vector of full data. And

$$p(\boldsymbol{\Theta}_m|\boldsymbol{\Theta}_{m'\neq m}, M, \mathbf{y}) \propto \begin{cases} p(\mathbf{y}|\boldsymbol{\Theta}_m,\ M = m)p(\boldsymbol{\Theta}_m|M = m) & \text{if } M = m, \\ p(\boldsymbol{\Theta}_m|M = m) & \text{if } M \neq m. \end{cases} \tag{7}$$

The densities $p(\boldsymbol{\Theta}_m|M = m)$ are denominated as the link functions which can be taken as the prior distribution of $\boldsymbol{\Theta}_m$.

In the context of the problem of identification of the autoregressive orders, the model indicator M is determined jointly by the values of variables $K_1, \ldots, K_l$. In this way, computing the density (6) is equivalent to computing the densities $p(k_j|\boldsymbol{\Theta}, k_{i,i\neq j}, \mathbf{y})$ with $j = 1, \ldots, l$, because when we know the conditional distribution of each k_j, we can sample values of k_j by using a Gibbs sampler and, thus, we can sample values of the model indicator M. In order to compute these densities, Nieto (2005) found that

$$p(k_j|\boldsymbol{\Theta}, k_{i,i\neq j}, l, \mathbf{x}, \mathbf{z}) = \frac{p(\mathbf{x}|\mathbf{z}, \boldsymbol{\Theta}, \mathbf{h}, \mathbf{k}, l)p(k_j)}{\sum_{k'_j=0}^{k_j} p(\mathbf{x}|\mathbf{z}, \boldsymbol{\theta}, \mathbf{h}, \mathbf{k}', l)p(k'_j)}, \tag{8}$$

where $\mathbf{k} = (k_1, \ldots, k_l)$, and $\mathbf{k}'$ is obtained by replacing the component k_j of the vector $\mathbf{k}$ by k'_j for all $j = 1, \ldots, l$.

In summary, using the densities (3), (4), (5) and (8), a Gibbs sampler can be implemented in order to obtain the estimations of the probabilities of all the possible values for each K_j with $j = 1, \ldots, l$. Denoting these estimated probabilities as $\hat{p}_{0j}, \hat{p}_{1j}, \ldots, \hat{p}_{k_{\max}j}$, we can choose the value of K_j for which the highest probability is associated.

4.2. Estimation of the Number of Regimes l

In order to estimate the number of regimes, we use again the approach developed in Nieto (2005) adapting the methodology of Carlin & Chib (1995). Suppose that the number of regimes l is the realization of a discrete random variable L which takes values in the set $\{2, \ldots, l_{\max}\}$, and the prior distribution of L is denoted by $p(l)$.

Clearly, when the value of l changes, the model specification also changes; we have $l_{\max} - 1$ possible models. Suppose that M is the discrete random variable indexing the model, then M takes values 2, ..., $l_{\max}$, and for each possible model, $M = j$, $\boldsymbol{\Theta}_j$ denotes the vector of the parameters in this model, that is:

$$\boldsymbol{\Theta}'_j = (\boldsymbol{\theta}'_1, \ldots, \boldsymbol{\theta}'_j, \mathbf{h}'_j, \mathbf{k}'_j, n),$$

with $\mathbf{k}'_j = (k_{1j}, \ldots, k_{jj})'$, where k_{ij} denotes the autoregressive order in the i-th regime in the model $M = j$, $\mathbf{h}'_j = (h^{(1)}, \ldots, h^{(j)})'$. Finally, we define $\boldsymbol{\Theta}' = (\boldsymbol{\Theta}'_2, \ldots, \boldsymbol{\Theta}'_{l_{\max}})$, the vector containing all the parameters for all the possible models.

Nieto (2005) found the following conditional densities

$$p(M = j|\boldsymbol{\Theta}, \mathbf{y}) = p(l|\boldsymbol{\Theta}, \mathbf{y}) \propto p(\mathbf{x}|\mathbf{z}, \boldsymbol{\Theta}_l, l)p(l) \qquad \text{for } l = 2, \ldots, l_{\max}, \tag{9}$$

$$p(k_{ij}|\boldsymbol{\Theta}_{-k_{ij}}, l, \mathbf{y}) = \begin{cases} \dfrac{p(\mathbf{x}|\mathbf{z}, \boldsymbol{\Theta}_l, l)p(k_{ij})}{\sum_{k'_{il}=0}^{k_{\max}} p(\mathbf{x}|\mathbf{z}, \boldsymbol{\Theta}_l, l)p(k'_{il})} & \text{if } j = l, \\ p(k_{ij}) & \text{if } j \neq l, \end{cases} \tag{10}$$

where $\boldsymbol{\Theta}_{-k_{ij}}$ denotes the vector $\boldsymbol{\Theta}$ without the element k_{ij}, and

$$p(\boldsymbol{\theta}_j, h^{(j)}|\boldsymbol{\Theta}_{-\boldsymbol{\theta}_j, h^{(j)}}, l, \mathbf{y}) \propto \begin{cases} p(\mathbf{y}|\boldsymbol{\Theta}_l, l)p(\boldsymbol{\Theta}_l) & \text{if } j = l, \\ p(\boldsymbol{\Theta}_j) & \text{if } j \neq l, \end{cases} \tag{11}$$

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

Jointly using the conditional densities (9), (10), (11) and (5), we can implement a Gibbs sampler and obtain the posterior probabilities for all possible values of L. The estimation of the number of regimes l could be the value with major posterior probability or the mode of the value of L in the iterations of the Gibbs sampler.

4.3. Estimation of the Number of Regimes l and the Thresholds

Finally, we assume that the $l-1$ thresholds are also unknown, and they need to be estimated jointly with the number of regimes l. Following the approach of Carlin & Chib (1995), the model is indexed by a discrete variable M, which takes values $2,\ldots,l_{\max}$ according to the value of the variable L. For each possible model $M=j$, the thresholds are denoted as $\mathbf{r}_j=(r_1,\ldots,r_{j-1})'$ with $j=2,\ldots,l_{\max}$.

It is straightforward to obtain the posterior conditional density of $\mathbf{r}_j$ given the values of other structural and non-structural parameters, given by

$$p(\mathbf{r}_j|l,\boldsymbol{\Theta}_{-\mathbf{r}_j},\mathbf{y}) \propto \begin{cases} \prod_{t=k+1}^{T}[h^{(j_t)}]^{-1}\prod_{t=1}^{T}\left(1+\frac{\left[x_t-a_0^{(j_t)}-\sum_{i=1}^{k_{j_t}}a_i^{(j_t)}x_{t-i}\right]^2}{(h^{(j_t)})^2(n-2)}\right)^{-\frac{n+1}{2}} & \text{if } j=l, \\ p(\mathbf{r}_j) & \text{if } j\neq l, \end{cases} \tag{12}$$

where $j_t=j$ if $Z_t\in R_j=(r_{j-1},r_j]$ for some $j=1,\ldots,l$. The posterior conditional density of l is given by (9). Note that the expression in (12) depends on the thresholds $\mathbf{r}_j$ since $j_t=j$ if $Z_t\in R_j=(r_{j-1},r_j]$, so R_j and j_t depend on the thresholds, and so does the expression (12). In this way, using the posterior conditional density of $\mathbf{r}_j$, l, and $\boldsymbol{\Theta}_j$, we can implement a Gibbs sampler and obtain the estimation of the number of regimes and thresholds. In order to extract samples from this density, we use the grid method where the parameter space consists of all possible ordered values of Z_t.

With respect to the prior density of $\mathbf{r}_j$, we recall that the values of the thresholds are based on the values of the process $\{Z_t\}$, so we can assume that the thresholds take values in a interval (a,b), appropriately specified; furthermore, we assume a uniform distribution for the thresholds r_1, ..., r_{j-1}, that is

$$p(\mathbf{r}_j)=p(r_1,\ldots,r_{j-1})\propto k \qquad \text{if } a<r_1<\cdots<r_{j-1}<b,$$

for $j=2,\ldots,l_{\max}$.

4.4. Proposed Algorithm

In conclusion, a three-stage process is proposed for the identification and estimation of TAR models with t-distributed noise with no missing data. This algorithm consists of the following steps:

1. The number of regimes and thresholds are estimated using a Gibbs sampler based on the densities (9), (10), (11), (12) and (5).

2. The number of regimes and thresholds are fixed and the autoregressive orders are estimated using a Gibbs sampler based on the densities (7), (8) and (5).

3. Finally, conditioned upon the estimated structural parameters, we estimate the non-structural parameters using a Gibbs sampler with densities (3), (4) and (5).

5. Forecasting

In order to develop the predictive inference, we focus on finding the posterior predictive distribution of the variable X_{T+h} conditional on the observed $\mathbf{x}_T = (x_1, \ldots, x_T)$ and $\mathbf{z}_T = (z_1, \ldots, z_T)$ with $h > 0$. Vargas (2012) worked on the formal Bayesian approach to find the predictive density of X_{T+h} involving the variability in the parameters of the model; this predictive density is given by:

$$p(x_{T+h}|\mathbf{x}_T, \mathbf{z}_T) = \sum_{j=1}^{l} p(x_{T+h}|R_j, \mathbf{x}_T, \mathbf{z}_T) p_j(h),$$

where $p_j(h) = P(Z_{T+h} \in R_j|\mathbf{x}_T, \mathbf{z}_T)$, for $h = 1, 2, \ldots$, and $j = 1, \ldots, l$, and

$$\begin{aligned} p(x_{T+h}|R_j, \mathbf{x}_T, \mathbf{z}_T) &= \int_{\boldsymbol{\Theta}_j} p(x_{T+h}, \boldsymbol{\theta}^{(j)}|R_j, \mathbf{x}_T, \mathbf{z}_T) d\boldsymbol{\theta}^{(j)} \\ &= \int_{\boldsymbol{\Theta}_j} p(x_{T+h}|\boldsymbol{\theta}^{(j)}, R_j, \mathbf{x}_T, \mathbf{z}_T) p(\boldsymbol{\theta}^{(j)}|R_j, \mathbf{x}_T, \mathbf{z}_T) d\boldsymbol{\theta}^{(j)}, \end{aligned} \tag{13}$$

where $\boldsymbol{\theta}^{(j)}$ denotes the vector of the non-structural parameters in the regime j, and

$$\begin{aligned} p(x_{T+h}|\boldsymbol{\theta}^{(j)}, R_j, \mathbf{x}_T, \mathbf{z}_T) = \int \cdots \int & p(x_{T+h}|\boldsymbol{\theta}^{(j)}, R_j, \mathbf{x}_{T+h-1}) \\ & \times p(x_{T+h-1}|\boldsymbol{\theta}^{(j)}, R_j, \mathbf{x}_{T+h-2}) \\ & \times \cdots \times p(x_{T+1}|\boldsymbol{\theta}^{(j)}, R_j, \mathbf{x}_T) dx_{T+1} \cdots dx_{T+h-1}. \end{aligned} \tag{14}$$

On the other hand, in order to forecast the threshold variable Z_{T+h}, Nieto (2008) found that:

$$\begin{aligned} p(z_{T+h}|\mathbf{z}_T) = \int \cdots \int & p(z_{T+h}|z_{T+h-1}, z_{T+1}, \mathbf{z}_T) \\ & \times p(z_{T+h-1}|z_{T+h-2}, z_{T+1}, \mathbf{z}_T) \cdots p(z_{T+1}|\mathbf{z}_T) dz_{T+1} \cdots dz_{T+h-1}. \end{aligned} \tag{15}$$

Based on the equations (13), (14) and (15), we can compute forecasts for both processes $\{X_t\}$ and $\{Z_t\}$. In order to draw values from $p(z_{T+h}|\mathbf{z}_T)$ with $h = 1, 2, \ldots$, Congdon (2001) suggests to draw a value for z_{T+1} from $p(z_{T+1}|\mathbf{z}_T)$, then draw value for z_{T+2} from $p(z_{T+2}|z_{T+1}, \mathbf{z}_T)$ and so on.

On the other hand, in order to forecast the process $\{X_t\}$, notice that $p(x_{T+h}|\mathbf{x}_T, \mathbf{z}_T)$ is a mixture density, so we just draw a value from $p(x_{T+h}|R_j, \mathbf{x}_T, \mathbf{z}_T)$ with probability $p_j(h)$. Secondly, we note that each term $p(x_{T+m}|\boldsymbol{\theta}^{(j)}, R_j, \mathbf{x}_{T+m-1},)$ for

$m = 1, \ldots, h$ corresponds to the density of a non-standarized Student's t-distribution with n degrees of freedom, location parameter $a_0^{(j)} + \sum_{i=1}^{k_j} a_i^{(j)} x_{T+m-i}$ and scale parameter $\frac{h^{(j)}}{\sqrt{n/(n-2)}}$. This concludes the forecasting process.

6. Estimation of Missing Data

We assume that there are missing observations in both processes $\{X_t\}$ and $\{Z_t\}$ and that the observed data of $\{X_t\}$ are located in time points $t_1, \ldots, t_N$ with $1 \leq t_1 \leq \cdots \leq t_N \leq T$; similarly, the observed data of $\{Z_t\}$ are located in time points $s_1, \ldots, s_M$ with $1 \leq s_1 \leq \cdots \leq s_M \leq T$. The estimation of these missing data can be carried out using the approach of Nieto (2005) as shown below.

The TAR model without missing data can be put in state space form taking the state vector as $\boldsymbol{\alpha}_t = (X_t, X_{t-1}, \ldots, X_{t-k+1})'$, with $k = \max\{k_1, \ldots, k_l\}$, as:

$$X_t = \boldsymbol{H}\boldsymbol{\alpha}_t, \tag{16}$$

$$\boldsymbol{\alpha}_t = \boldsymbol{C}_{J_t} + \boldsymbol{A}_{J_t}\boldsymbol{\alpha}_{t-1} + \boldsymbol{R}_{J_t}\boldsymbol{\omega}_t, \tag{17}$$

where $\boldsymbol{H} = (1, 0, \ldots, 0)$, $\boldsymbol{\omega}_t = (e_t, 0, \ldots, 0)'$ and $J_t = j$ if $Z_t \in R_j$. For each $j = 1, \ldots, l$, $\boldsymbol{C}_j = (a_0^{(j)}, 0, \ldots, 0)'$,

$$\boldsymbol{R}_j = \begin{pmatrix} h^{(j)} & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

and

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

where $a_i^{(j)} = 0$ for $i > k$ and $\mathbf{I}_{k-1}$ denote the identity matrix of order $k-1$. The equation (16) is the observation equation and the equation (17) is the state equation. As pointed by Nieto (2005), this state space form corresponds to a state space model with regime switching and can be analysed efficiently using MCMC simulation procedure.

When there are missing data, the state space form in (16) can be modified to include such missing data; the new observation equation is:

$$X_t = \boldsymbol{H}_t\boldsymbol{\alpha}_t + \delta_t W,$$

where $\boldsymbol{H}_t = \boldsymbol{H}$ and $\delta_t = 0$ if $t \in \{t_1, \ldots, t_N\}$ and $\boldsymbol{H}_t = \mathbf{0}'$ and $\delta_t = 1$, otherwise, W is a discrete random variable with $Pr(W = w_0) = 1$ for some point w_0 in the support of X_t. The state equation remains the same.

Since the optimal estimates of the missing data, in the sense of minimum mean square error criterion, are the conditional expectations of the missing data given observed data, we need to sample from the density $p(\mathbf{x}_m, \mathbf{z}_m | \mathbf{x}_o, \mathbf{z}_o)$, where $\mathbf{x}_m$ and $\mathbf{z}_m$ denote the missing data set, and $\mathbf{x}_o$ and $\mathbf{z}_o$ denote the observed data set.

Nieto (2005) states that this goal can be achieved by sampling from $p(\boldsymbol{\alpha}, \mathbf{z}|\mathbf{x})$, where $\mathbf{x}$ and $\mathbf{z}$ are constituted by full data $x_1, \ldots, x_N$ and $z_1, \ldots, z_N$, and the missing data are replaced by artificial data, for example, the median of $\{x_t\}$ and $\{z_t\}$ and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_T)'$.

Nieto (2005) propose the use of a Gibbs sampler in order to draw samples from $p(\mathbf{z}|\boldsymbol{\alpha}, \mathbf{x})$ and $p(\boldsymbol{\alpha}|\mathbf{z}, \mathbf{x})$. The density $p(\mathbf{z}|\boldsymbol{\alpha}, \mathbf{x})$ it is found to be:

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

where $\mathbf{z}_t = (z_{t-p+1}, \ldots, z_t)$, $\boldsymbol{\alpha}_t$ and $\mathbf{x}_t$ are similarly defined, and

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

and for $t = T - p, \ldots, 1$

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

Finally, in order to sample values from the joint distribution of $p(\boldsymbol{\alpha}|\mathbf{z}, \mathbf{x})$, note that

$$p(\boldsymbol{\alpha}|\mathbf{z}, \mathbf{x}) = p(\boldsymbol{\alpha}_T|\boldsymbol{\alpha}_{T-1}, \ldots, \boldsymbol{\alpha}_1, \mathbf{z}, \mathbf{x}) p(\boldsymbol{\alpha}_{T-1}|\boldsymbol{\alpha}_{T-2}, \ldots, \boldsymbol{\alpha}_1, \mathbf{z}, \mathbf{x}) \cdots p(\boldsymbol{\alpha}_2|\boldsymbol{\alpha}_1, \mathbf{z}, \mathbf{x}),$$

where sampling each term $p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}, \ldots, \boldsymbol{\alpha}_1, \mathbf{z}, \mathbf{x})$ is equivalent to sample values from $\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}, Z_t$ since $\boldsymbol{\alpha}_t = (X_t, X_{t-1}, \ldots, X_{t-k+1})'$; and this is equivalent to sample values from the density $p(X_t|X_{t-1}, \ldots, X_{t-k}, Z_t)$, so we just need to sample values from the density of a non-standardized Student's t-distribution with n degrees of freedom, location parameter $a_0^{(j)} + \sum_{i=1}^{k_j} a_i^{(j)} x_{T+m-i}$ and scale parameter $\frac{h^{(j)}}{\sqrt{n/(n-2)}}$.

In summary, the estimation of missing data in TAR models can be carried out as follows:

1. Completion of the time series replacing the missing data $\{x_t\}$ and $\{z_t\}$, with their respective median.

2. Identification and estimation of the TAR model using the completed time series following the algorithm presented in subsection 2.4.

3. Estimation of the missing data by means of a Gibbs sampler using the above methodology.

4. Re-estimation of the TAR model with the missing data replaced by their estimates.

7. Illustrations

7.1. Simulated Examples

In this section we present two simulation examples in order to illustrate the performance of the proposed methodology.

7.1.1. Example 1

We simulated a series $\{x_t\}$ of 100 observations from the model:

$$X_t = \begin{cases} 1 + 0.5X_{t-1} - 0.3X_{t-2} + e_t & \text{if } Z_t \leq 0 \\ -0.5 - 0.7X_{t-1} + 1.5e_t & \text{if } Z_t > 0, \end{cases} \tag{18}$$

with $e_t \sim \frac{t_5}{\sqrt{5/3}}$, $Z_t = 0.5Z_{t-1} + \epsilon_t$ and ϵ_t is a Gaussian white noise process of mean 0 and variance 1 (GWN(0,1)). The simulated series are shown in Figure 1.

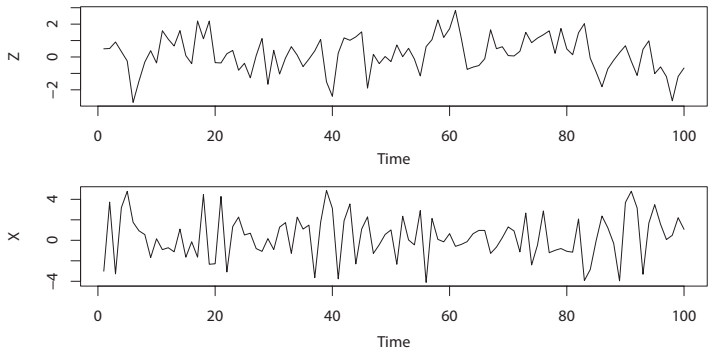

FIGURE 1: Simulated data in example1.

In the first stage, we identified the number of regimes and the thresholds. Following Nieto, Zhang & Li (2013), the prior distribution for l is the Poisson distribution truncated[1] in the set $\{2, 3, 4\}$ with parameter 3, and the prior distribution of the thresholds is as described above. We run a Gibbs sampler of 1,000 iterations with a burn-in period of 200. In order to ensure that for each parameter, the draws have converged to the posterior distribution, we use the Geweke's Z-score plot (Geweke 1992) from the package `coda` in `R` (Plummer, Best, Cowles & Vines 2006). Geweke's Z-score computes the difference between the means of the first and last part of the draws of a Markov chain, and the plot shows the values of the Z-score where successively larger numbers (at most half of the chain) of iterations are removed from the beginning of the chain. For the number of regimes and autoregressive orders, we monitor the corresponding posterior probabilities. Since there are a lot of parameters in the model, we cannot show all the plots, only

[1]Note that we have excluded the value 1 which corresponds to a linear AR model, in real applications, non-linear should be carried out.

a few where we consider that the burn-in period of 200 iterations is appropriate (Figure 2).

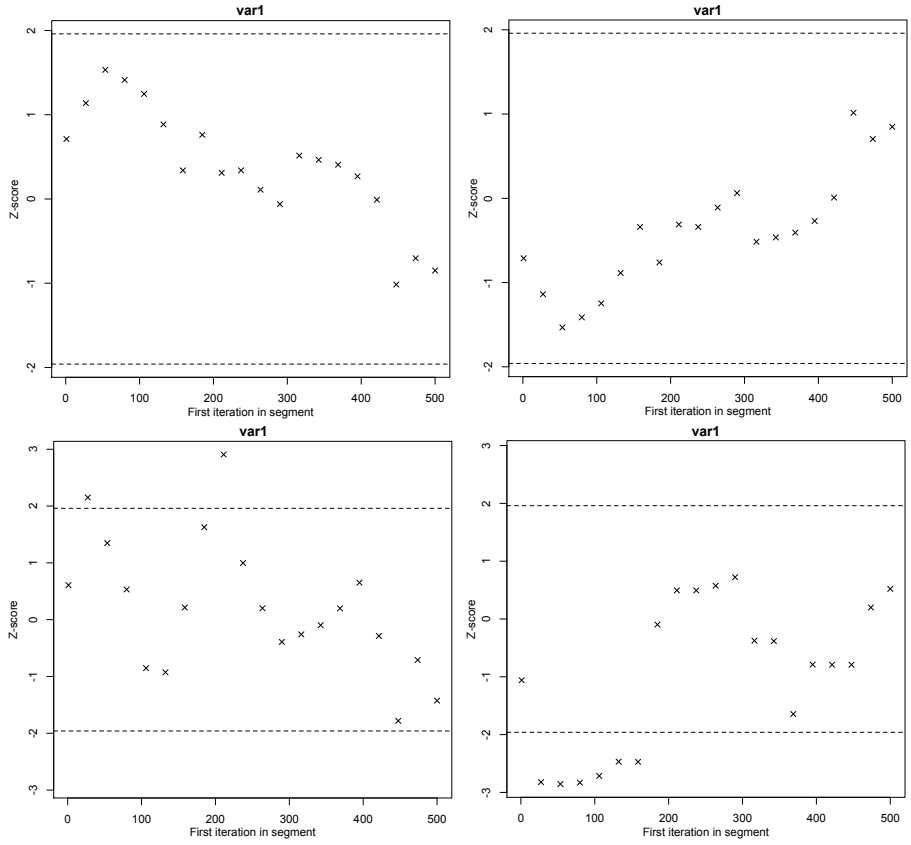

FIGURE 2: Geweke's Z-score for some of the parameters in the Gibbs sampler.

The posterior probability of the number of regimes is given in Table 1, where we can see that the number of regimes associated with the largest posterior probability is 2.

TABLE 1: Posterior probability for the number of regimes L in example 1.

l	2	3	4
Posterior probability	0.60	0.40	0

The estimation of the threshold is 0.0846[2]. The 95% credible interval for the threshold is given by (-0.2892, 0.6737) containing the real threshold 0.

[2]For a certain model $l = j$, the possible values of the thresholds $\mathbf{r}_j$ are the quantiles of the process $\{Z_t\}$, after removing the thresholds that induce regimes with too little data; in this case, we eliminate the thresholds that induce any regime with less than 20 data.

In the second stage, we estimated the autoregressive order in each of the two regimes where the number of regimes is fixed to be 2 and the value of the threshold to be 0.0846. The prior distribution for k_j is the truncated Poisson distribution with parameter 2 in the set $\{0, 1, 2, 3\}$[3] for each $j = 1, 2$. We run a Gibbs sampler of 1,000 iterations, and obtained the posterior probabilities for k_1 and k_2 displayed in Table 2. We can see that the identified autoregressive orders are $\hat{k}_1 = 2$ and $\hat{k}_2 = 1$, corresponding to the real autoregressive orders.

TABLE 2: Posterior probabilities of the variables K_1 and K_2 in example 1.

Autoregressive order	Regime	
	1	2
0	0.00	0.00
1	0.00	**0.72**
2	**0.51**	0.17
3	0.49	0.11

Finally, we estimated the non-structural parameters: autoregressive coefficients, the variance weights and the degrees of freedom of the process of error. The prior distribution for these parameters is: $N(0, 10)$ for the autoregressive coefficients a_i^j with $i = 1, \cdots, k_j$ and $j = 1, 2$; distribution $IG(2, 3)$ for the variance weights $(h^{(1)})^2$ and $(h^{(2)})^2$; and distribution $Gamma(1, 0.1)$ for the degrees of freedom n. In this way, the prior mean of n is 10 and the prior variance is 100, which can be considered as a non-informative prior distribution.

We run another Gibbs sampler of 1,000 iterations; the estimation and the 95% credible intervals of the autoregressive coefficients and the variance weights are given in Table 3. These estimations are close to the true parameters and all the 95% credible intervals contain the true parameters.

TABLE 3: Estimation and 95% credible intervals for the non-structural parameters for the simulated data in example 1.

Regime	$a_i^{(j)}$			$h^{(j)}$
1	0.89	0.55	-0.41	0.95
	(0.58, 1.20)	(0.41, 0.68)	(-0.52, -0.28)	(0.74, 1.25)
2	-0.38	-0.67		1.49
	(-0.74, 0.00)	(-0.84, -0.52)		(1.21, 1.89)

With respect to the degrees of freedom n, the results obtained from the Gibbs sampler are displayed in Figure 4, where we noted that the values of n with large posterior probability is around the true parameter 5. The posterior mean of n is given by 7.25, and a 95% credible interval of n is given by (3.3, 21.56). In order to check the appropriateness of the model, we use the CUSUM and CUSUMSQ plot of the standardized residuals since to our knowledge, there is no investigation about the distribution of the usual statistical tests in TAR models. These plots are shown in the Figure (3) where we can see that the overall performance is good.

[3]The maximum autoregressive order is chosen to be the autoregressive order p of the linear model $AR(p)$ that best fitted the data, which is 3

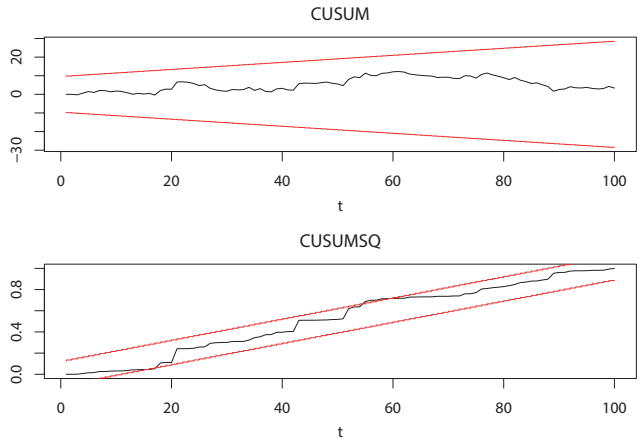

FIGURE 3: CUSUM and CUSUMSQ plot of the standardized residuals of the estimated model in example 1.

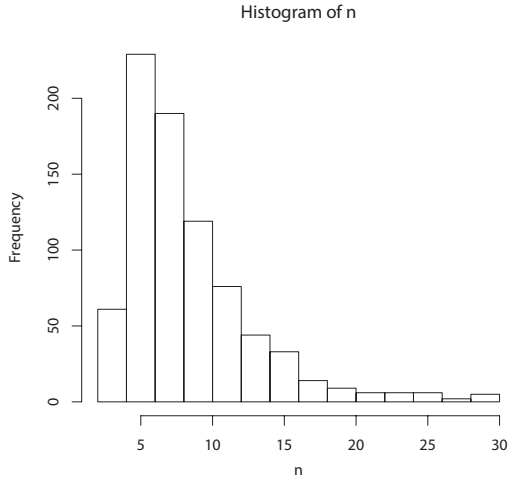

FIGURE 4: Histogram of the simulated values of the degrees of freedom n in example 1.

In conclusion, identification and estimation results were satisfactory, and we proceeded with the illustration of the estimation of the missing data. We set the number of missing data in the processes $\{Z_t\}$ and $\{X_t\}$ to be 4 and 6, respectively, and placed the missing data randomly. The resulting missing data for $\{Z_t\}$ and $\{X_t\}$ were situated at time points 8, 55, 63, 83 and 2, 13, 37, 41, 77, 80, respectively. The estimation and the credible intervals for the missing data after 5000 iterations are shown in Table 4. We can see that the overall performance of the procedure is satisfactory except at time 63 for $\{Z_t\}$, where the observed value lays beyond the 95% credible interval.

TABLE 4: Estimation and 95% credible intervals for the missing data in example 1.

Process $\{Z_t\}$			
Time	Observed	Estimated	Credible interval
8	-0.86	-0.37	(-2.0, 1.14)
55	-0.63	-0.33	(-1.8, 1.24)
63	-2.13	-0.14	(-1.67, 1.43)
83	-0.05	-0.42	(-2.21, 1.03)
Process $\{X_t\}$			
Time	Observed	Estimated	Credible interval
2	1.17	-0.19	(-3.10, 2.69)
13	-0.03	-0.95	(-3.81, 1.80)
37	-0.17	-0.35	(-3.15, 2.75)
41	-1.30	-1.22	(-4.05, 1.75)
77	-0.51	-1.73	(-4.72, 1.24)
80	0.65	0.16	(-2.697, 3.092)

Finally, we illustrate the forecast procedure where the sample period considered is 1-92, and the forecast horizon is set to be 8. We assume that non-structural parameters are known and for each horizon we simulate 100 series from the model (18). For each we estimate the structural parameters and compute the forecasting value and the respective credible interval. In order to illustrate the results we compute the percentage of these 100 credible intervals containing the true observation. In Table 5 we show these percentages.

TABLE 5: Coverage of the credible intervals of forecasting results for the simulated $\{X_t\}$ in example 1.

	Forecasting horizon							
Coverage	1	2	3	4	5	6	7	8
X_t	1	1	1	1	1	1	1	1

Also we illustrate the predictive density using kernel density for one of the 100 iterations described before (see Figure 5).

7.1.2. Example 2

We simulated a series $\{x_t\}$ of 300 observations from the model

$$X_t = \begin{cases} 1 + 0.5X_{t-1} + e_t & \text{if } Z_t \leq -0.6 \\ 0.5 + 0.2X_{t-1} + 0.5X_{t-2} + 1.5e_t & \text{if } -0.6 < Z_t \leq 0.6 \\ -0.5 - 0.7X_{t-1} + 2e_t & \text{if } Z_t > 0.6 \end{cases},$$

with $e_t \sim \frac{t_5}{\sqrt{5/3}}$, $Z_t = 0.5Z_{t-1} + \epsilon_t$ and $\epsilon_t \sim GWN(0,1)$. The simulated series are shown in Figure 6.

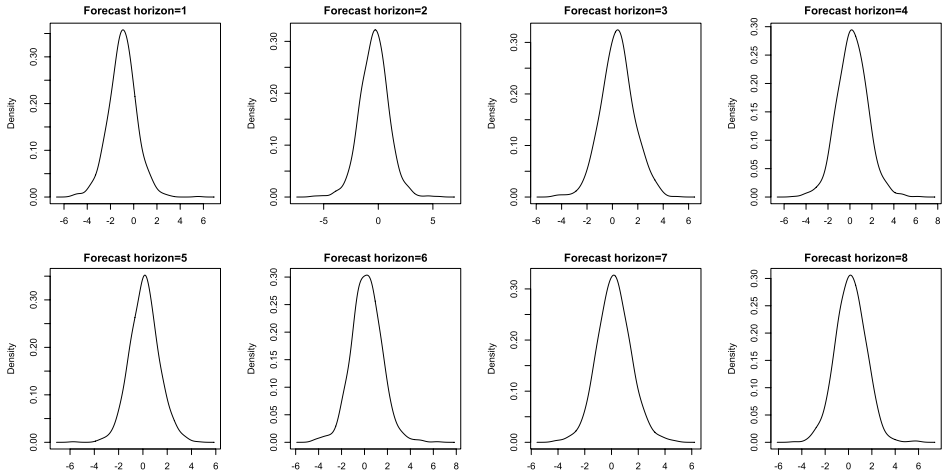

FIGURE 5: Kernel density of the predictive densities with forecast horizon 1, ..., 8.

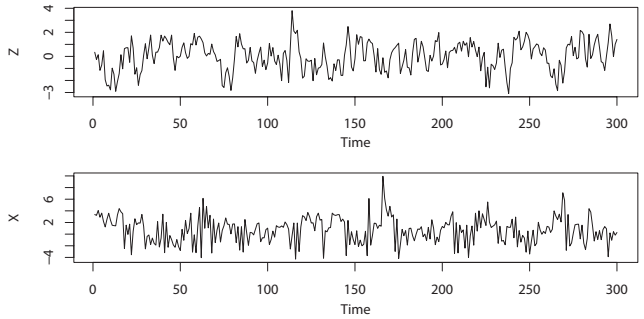

FIGURE 6: Simulated data in example 2.

In the first stage, we identified the number of regimes and the thresholds. The prior distribution for l is the Poisson distribution truncated in the set $\{2, 3, 4\}$ with parameter 3, and the prior distribution of the thresholds is as described before. We run a Gibbs sampler of 1,000 iterations and the posterior probability for all the possible values of the number of regimes are shown in Table 6 suggesting that $\hat{l} = 3$.

TABLE 6: Posterior probability for the number of regimes L in example 2.

l	2	3	4
Posterior probability	0.25	0.75	0

The estimation of the two thresholds are -0.6394 and 0.5205, respectively. In Figure (7), we present the histogram of the simulated values for the threshold. The 95% credible interval for the threshold is given by $(-1.1701, 0.6956)$ and $(-0.3848, 1.1779)$, respectively, containing the real values of the two thresholds. However, note that the histogram of the simulated thresholds seems to be bimodal,

so we computed the median given by -1.08 and 0.723, which are slightly away from the simulated values. In future research, we will do more simulations in order to validate that , in the cases, the median works well.

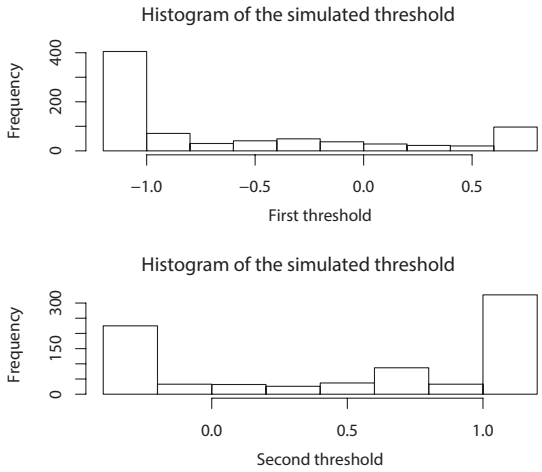

FIGURE 7: Histogram of the simulated values of the thresholds for three regimes in example 2.

In the second stage, we estimated the autoregressive order in each of the two regimes. The prior distribution for k_j is the truncated Poisson distribution with parameter 2 in the set $\{0, 1, 2, 3\}$ for each $j = 1, 2, 3$. We run a Gibbs sampler of 1,000 iterations, and obtained the posterior probabilities for K_1, K_2 and K_3, displayed in Table 7. We can see that the identified autoregressive orders are $\hat{k}_1 = 1$, $\hat{k}_2 = 2$ and $\hat{k}_3 = 1$, corresponding to the real autoregressive orders.

TABLE 7: Posterior probabilities of the variables K_1 and K_2 in example 2.

Autoregressive order	Regime		
	1	2	3
0	0	0	0
1	**0.37**	0	**0.38**
2	0.35	**0.75**	0.38
3	0.28	0.25	0.24

Finally, we estimated non-structural parameters: autoregressive coefficients, variance weights and degrees of freedom of the process of error. The prior distribution for these parameters is: $N(0, 10)$ for the autoregressive coefficients a_i^j with $i = 1, \ldots, k_j$ and $j = 1, 2, 3$; distribution $IG(2, 3)$ for the variance weights $(h^{(1)})^2$ and $(h^{(2)})^2$; distribution $Gamma(1, 0.1)$ for the degrees of freedom n, which can be considered a non-informative prior distribution as discussed in the example 1.

We run another Gibbs sampler of 1,000 iterations; the estimation of the autoregressive coefficients and the variance weights are given in Table 8. These estimations are close to the true parameters and, except for paramter $a_1^{(2)}$ whose value is 0.2, all the 95% credible intervals contain the true parameters.

TABLE 8: Estimation and 95% credible intervals for the non-structural parameters in example 2.

Regime	$a_i^{(j)}$			$h^{(j)}$
1	0.94	0.53		1.21
	(0.66, 1.22)	(0.43, 0.63)		(0.99, 1.47)
2	0.40	0.32	0.52	1.43
	(0.19, 0.66)	(0.21, 0.44)	(0.39, 0.64)	(1.19, 1.76)
3	-0.68	-0.61		1.82
	(-0.97, -0.38)	(-0.71, -0.49)		(1.49, 2.24)

With respect to the degrees of freedom n, the results obtained from the Gibbs sampler are displayed in Figure 8, where we noted that the values of n with large posterior probability are around the true parameter 5. The posterior mean of n is given by 5.11 and a 95% credible interval of n is given by (3.17, 8.92).

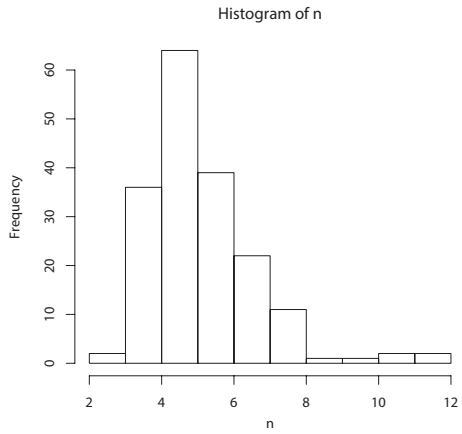

FIGURE 8: Histogram of the simulated values of the degrees of freedom n in example 2.

The CUSUM and CUSUMSQ plot of standardized residuals of the estimated model are shown in the Figure (9).

Since identification and estimation results are, generally speaking, satisfactory, we proceeded with the estimation of the missing data. We set the number of missing data in the processes $\{Z_t\}$ and $\{X_t\}$ to be 10 in both series, and placed the missing data randomly. The estimation and the credible intervals after 5000 iterations are shown in Table 9. We can see that the overall performance of the procedure is satisfactory in the sense that all the observed data are within the 95% credible interval.

Finally, we illustrate the forecast procedure where the sample period considered is 1-290, and the forecast horizon is set to be 10. We assume that the non-structural parameters are known and for each horizon we simulate 100 series from the model (18). Like the previous example, we compute the percentage of these 100 credible intervals containing the true observation. In Table 10 we show these percentages.

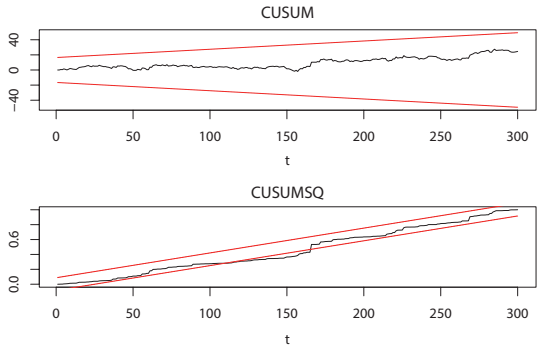

FIGURE 9: CUSUM and CUSUMSQ plot of the standardized residuals of the estimated model in example 2.

TABLE 9: Estimation and 95% credible intervals for the missing data in example 2.

Process $\{Z_t\}$			
Time	Observed	Estimated	Credible interval
32	0.71	0.48	(-0.88, 2.03)
88	-0.55	-0.14	(-1.68, 1.32)
95	0.25	0.18	(-1.44, 1.76)
105	-0.67	0.01	(-1.64, 1.68)
116	1.90	0.36	(-1.08, 2.26)
180	-0.90	-0.12	(-1.76, 1.53)
181	-0.95	-0.31	(-2.11, 1.20)
213	0.92	0.06	(-2.09, 2.36)
292	-0.53	-0.07	(-2.06, 1.93)
293	-1.17	0.12	(-2.15, 2.22)
Process $\{X_t\}$			
Time	Observed	Estimated	Credible interval
3	4.05	2.62	(0.11, 5.08)
46	-1.85	-0.50	(-2.66, 1.80)
82	-1.46	-1.92	(-2.50, 0.36)
86	-3.08	-1.55	(-4.77, -1.89)
183	-0.29	1.48	(-1.15, 3.84)
189	-0.44	0.62	(-2.94, 2.03)
216	-0.30	-1.15	(-3.51, 1.22)
262	0.30	1.48	(-1.10, 3.69)
290	0.60	0.42	(-1.94, 2.88)
294	0.92	1.59	(-0.66, 3.98)

TABLE 10: Coverage of the credible intervals of forecasting results for the simulated $\{X_t\}$ in example 2.

	Forecasting horizon									
Coverage	1	2	3	4	5	6	7	8	9	10
X_t	0.97	1	1	1	1	0.98	1	1	0.98	1

Also we illustrate the predictive density using kernel density for one of the 100 iterations described before (see Figure 10).

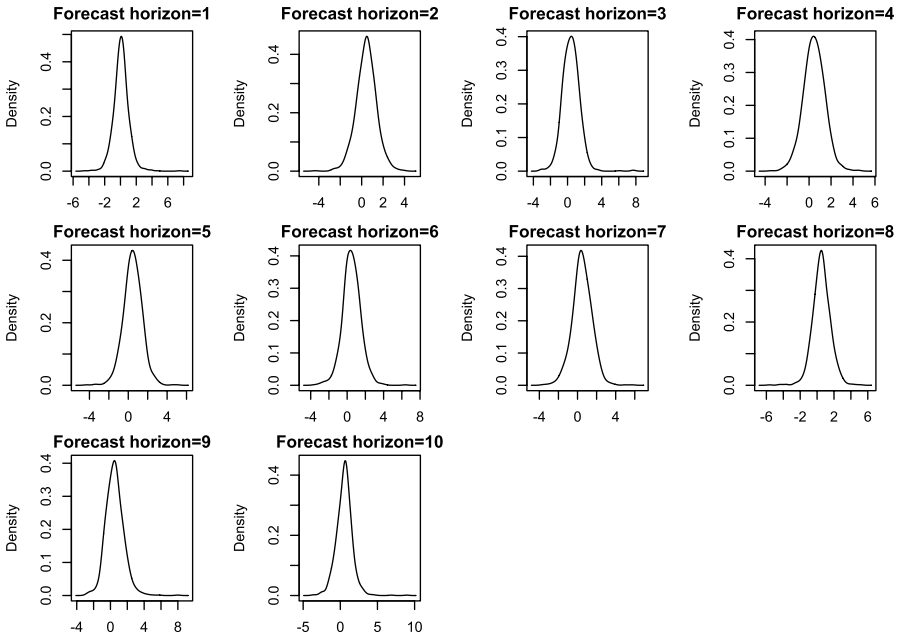

FIGURE 10: Kernel density of the predictive densities with forecast horizon 1, ..., 10.

7.2. An Application in Finance

In this section, we applied the proposed algorithm to financial time series to illustrate the methodology. Specifically, we used the daily log return of the Dow Jones industrial average as the threshold series, and the daily log return of the BOVESPA index (Brasil Sao Paulo Stock Exchange Index) as the series of interest, from December 12th, 2000 to June 2nd, 2010. Moreno (2010) tested the non-linearity of the data using the test of Tsay (1998) with lag up to 4 for the log return of the Dow Jones index, and found that the appropriate lag is 0. In this way, we defined $X_t = \ln(BOVESPA_t) - \ln(BOVESPA_{t-1})$ and $Z_t = \ln(DOWJONES_t) - \ln(DOWJONES_{t-1})$. The log return of these series is displayed in Figure 11.

In the first stage of the algorithm, the identified number of regimes is 3 with probability 1, that is, in all of the 1,000 iterations of the Gibbs sampler, the sampled value of L is 3. The estimated thresholds are -0.0051 and 0.0054, with respective credible intervals (-0.0242, 0.0099) and (-0.0081, 0.0226). The histograms of the sampled thresholds are shown in Figure 12. Observing the values of the two thresholds, we could name the three regimes as *large negative return in Dow Jones*, *small return in Dow Jones* and *large positive return in Dow Jones*.

Once the number of regimes and thresholds were identified, we proceeded with the identification of the autoregressive orders using another Gibbs sampler. In Table 11, we show the posterior probabilities for all possible values of the autoregressive orders, suggesting that $\hat{k}_1 = \hat{k}_2 = 1$ and $\hat{k}_3 = 3$.

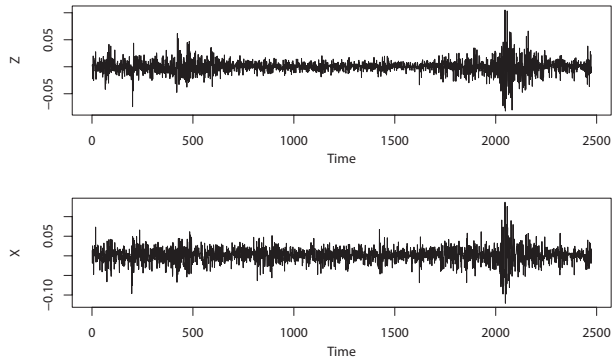

FIGURE 11: Finance data. X: daily log return of the Dow Jones industrial average and Z: daily log return of the BOVESPA index.

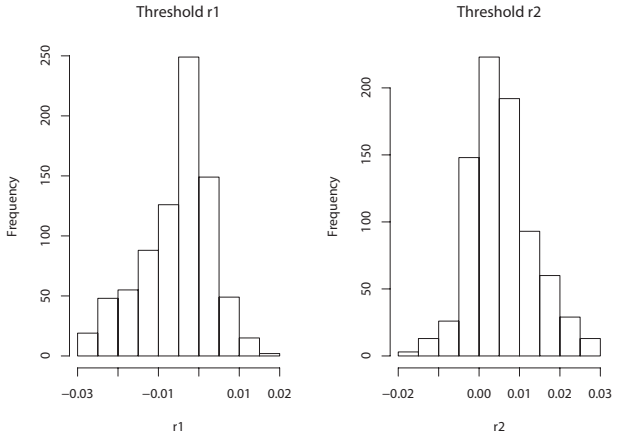

FIGURE 12: Histograms of the sampled thresholds for the finance data.

TABLE 11: Posterior probability density of the autoregressive orders in each regime.

	Regime		
autoregressive order	1	2	3
0	0	0	0
1	1	1	0
2	0	0	0
3	0	0	1

Finally with all the structural parameters identified, we estimated the non-structural parameters, leading to the following model for the data:

$$X_t = \begin{cases} -0.0106 + 0.1296X_{t-1} + 0.0355e_t & \text{if } Z_t < -0.0051 \\ 0.0009 + 0.0099X_{t-1} + 0.0259e_t & \text{if } -0.0051 \leq Z_t < 0.0054 \\ 0.0128 - 0.0054X_{t-1} - 0.0201X_{t-2} & \\ \qquad -0.0917X_{t-3} + 0.0354e_t & \text{if } Z_t > 0.0054 \end{cases}, \quad (19)$$

where the degrees of freedom of the process $\{e_t\}$ are estimated to be 2.3.

The credible intervals of the parameters in (19) are given in

TABLE 12: 95% credible intervals for the parameters in the model (19).

Regime	$a_i^{(j)}$				$h^{(j)}$
1	(-0.012, -0.009)	(0.062,0.189)			(0.033,0.037)
2	(0.000, 0.001)	(-0.030,0.052)			(0.025,0.027)
3	(0.012, 0.014)	(-0.068,0.062)	(-0.085,0.040)	(-0.152,-0.023)	(0.033,0.037)

Moreno (2010) found a similar TAR model for the same data using the approach of Nieto (2005), that is, assuming the Gaussian distribution for the noise process. The TAR model found in Moreno (2010) is:

$$X_t = \begin{cases} -0.0127 + 0.111X_{t-1} - 0.068X_{t-2} + 0.0198e_t & \text{if } Z_t < -0.0054 \\ 0.00068 + 0.0137e_t & \text{if } -0.0054 \leq Z_t < 0.0057 \\ 0.0135 - 0.0837X_{t-1} - 0.0684X_{t-2} - 0.1687X_{t-3} & \\ \qquad -0.0633X_{t-4} + 0.0191e_t & \text{if } Z_t > 0.0057 \end{cases} \tag{20}$$

We can observe that the number of regimes is the same and that the two thresholds are quite similar. Also, the type II conditional variance in the first and third regimes are similar and larger than the conditional variance in the second regime, that is, the series of log return of BOVESPA is more stable when the Dow Jones index is relatively stable. On the other hand, in spite of the fact that the autoregressive orders are different in the two models, the autoregressive coefficients in common are also similar. However, we note a better fit with t noise since the DIC (deviance information criterion) is 5,485.377 for the TAR model with Gaussian error and 4338.799 for TAR with t error (up to a constant).

In order to check the appropriateness of the model, we use the CUSUM and CUSUMSQ plot of standardized residuals, shown in Figure 13, which suggests that the fitted model (19) is appropriate.

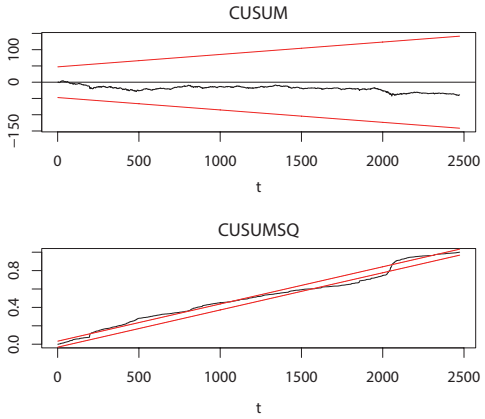

FIGURE 13: CUSUM and CUSUMSQ plot of the standardized residuals of the model (19).

As shown in the work of Moreno (2010), the TAR model with Gaussian noise (20), in spite of showing good performance in the CUSUM and CUSUMSQ plots

of the residuals, the squared residuals show large autocorrelations, which is a disadvantage compared to the family of GARCH models where the residuals show strong evidence of independence (see the ACF of residuals and squared residuals of a GARCH(1,1) model in Figure 14). In Figure 15, we can observe the ACF of the residuals and squared residuals, and obviously the squared residuals of the TAR model with t distributed noise still exhibit the same problem as the TAR model with Gaussian noise. Although the TAR model seems to fail in capturing all the structure of dependence in the data, Nieto & Moreno (2013) found the expression for the conditional variance $Var(X_t|x_{t-1}, \ldots, x_1)$ in a TAR model, making this class of model comparable with the GARCH models. In Figure 16, we show the conditional variance $Var(X_t|x_{t-1}, \ldots, x_1)$ in the TAR model 19, as well as the GARCH(1,1) model; we can see that the general behaviour is similar for the two models, although the bottom line in the TAR model is around 0.0076, while in the GARCH model it around 0.0003.

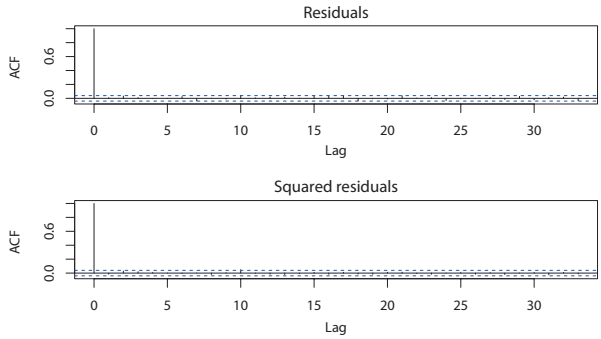

FIGURE 14: ACF of residuals and squared residuals of the model GARCH(1,1).

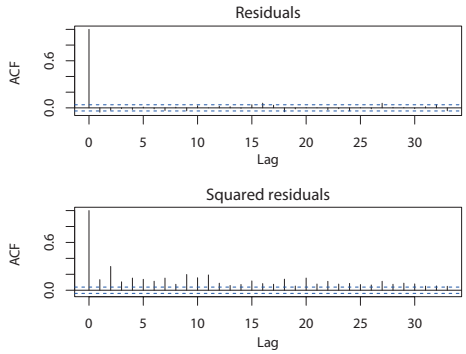

FIGURE 15: ACF of residuals and squared residuals of the model (19).

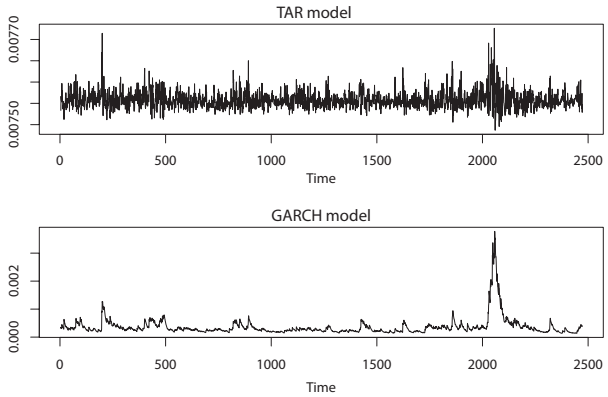

FIGURE 16: Conditional variance of the TAR model and the GARCH model.

8. Conclusion

In this work, we proposed a new family of TAR models: the TAR models with t-distributed noise process with a three-stage procedure consisting of: (1) identifying the number of regimes and the corresponding thresholds, (2) identifying the autoregressive order in each regime, and (3) estimating non-structural parameters, i.e., the autoregressive coefficients and the type II conditional variance in each regime, and other parameters that each particular model may contain. The performance of the developed methodology is satisfactory in simulated data, however, the GARCH models seems to better capture the heterocedastic aspect contained in financial data. In future investigation the GARCH models may be used together with the TAR model.

[Received: October 2013 — Accepted: November 2014]

References

Briñez, A. & Nieto, F. (2005), 'Fitting a nonlinear model to the precipitation variable in a Colombian Hydrological/Meteorological station', *Revista Colombiana de Estadística* **28**, 113–124.

Carlin, B. P. & Chib, S. (1995), 'Bayesian model choice via Markov Chain Monte Carlo Methods', *Journal of the Royal Statistical Society. Serie B* **37**(3), 473–484.

Chen, H., Chong, T. T. & Bai, J. (2012), 'Theory and applications of TAR model with two threshold variables', *Econometric Reviews* **31**, 142–170.

Congdon, P. (2001), *Bayesian Statistical Modeling*, John Wiley & Sons, New York.

Geweke, J. (1992), Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, *in* 'Bayesian Statistics', University Press, pp. 169–193.

Moreno, E. (2010), Modelos TAR en series de tiempo financieras, Master's thesis, Universidad Nacional de Colombia.

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

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

Nieto, F. H. & Moreno, E. (2013), A note on the specification of conditional heteroscedasticity using a TAR model, Technical Report RI21, Universidad Nacional de Colombia.

Nieto, F. H., 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.

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

Plummer, M., Best, N., Cowles, K. & Vines, K. (2006), 'Coda: Convergence diagnosis and output analysis for mcmc', *R News* **6**(1), 7–11. *http://CRAN.R-project.org/doc/Rnews/

Sáfadi, T. & Morettin, P. (2000), 'Bayesian analysis of thresholds autoregressive moving average models', *The Indian Journal of Statistics* **62**, 353–371.

Tong, H. (1978), On a Threshold Model, *in* C. H. Chen, ed., 'Pattern Recognition and Signal Processing', Sijthoff & Noordhoff, Netherlands, pp. 575–586.

Tsay, R. S. (1989), 'Testing and modeling threshold autoregressive processes', *Journal of American Statistical Association* **84**, 231–240.

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

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

Watanabe, T. (2001), 'On sampling the degree-of-freedom of Student's-t disturbances', *Statistics & Probability Letters* **52**, 177–181.

Xia, Q., Liu, L., Pan, J. & Liang, R. (2012), 'Bayesian analysis of two-regime threshold autoregressive moving average model with exogenous inputs', *Communications in Statistics - Theory and Methods* **41**, 1089–1104.

Zhang, H. (2012), 'Estimación de los modelos TAR cuando el proceso del ruido sigue una distribución t', *Comunicaciones en Estadística* **4**(2), 109–119.