Statistica Sinica **31** (2021), 2257-2274
doi:https://doi.org/10.5705/ss.202020.0411

MULTIVARIATE HYSTERETIC AUTOREGRESSIVE MODELS

Ruey S. Tsay

University of Chicago

Abstract: This paper proposes a multivariate hysteretic autoregressive model with multiple threshold variables for modeling nonlinear time series. The proposed model encompasses the two-regime multivariate threshold autoregressive model and the hysteretic autoregressive model as special cases. A special feature of the proposed model is that it employs multiple threshold variables, each with a single threshold value. The resulting model is more flexible, yet parsimonious, than several multivariate nonlinear time series models available in the literature. The paper also studies some basic properties of the proposed model, uses a conditional least squares estimation, and proposes a modeling procedure. Finally, we demonstrate applications of the proposed model using simulated and real examples.

Key words and phrases: Hysteresis, least squares estimation, Markov chain, nonlinear model, threshold variable.

1. Introduction

The need to analyze multivariate nonlinear time series increases with the availability of big data. Yet research on such time series is relatively scarce compared with that on its scalar counterpart. This lack of research can be attributed to multiple reasons, including the complexity of the dynamic dependence in multiple series and high demand in terms of computation. It is then highly desirable to consider a widely applicable, yet relatively simple multivariate nonlinear time series model. This study marks some progress in this direction, focusing on parametric models.

For multiple nonlinear time series, multivariate threshold autoregressive (MT AR) models and Markov-switching models (MSMs) are perhaps the most commonly used parametric models in the literature; see, for instance, Tsay and Chen (2019) and the references therein. These two classes of models are mixtures of linear models, but they differ in terms of their mixing mechanisms. A MTAR model uses a threshold variable to govern the mixture. As such, the switching

Corresponding author: Ruey S. Tsay, Booth School of Business, University of Chicago, 5807 S. Woodlawn Avenue, Chicago, IL, 60637, USA. E-mail: ruey.tsay@chicagobooth.edu.

between linear models is deterministic once the threshold variable is observed; see Tong (1978) and Tong (1990). On the other hand, the switching between the linear models of an MSM is stochastic, driven by the transition probability matrix of an underlying Markov chain; see Hamilton (1989), among others. There are advantages and disadvantages for each type of model. For instance, threshold models are easier to estimate once the threshold variable is given, but an MSM may enjoy good interpretations in economic applications. This study examines threshold models.

A criticism of threshold autoregressive (TAR) models is that the threshold often marks a discontinuity in their conditional expectation. To address this weakness, Li et al. (2015) propose a hysteretic autoregressive (HAR) model that explores using hysteresis to govern the switching between linear models. Here, we extend the HAR model in two directions. First, we consider multivariate time series, and second, we employ multiple threshold variables. In TAR modeling, the choice of a proper threshold variable is always challenging. Tsay (1989) uses the p-value of a threshold nonlinearity test to select the delay d of a self-exciting TAR (SETAR) model. Xia, Li and Tong (2007) propose nonparametric methods for selecting the threshold variable, and Wu and Chen (2007) combine Markov switching with generalized linear models to estimate the threshold variable, which is a linear combination of several variables. These works all assume *a priori* that there is a single threshold variable. However, the use of multiple threshold variables has been found to be useful in empirical studies, such as that of Tiao and Tsay (1994). Thus, our goal is to investigate the use of multiple threshold variables. In addition, to keep the model relatively simple, we consider only a single threshold for each threshold variable.

The main difficulty when using multiple threshold variables is that they often result in many regimes, which means the resulting TAR models may contain many parameters. This is particularly so for multivariate time series. For example, consider a k-dimensional MTAR model of order p. An increase of one regime would add $k(kp+1)+k(k+1)/2$ parameters in the mean equation and residual covariance matrix. For a moderate $k=5$ with $p=2$, introducing an additional threshold would add 70 parameters. Here, we show that the proposed multivariate hysteretic autoregressive (MHAR) model overcomes this over-parameterization problem because it has only two regimes, regardless of the number of threshold variables used. This is achieved by adopting the idea of hysteresis, while reducing the number of thresholds for each threshold variable to one. The HAR model of Li et al. (2015) uses a single threshold variable with lower and upper thresholds. Consequently, our generalization of the HAR model employs a simplifying switching mechanism.

The remainder of the paper is organized as follows. We introduce the proposed model and study its properties in Section 2. Section 3 focuses on estimation and the limiting properties of the conditional least squares estimates. Section 4 considers a modeling procedure, and Section 5 contains simulation studies and an empirical application. Section 6 concludes the paper.

2. The Proposed Model

Let $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{kt})'$ be a k-dimensional time series of interest and $\boldsymbol{x}_t = (x_{1t}, \ldots, x_{mt})'$ be m-dimensional observable threshold variables. For simplicity, we assume that both $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$ are continuous random vectors and both k and m are finite. In addition, the threshold vector $\boldsymbol{x}_t$ is stationary. For each threshold variable x_{it}, let d_i be a positive integer denoting the delay and r_i be the threshold. That is, $x_{i,t-d_i}$ is the observed value used in model switching. Details are given below. Let $\boldsymbol{d} = (d_1, \ldots, d_m)'$ be the vector of delays. For a given time index t, denote the threshold vector as $\boldsymbol{x}_t(\boldsymbol{d}) = (x_{1,t-d_1}, \ldots, x_{m,t-d_m})'$, and partition the m-dimensional Euclidean space $\Re^m$ as follows: Let $\Omega_{1t} = \{\boldsymbol{x}_t(\boldsymbol{d}) | x_{i,t-d_i} \le r_i, 1 \le i \le m\}$, $\Omega_{2t} = \{\boldsymbol{x}_t(\boldsymbol{d}) | x_{i,t-d_i} > r_i, 1 \le i \le m\}$, and $\Omega_{3t} = \Re^m - \Omega_{1t} - \Omega_{2t} = \{\boldsymbol{x}_t(\boldsymbol{d}) | (x_{i,t-d_i} > r_i) \cap (x_{j,t-d_j} \le r_j),\ 1 \le i, j \le m\}$. Thus, Ω_{1t} is the region in which all threshold variables are less than their thresholds, Ω_{2t} denotes the region in which all threshold variables exceed their thresholds, and Ω_{3t} is the complement of $\Omega_{1t} \cup \Omega_{2t}$. The proposed MHAR model with regime indicator R_t is given by

$$\boldsymbol{y}_t = \begin{cases} \boldsymbol{\phi}_0 + \sum_{i=1}^{p_1} \boldsymbol{\phi}_i \boldsymbol{y}_{t-i} + \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\epsilon}_t, & R_t = 1, \\ \boldsymbol{\psi}_0 + \sum_{i=1}^{p_2} \boldsymbol{\psi}_i \boldsymbol{y}_{t-i} + \boldsymbol{\Sigma}_2^{1/2} \boldsymbol{\epsilon}_t, & R_t = 0, \end{cases} \quad R_t = \begin{cases} 1, & \boldsymbol{x}_t(\boldsymbol{d}) \in \Omega_{1t}, \\ 0, & \boldsymbol{x}_t(\boldsymbol{d}) \in \Omega_{2t}, \\ R_{t-1}, & \text{otherwise}, \end{cases} \tag{2.1}$$

where p_1 and p_2 are non-negative integers, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are $k \times k$ positive-definite matrices, $\{\boldsymbol{\epsilon}_t\}$ is a sequence of independent and identically distributed (i.i.d.) random vectors with mean zero and $\text{cov}(\boldsymbol{\epsilon}_t) = \boldsymbol{I}_k$, the $k \times k$ identity matrix, $\boldsymbol{\phi}_0$ and $\boldsymbol{\psi}_0$ are k-dimensional constant vectors, and $\boldsymbol{\phi}_i$ and $\boldsymbol{\psi}_j$ are $k \times k$ real-valued matrices. In model (2.1), $\boldsymbol{\epsilon}_t$ is independent of $\boldsymbol{y}_{t-j}$ and $\boldsymbol{x}_{t-j}$, for $j > 0$. If $\boldsymbol{x}_t$ is a subset of $\boldsymbol{y}_t$, then model (2.1) is a self-exciting MHAR model. Clearly, model (2.1) reduces to the conventional two-regime MTAR model if $m = 1$, that is, there is a single threshold variable. If $k = 1$, $m = 2$ with $x_{1t} = x_{2t} = y_{1t}$, $d_1 = d_2$, but $r_1 < r_2$, then model (2.1) reduces to the HAR model of Li et al. (2015). Thus, the proposed model encompasses both conventional two-regime MTAR and HAR models as special cases. It also shares a similar switching mechanism to that of the HAR model. On the other hand, model (2.1) is not a straightforward generalization of HAR models, because the latter uses a single threshold variable with two thresholds, whereas the former uses multiple threshold variables, each with a single threshold. From a hysteresis point of view, Ω_{1t} and Ω_{2t} of model (2.1) represent the two well-defined regimes (one or zero), whereas Ω_{3t} signifies a transition period. Under the proposed model, the transition between regimes one and zero occurs only when all the signals switch. In other words, under the proposed MHAR model, switching is a unanimous decision of the threshold variables.

The use of multiple threshold variables can be justified in many ways. Consider, for instance, that $\boldsymbol{y}_t$ consists of the quarterly gross domestic product (GDP) and unemployment rates of an economy. It is then understandable that the dynamic structure of $\boldsymbol{y}_t$ would depend on the status of the economy, which is often determined jointly by multiple economic indicators, such as the growth rate of GDP, inflation rate, and change in productivity.

Some remarks on the proposed MHAR model are in order. First, under the proposed model, the threshold vector $\boldsymbol{x}_t$ may contain a threshold variable twice with the same delay, for example, $x_{1t} = x_{2t}$ and $d_1 = d_2$, but it cannot do so more than twice. This is due to the inherent nature of the HAR model, which cannot have more than two different thresholds. Second, for the MHAR model to be useful in practice, the number of threshold variables (i.e., the dimension m of $\boldsymbol{x}_t$) cannot be too large, because the model uses a unanimous decision in model switching, which becomes increasingly difficult to achieve as m increases.

2.1. Ergodicity of the proposed model

Similarly to the HAR model, the regime indicator R_t of model (2.1) follows the model

$$
\begin{aligned}
R_t &= I(\Omega_{1t}) + I(\Omega_{3t})R_{t-1} \\
&= I(\Omega_{1t}) + \sum_{j=0}^{\infty}\prod_{i=0}^{j} I(\Omega_{3,t-i})I(\Omega_{1,t-j-1}),
\end{aligned}
$$

almost surely, where $I(A)$ is the indicator for the set A. Thus, for a well-defined nondegenerated MHAR model, the regime indicator R_t depends on all past values of the threshold variables.

Let $p = \max\{p_1, p_2\}$ and $\boldsymbol{\phi}_i = \mathbf{0}$ if $i > p_1$, and $\boldsymbol{\psi}_i = \mathbf{0}$ if $i > p_2$. Assume $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$. Define $\boldsymbol{Y}_t = (\boldsymbol{y}_t', \ldots, \boldsymbol{y}_{t-p+1}', R_t)'$, and $\boldsymbol{e}_t = [(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t)', 0, \ldots, 0]'$, and let $\boldsymbol{M}_{0t} = [\boldsymbol{m}'_{0t}, 0, \ldots, 0, I(\Omega_{1t})]'$ be $(kp+1)$-dimensional random vectors, where $\boldsymbol{m}_{0t} = \boldsymbol{\phi}_0 I(D_t) + \boldsymbol{\psi}_0 I(D_t^c)$, with $D_t = \Omega_{1t} \cup \{\Omega_{3t} \cap (R_{t-1} = 1)\}$ and D_t^c being the complement of D_t. Furthermore, let $\boldsymbol{\Phi}$ be the $kp \times kp$ comparison matrix of the matrix polynomial of Regime 1 of model (2.1), that is,

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_{p-1} & \boldsymbol{\phi}_p \\ \boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I} & \boldsymbol{0} \end{bmatrix}, \tag{2.2}$$

where $\boldsymbol{0}$ denotes the $k \times k$ zero matrix. Similarly, let $\boldsymbol{\Psi}$ be the $kp \times kp$ companion matrix of the polynomial matrix of regime 0 of model (2.1). Define $\boldsymbol{M}_t = \boldsymbol{\Phi} I(D_t) + \boldsymbol{\Psi} I(D_t^c)$ and

$$\boldsymbol{M}_{1t} = \begin{bmatrix} \boldsymbol{M}_t & \boldsymbol{0}_1 \\ \boldsymbol{0}'_1 & I(\Omega_{3t}) \end{bmatrix},$$

where $\boldsymbol{0}_1$ is the kp-dimensional vector of zeros. It can be verified that $\boldsymbol{Y}_t = \boldsymbol{M}_{0t} + \boldsymbol{M}_{1t}\boldsymbol{Y}_{t-1} + \boldsymbol{e}_t$. Therefore, in this case, $\boldsymbol{Y}_t$ is a Markov chain, and one can apply the results of Chan and Tong (1985) to derive a sufficient condition for its geometric ergodicity. Let a_{ij} be the (i,j)th element of the real-valued matrix $\boldsymbol{A}$.

Theorem 1. *Consider model* (2.1). *Suppose that $\boldsymbol{\epsilon}_t$ has a continuous density function that is positive everywhere in $\Re^k$ and $E(\|\boldsymbol{\epsilon}_t\|) < \infty$. If there exists a $kp \times kp$ matrix $\boldsymbol{C} = [c_{ij}]$ satisfying $c_{ij} \geq \max\{|\Phi_{ij}|, |\Psi_{ij}|\}$ for all (i,j), such that all eigenvalues of $\boldsymbol{C}$ are less than one in modulus, then $\boldsymbol{Y}_t$ is geometrically ergodic and, hence, $\boldsymbol{y}_t$ is geometrically ergodic.*

This theorem is given in Example 4 of Chan and Tong (1985). The case of $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$ can also be obtained using the results of the aforementioned reference under the same condition as Theorem 1.

3. Estimation

We apply the conditional least squares method to estimate the proposed MHAR model. For ease of notation, we assume $p_1 = p_2$ and rewrite model (2.1) as

$$\boldsymbol{y}_t = \begin{cases} \boldsymbol{\phi}'\boldsymbol{z}_t + \boldsymbol{e}_{1t}, & R_t = 1, \\ \boldsymbol{\psi}'\boldsymbol{z}_t + \boldsymbol{e}_{2t}, & R_t = 0, \end{cases} \tag{3.1}$$

where $\boldsymbol{z}_t = (1, \boldsymbol{y}'_{t-1}, \ldots, \boldsymbol{y}'_{t-p})'$, $\boldsymbol{\phi}' = [\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_p]$, $\boldsymbol{\psi}' = [\boldsymbol{\psi}_0, \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_p]$, and $\boldsymbol{e}_{it} = \boldsymbol{\Sigma}_i^{1/2}\boldsymbol{\epsilon}_t$, for $i = 1, 2$. For simplicity, we consider only the case that the dimensions k and m are finite and fixed, and assume that for each threshold variable x_{it}, there exists a bounded interval $[a_i, b_i]$ such that $a_i < r_i < b_i$ and $0 < P(x_{it} \in [a_i, b_i]) < 1$, for $i = 1, \ldots, m$. Let $\boldsymbol{d} = (d_1, \ldots, d_m)'$ be the m-dimensional vector of delays and $\boldsymbol{r} = (r_1, \ldots, r_m)'$ be the vector of thresholds. Let $d_{max} = \max\{d_i | i = 1, \ldots, m\}$. We further assume that d_{max} is a known positive integer so that the delay vector $\boldsymbol{d} \in \{1, \ldots, d_{max}\}^m$, which is a finite set.

Let $\boldsymbol{\omega} = (\boldsymbol{\theta}', \boldsymbol{r}', \boldsymbol{d}')'$ be the parameter vector of model (3.1), except $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, where $\boldsymbol{\theta} = (\text{vec}(\boldsymbol{\phi})', \text{vec}(\boldsymbol{\psi})')'$ is the collection of coefficient parameters, and let $\boldsymbol{\Theta}$ be a compact subset of $R^{2k(kp+1)}$. Denote the true parameters by $\boldsymbol{\omega}_o = (\boldsymbol{\theta}_o', \boldsymbol{r}_o', \boldsymbol{d}_o')'$. We assume that $\boldsymbol{\theta}_o$ is an interior point of $\boldsymbol{\Theta}$, and that $a_i < r_{io} < b_i$ and $d_{io} \in \{1, \ldots, d_{max}\}$, for $i = 1, \ldots, m$.

Let $n_0 = \max\{p, d_{max}\}$. Given the data $\{\boldsymbol{y}_t, \boldsymbol{x}_t | t = 1, \ldots, T\}$, we assume, for simplicity, that R_{n_0} is known, and consider the sum of the squared errors $L_T(\boldsymbol{\omega}) = \sum_{t=n_0+1}^{T} \boldsymbol{e}_t(\boldsymbol{\omega})' \boldsymbol{e}_t(\boldsymbol{\omega})$, where

$$\boldsymbol{e}_t(\boldsymbol{\omega}) = (\boldsymbol{y}_t - \boldsymbol{\phi}' \boldsymbol{z}_t) R_t(\boldsymbol{r}, \boldsymbol{d}) + (\boldsymbol{y}_t - \boldsymbol{\psi}' \boldsymbol{z}_t)[1 - R_t(\boldsymbol{r}, \boldsymbol{d})], \tag{3.2}$$

where $R_t(\boldsymbol{r}, \boldsymbol{d})$ signifies that the regime indicator depends on $\boldsymbol{r}$ and $\boldsymbol{d}$. The conditional least squares estimate (CLSE) of $\boldsymbol{\omega}$ is defined as

$$\widehat{\boldsymbol{\omega}} = \operatorname*{argmin}_{\boldsymbol{\omega}} L_T(\boldsymbol{\omega}).$$

In practice, T is fixed and the ith threshold r_i assumes only some ordered statistics of $\{x_{i,t-d_i} | t = n_0 + 1, \ldots, T\}$; as such, the parameter space for $\boldsymbol{r}$ and $\boldsymbol{d}$ contains only a finite number of possible values. Therefore, $\widehat{\boldsymbol{\omega}}$ can be obtained using the following steps:

1. For given $\boldsymbol{r}$ and $\boldsymbol{d}$ and under the assumption of knowing the initial value $R_{n_0}(\boldsymbol{r}, \boldsymbol{d})$, obtain the estimate of $\boldsymbol{\theta}$ as

$$\widehat{\theta}(\boldsymbol{r}, \boldsymbol{d}) = \operatorname*{argmin}_{\boldsymbol{\theta}} L_T(\boldsymbol{\theta}, \boldsymbol{r}, \boldsymbol{d}).$$

   Denote the resulting sum of the squared errors by $L_T(\widehat{\boldsymbol{\theta}}, \boldsymbol{r}, \boldsymbol{d})$.

2. Because $L_T(\widehat{\boldsymbol{\theta}}, \boldsymbol{r}, \boldsymbol{d})$ assumes only a finite number of possible values, apply a grid search to obtain estimates of $\boldsymbol{r}$ and $\boldsymbol{d}$. That is, $(\widehat{\boldsymbol{r}}, \widehat{\boldsymbol{d}}) = \arg\min_{\boldsymbol{r},\boldsymbol{d}} L_T(\widehat{\boldsymbol{\theta}}, \boldsymbol{r}, \boldsymbol{d})$.

3. Applying the plug-in method, one has $\widehat{\boldsymbol{\omega}} = [\widehat{\boldsymbol{\theta}}(\widehat{\boldsymbol{r}}, \widehat{\boldsymbol{d}})', \widehat{\boldsymbol{r}}', \widehat{\boldsymbol{d}}']'$.

Denoting $\widehat{R}_t$ as the estimated regime indicator based on the CLSE, we define the residuals as $\hat{\boldsymbol{e}}_{1t} = (\boldsymbol{y}_t - \widehat{\boldsymbol{\phi}}'\boldsymbol{z}_t)\widehat{R}_t$ and $\hat{e}_{2t} = (\boldsymbol{y}_t - \widehat{\boldsymbol{\psi}}'\boldsymbol{z}_t)(1 - \widehat{R}_t)$, and estimate the covariance matrices $\boldsymbol{\Sigma}_i$ $(i = 1, 2)$ using

$$\widehat{\boldsymbol{\Sigma}}_1 = \frac{1}{T_1}\sum_{t=n_0+1}^{T} \hat{\boldsymbol{e}}_{1t}\hat{\boldsymbol{e}}_{1t}', \quad \widehat{\boldsymbol{\Sigma}}_2 = \frac{1}{T_2}\sum_{t=n_0+1}^{T} \hat{\boldsymbol{e}}_{2t}\hat{\boldsymbol{e}}_{2t}', \tag{3.3}$$

where $T_1 = \sum_{t=n_0+1}^{T} \widehat{R}_t$ and $T_2 = T - n_0 - T_1$.

Some remarks are in order. First, the assumption that R_{n_0} is known should have negligible impact when the sample size T is sufficiently large. As a matter of fact, because R_{n_0} is either one or zero, one can entertain both cases if needed, and choose the final CLSE based on the resulting smaller sum of the squared errors. Second, for given $\boldsymbol{r}$ and $\boldsymbol{d}$, define $\boldsymbol{X}_t = [\boldsymbol{z}_t' R_t(\boldsymbol{r}, \boldsymbol{d}), \boldsymbol{z}_t'\{1 - R_t(\boldsymbol{r}, \boldsymbol{d})\}]'$. Then, taking the transpose of model (3.1), we see that the CLSE of the coefficient matrices $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ can be written as

$$\begin{bmatrix} \widehat{\boldsymbol{\phi}}(\boldsymbol{r}, \boldsymbol{d}) \\ \widehat{\boldsymbol{\psi}}(\boldsymbol{r}, \boldsymbol{d}) \end{bmatrix} = \left(\sum_{t=n_0+1}^{T} \boldsymbol{X}_t\boldsymbol{X}_t'\right)^{-1} \left(\sum_{t=n_0+1}^{T} \boldsymbol{X}_t\boldsymbol{y}_t\right). \tag{3.4}$$

To study the asymptotic properties of the CLSE, we need some additional assumptions. For simplicity, we focus on the case that $\boldsymbol{x}_t = \boldsymbol{y}_t$, so that $m = k$ (fixed). In other words, we study the asymptotic properties of a finite-dimensional self-exciting MHAR model in this section.

Assumption 1. *Assume that $\boldsymbol{\phi} \neq \boldsymbol{\psi}$ and that for each threshold variable y_{it}, there exists a bounded interval $[a_i, b_i]$ such that $0 < P(y_{it} \in [a_i, b_i]) < 1$. In addition, the innovation process $\boldsymbol{\epsilon}_t$ has a bounded, continuous, and positive density function on $\Re^k$.*

Theorem 2. *Assume that the process $\boldsymbol{y}_t$ of model (2.1) is strictly stationary and ergodic with $E(|y_{it}y_{jt}|^{1+\delta}) < \infty$, for some $\delta > 0$ and for $i, j = 1, \ldots, k$. In addition, Assumption 1 holds. Then, $\widehat{\boldsymbol{\omega}} \to \boldsymbol{\omega}_o$ and $\widehat{\boldsymbol{\Sigma}}_i \to \boldsymbol{\Sigma}_{io}$, for $i = 1, 2$, almost surely, as $T \to \infty$, where $\boldsymbol{\omega}_o$ and $\boldsymbol{\Sigma}_{io}$ denote the true parameter values.*

This theorem is a generalization of Theorem 2 of Li et al. (2015) and can be shown by the standard argument for strong consistency. With $m = k$ fixed, the choices of $\boldsymbol{d}$ are finite, so that $\widehat{\boldsymbol{d}}$ obtained by a grid search will equal $\boldsymbol{d}_o$ when the sample size T is sufficiently large. Therefore, we assume, for simplicity, that $\boldsymbol{d}$ is given and focus on the other parameters. We recommend a simpler procedure to select $\boldsymbol{d}$ in the next section.

Assumption 2. *The process $\boldsymbol{y}_t$ of model (2.1) is strictly stationary with $E(|y_{it}y_{jt}$*

$y_{ut}y_{vt}|^{1+\delta}) < \infty$, *for some* $\delta > 0$ *and for* $i, j, u, v = 1, \ldots, k$, *and the process* $\boldsymbol{\epsilon}_t$ *satisfies* $E(|\epsilon_{it}\epsilon_{jt}\epsilon_{ut}\epsilon_{vt}|) < \infty$, *for* $i, j, u, v = 1, \ldots, k$.

For a given time index t, we say that $\boldsymbol{y}_t(\boldsymbol{d})$ is in the hysteresis zone if $(y_{1,t-d_1}, \ldots, y_{k,t-d_k})' \in \Omega_{3t} = \Re^k - \Omega_{1t} - \Omega_{2t}$, where Ω_{1t} consists of $y_{i,t-d_i} \le r_i$ for $i = 1, \ldots, k$, and Ω_{2t} consists of $y_{i,t-d_i} > r_i$ for $i = 1, \ldots, k$. Next, without loss of generality, assume all delays satisfy $d_i \le p$.

Assumption 3. *The vector autoregressive function of* $\boldsymbol{y}_t$ *is discontinuous in the hysteresis zone. Specifically, there exist p k-dimensional vectors, given below: For* $j = p-1, \ldots, 0$,

i. $\boldsymbol{y}_j^*$ *is a constant vector in* $\Re^k$ *if* $j \ne d_i$, *for all* i.

ii. $\boldsymbol{y}_j^*$ *is a partially random vector if* $j = d_i$, *for some* i. *In this case,* $y_{i,j}^*$ *is a random variable with its value in the hysteresis zone,*

such that $\boldsymbol{z}'(\boldsymbol{\phi}_o - \boldsymbol{\psi}_o) \ne \boldsymbol{0}$, *where* $\boldsymbol{z} = [1, (\boldsymbol{y}_{p-1}^*)', \ldots, (\boldsymbol{y}_0^*)']'$ *and* $\boldsymbol{\phi}_o$ *and* $\boldsymbol{\psi}_o$ *are the true parameter values.*

Consider the Markov chain process $\boldsymbol{Y}_t = (\boldsymbol{y}_t', \ldots, \boldsymbol{y}_{t-p+1}', R_t)'$ of Section 2. Denote its n-step transition probability function by $P^n(\boldsymbol{y}, A)$, where $\boldsymbol{y} \in \Re^{kp} \times \{0,1\}$, $A \in B_{kp} \times U$, B_{kp} is the class of Borel sets of $\Re^{kp}$, and $U = \{\emptyset, \{0\}, \{1\}, \{0,1\}\}$.

Assumption 4. *The process* $\boldsymbol{Y}_t$ *admits a unique invariant measure* $\pi(.)$ *such that there exist* $K > 0$ *and* $0 \le \rho < 1$, *for any* $\boldsymbol{y} \in \Re^{kp} \times \{0,1\}$ *and any* n, $\|P^n(\boldsymbol{y}, \cdot) - \pi(\cdot)\|_v \le K(1 + \|\boldsymbol{y}\|)\rho^n$, *where* $\|\cdot\|_v$ *and* $\|\cdot\|$ *are, respectively, the total variation norm and the Euclidean norm.*

Under Assumption 4, $\{\boldsymbol{Y}_t\}$ is said to be V-uniformly ergodic with $V(\boldsymbol{y}) = K(1 + \|\boldsymbol{y}\|)$, which is stronger than geometric ergodicity; see Meyn and Tweedie (1993).

Theorem 3. *Consider the self-exciting MHAR model* $\boldsymbol{y}_t$ *in* (2.1). *If Assumptions* 1 – 4 *hold, then*

i. $T(\widehat{\boldsymbol{r}} - \boldsymbol{r}) = O_p(1)$,

ii. $T^{1/2} \sup_{T\|\boldsymbol{r}-\boldsymbol{r}_o\| \le C} \|\widehat{\boldsymbol{\theta}}(\boldsymbol{r}) - \widehat{\boldsymbol{\theta}}(\boldsymbol{r}_o)\| = o_p(1)$ *for any fixed* $0 < C < \infty$, *where* $\boldsymbol{\theta} = [vec(\boldsymbol{\phi})', vec(\boldsymbol{\psi})']'$ *and* $\widehat{\boldsymbol{\theta}}(\boldsymbol{r})$ *is given by Equation* (3.4).

In addition,

$$T^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o) \to_D N[\boldsymbol{0}, diag(\boldsymbol{\Sigma}_1 \otimes \boldsymbol{G}_1^{-1}, \boldsymbol{\Sigma}_2 \otimes \boldsymbol{G}_2^{-1})],$$

where $\to_D$ *denotes convergence in distribution,* $\boldsymbol{G}_1 = E(\boldsymbol{X}_t\boldsymbol{X}_t'R_t)$*, and* $\boldsymbol{G}_2 = E[\boldsymbol{X}_t\boldsymbol{X}_t'(1-R_t)]$*, with* $\boldsymbol{X}_t$ *given in Equation* (3.4)*.*

Consider Equation (3.3). Let $\widehat{\mathbf{S}}_i = T_i\widehat{\boldsymbol{\Sigma}}_i$. The limiting properties of the residual covariance matrices are given below.

Theorem 4. *Consider the self-exciting MHAR model* $\boldsymbol{y}_t$ *in* (2.1)*. (a) If Assumptions* 1–4 *hold, then Var[vec(*$\widehat{\boldsymbol{\Sigma}}_i$*)]* $\to 0$ *as* $T \to \infty$*. (b) In addition, if the innovations* $\boldsymbol{\epsilon}_t$ *are multivariate Gaussian, then* $T_i\widehat{\boldsymbol{\Sigma}}_i$*, conditioned on* $\boldsymbol{\Sigma}_i$*, follows asymptotically a Wishart distribution with degrees of freedom* $T_i - kp - 1$*.*

Part (a) of Theorem 4 can be shown using the same method as that for Corollary 3.2.1 of Lütkepohl (2005). Part (b) follows from the consistency results of Theorems 2 and 3 and the properties of the multivariate Gaussian distribution.

Finally, similarly to Li et al. (2015), under Assumptions $1-4$, one can derive the limiting distribution of $\widehat{\boldsymbol{r}}$ and the asymptotic independence between $T(\widehat{\boldsymbol{r}}-\boldsymbol{r}_o)$ and $T^{1/2}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_o)$. The details are rather complicated and, hence, omitted.

4. A Modeling Procedure

In this section, we propose a modeling procedure for the MHAR model by leveraging the computationally efficient algorithm of Li and Tong (2016) and the procedure for modeling the multivariate TAR model in Tsay (1998). Given observations $\{\boldsymbol{y}_t|t = 1,\ldots,T\}$, we assume that $\boldsymbol{y}_t$ has no unit root; otherwise, some co-integration procedure should be used to transform the series into unit-root stationarity. See, for instance, the methods discussed in Tsay (2014), and the references therein. The proposed modeling procedure is as follows:

1. Preliminaries: Use information criteria, such as the Akaike information criterion (AIC), to select the order p_o of a VAR model for $\boldsymbol{y}_t$, and set the maximum possible delay d_{max}.
2. Testing threshold nonlinearity and selecting threshold variables: Set $m = 0$ and the threshold variable $\boldsymbol{x}_t$ to null.
   - For $i = 1,\ldots,k$, apply the multivariate threshold nonlinearity test of Tsay (1998) to $\boldsymbol{y}_t$ by fitting a VAR(p_o) model, with y_{it} as the threshold variable and delay $d_i \in \{1,\ldots,d_{max}\}$. Compute the test statistic F_i and its p-value for each delay d_i. Denote the largest test statistic and the associated p-value by $F(i)$ and $\alpha(i)$, respectively.
   - If $\alpha(i) < \alpha_o$, increase m by one and add y_{it} to $\boldsymbol{x}_t$, where α_o is a prespecified type-I error, for example, $\alpha_o = 0.05$.

If $\boldsymbol{x}_t$ remains null, then $\boldsymbol{y}_t$ follows a linear model and the modeling procedure stops. Otherwise, $\boldsymbol{y}_t$ is threshold nonlinear with threshold variable $\boldsymbol{x}_t \subset \boldsymbol{y}_t$.

3. Estimation of delays and initial thresholds: Given d_{max} and $m > 0$, denote the space of the possible delays by $D_m = \{(i_1, \ldots, i_m)|i_j = 1, \ldots, d_{max}; j = 1, \ldots, m\}$.

   3a. Preliminary threshold estimates: For $i = 1, \ldots, m$ and $d = 1, \ldots, d_{max}$, fit a two-regime MTAR model of order p_o to $\boldsymbol{y}_t$, with threshold x_{it} and delay d. Denote the threshold estimate by $\bar{r}(i, d)$. Let $H(m, d_{\max}) = \{\bar{r}(i, d)|d = 1, \ldots, d_{max}; i = 1, \ldots, m\}$ be the collection of threshold estimates.

   3b. Selection of delay vector: For each element $\boldsymbol{d} \in D_m$, fit an MHAR model of order p_o with threshold vector $\boldsymbol{x}_t$, delay $\boldsymbol{d}$, and the corresponding thresholds in $H(m, d_{max})$. Let AIC($\boldsymbol{d}$) be the resulting AIC of the fitted model. Select $\widehat{\boldsymbol{d}} = \arg\min_{\boldsymbol{d} \in D_m} \text{AIC}(\boldsymbol{d})$, and denote the corresponding threshold vector by $\tilde{\boldsymbol{r}} = (\tilde{r}_1, \ldots, \tilde{r}_m)'$.

4. Estimation of thresholds: Given $\widehat{\boldsymbol{d}}$ and the corresponding initial threshold estimates of step 3b, refine the thresholds as follows:

   4a. For $i = 1, \ldots, m$, the preliminary threshold $\tilde{r}_i$ assumes an ordered statistic of $\{x_{it}\}$, say, $\tilde{r}_i = x_{i,(t_o)}$. Consider the neighborhood $I_i = \{x_{i,(t_o - n_*)}, x_{(t_o - n_* + 1)}, \ldots, x_{i,(t_o + n_*)}\}$ of $x_{i,(t_o)}$ as candidates for r_i, where n_* is a prespecified positive integer.

   4b. Consider $II = \prod_{i=1}^{m} I_i$ as possible candidates for the threshold vector $\boldsymbol{r}$, and perform a grid search by fitting MHAR models with order p_o, delay $\widehat{\boldsymbol{d}}$, and threshold vectors in II. Select $\widehat{\boldsymbol{r}}$ as the element in II that gives the minimum AIC of the fitted MHAR model.

   4c. Iterate steps 4a and 4b until the estimated thresholds converge.

5. Final estimation and model checking: With $\widehat{\boldsymbol{d}}$ of step 3 and $\widehat{\boldsymbol{r}}$ of step 4, estimate an MHAR model of order p_o. Perform model checking to verify that the fitted model is adequate, and refine the model if needed.

Step 3a is carried out using the nested sub-sampling search (NeSS) algorithm of Li and Tong (2016), which is an efficient way to search for the threshold for a two-regime TAR model, especially when the sample size is large. See the comparison in Liu, Chen and Tsay (2020). We apply multivariate Ljung−Box test statistics to check the serial correlations of the standardized residuals in step 5.

The proposed modeling procedure is easy to implement. In particular, the VAR order p_o for $\boldsymbol{y}_t$ and the NeSS algorithm for multivariate time series are available in the R packages **MTS** and **NTS**, respectively. The threshold nonlinearity test of Tsay (1998) and the CLSE of the MHAR model are also easy to perform.

The proposed modeling procedure is based on the following considerations. It starts by assuming a linear VAR model for $\boldsymbol{y}_t$, and performs multivariate threshold nonlinearity tests to verify that $\boldsymbol{y}_t$ indeed has threshold nonlinearity. The procedure then considers ways to simplify the computation, while maintaining its effectiveness. Like the conventional TAR models, the thresholds $\boldsymbol{r}$ assume ordered statistics of the threshold variables. For a large sample size T and multiple threshold variables, it would be computationally expensive to search over all possible combinations of ordered statistics. The NeSS algorithm of Li and Tong (2016) is an efficient algorithm for a given threshold variable with a single threshold. In the proposed procedure, we use this algorithm to obtain an initial estimate of the threshold. It is possible that this initial estimate is biased because the model is misspecified. Therefore, we consider a neighborhood around the initial estimate. In addition, we conduct a grid search among possible configurations of the product space of the selected neighborhood of the initial thresholds. For m threshold variables, the grid search searches over $(2n_* + 1)^m$ possible thresholds. Therefore, we keep n_* relatively small, say $n_* = 10$, for large m. If needed, one can iterate this grid search procedure to refine the estimation of the thresholds.

5. Examples

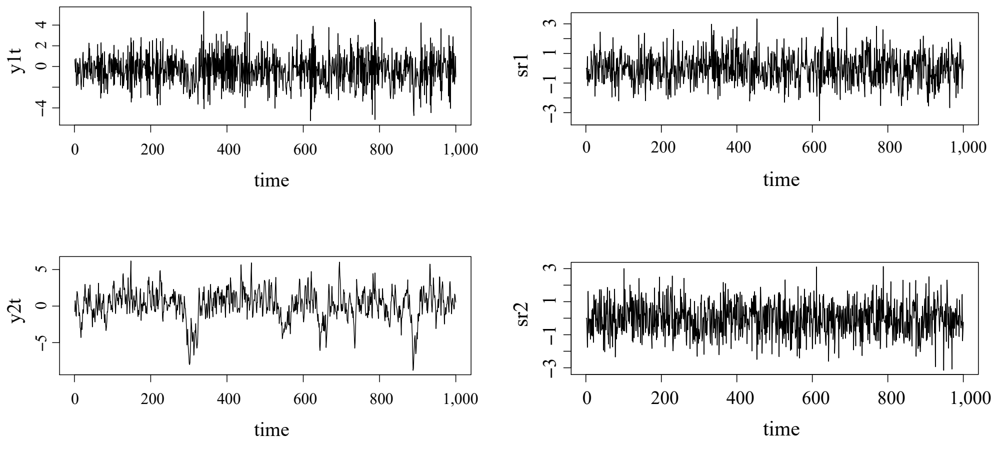
We demonstrate the efficacy of the proposed modeling procedure using simulation studies, and show the applicability of the proposed MHAR model using a real example.

5.1. Simulation

Example 1. Consider the simple MHAR model

$$\boldsymbol{y}_t = \begin{cases} \boldsymbol{\phi}\boldsymbol{y}_{t-1} + \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\epsilon}_t, & \text{if } R_t = 1, \\ \boldsymbol{\psi}\boldsymbol{y}_{t-1} + \boldsymbol{\Sigma}_2^{1/2}\boldsymbol{\epsilon}_t, & \text{if } R_t = 0, \end{cases} \qquad R_t = \begin{cases} 1, & \boldsymbol{x}_t(\boldsymbol{d}) \in \Omega_{1t}, \\ 0, & \boldsymbol{x}_t(\boldsymbol{d}) \in \Omega_{2t}, \\ R_{t-1}, & \text{otherwise}, \end{cases} \tag{5.1}$$

where $\boldsymbol{x}_t = \boldsymbol{y}_t$, $\boldsymbol{d} = (1, 2)$, $\boldsymbol{\epsilon}_t$ are i.i.d. as bivariate normal with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{I}$, and the other parameters are given below:



(a) (b)

Figure 1. Time plots of the series in Example 1. Part (a) is the original time series, and part (b) is the standardized residuals of a fitted MHAR model.

$$\boldsymbol{\phi} = \begin{bmatrix} 0.2 & 0.3 \\ -0.6 & 1.1 \end{bmatrix}, \ \boldsymbol{\psi} = \begin{bmatrix} -0.7 & -0.2 \\ 0.2 & 0.6 \end{bmatrix},$$
$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}, \ \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1.6 & -0.2 \\ -0.2 & 1 \end{bmatrix}.$$

The thresholds used are $(0,0)'$, so that we have $\Omega_{1t} = \{\boldsymbol{y}_t | (y_{1,t-1} \leq 0) \cap (y_{2,t-2} \leq 0)\}$ and $\Omega_{2t} = \{\boldsymbol{y}_t | (y_{1,t-1} > 0) \cap (y_{2,t-2} > 0)\}$. In this particular case, the two eigenvalues of $\boldsymbol{\phi}$ are 0.8 and 0.5, whereas those of $\boldsymbol{\psi}$ are approximately -0.67 and 0.57. Part (a) of Figure 1 shows time plots of 1,000 observations generated by model (5.1). We used $R_0 = 1$ and $\boldsymbol{y}_0 = \boldsymbol{\epsilon}_0$ in the simulation, but dropped the first 50 observations to remove the impact of the initial values. As expected, the time plots exhibit a stationary vector series.

Following the proposed modeling procedure, we selected a VAR(2) model for $\boldsymbol{y}_t$ using the Bayesian information criterion. Table 1 shows the results of the multivariate threshold nonlinearity test, using marginal series as its threshold variable and $d \in \{1,2,3,4\}$. From the table, we see that (a) the threshold nonlinearity tests confirm that $\boldsymbol{y}_t$ is indeed nonlinear, and (b) both y_{1t} and y_{2t} are selected as threshold variables, that is, $\boldsymbol{x}_t = \boldsymbol{y}_t$. With $d_{max} = 4$, step 3 of the proposed procedure selects $\widehat{\boldsymbol{d}} = (1,2)'$, with an initial threshold estimate $\tilde{\boldsymbol{r}} \approx (-0.045, -0.052)'$. These results indicate that the proposed procedure is effective in selecting the delays and obtaining the initial threshold estimates.

Table 1. Multivariate Threshold Nonlinearity Tests of Example 1. The values in parentheses are p-values

Threshold	Delay: d			
Variable	1	2	3	4
y_{1t}	136.04 (0)	42.41 (6.34×10^{-6})	25.81 (0.004)	13.11 (0.22)
y_{2t}	104.84 (0)	287.30 (0)	246.43 (0)	156.80 (0)

Next, we follow step 4 to estimate the thresholds. Using $n_* = 10$, $\widehat{\boldsymbol{d}} = (1,2)'$, and the initial thresholds $\tilde{\boldsymbol{r}}$, we refine the estimates of the thresholds. The refinement stops after two iterations, and the final estimate of the threshold is $\widehat{\boldsymbol{r}} \approx (-0.0179, -0.006)'$, which is close to the true value $(0,0)'$. If one uses $n_* = 20$, then step 4 requires only one iteration to select the final threshold estimates.

Finally, using $\widehat{\boldsymbol{d}} = (1,2)'$ and $\widehat{\boldsymbol{r}} = (-0.0179, -0.006)'$, we estimate the MHAR model. It turns out that the lag-2 coefficients for each regime are not statistically significant; thus, we use order 1 for each regime. The final estimates of the coefficients and residual covariance matrices are

$$\widehat{\boldsymbol{\phi}} = \begin{bmatrix} 0.194(0.036) & 0.288(0.022) \\ -0.618(0.036) & 1.107(0.022) \end{bmatrix}, \quad \widehat{\boldsymbol{\psi}} = \begin{bmatrix} -0.706(0.033) & -0.154(0.030) \\ 0.203(0.025) & 0.614(0.023) \end{bmatrix},$$

$$\widehat{\boldsymbol{\Sigma}_1} = \begin{bmatrix} 1.05 & 0.15 \\ 0.15 & 1.04 \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}_2} = \begin{bmatrix} 1.61 & -0.19 \\ -0.19 & 0.94 \end{bmatrix},$$

where the values in parentheses denote asymptoic standard errors based on Theorem 3. These estimates are close to their true values. Figure 1(b) shows time plots of the standard residuals of the fitted model. The multivariate Ljung−Box statistics of the standardized residuals give Q(10) = 42.86(0.10) and Q(20) = 72.37 (0.49), indicating that, as expected, one cannot reject the null hypothesis that the standardized residuals have no serial correlations, where the values in parentheses denote p-values. The p-values are based on χ^2 distributions with degrees of freedom 32 and 72, respectively, after adjusting for parameter estimates. This simple example demonstrates that the proposed modeling procedure works well.

Based on the recommendation of a referee, we repeated the above analysis for 3,000 iterations. That is, we generated 3,000 data sets, each with 1,000 observations, and examined the performance of the proposed modeling procedure. Because the sample size is relatively large, the selected delay is $\widehat{\boldsymbol{d}} = (1,2)'$ for every iteration, confirming the strong consistency of the delay estimation. Using

Table 2. Probabilities of Correct Selection of the Delay Vector $\boldsymbol{d}$ Based on 3,000 Iterations and Sample Sizes 500 and 10,000.

	$\boldsymbol{d}=(1,1)'$		$\boldsymbol{d}=(2,1)'$		$\boldsymbol{d}=(1,2)'$	
Sample size	500	1,000	500	1,000	500	1,000
Probability	0.908	0.997	0.753	0.997	0.956	1.000

$n_* = 10$, the average and median numbers of iterations of step 4 are 2.46 and 2, respectively. The associated sample standard deviation is 1.96. Therefore, the proposed procedure can locate the estimates of the thresholds effectively. The average of the estimated thresholds and their sample standard errors are $(-0.040, -0.008)'$ and $(0.148, 0.035)'$, respectively, showing that the thresholds can be estimated consistently. Finally, the average estimates of the parameters and their sample standard deviations, in parentheses, over the 3,000 iterations are given below:

$$\widehat{\boldsymbol{\phi}} = \begin{bmatrix} 0.194(0.038) & 0.303(0.026) \\ -0.593(0.044) & 1.091(0.029) \end{bmatrix}, \quad \widehat{\boldsymbol{\psi}} = \begin{bmatrix} -0.696(0.035) & -0.194(0.034) \\ 0.198(0.028) & 0.607(0.031) \end{bmatrix},$$

$$\widehat{\boldsymbol{\Sigma}_1} = \begin{bmatrix} 1.01(0.081) & 0.19(0.054) \\ 0.19(0.054) & 1.01(0.081) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}_2} = \begin{bmatrix} 1.59(0.099) & -0.19(0.060) \\ -0.19(0.060) & 1.01(0.071) \end{bmatrix}.$$

As expected, these estimates confirm the asymptotic behavior of the CLSE.

Example 2. In this simulation, we use the same parameters as those of Example 1, except for the delay vector $\boldsymbol{d}$. Our goal is to study the performance of the proposed procedure in selecting the delay vector. The delay vectors considered are $(1,1)'$, $(2,1)'$, and $(1,2)'$, and the sample sizes used are 500 and 1,000. For each $\boldsymbol{d}$ and sample size, we generate $\boldsymbol{y}_t$ using the same procedure as that in Example 1. We repeat the process for 3,000 iterations, and tabulate the performance of the multivariate threshold nonlinearity test and the percentages of correct selection of delays. For all but two realizations with sample size 500, the threshold nonlinearity test confirms that the simulated series is indeed nonlinear. Thus, the proposed procedure can select the dimension of the threshold vector $\boldsymbol{x}_t$ effectively. Table 2 tabulates the probabilities of correct selection of the delay vector $\boldsymbol{d}$. From the table, we see that the proposed modeling procedure is effective in selecting the delay vector $\boldsymbol{d}$, especially when the sample size is 1,000.

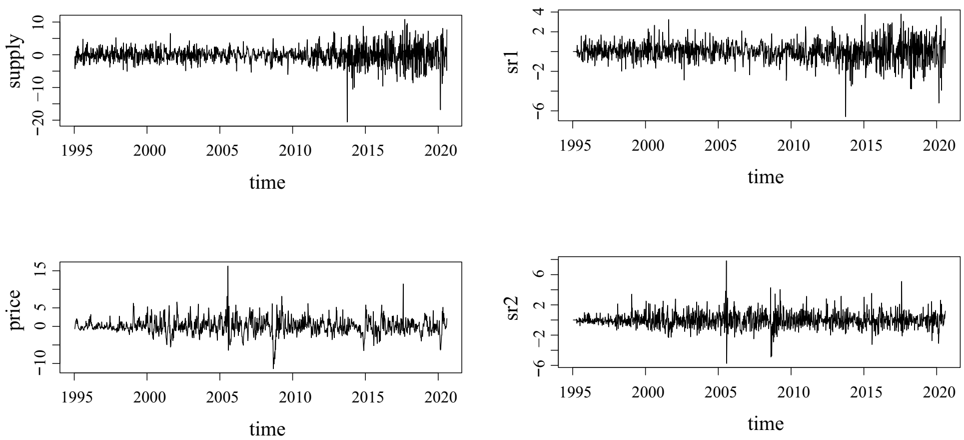
5.2. Real-Data Analysis

In this section, we analyze the weekly growth rates (first difference of log) of U.S. conventional regular gasoline stocks and prices. The growth rates are given as percentages. The data are downloaded from the U.S. Energy Information Administration at `www.eia.gov`. The conventional motor gasoline stocks are from March 3, 1995, to August 28, 2020, and are in thousand barrels. The regular gasoline prices are from March 6, 1995, to August 31, 2020, and are in dollars per gallon. Here, the stocks serve as the supply, and we are interested in finding the relationship between the supply and the price of weekly regular gasoline. Let $\boldsymbol{y}_t$ be the weekly growth rates of the series, that is, $\boldsymbol{y}_t = (s_t, p_t)'$, with s_t being the growth rate of the gasoline supply. Figure 2(a) shows time plots of the two growth rates. The volatility of the first series seems to increase in recent years. For simplicity, we do not consider conditional heteroscedasticity in our analysis, but the CLSEs remain consistent.

Following the proposed modeling procedure, we identify a VAR model for $\boldsymbol{y}_t$. The AIC selects $p_o = 12$. We then perform multivariate threshold nonlinearity tests with $d_{max} = 4$. The tests confirm that both components of $\boldsymbol{y}_t$ can serve as threshold variables. For instance, the threshold test of VAR(12) with y_{1t} being the threshold variable and delay 2 is 105.4, with p-value 7.9×10^{-6}. The test of VAR(12) with y_{2t} as the threshold variable and delay 1 is 140.6, with p-value 1.5×10^{-10}. Therefore, in this instance, $\boldsymbol{x}_t = \boldsymbol{y}_t$ and $m = 2$.

Next, we apply step 3 of the proposed procedure to select the delay, using order $p_o = 12$ and $d_{max} = 4$. It selects $\widehat{\boldsymbol{d}} = (2, 1)'$, with an initial estimate of the threshold vector $\tilde{\boldsymbol{r}} = (-2.224, 1.252)'$. We then apply step 4 with $n_* = 20$ to refine the estimation of the thresholds. The step iterates three times for the threshold estimates to converge, resulting in an estimated threshold vector $\widehat{\boldsymbol{r}} = (-2.345, 1.037)'$. Finally, using $\widehat{\mathbf{d}}$ and $\widehat{\mathbf{r}}$, we estimate the MHAR model of order $p = 12$. The AIC of the fitted model is 3,767.202. Model checking indicates that the residuals have no significant serial correlations, but there exist a couple of outlying values in the standardized residuals. See Figure 2(b). The multivariate Ljung−Box statistics of the standardized residuals give Q(24) = 64.97 (0.052) and Q(36) = 111.24 (0.14), where the values in parentheses denote p-values. For the fitted model, the sample sizes for $R_t = 1$ and $R_t = 0$ are 513 and 805, respectively.

For comparison, the linear VAR(12) model gives AIC = 3,961.64. A two-regime MTAR(12) model with y_{1t} as the threshold variable and delay 2 gives an estimated threshold −2.792 and AIC = 3,877.44. The sample sizes for the two



(a) (b)

Figure 2. Time plots of the series of weekly growth rates of conventional motor gasoline stocks and regular gasoline prices. Part (a) is the growth rate series and part (b) is the standardized residuals of a fitted MHAR model.

regimes are 158 and 1,160, respectively. A two-regime MTAR(12) model with y_{2t} as the threshold variable and delay 1 gives an estimated threshold 1.252 and AIC = 3,864.25. The sample sizes of the two regimes are 996 and 322, respectively. Therefore, based on the AIC, the proposed MHAR model fares better than the linear VAR(12) and the two-regime MTAR(12) models.

Finally, keeping only coefficient estimates with t-ratios greater than 1.645, the fitted model for regime $\widehat{R}_t = 1$ can be written as

$$s_t \approx -0.13 s_{t-1} - 0.18 p_{t-2} + 0.08 s_{t-4} - 0.12 s_{t-11} - 0.10 s_{t-12} + 0.12 p_{t-12} + e_{11,t},$$
$$p_t \approx -0.06 s_{t-1} + 0.62 p_{t-1} + 0.18 p_{t-2} - 0.09 p_{t-4} - 0.09 p_{t-6} + 0.04 p_{t-8} + e_{12,t},$$

whereas that for the regime $\widehat{R}_t = 0$ is

$$\begin{aligned} s_t \approx & -0.21 - 0.16 s_{t-1} - 0.10 p_{t-1} + 0.11 s_{t-2} + 0.07 p_{t-2} + 0.11 s_{t-3} - 0.09 p_{t-8} \\ & -0.11 s_{t-9} - 0.14 s_{t-10} - 0.09 p_{t-10} - 0.13 s_{t-11} - 0.11 s_{t-12} + e_{21,t}, \\ p_t \approx & \ 0.20 - 0.11 s_{t-1} + 0.42 p_{t-1} - 0.07 s_{t-2} - 0.09 s_{t-3} + 0.11 p_{t-3} + 0.10 p_{t-8} \\ & -0.13 p_{t-9} + 0.06 s_{t-10} + e_{22,t}. \end{aligned}$$

From the fitted model, we see several differences between the two regimes. First, in the regime $\widehat{R}_t = 0$, the two constant terms are significant, but with different signs, indicating that the gasoline stock (or supply) continues to drop, but the gasoline price keeps increasing. This feature does not show up in regime $\widehat{R}_t = 1$. Second, as expected, there is a feedback relationship between the two growth rate series in both regimes, but the dependence of the price growth rate on the supply is relatively weak in the regime $\widehat{R}_t = 1$. See the only significant coefficient -0.06 at s_{t-1}. This negative sign is expected because the price and supply of gasoline are negatively related. Third, the gasoline supply seems to depend not only on its own past lag at $t-12$, but also on higher-order lags of the gasoline price; see lag $t-12$ in regime $\widehat{R}_t = 1$ and lag $t-10$ in regime $\widehat{R}_t = 0$. Fourth, the gasoline price seems to be sticky, because the coefficients of its own lag-1 are relatively large in both regimes. This is reasonable and agrees with the common sense that the gasoline price decreases slowly.

6. Conclusion

We have proposed a multivariate hysteretic autoregressive model for time series analysis. We briefly studied the properties of the proposed model, and used the conditional least squares method for the estimation. We also proposed a modeling procedure and demonstrated its efficacy using simulation studies and a real example. The results show that the proposed model and modeling procedure are useful in some applications of multivariate time series analysis. Finally, this study is concerned with a dynamic model to which Professor T. L. Lai has made fundamental contributions. Indeed, the proposed model can be extended to include exogenous variables using the results of Lai and Wei (1982). The details are similar to those of Tsay (1998) and, hence, are omitted.

Acknowledgments

This research was supported in part by the Booth School of Business, University of Chicago. The author thanks W.K. Li for constructive comments on an early draft, and K.S. Chan for discussions on the ergodic properties.

References

Chan, K. S. and Tong, H. (1985). On the use of the Deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Advances in Applied Probability* **17**, 666–678.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycles. *Econometrica* **57**, 331–339.

Lai, T. L. and Wei, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Annals of Statistics* **10**, 154–166.

Li, G., Guan, B., Li, W. K. and Yu, P. L. H. (2015). Hysteretic autoregressive time series models.

Biometrika **102**, 717–723.

Li, D. and Tong, H. (2016). Nested sub-sampling search algorithm for estimation of threshold models. *Statistica Sinica* **26**, 1543–1554.

Liu, X., Chen, R. and Tsay, R. S. (2020). NTS: An R package for nonlinear time series analysis. Working paper, San Diego State University.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, New York.

Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, New York.

Tiao, G. C. and Tsay, R. S. (1994). Some advances in nonlinear and adaptive modeling in time series analysis. *Journal of Forecasting* **13**, 109–140.

Tong, H. (1978). On a threshold model. In *Pattern Recognition and Signal Processing* (Edited by C. H. Chen). Sijthoff and Noordhoff, Amsterdam, NL.

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

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

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

Tsay, R. S. (2014). *Multivariate Time Series Analysis*. Wiley, Hoboken, NJ.

Tsay, R. S. and Chen, R. (2019). *Nonlinear Time Series Analysis*. Wiley, Hoboken, N.J.

Wu, S. and Chen, R. (2007). Threshold variable determination and threshold variable driven switching autoregressive models. *Statistica Sinica* **17**, 241–264.

Xia, Y., Li, W. K. and Tong, H. (2007). Threshold variable selection using nonparametric methods. *Statistica Sinica* **17**, 265–287.

Ruey S. Tsay

Booth School of Business, University of Chicago, 5807 S. Woodlawn Avenue, Chicago, IL 60637, USA.

E-mail: ruey.tsay@chicagobooth.edu

(Received October 2020; accepted December 2020)