On the least squares estimation of multiple-regime threshold autoregressive models

Dong Li, Shiqing Ling*

Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

---

Abstract

This paper studies the least squares estimator (LSE) of the multiple-regime threshold autoregressive model and establishes its large sample theory. It is shown that the LSE is strongly consistent. When the autoregressive function is discontinuous, the estimated thresholds are n-consistent and asymptotically independent, each of which converges weakly to the smallest minimizer of a one-dimensional two-sided compound Poisson process. The remaining parameters are $\sqrt{n}$-consistent and asymptotically normal. A simulation method is proposed to implement the limiting distributions of the estimated thresholds. Simulation studies are conducted to assess the performance of the LSE in finite samples. The results are illustrated with an application to the quarterly U.S. real GNP data over the period 1947-2009.

Keywords: Asymptotic distribution, Compound Poisson process, Least squares estimation, Multiple-regime threshold autoregressive model.
JEL classification: C13, C22.

---

1. Introduction

The threshold autoregressive (TAR) model, proposed by Tong (1978), is one of the mature nonlinear time series models in the literature. It has attracted considerable attention and has been widely used in diverse areas, including biological sciences, econometrics, environmental sciences, finance, hydrology, physics, population dynamics, and among others. The TAR model

*Corresponding author. Tel.: +852 23587459; fax: +852 23591016
Email address: maling@ust.hk (Shiqing Ling)

is capable of producing and modeling many nonlinear phenomena such as amplitude dependent frequencies, asymmetric limit cycle, chaos, harmonic distortion, jump resonance and so on. A very all-sided survey on TAR models is available in Tong (1990) and a selective survey of the history of threshold models is given by Tong (2010). In comparison with many other nonlinear time series models, the success of TAR models is partially due to the fact that it may typically produce a simple and easy-to-handle approximation to complicate dynamic functions, perhaps more importantly, it can offer a reasonable model-interpretation. The numerous applied econometrics literature has also witnessed a growing interest in TAR models. For example, Koop and Potter (1999) used a three-regime TAR model to capture the nonlinearity in the U.S. unemployment rate of the period 1959–1996, and Tiao and Tsay (1994) constructed a four-regime TAR model to fit the quarterly U.S. real GNP data from February 1947 to January 1991.

Probabilistic structures of TAR models were studied by Chan et al.(1985), Chan and Tong (1985) and Tong (1990). More related results can be found in An and Huang (1996), Brockwell et al.(1992), Chen and Tsay (1991), Cline and Pu (2004), Ling (1999), Liu and Suskov (1992) and so on. The large sample theory of the LSE of two-regime TAR models was established by Chan (1993) and Chan and Tsay (1998), see also Petruccelli (1985), Qian (1998) and Tsay (1998). The most difficult part in TAR models is to study the asymptotic properties of the estimated threshold. Chan (1993) showed that the estimated threshold is n-consistent and its limiting distribution is the smallest minimizer of a one-dimensional two-sided compound Poisson process for two-regime TAR models when the autoregressive function is discontinuous. Hansen (1997, 2000) also studied the two-regime threshold AR/regression model. Under the assumption that the threshold effect is vanishingly small, he obtained the distribution- and parameter-free limit of the estimated threshold. Seo and Linton (2007) proposed a smoothed least squares estimation for the two-regime threshold AR/regression model. They showed that the estimated threshold is asymptotically normal but its convergence rate is less than n and depends on the bandwidth. Gonzalo and Pitarakis (2002) considered the sequential estimation of a multiple-regime TAR model and only obtained the consistency and n-convergence rate of the estimated thresholds. Up to now, however, the large sample theory of the estimator, particularly the limiting distributions of the estimated thresholds, is still an open problem in the multiple-regime TAR model.

This paper studies the LSE of multiple-regime TAR models and develops its large sample theory. Under some suitable conditions, it is shown that the LSE is strongly consistent. More importantly, when the autoregressive function is discontinuous over each threshold, the estimated thresholds are n-consistent, and after a normalization, they are asymptotically independent and each of them converges weakly to the smallest minimizer of a one-dimensional two-sided compound Poisson process. The remaining parameters are $\sqrt{n}$-consistent and asymptotically normal. A simulation method is proposed to implement the limiting distributions of the estimated thresholds. Simulation studies are conducted to assess the performance of the LSE in finite samples. To illustrate the results, an application to the quarterly U.S. real GNP data over the period 1947-2009 is given.

The rest of the paper is organized as follows. Section 2 presents the model and its estimation procedure. Section 3 states our main results. Section 4 provides a numerical method to implement the limiting distribution of the estimated thresholds. Section 5 reports simulation results. Section 6 gives an empirical example. All proofs of Theorems are given in Section 7.

2. Model and least squares estimation

A time series $\{y_t\}$ is said to be an m-regime TAR model ($m \geq 2$) with order p if it satisfies the equation

$$y_t = \sum_{j=1}^{m} (\mathbf{Y}_{t-1}' \boldsymbol{\beta}_j + \sigma_j \varepsilon_t) \mathbb{1}(r_{j-1} < y_{t-d} \leq r_j), \tag{2.1}$$

where $\mathbf{Y}_{t-1} = (1, y_{t-1}, ..., y_{t-p})'$, $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, ..., \beta_{jp})' \in \mathbb{R}^{p+1}$, $j = 1, ..., m$, $-\infty = r_0 < r_1 < ... < r_m = \infty$; σ_j's are positive numbers. The number m of regimes and the order p of model (2.1) are positive integers. d is a positive integer called the delay parameter. $\{r_1, ..., r_{m-1}\}$ are threshold parameters. The errors $\{\varepsilon_t\}$ are independent and identically distributed (i.i.d.) random variables with zero mean and unit variance, and ε_t is independent of the past information $\{y_{t-j} : j \geq 1\}$. Throughout the paper, we assume that m and p are known[1].

[1] In practice, m and p are needed to be identified for a given dataset. For the specification of m, see Gonzalo and Pitarakis (2002). When m is given, we can use the AIC to select p, see Tsay (1998).

Let $\mathbf{r} = (r_1, ..., r_{m-1})' \in \mathbb{R}^{m-1}$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mathbf{r}', d)' = (\boldsymbol{\beta}'_1, ..., \boldsymbol{\beta}'_m, \mathbf{r}', d)' \in \mathbb{R}^{m(p+1)+(m-1)} \times \{1, ..., D_0\}$, where D_0 is a known positive integer. Suppose that a sample $\{y_1, ..., y_n\}$ is from model (2.1) with true value $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}'_{10}, ..., \boldsymbol{\beta}'_{m0}, \mathbf{r}'_0, d_0)'$. Given the initial values $\{y_0, ..., y_{1-p}\}$, the sum of square errors function $L_n(\boldsymbol{\theta})$ is defined as

$$L_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} [y_t - \mathbb{E}_{\boldsymbol{\theta}}(y_t|\mathcal{F}_{t-1})]^2,$$

where $\mathcal{F}_t$ is the σ-algebra generated by $\{y_{1-p}, ..., y_t\}$ and $\mathbb{E}_{\boldsymbol{\theta}}(\cdot|\cdot)$ denotes the conditional expectation assuming $\boldsymbol{\theta}$ to be the true parameter. The minimizer $\widehat{\boldsymbol{\theta}}_n$ of $L_n(\boldsymbol{\theta})$ is called a LSE of $\boldsymbol{\theta}_0$, that is,

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}).$$

Since $L_n(\boldsymbol{\theta})$ is discontinuous in $\mathbf{r}$ and d, a multi-parameter grid-search algorithm is needed. The way to obtain $\widehat{\boldsymbol{\theta}}_n$ is as follows.

- Fix $\mathbf{r} \in \mathbb{R}^{m-1}$ and $d \in \{1, ..., D_0\}$, then minimize $L_n(\boldsymbol{\theta})$ and get its minimizer $\widehat{\boldsymbol{\beta}}_n(\mathbf{r}, d)$ and minimum $L_n^*(\mathbf{r}, d) \equiv L_n(\boldsymbol{\theta})|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_n(\mathbf{r},d)}$.

- Since $L_n^*(\mathbf{r}, d)$ only takes finite possible values, one can get the minimizer $(\widehat{\mathbf{r}}'_n, \widehat{d}_n)'$ of $L_n^*(\mathbf{r}, d)$ by the enumeration approach.

- Use a plug-in method, one can finally get $\widehat{\boldsymbol{\beta}}_n(\widehat{\mathbf{r}}_n, \widehat{d}_n)$ and $\widehat{\boldsymbol{\theta}}_n$.

Generally, $\widehat{\mathbf{r}}_n$ is taken as the form $(y_{(i_1)}, ..., y_{(i_{m-1})})'$, where $i_1 < \cdots < i_{m-1}$ and $\{y_{(1)}, ..., y_{(n)}\}$ is the order statistics of the sample $\{y_1, ..., y_n\}$. If $(y_{(j_1)}, ..., y_{(j_{m-1})})'$ is an estimator of $\mathbf{r}_0$, then $L_n^*(\mathbf{r}, \widehat{d}_n)$ is a constant over $\mathcal{R}$, where

$$\mathcal{R} = \{\mathbf{r} = (r_1, ..., r_{m-1})' : r_i \in [y_{(j_i)}, y_{(j_i+1)}), i = 1, ..., m-1\}.$$

Thus, there exist infinitely many $\mathbf{r}$ such that $L_n(\cdot)$ can achieves its global minimum and each $\mathbf{r} \in \mathcal{R}$ can be considered as an estimator of $\mathbf{r}_0$. In this case, we choose $(y_{(j_1)}, ..., y_{(j_{m-1})})'$ as a representative of $\mathcal{R}$ and denote it as the estimator of $\mathbf{r}_0$. According to the procedure for obtaining $\widehat{\boldsymbol{\theta}}_n$, it is not hard to show that $\widehat{\boldsymbol{\theta}}_n$ is the LSE of $\boldsymbol{\theta}_0$.

Let σ_{j0} be the true value of σ_j for $j = 1, ..., m$. Once $\widehat{\boldsymbol{\theta}}_n$ is obtained, we then can estimate σ_{j0}^2 by

$$\widehat{\sigma}_{jn}^2 = \frac{1}{n_j}\sum_{t=1}^{n}(y_t - \mathbf{Y}_{t-1}'\widehat{\boldsymbol{\beta}}_{jn})^2 \mathbb{1}(\widehat{r}_{j-1,n} < y_{t-\widehat{d}_n} \le \widehat{r}_{jn}), \tag{2.2}$$

where $n_j = \sum_{t=1}^{n} \mathbb{1}(\widehat{r}_{j-1,n} < y_{t-\widehat{d}_n} \le \widehat{r}_{jn})$.

In order to get the global minimum of $L_n(\cdot)$ with m regimes and sample size n, the required number of calculations is $O(n^{m-1}/(m-1)!)$. When m is large, however, the computational burden becomes substantial, requiring multi-parameter grid-based search over all possible values of all threshold parameters taken together, and hence this algorithm is very time-consuming. For a fixed m, the consumed time soars at an exponential rate as the sample size n increases. This problem is similar to the computational problem arising from multiple change-point models investigated by Bai and Perron (2003, 2006). Tsay (1989) transforms model (2.1) into a change-point model and use the rearranged technique to localize possible positions of threshold parameters. Similarly, using same rearranged technique, Coakley, Fuertes and Pérez (2003) provides an efficient estimation approach which relies on the computational advantages of QR factorizations of matrices. When m is small, the grid-based search algorithm is an easy way to obtain the global minimum of $L_n(\cdot)$.

3. Main results

Let $\Theta \times \{1, ..., D_0\}$ be the parameter space, where $\Theta = \Theta_\beta \times \Theta_r$ is a compact subset of $\mathbb{R}^{m(p+1)} \times \mathfrak{R}^{m-1}$ and $\mathfrak{R}^{m-1} = \{(r_1, ..., r_{m-1}) : -\infty < r_1 < ... < r_{m-1} < \infty\}$. The following result states the strong consistency of the estimator $\widehat{\boldsymbol{\theta}}_n$.

Theorem 3.1. *Suppose that (i) $\{y_t\}$ satisfying (2.1) is strictly stationary and ergodic, having finite second moments, (ii) $\boldsymbol{\beta}_{j0} \neq \boldsymbol{\beta}_{j+1,0}$ for $j = 1, ..., m-1$, and (iii) ε_1 admits a bounded, continuous and positive density on $\mathbb{R}$. Then, $\widehat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ a.s. as $n \to \infty$ and so are $\widehat{\sigma}_{jn}^2$'s.*

The condition (ii) in Theorem 3.1 is required to guarantee the identification of $\mathbf{r}_0$. The strong consistency of $\widehat{\boldsymbol{\theta}}_n$ holds regardless if the autoregressive function is continuous over the thresholds or not. From Theorem 3.1, we know that $\widehat{d}_n$ equals d_0 eventually. Thus, without loss of generality, we assume that the delay parameter d_0 is known for the remainder of this paper and it is deleted from $\boldsymbol{\theta}_0$, i.e., $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0', \mathbf{r}_0')'$, and so is $\widehat{d}_n$ from $\widehat{\boldsymbol{\theta}}_n$. The parameter space becomes Θ, accordingly, and we write d for d_0 in what follows.

To obtain the convergence rate of $\widehat{\mathbf{r}}_n$ and the asymptotic normality of $\widehat{\boldsymbol{\beta}}_n \equiv \widehat{\boldsymbol{\beta}}_n(\widehat{\mathbf{r}}_n)$, we first give three assumptions as follows.

Assumption 3.1. *$\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with mean 0 and $\mathbb{E}\varepsilon_t^4 < \infty$. ε_1 has a bounded, continuous and positive density function on $\mathbb{R}$.*

Assumption 3.2. *$\{y_t\}$ is strictly stationary with $\mathbb{E}y_t^4 < \infty$.*

Let $\mathbf{Z}_t = (y_t, ..., y_{t-p+1})'$. Then $\{\mathbf{Z}_t\}$ is a Markov chain. Denote its l-step transition probability by $\mathcal{P}^l(\mathbf{z}, A)$, where $\mathbf{z} \in \mathbb{R}^p$ and A is a Borel set.

Assumption 3.3. *$\{\mathbf{Z}_t\}$ admits a unique invariant measure $\Pi(\cdot)$ such that there exist $K > 0$ and $\rho \in [0, 1)$, for any $\mathbf{z} \in \mathbb{R}^p$ and any n, $\|\mathcal{P}^n(\mathbf{z}, \cdot) - \Pi(\cdot)\|_{\mathrm{v}} \le K(1 + \|\mathbf{z}\|)\rho^n$, where $\|\cdot\|_{\mathrm{v}}$ denotes the total variation norm.*

Under Assumption 3.3, $\{\mathbf{Z}_t\}$ is V-uniformly ergodic with $V(\mathbf{z}) = K(1 + \|\mathbf{z}\|)$, which is stronger than geometric ergodicity. For the concept of V-uniform ergodicity, see Meyn and Tweedie (1993). If Assumption 3.1 holds and $\max_{1\le i\le m} \sum_{j=1}^p |\beta_{ij}| < 1$, then Assumption 3.3 holds and $\mathbb{E}y_t^4 < \infty$, see Chan (1989) and Chan and Tong (1985). If the initial value $\mathbf{Z}_0$ is from the distribution $\Pi(\cdot)$, then Assumption 3.3 implies that $\{y_t\}$ is strictly stationary.

In order to obtain the n-convergence rate of $\widehat{\mathbf{r}}_n$ and the limiting distribution of $n(\widehat{\mathbf{r}}_n - \mathbf{r}_0)$, we need another assumption.

Assumption 3.4. *There exist nonrandom vectors $\mathbf{w}_i^* = (1, w_{i1}, ..., w_{ip})'$ with $w_{id} = r_{i0}$ such that $(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{i+1,0})'\mathbf{w}_i^* \neq 0$ for $i = 1, ..., m-1$.*

When $p = 1$, Assumption 3.4 implies that the autoregressive mean function is discontinuous at all thresholds $\{r_1, ..., r_{m-1}\}$. Assumption 3.4 in the general case implies that $\|\mathbf{Y}_{t-1}'(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{i+1,0})\|$ is bigger than a positive constant with a positive probability and plays a key role in obtaining the n-convergence rate of $\widehat{\mathbf{r}}_n$ and its limiting distribution.

Theorem 3.2. *If Assumptions 3.1-3.4 hold. Then,*

(i). $n\|\widehat{\mathbf{r}}_n - \mathbf{r}_0\| = O_p(1)$;

(ii). $\sqrt{n} \sup_{\|\mathbf{r}-\mathbf{r}_0\|<B/n} \|\widehat{\boldsymbol{\beta}}_n(\mathbf{r}) - \widehat{\boldsymbol{\beta}}_n(\mathbf{r}_0)\| = o_p(1)$ *for any fixed* $B \in (0, +\infty)$.

Furthermore,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = \sqrt{n}(\widehat{\boldsymbol{\beta}}_n(\mathbf{r}_0) - \boldsymbol{\beta}_0) + o_p(1) \Longrightarrow \mathcal{N}(0, \Sigma) \quad \textit{as } n \to \infty,$$

where $\Sigma = \mathrm{diag}(\sigma_{10}^2\Sigma_1, ..., \sigma_{m0}^2\Sigma_m)$ *and*

$$\Sigma_j^{-1} = \mathbb{E}[\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}'\mathbb{1}(r_{j-1,0} < y_{t-d} \le r_{j0})], \quad j = 1, ..., m.$$

From Theorem 3.2(i), we know that the convergence rate of $\widehat{\mathbf{r}}_n$ is n. To study the limiting distribution of $n(\widehat{\mathbf{r}}_n - \mathbf{r}_0)$, we consider the following profile sum of squares errors function:

$$\widetilde{L}_n(\mathbf{s}) = L_n\Big(\widehat{\boldsymbol{\beta}}_n(\mathbf{r}_0 + \frac{\mathbf{s}}{n}),\, \mathbf{r}_0 + \frac{\mathbf{s}}{n}\Big) - L_n\Big(\widehat{\boldsymbol{\beta}}_n(\mathbf{r}_0),\, \mathbf{r}_0\Big), \quad \mathbf{s} \in \mathbb{R}^{m-1}. \tag{3.1}$$

Using Theorem 3.2 and Taylor's expansion, we can show that $\widetilde{L}_n(\mathbf{s}) = \wp_n(\mathbf{s}) + o_p(1)$, where

$$\begin{aligned}
\wp_n(\mathbf{s}) &= L_n(\boldsymbol{\beta}_0, \mathbf{r}_0 + \frac{\mathbf{s}}{n}) - L_n(\boldsymbol{\beta}_0, \mathbf{r}_0) \\
&= \sum_{i=1}^{m-1}\sum_{t=1}^{n} \Big[\big\{[\mathbf{Y}_{t-1}'(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{i+1,0})]^2 + 2\sigma_{i0}\varepsilon_t[\mathbf{Y}_{t-1}'(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{i+1,0})]\big\} \\
&\qquad\qquad \times \mathbb{1}\Big(r_{i0} + \frac{s_i}{n} < y_{t-d} \le r_{i0}\Big)\mathbb{1}(s_i < 0) \\
&\qquad + \big\{[\mathbf{Y}_{t-1}'(\boldsymbol{\beta}_{i+1,0} - \boldsymbol{\beta}_{i0})]^2 + 2\sigma_{i+1,0}\varepsilon_t[\mathbf{Y}_{t-1}'(\boldsymbol{\beta}_{i+1,0} - \boldsymbol{\beta}_{i0})]\big\} \\
&\qquad\qquad \times \mathbb{1}\Big(r_{i0} < y_{t-d} \le r_{i0} + \frac{s_i}{n}\Big)\mathbb{1}(s_i \ge 0)\Big] \\
&= \sum_{i=1}^{m-1}\sum_{t=1}^{n} \Big[\xi_t^{(i,i+1)}\mathbb{1}\Big(r_{i0} + \frac{s_i}{n} < y_{t-d} \le r_{i0}\Big)\mathbb{1}(s_i < 0) \\
&\qquad + \xi_t^{(i+1,i)}\mathbb{1}\Big(r_{i0} < y_{t-d} \le r_{i0} + \frac{s_i}{n}\Big)\mathbb{1}(s_i \ge 0)\Big],
\end{aligned}$$

where

$$\xi_t^{(i,j)} = \big[\mathbf{Y}_{t-1}'(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{j0})\big]^2 + 2\sigma_{i0}\varepsilon_t\mathbf{Y}_{t-1}'(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{j0}), \quad i, j = 1, ..., m. \tag{3.2}$$

Let $F_{(i,j)}(\cdot|r)$ be the conditional distribution function of $\xi_t^{(i,j)}$ given $y_{t-d} = r$. We define a sequence of independent one-dimensional two-sided compound Poisson processes $\{\mathfrak{P}_j(z), z \in \mathbb{R}\}_{j=1}^{m-1}$ as

$$\mathfrak{P}_j(z) = \mathbb{1}(z<0)\left[\sum_{k=1}^{\mathfrak{N}_1^{(j)}(-z)} Y_k^{(j,j+1)}\right] + \mathbb{1}(z \geq 0)\left[\sum_{k=1}^{\mathfrak{N}_2^{(j)}(z)} Z_k^{(j+1,j)}\right], \qquad (3.3)$$

for $j = 1, ..., m-1$, where $\{\mathfrak{N}_1^{(j)}(z), z \geq 0\}$ and $\{\mathfrak{N}_2^{(j)}(z), z \geq 0\}$ are two independent Poisson processes with $\mathfrak{N}_1^{(j)}(0) = \mathfrak{N}_2^{(j)}(0) = 0$ a.s. and with the same jump rate $\pi(r_{j0})$, where $\pi(\cdot)$ is the density function of y_0. $\{Y_k^{(j,j+1)}\}_{k=1}^{\infty}$ are i.i.d. random variables with the distribution $F_{(j,j+1)}(\cdot|r_{j0})$, and $\{Z_k^{(j+1,j)}\}_{k=1}^{\infty}$ are i.i.d. with the distribution $F_{(j+1,j)}(\cdot|r_{j0})$. $\{Y_k^{(j,j+1)}\}_{k=1}^{\infty}$ and $\{Z_k^{(j+1,j)}\}_{k=1}^{\infty}$ are mutually independent. Here, we work with the left continuous version for $\{\mathfrak{N}_1^{(j)}(\cdot)\}_{j=1}^{m-1}$ and the right continuous version for $\{\mathfrak{N}_2^{(j)}(\cdot)\}_{j=1}^{m-1}$.

We further define a spatial compound Poisson process $\wp(\mathbf{s})$ as follows,

$$\wp(\mathbf{s}) = \sum_{j=1}^{m-1} \mathfrak{P}_j(s_j), \quad \mathbf{s} = (s_1, ..., s_{m-1})' \in \mathbb{R}^{m-1}. \qquad (3.4)$$

Clearly, $\wp(\mathbf{s})$ goes to $+\infty$ a.s. when $\|\mathbf{s}\| \to \infty$ since $\mathbb{E}Y_t^{(i,i+1)} = \mathbb{E}Z_t^{(i+1,i)} > 0$ by Assumption 3.4. Therefore, there exists a unique random $(m-1)$-dimensional cube $[\mathbf{M}_-, \mathbf{M}_+) \equiv [M_-^{(1)}, M_+^{(1)}) \times \cdots \times [M_-^{(m-1)}, M_+^{(m-1)})$ on which the process $\{\wp(\mathbf{s}), \mathbf{s} \in \mathbb{R}^{m-1}\}$ attains its global minimum a.s. That is,

$$[\mathbf{M}_-, \mathbf{M}_+) = \arg\min_{\mathbf{s} \in \mathbb{R}^{m-1}} \wp(\mathbf{s}).$$

From (3.4), the minimization above is equivalent to

$$[M_-^{(j)}, M_+^{(j)}) = \arg\min_{z \in \mathbb{R}} \mathfrak{P}_j(z), \quad j = 1, ..., m-1.$$

Note that the processes $\{\mathfrak{P}_j(z)\}_{j=1}^{m-1}$ are independent, so are $\{M_-^{(j)}\}_{j=1}^{m-1}$. Now, we can state our another result as follows.

Theorem 3.3. *If Assumptions 3.1-3.4 hold, then $n(\widehat{\mathbf{r}}_n - \mathbf{r}_0)$ converges weakly to $\mathbf{M}_-$ and its components are asymptotically independent as $n \to \infty$. Furthermore, $n(\widehat{\mathbf{r}}_n - \mathbf{r}_0)$ is asymptotically independent of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ which is always asymptotically normal.*

When $m = 2$, Theorem 3.3 reduces to Theorem 2.2 of Chan (1993). The limit distribution of M_- does not have a closed form and depends on the nuisance parameters and the distribution of ε_t. In next section, we will describe how to do inference for $\mathbf{r}_0$ via a simulation method.

4. Numerical implementation of $\mathbf{M}_-$

From Theorem 3.3, we know that obtaining $\mathbf{M}_-$ is equivalent to obtaining each $M_-^{(j)}$ separately for $j = 1, ..., m-1$. From (3.3), we know that two factors determine $M_-^{(j)}$, that is, the jump rate $\pi(r_{j0})$ and the jump distributions $F_{(j,j+1)}(\cdot|r_{j0})$ and $F_{(j+1,j)}(\cdot|r_{j0})$. We can simulate $M_-^{(j)}$ via simulating the two-sided compound Poisson process (3.3) on the interval $[-T, T]$ for any given $T > 0$ large enough. The algorithm is as follows.

Algorithm A:

Step A.1. Generate two independent Poisson r.v.s $N_1^{(j)}$ and $N_2^{(j)}$ with the same parameter $\pi(r_{j0})T$ which are the total number of jumps on the intervals $[-T, 0]$ and $[0, T]$, respectively.

Step A.2. Generate two independent jump time sequences: $\{U_1, \cdots, U_{N_1^{(j)}}\}$ and $\{V_1, \cdots, V_{N_2^{(j)}}\}$, where U_i's and V_i's are independently and uniformly distributed on $[-T, 0]$ and $[0, T]$, respectively.

Step A.3. Generate two independent jump-size sequences: $\{Y_1, \cdots, Y_{N_1^{(j)}}\}$ and $\{Z_1, \cdots, Z_{N_2^{(j)}}\}$ from $F_{(j,j+1)}(\cdot|r_{j0})$ and $F_{(j+1,j)}(\cdot|r_{j0})$, respectively.

For $z \in [-T, T]$, the trajectory of (3.3) is given by

$$\mathfrak{P}_j(z) = \mathbb{1}(z < 0) \sum_{i=1}^{N_1^{(j)}} \mathbb{1}(U_i > z) Y_i + \mathbb{1}(z \geq 0) \sum_{j=1}^{N_2^{(j)}} \mathbb{1}(V_j < z) Z_j.$$

Then, we take the smallest minimizer of $\mathfrak{P}_j(z)$ on $[-T, T]$ as one observed value of $M_-^{(j)}$. Repeat above algorithm B times and use the nonparametric kernel method, we can get the density of $M_-^{(j)}$ numerically for $j = 1, ..., m-1$.

In the above algorithm, the key is how to generate the jump-size sequences from $F_{(j,j+1)}(\cdot|r_{j0})$ and $F_{(j+1,j)}(\cdot|r_{j0})$ in Step A.3. When $p = 1$, it is easy because the conditional distributions $F_{(j,j+1)}(\cdot|r_{j0})$ and $F_{(j+1,j)}(\cdot|r_{j0})$ become the unconditional ones. When $p > 1$, it is more complicate. Note that

$$\begin{aligned}
F_{(j,j+1)}(x|r_{j0}) &= \int_{\mathbb{R}^p} \mathbb{P}(\xi_p^{(j,j+1)} \le x | y_{p-d} = r_{j0}, \mathbf{Z}_{p-d-1} = \mathbf{z}) \frac{\pi(r_{j0}|\mathbf{z})}{\pi(r_{j0})} G(d\mathbf{z}) \\
&\approx \frac{1}{K} \sum_{i=1}^{K} \mathbb{P}(\xi_p^{(j,j+1)} \le x | y_{p-d} = r_{j0}, \mathbf{Z}_{p-d-1} = \mathbf{z}_i) \frac{\pi(r_{j0}|\mathbf{z}_i)}{\pi(r_{j0})} \\
&\approx \sum_{i=1}^{K} \mathbb{P}(\xi_p^{(j,j+1)} \le x | y_{p-d} = r_{j0}, \mathbf{Z}_{p-d-1} = \mathbf{z}_i) \frac{\pi(r_{j0}|\mathbf{z}_i)}{\sum_{l=1}^{K} \pi(r_{j0}|\mathbf{z}_l)}
\end{aligned}$$

by the property of the conditional expectation, the law of large numbers (that is, $K \to \infty$), and $\mathbb{E}[\pi(r_{j0}|\mathbf{Z}_{p-d-1})] = \pi(r_{j0})$, where $\mathbf{Z}_{p-d-1} = (y_{p-d-1}, ..., y_{-d})'$, $\mathbf{z}_i \in \mathbb{R}^p$, $G(\cdot)$ is the distribution of $\mathbf{Z}_{p-d-1}$, and $\pi(r_{j0}|\mathbf{z})$ is the conditional density of y_{p-d} given $\mathbf{Z}_{p-d-1} = \mathbf{z}$. Let

$$\begin{aligned}
h(\mathbf{Y}_{t-1}, \boldsymbol{\theta}) &= \sum_{j=1}^{m} \mathbf{Y}'_{t-1} \boldsymbol{\beta}_j \mathbb{1}(r_{j-1} < y_{t-d} \le r_j), \\
\sigma(\mathbf{Y}_{t-1}, \boldsymbol{\theta}) &= \sum_{j=1}^{m} \sigma_{j0} \mathbb{1}(r_{j-1} < y_{t-d} \le r_j).
\end{aligned}$$

Then,

$$\pi(r_{j0}|\mathbf{z}_i) = [\sigma(\mathbf{y}_i, \boldsymbol{\theta}_0)]^{-1} f_\varepsilon([\sigma(\mathbf{y}_i, \boldsymbol{\theta}_0)]^{-1}[r_{j0} - h(\mathbf{y}_i, \boldsymbol{\theta}_0)]),$$

where $\mathbf{y}_i = (1, \mathbf{z}'_i)'$ and $f_\varepsilon(\cdot)$ is the density of ε_t.

When $\boldsymbol{\theta}_0$, σ_{j0}, $\pi(r_{j0})$, $f_\varepsilon(\cdot)$ and $G(\cdot)$ are known, the following algorithm describes how to sample an observation Y_1 from $F_{(j,j+1)}(x|r_{j0})$.

Algorithm B:

Step B.1. Choose a sufficiently large positive integer K and then draw a sample $\{\mathbf{z}_1, ..., \mathbf{z}_K\}$ from $G(\cdot)$.

Step B.2. Calculate $\pi(r_{j0}|\mathbf{z}_i)$'s, and generate a random variable U, independent of y_{p-d} and $\mathbf{Z}_{p-d-1}$, such that $\mathbb{P}(U = i) = \frac{\pi(r_{j0}|\mathbf{z}_i)}{\sum_{l=1}^{K} \pi(r_{j0}|\mathbf{z}_l)}$ for $i = 1, ..., K$.

Step B.3.[2] For each fixed $i \in \{1, ..., K\}$, given the initial values $y_{p-d} = r_{j0}$ and $\mathbf{Y}_{p-d-1} = (1, \mathbf{z}_i')'$, generate $\{y_{p-d+1}, ..., y_{p-1}\}$ by iterating equation (2.1) $(d-1)$-times and ε_p from $f_\varepsilon(\cdot)$, and then use them to calculate $\xi_p^{(j,j+1)}$ in (3.2) and write the value as ζ_i.

Step B.4. Obtain an observation $Y_1 = \zeta_U$.

In practice, however, since only one sample $\{y_1, ..., y_n\}$ is available, we can use $\widehat{\boldsymbol{\theta}}_n$, $\widehat{\sigma}_{j0}$, $\widehat{\pi}(\widehat{r}_{j0})$ and $\widehat{f}_\varepsilon(\cdot)$ to replace $\boldsymbol{\theta}_0$, σ_{j0}, $\pi(r_{j0})$ and $f_\varepsilon(\cdot)$, respectively, where $\widehat{\pi}(\cdot)$ and $\widehat{f}_\varepsilon(\cdot)$ are the kernel density estimators of $\pi(\cdot)$ and $f_\varepsilon(\cdot)$, respectively. In Step B.1, we can let $K = n - p + 1$ and $\mathbf{z}_i = (y_{i+p-1}, ..., y_i)'$. Then, by Algorithm B, we can get an observation Y_1 from $F_{(j,j+1)}(\cdot|r_{j0})$ approximately. Similarly, we draw an observation Z_1 from $F_{(j+1,j)}(\cdot|r_{j0})$. Simulation studies show that the simulation method above does work well for simulating the distribution of $\mathbf{M}_-$.

5. Simulation studies

To assess the performance of the LSE of $\boldsymbol{\theta}_0$ in finite samples, we use sample sizes $n = 300, 600, 900$ and 1200, each with replications 1000 for the following three-regime TAR model:

$$y_t = \begin{cases} \beta_{10} + \beta_{11} y_{t-1} + \varepsilon_t, & y_{t-1} \le r_1, \\ \beta_{20} + \beta_{21} y_{t-1} + \varepsilon_t, & r_1 < y_{t-1} \le r_2, \\ \beta_{30} + \beta_{31} y_{t-1} + \varepsilon_t, & y_{t-1} > r_2, \end{cases} \tag{5.1}$$

with the true value $\boldsymbol{\theta}_0 = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, \beta_{30}, \beta_{31}, r_1, r_2)' = (1, -0.4, 0.6, 1, -1, -0.2, -0.8, 0.5)'$ and $\varepsilon_t \sim i.i.d. \mathcal{N}(0, 1)$. Clearly, the autoregressive function is not continuous over two thresholds $\{-0.8, 0.5\}$.

Table 1 summarizes the bias, the empirical standard deviation (ESD) and the asymptotic standard deviation (ASD). Here, the ASD's of $\widehat{\boldsymbol{\beta}}_n$ are computed by using Σ in Theorem 3.2 and the ASD of $\widehat{\mathbf{r}}_n$ is obtained by the simulation method in Section 4. From Table 1, we can see that the larger the sample size, the closer the ESDs and ASDs on the whole. We also see that the ESDs of $\widehat{\mathbf{r}}_n$ are smaller than those of $\widehat{\boldsymbol{\beta}}_n$. This partially illustrates the n-consistency of the estimated thresholds, under which they would approach the true thresholds much faster than other estimated parameters.

[2]When $d = 1$, the iteration is not necessary since y_p is not needed.

[Table 1 about here.]

It is well known that the stationarity of the multiple-regime TAR(1) model is determined by the extreme left and the extreme right regimes, see Chan et al. (1985) and Chen and Tsay (1991). In model (5.1), although $\beta_{21} = 1$, $\{y_t\}$ is not a unit-root process even not a partial unit-root process in Liu et al. (2010). $\sqrt{n}(\widehat{\beta}_{21,n} - \beta_{21,0})$ is still asymptotically normal. Fig 1 displays the densities of $\sqrt{n}(\widehat{\beta}_{21,n} - \beta_{21,0})$ and $\mathcal{N}(0, 5.35^2)$ when $n = 300$ and 600, respectively. The number 5.35^2 is the estimator of the asymptotic variance of $\widehat{\beta}_{21,n}$ in Theorem 3.2 (ii). From Fig 1, we see that they are very close each other when $n = 600$.

[Figure 1 about here.]

We now study the coverage probabilities of r_{10} and r_{20}. Using the simulation method in Section 4, we first obtain the empirical quantiles of $M_-^{(1)}$ and $M_-^{(2)}$ by 20,000 replication. When the significance level α equals 0.5%, 1%, 2.5%, 5%, 95%, 97.5%, 99% and 99.5%, the values are given in Table 2.

[Table 2 about here.]

Based on the critical values in Table 2, the coverage probabilities of r_{10} and r_{20} are reported in Table 3 when $n = 300$, 600, 900 and 1200, respectively. It can be seen that the coverage probability is rather accurate.

[Table 3 about here.]

To see the overall approximation of the estimated thresholds, Fig 2 shows the density functions of $n(\widehat{r}_{jn} - r_{j0})$ and $M_-^{(j)}$, $j = 1, 2$, when $n = 900$. It is plotted by the software R. A nonparametric kernel method is used and the bandwidth is chosen automatically. From Fig 2, we see that the density functions of both $n(\widehat{r}_{jn} - r_{j0})$ and $M_-^{(j)}$ are very close. We also note that the density of $M_-^{(j)}$ is leptokurtic and asymmetric, skewing towards the left hand side of the origin. In fact, the skewness is -0.04 and the kurtosis is 12.51 for $M_-^{(1)}$, and the skewness is -0.18 and the kurtosis is 9.37 for $M_-^{(2)}$. Owing to the skewness, an extreme caution should be taken in constructing confidence intervals of thresholds in practice.

[Figure 2 about here.]

By Theorem 3.3, $n(\widehat{r}_{1n} - r_{10})$ and $n(\widehat{r}_{2n} - r_{20})$ are asymptotically independent as $n \to \infty$. To check this fact empirically in finite samples, the multivariate independence test is used, which is based on the empirical copula process and is proposed by Genest and Rémillard (2004). This test can be implemented by the functions "`indepTestSim`" and "`indepTest`" contained within the package *copula* in the software R. The p-value of the test is summarized in Table 4 when $n = 300, 600, 900$ and 1200. From Table 4, one can see that $n(\widehat{r}_{1n} - r_{10})$ and $n(\widehat{r}_{2n} - r_{20})$ are indeed independent at 5% significant level. This is a further evidence of Theorem 3.3 to a certain extent.

[Table 4 about here.]

Finally, we see whether the simulation method in Section 4 works well. We generate a single sample $\{y_1, ..., y_{300}\}$ from model (5.1). Fig 3 exhibits the densities of $M_{-}^{(j)}$ obtained when all parameters used in the algorithm are known and unknown, respectively. It shows that two curves are almost identical.

[Figure 3 about here.]

6. An empirical example

In economics, to characterize the dynamics of macroeconomic variables, some researchers suggested that two-regime TAR models may be appropriate for expansion and recession, see Tiao and Tsay (1994). Others (e.g., Koop and Potter 1999), however, argued that perhaps three-regime TAR models, encompassing bad times, good times and normal times, should be more reasonable. Following this suggestion, we use a three-regime TAR model (2.1) to fit the quarterly U.S. real GNP data over the period 1947-2009 with a total of 252 observations.

Let $y_1, ..., y_{252}$ denote the original data. We consider the growth rate series

$$x_t = 100(\log y_t - \log y_{t-1}), \quad t = 2, ..., 252.$$

The data $\{y_t\}$ and the growth rate series $\{x_t\}$ are plotted in Fig 4.

[Figure 4 about here.]

Setting $m_0 = \max\{p_1, p_2, p_3\} \le 10$ and $1 \le d \le \max\{m_0, 1\}$, the AIC selects the following TAR model:

$$x_t = \begin{cases} \beta_{10} + \sum_{i=1}^{6} \beta_{1i} x_{t-i} + \varepsilon_t, & x_{t-6} \le 1.2029, \\ \beta_{20} + \sum_{i=1}^{7} \beta_{2i} x_{t-i} + \varepsilon_t, & 1.2029 < x_{t-6} \le 2.4266, \\ \beta_{30} + \sum_{i=1}^{10} \beta_{3i} x_{t-i} + \varepsilon_t, & x_{t-6} > 2.4266. \end{cases} \tag{6.1}$$

The coefficients $\{\beta_{ij}\}$ with their standard deviations are summarized in Table 5.

[Table 5 about here.]

$\beta_{14}, \beta_{22}, \beta_{31}$ and β_{37} are not significant at the 5% level. The standard deviation of ε_t is 0.7976. The Ljung-Box test statistics $Q(6) = 0.998$ and the Li-Mak test statistics $Q^2(6) = 0.518$, which suggest that model (6.1) is adequate for $\{x_t\}$. The number of data $\{x_t\}$ in three regimes are 79, 121 and 51, respectively. The 95% confidence intervals of r_{10} and r_{20} are $(1.0546, 1.3388)$ and $(2.2719, 2.5432)$, respectively. Based on the procedure in Section 4, the densities of the estimated thresholds are plotted Fig 5.

[Figure 5 about here.]

According to model (6.1), the normal growth rate of the U.S. GNP is in the interval $(1.2029, 2.4266]$. The growth rate is considered as the high one if it is bigger than 2.4266. Otherwise, it is regarded as the low one if it is smaller than 1.2029. In each regime, the growth rate can be fitted by different AR models, respectively.

7. Proofs of Theorems 3.1-3.3

7.1. Proof of Theorem 3.1

The proof of Theorem 3.1 is similar to that of Theorem 1 in Chan (1993) by the following lemma and hence it is omitted.

Lemma 7.1. *If the conditions in Theorem 3.1 hold, then* $\mathbb{E}[y_t - \mathbb{E}_{\boldsymbol{\theta}}(y_t|\mathcal{F}_{t-1})]^2 \ge \mathbb{E}[y_t - \mathbb{E}_{\boldsymbol{\theta}_0}(y_t|\mathcal{F}_{t-1})]^2$ *for all* $\boldsymbol{\theta} \in \Theta \times \{1, ..., D_0\}$ *and the equality holds if and only if* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Proof. For simplicity, we consider the case when $m = 3$ and the proof is similar when $m > 3$. Clearly,

$$\begin{aligned}
\mathbb{E}[y_t - \mathbb{E}_{\boldsymbol{\theta}}(y_t|\mathcal{F}_{t-1})]^2 &= \mathbb{E}[y_t - \mathbb{E}_{\boldsymbol{\theta}_0}(y_t|\mathcal{F}_{t-1}) + \mathbb{E}_{\boldsymbol{\theta}_0}(y_t|\mathcal{F}_{t-1}) - \mathbb{E}_{\boldsymbol{\theta}}(y_t|\mathcal{F}_{t-1})]^2 \\
&= \mathbb{E}[y_t - \mathbb{E}_{\boldsymbol{\theta}_0}(y_t|\mathcal{F}_{t-1})]^2 + \mathbb{E}\{[\mathbb{E}_{\boldsymbol{\theta}_0}(y_t|\mathcal{F}_{t-1}) - \mathbb{E}_{\boldsymbol{\theta}}(y_t|\mathcal{F}_{t-1})]^2\} \\
&\geq \mathbb{E}[y_t - \mathbb{E}_{\boldsymbol{\theta}_0}(y_t|\mathcal{F}_{t-1})]^2.
\end{aligned}$$

If the equality holds for some $\boldsymbol{\theta}^*$, then, $\mathbb{E}_{\boldsymbol{\theta}_0}(y_t|\mathcal{F}_{t-1}) = \mathbb{E}_{\boldsymbol{\theta}^*}(y_t|\mathcal{F}_{t-1})$ a.s., or

$$\mathbf{Y}'_{t-1}\Big[\sum_{j=1}^{3}\boldsymbol{\beta}_j^*\mathbb{1}(r_{j-1}^* < y_{t-d^*} \leq r_j^*) - \sum_{j=1}^{3}\boldsymbol{\beta}_{j0}\mathbb{1}(r_{j-1,0} < y_{t-d_0} \leq r_{j0})\Big] = 0 \text{ a.s.}$$

which is equivalent to

$$\sum_{i=1}^{3}\sum_{j=1}^{3}(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_j^*)\mathbb{1}(r_{j-1,0} < y_{t-d_0} \leq r_{j0}, r_{j-1}^* < y_{t-d^*} \leq r_j^*) = \mathbf{0}$$

with $r_0^* = -\infty$ and $r_3^* = \infty$. Using the orthogonality among the indicator functions above, we have

$$\sum_{i=1}^{3}\sum_{j=1}^{3}\|\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_j^*\|\mathbb{P}(r_{i-1,0} < y_{t-d_0} \leq r_{i0}, r_{j-1}^* < y_{t-d^*} \leq r_j^*) = 0. \qquad (7.1)$$

The *first step* is to prove that $d^* = d_0$. If not, without loss of generality, suppose that $d^* > d_0$. Using the independence of ε_{t-d_0} and $\{\mathbf{Y}_{t-d_0-1}, y_{t-d^*}\}$ and the condition (iii) in Theorem 3.1, then, for any $i, j = 1, 2, 3$,

$$\begin{aligned}
&\mathbb{P}(r_{i-1,0} < y_{t-d_0} \leq r_{i0}, r_{j-1}^* < y_{t-d^*} \leq r_j^*) \\
&= \mathbb{P}(r_{i-1,0} < \varepsilon_{t-d_0} + h(\mathbf{Y}_{t-d_0-1}, \boldsymbol{\theta}_0) \leq r_{i0}, r_{j-1}^* < y_{t-d^*} \leq r_j^*) > 0.
\end{aligned}$$

Thus, by (7.1), $\boldsymbol{\beta}_{10} = \boldsymbol{\beta}_{20} = \boldsymbol{\beta}_{30} = \boldsymbol{\beta}_1^* = \boldsymbol{\beta}_2^* = \boldsymbol{\beta}_3^*$, which is a contradiction with the condition (ii). Thus, $d^* \leq d_0$. Similarly, one can prove that $d^* \geq d_0$. Therefore, $d^* = d_0$.

The *second step* is to show that $r_1^* = r_{10}$. Suppose that $r_1^* < r_{10}$. If $r_2^* \leq r_{10}$, then we have

$$\begin{aligned}
&\mathbb{P}(r_{20} < y_{t-d_0} \leq r_{30}, r_2^* < y_{t-d_0} \leq r_3^*) = \mathbb{P}(y_{t-d_0} > r_{20}) > 0, \\
&\mathbb{P}(r_{10} < y_{t-d_0} \leq r_{20}, r_2^* < y_{t-d_0} \leq r_3^*) = \mathbb{P}(r_{10} < y_{t-d_0} \leq r_{20}) > 0,
\end{aligned}$$

since the density of y_{t-d_0} is continuous and positive on $\mathbb{R}$. Thus, by (7.1), we have $\boldsymbol{\beta}_{20} = \boldsymbol{\beta}_3^* = \boldsymbol{\beta}_{30}$. Similarly, if $r_2^* > r_{10}$, then we can show that $\boldsymbol{\beta}_{10} = \boldsymbol{\beta}_{20}$. Both cases result in a contradiction with the condition (ii). Hence, $r_1^* \geq r_{10}$. Similarly, one can show $r_1^* \leq r_{10}$. Thus, $r_1^* = r_{10}$. Using the same way, one can get $r_2^* = r_{20}$ and in turn get $\boldsymbol{\beta}_i^* = \boldsymbol{\beta}_{i0}$ for $i = 1, 2, 3$. Thus, $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$. $\square$

7.2. Proof of Theorem 3.2

(i). Since $\widehat{\boldsymbol{\theta}}_n$ is consistent by Theorem 3.1, we restrict the parameter space to an open neighborhood of $\boldsymbol{\theta}_0$. To this end, define $V_\delta = \{\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| < \delta,\ |r_j - r_{j0}| < \delta, j = 1, ..., m-1\}$ for some $0 < \delta < 1$ to be determined later. Choose δ small enough so that $\{r : |r - r_{j-1,0}| < \delta\} \cap \{r : |r - r_{j0}| < \delta\} = \varnothing$ for $j = 2, ..., m-1$. It suffices to prove that for any $\epsilon > 0$, there exists a $B > 0$ such that with probability greater than $1 - \epsilon$,

$$L_n(\boldsymbol{\beta}, \mathbf{r}) - L_n(\boldsymbol{\beta}, \mathbf{r}_0) > 0 \quad \text{for } \|\mathbf{r} - \mathbf{r}_0\| > B/n \text{ and } \boldsymbol{\theta} \in V_\delta.$$

By a simple calculation, we have the following decomposition, which plays a key role in the proof so that we can use the results about two-regime TAR models to obtain counterparts of multiple-regime ones,

$$L_n(\boldsymbol{\beta}, \mathbf{r}) - L_n(\boldsymbol{\beta}, \mathbf{r}_0) = \sum_{j=1}^{m-1} L_n^{(j)}(\boldsymbol{\beta}, r_j),$$

where $\mathbf{r} = (r_1, ..., r_{m-1})' \in \mathbb{R}^{m-1}$ and

$$\begin{aligned} L_n^{(j)}(\boldsymbol{\beta}, r_j) = \sum_{t=1}^{n} \Big\{ &\operatorname{sign}(r_j - r_{j0}) \left[(y_t - \mathbf{Y}_{t-1}'\boldsymbol{\beta}_j)^2 - (y_t - \mathbf{Y}_{t-1}'\boldsymbol{\beta}_{j+1})^2 \right] \\ &\times \mathbb{1}(r_j \wedge r_{j0} < y_{t-d} \leq r_j \vee r_{j0}) \Big\}. \end{aligned}$$

For each $L_n^{(j)}(\boldsymbol{\beta}, r_j)$, it suffices to prove that for any $\epsilon_j > 0$, there exists a $B_j > 0$ such that with probability greater than $1 - \epsilon_j$,

$$L_n^{(j)}(\boldsymbol{\beta}, r_j) > 0 \quad \text{for } |r_j - r_{j0}| > B_j/n \text{ and } \boldsymbol{\theta} \in V_\delta. \tag{7.2}$$

Following the idea of the proof of Proposition 1 in Chan (1993) and we only need to verify same inequalities as (4.4a)-(4.4c) in (1993). These inequalities still hold under Assumptions 3.1-3.4 since their proofs only require the V-uniform ergodicity of $\{\mathbf{Z}_t\}$ and discontinuity of the autoregressive mean function $h(\mathbf{x}, \boldsymbol{\theta}_0) = \sum_{j=1}^{m} \mathbf{x}'\boldsymbol{\beta}_{j0}\mathbb{1}(r_{j-1,0} < x_d \le r_{j0})$ over each hyperplane $x_d = r_{i0}$ for $j = 1, ..., m-1$. Hence, (7.2) holds and $n|\widehat{r}_{jn} - r_{j0}| = O_p(1)$, $j = 1, ..., m-1$, i.e., $n\|\widehat{\mathbf{r}}_n - \mathbf{r}_0\| = O_p(1)$. This completes the proof of (i).

As for (ii), the proof is similar to that of Theorem 4 in Qian (1998), and hence it is omitted. □

7.3. Proof of Theorem 3.3

For simplicity, we only deal with the case when $m = 3$ and the proof is similar when $m > 3$. First of all, we consider the case $s_1 \ge 0, s_2 \ge 0$. By a calculation, it follows that

$$\begin{aligned}\wp_n(\mathbf{s}) &= L_n(\boldsymbol{\beta}_0, \mathbf{r}_0 + \frac{\mathbf{s}}{n}) - L_n(\boldsymbol{\beta}_0, \mathbf{r}_0) \\ &= \sum_{t=1}^{n} \left[\xi_t^{(2,1)}\mathbb{1}\left(r_{10} < y_{t-d} \le r_{10} + \frac{s_1}{n}\right) + \xi_t^{(3,2)}\mathbb{1}\left(r_{20} < y_{t-d} \le r_{20} + \frac{s_2}{n}\right)\right],\end{aligned}$$

where $\mathbf{s} = (s_1, s_2)' \in \mathbb{R}^2$ and $\xi_t^{(i,j)}$ is defined in (3.2). Without loss of generality, we assume $\xi_t^{(i,j)}$ is bounded. Otherwise, use the truncating technique in Li et al. (2010) to truncate $\xi_t^{(i,j)}$ and then consider a truncated process.

Consider the weak convergence of the process $\wp_n(\mathbf{s})$ on the rectangle $\mathbf{T} = [0, T] \times [0, T]$. The tightness of $\wp_n(\mathbf{s})$ can be easily shown by Theorem 5 in Kushner (1984, page 32) by Assumption 3.1. The key step is to describe convergence of finite dimensional distributions. To do this, for any $\mathbf{s}_i = (s_{i1}, s_{i2})' \in \mathbf{T}$, satisfying $s_{1j} \le s_{2j} < s_{3j} \le s_{4j}$, $i = 1, ..., 4$, $j = 1, 2$, and for any constants c_1 and c_2, the linear combination of the increments of $\wp_n(\mathbf{s})$ is

$$\begin{aligned}S_n &\equiv c_1\{\wp_n(\mathbf{s}_2) - \wp_n(\mathbf{s}_1)\} + c_2\{\wp_n(\mathbf{s}_4) - \wp_n(\mathbf{s}_3)\} \\ &= \sum_{t=1}^{n} \left\{\xi_t^{(2,1)}\left[c_1\mathbf{1}_t^{(1,1)} + c_2\mathbf{1}_t^{(1,3)}\right] + \xi_t^{(3,2)}\left[c_1\mathbf{1}_t^{(2,1)} + c_2\mathbf{1}_t^{(2,3)}\right]\right\},\end{aligned}$$

where

$$\mathbf{1}_t^{(i,j)} = \mathbb{1}\left(r_{i0} + \frac{s_{ji}}{n} < y_{t-d} \le r_{i0} + \frac{s_{j+1,i}}{n}\right), \quad i = 1, 2, \; j = 1, 3.$$

Let

$$J_t = \xi_t^{(2,1)}\left[c_1\mathbf{1}_t^{(1,1)} + c_2\mathbf{1}_t^{(1,3)}\right] + \xi_t^{(3,2)}\left[c_1\mathbf{1}_t^{(2,1)} + c_2\mathbf{1}_t^{(2,3)}\right].$$

We first verify Assumptions A.1-A.3 in Appendix A in Li et al. (2010) for J_t. By Assumption 3.3, we have

$$\lambda = \lim_{\substack{\varepsilon\to 0\\ n\to\infty}} \varepsilon^{-1}\mathbb{P}_k^{\varepsilon}(J_n^{\varepsilon}\neq 0) = \pi(r_{10})[(s_{21}-s_{11})+(s_{41}-s_{31})] \\ + \pi(r_{20})[(s_{22}-s_{12})+(s_{42}-s_{32})]. \tag{7.3}$$

By the stationarity of $\{y_t\}$ and Assumption 3.3 again, for any Borel set B, it follows that

$$\mathbb{Q}^*(B) = \sum_{i=1}^{4} w_i \mathbb{Q}_i^*(B), \tag{7.4}$$

where

$$w_1 = \pi(r_{10})(s_{21}-s_{11})/\lambda, \quad w_2 = \pi(r_{10})(s_{41}-s_{31})/\lambda,$$
$$w_3 = \pi(r_{20})(s_{22}-s_{12})/\lambda, \quad w_4 = \pi(r_{20})(s_{42}-s_{32})/\lambda,$$

and

$$\mathbb{Q}_1^*(B) = \mathbb{P}(c_1\xi_t^{(2,1)}\in B|y_{t-d}=r_{10}),\ \mathbb{Q}_2^*(B) = \mathbb{P}(c_2\xi_t^{(2,1)}\in B|y_{t-d}=r_{10}),$$
$$\mathbb{Q}_3^*(B) = \mathbb{P}(c_1\xi_t^{(3,2)}\in B|y_{t-d}=r_{20}),\ \mathbb{Q}_4^*(B) = \mathbb{P}(c_2\xi_t^{(3,2)}\in B|y_{t-d}=r_{20}).$$

Similarly, we can verify that, for any $f\in\widehat{C}_0^2$, a space of functions with compact support and continuous second derivative, and a scalar x,

$$\lim_{n\to\infty}\mathbb{E}_k^{\varepsilon}\{f(x+J_n^{\varepsilon})-f(x)|J_n^{\varepsilon}\neq 0\} = \lim_{n\to\infty}\mathbb{E}\{f(x+J_n^{\varepsilon})-f(x)|J_n^{\varepsilon}\neq 0\} \\ = \int [f(x+u)-f(x)]\mathbb{Q}^*(du). \tag{7.5}$$

By (7.3)-(7.5), Assumptions A.1-A.3 in Li et al. (2010) hold. Furthermore, by their Theorem A.1, we claim that S_n weakly converges to a compound Poisson random variable J with jump rate λ and the jump distribution $\mathbb{Q}^*$. The characteristic function $f_J(t)$ of J can be written as

$$f_J(t) = \exp\left\{-\lambda\left[1-\int_{\mathbb{R}} e^{itx}\mathbb{Q}^*(dx)\right]\right\} = \prod_{i=1}^{4}\exp\left\{-\lambda w_i\left[1-\int_{\mathbb{R}} e^{itx}\mathbb{Q}_i^*(dx)\right]\right\},$$

which is equal to that of the linear combination $c_1\{\mathscr{P}(\mathbf{s}_2)-\mathscr{P}(\mathbf{s}_1)\}+c_2\{\mathscr{P}(\mathbf{s}_4)-\mathscr{P}(\mathbf{s}_3)\}$ of the independent increments of a spatial compound Poisson process

$$\mathscr{P}(\mathbf{s}) = \sum_{i=1}^{N^{(1)}_{(2,1)}(s_1)} \zeta^{(2,1)}_{1,i} + \sum_{i=1}^{N^{(1)}_{(3,2)}(s_2)} \zeta^{(3,2)}_{1,i}, \quad s_1 \geq 0, s_2 \geq 0,$$

that is, the finite dimensional distribution of $\wp_n(\mathbf{s})$ converges weakly to those of $\mathscr{P}(\mathbf{s})$ as $s_1 \geq 0$ and $s_2 \geq 0$.

Similarly, we can deal with other cases, respectively. Thus, as $n \to \infty$,

$$\begin{aligned}\wp_n(\mathbf{s}) \Longrightarrow & \left\{ \sum_{i=1}^{N^{(1)}_{(2,1)}(s_1)} \zeta^{(2,1)}_{1,i} + \sum_{i=1}^{N^{(1)}_{(3,2)}(s_2)} \zeta^{(3,2)}_{1,i} \right\} \mathbb{1}(s_1 \geq 0, s_2 \geq 0) \\ & + \left\{ \sum_{i=1}^{N^{(2)}_{(2,1)}(s_1)} \zeta^{(2,1)}_{2,i} + \sum_{i=1}^{N^{(1)}_{(2,3)}(-s_2)} \zeta^{(2,3)}_{1,i} \right\} \mathbb{1}(s_1 \geq 0, s_2 < 0) \\ & + \left\{ \sum_{i=1}^{N^{(1)}_{(1,2)}(-s_1)} \zeta^{(1,2)}_{1,i} + \sum_{i=1}^{N^{(2)}_{(3,2)}(s_2)} \zeta^{(3,2)}_{2,i} \right\} \mathbb{1}(s_1 < 0, s_2 \geq 0) \\ & + \left\{ \sum_{i=1}^{N^{(2)}_{(1,2)}(-s_1)} \zeta^{(1,2)}_{2,i} + \sum_{i=1}^{N^{(2)}_{(2,3)}(-s_2)} \zeta^{(2,3)}_{2,i} \right\} \mathbb{1}(s_1 < 0, s_2 < 0),\end{aligned}$$

where $\{N^{(k)}_{(i+1,i)}(z)\}$ and $\{N^{(k)}_{(i,i+1)}(z)\}$ are independent Poisson processes with the jump rate $\pi(r_{i0})$ for $k, i = 1, 2$, $\{\zeta^{(i,j)}_{k,l}\}_{l=1}^{\infty}$ are i.i.d random variables from $F_{(i,j)}(\cdot|r_{i\wedge j,0})$. All random variables and processes defined above are independent. Since $\{N^{(1)}_{(2,1)}(z)\}$ and $\{N^{(2)}_{(2,1)}(z)\}$ have the same jump rate $\pi(r_{10})$ and $\zeta^{(2,1)}_{1,i}$ and $\zeta^{(2,1)}_{2,i}$ have the same distribution, there exist the process $\mathfrak{N}^{(1)}_2(z)$ and the i.i.d. random variables $\{V^{(2,1)}_k\}$, defined by (3.3), such that

$$\begin{aligned}& \mathbb{1}(s_1 \geq 0, s_2 \geq 0) \sum_{i=1}^{N^{(1)}_{(2,1)}(s_1)} \zeta^{(2,1)}_{1,i} + \mathbb{1}(s_1 \geq 0, s_2 < 0) \sum_{i=1}^{N^{(2)}_{(2,1)}(s_1)} \zeta^{(2,1)}_{2,i} \\ & \stackrel{d}{=} \mathbb{1}(s_1 \geq 0) \sum_{k=1}^{\mathfrak{N}^{(1)}_2(s_1)} V^{(2,1)}_k,\end{aligned}$$

where "$X \stackrel{d}{=} Y$" indicates that the random elements X and Y have the same distribution. Similarly, using the same way to combine the other terms, we can get

$$\wp_n(\mathbf{s}) \Longrightarrow \wp(\mathbf{s}), \quad \text{as } n \to \infty,$$

where $\wp(\mathbf{s})$ is defined in (3.4). Thus, $\widetilde{L}_n(\mathbf{s})$, defined in (3.1), converges weakly to $\wp(\mathbf{s})$ as $n \to \infty$. Using Skorohod embedding, we may assume for simplicity that the convergence is almost sure convergence. Since $\widehat{\mathbf{r}}_n = \mathbf{r}_0 + O_p(n^{-1})$, it is readily seen that $n(\widehat{\mathbf{r}}_n - \mathbf{r}_0)$ converges weakly to $\mathbf{M}_-$, where $[\mathbf{M}_-, \mathbf{M}_+)$ is the unique $(m-1)$-dimensional random cube over which $\wp(\mathbf{s})$ attains its global minimum. The remainder of the proof is similar to that of Theorem 2 in Chan (1993). □

Acknowledgements

We thank two anonymous referees, the Associate Editor and the Editor for their very helpful comments and suggestions. The research was partially supported by Hong Kong Research Grants Commission Grants HKUST601607 and HKUST602609..

References

An, H. Z. and Huang, F.C. (1996). The geometrical ergodicity of nonlinear autoregressive models. *Statist. Sinica* **6** 943–956.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *J. Appl. Econometrics* **18** 1–22.

Bai, J. and Perron, P. (2006). Multiple structural change models: a simulation analysis. In *Econometric theory and practice* (edited by Corbae, D., Durlauf, S.N. and Hansen, B.E.), Cambridge Univ. Press, Cambridge. 212–237.

Brockwell, P.J., Liu, J. and Tweedie, R.L. (1992). On the existence of stationary threshold autoregressive moving-average processes. *J. Time Ser. Anal.* **13** 95–107.

Caner, M. and Hansen, B. E. (2001). Threshold autoregression with a unit root. *Econometrica* **69** 1555–1596.

Chan, K. S. (1989). A note on the geometric ergodicity of a Markov chain. *Adv. in Appl. Probab.* **21** 702–704.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist.* **21** 520–533.

Chan, K.S., Petruccelli, J.D., Tong, H. and Woolford, S.W. (1985). A multiple-threshold AR(1) model. *J. Appl. Probab.* **22** 267–279.

Chan, K. S. and Tong, H. (1985). On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. in Appl. Probab.* **17** 666–678.

Chan, K. S. and Tsay, R.S. (1998). Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika* **85** 413–426.

Chen, R. and Tsay, R.S. (1991). On the ergodicity of TAR(1) processes. *Ann. Appl. Probab.* **1** 613–634.

Cline, D.B.H. and Pu, H.H. (2004). Stability and the Lyapounov exponent of threshold AR-ARCH models. *Ann. Appl. Probab.* **14** 1920–1949.

Coakley, J., Fuertes, A. and Pérez, M. (2003). Numerical issues in threshold autoregressive modeling of time series. *J. Econom. Dynam. Control* **27** 2219–2242.

Fan, J. and Yao, Q. (2003). *Nonlinear time series: Nonparametric and parametric methods.* Springer, New York.

Gonzalo, J. and Pitarakis, J.-Y. (2002). Estimation and model selection based inference in single and multiple threshold models. *J. Econometrics* **110** 319–352.

Genest, C. and Rémillard, B. (2004). Tests of independence and randomness based on the empirical copula process. *Test* **13** 335–369.

Hansen, B. E. (1997). Inference in TAR models. *Stud. Nonlinear Dyn. Econom.* **2** 1–14.

HANSEN, B. E. (2000). Sample splitting and threshold estimation. *Econometrica* **68** 575–603.

HUBER, P.J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–233. Univ. California Press, Berkeley.

KOOP, G. and POTTER, S.M. (1999). Dynamic asymmetries in U.S. unemployment. *J. Bus. Econom. Statist.* **17** 298–312.

KUSHNER, H.J. (1984). *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory.* MIT.

LI, D., LING, S. and LI, W.K. (2010). Least square estimation of the threshold moving-average model. *Working paper in HKUST.*

LING, S. (1999). On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model. *J. Appl. Probab.* **36** 688–705.

LIU, J. and SUSKO, E. (1992). On strict stationarity and ergodicity of a nonlinear ARMA model. *J. Appl. Probab.* **29** 363–373.

LIU, W., LING, S. and SHAO, Q. (2010). On non-stationary threshold autoregressive models. *To appear in Bernoulli*

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

PETRUCCELLI, J.D. (1986). On the consistency of least squares estimators for a threshold AR(1) model. *J. Time Ser. Anal.* **7** 269–278.

POTTER, S.M. (1995). A nonlinear approach to US GNP. *J. Appl. Econometrics* **2** 109–125.

QIAN, L. (1998). On maximum likelihood estimators for a threshold autoregression. *J. Statist. Plann. Inference* **75** 21–46.

SEO, M. H. and LINTON, O. (2007). A smoothed least squares estimator for threshold regression models. *J. Econometrics* **141** 704–735.

Tiao, G.C. and Tsay, R.S. (1994). Some advances in nonlinear and adaptive modeling in time series. *J. Forecast.* **13** 109–131.

Tong, H. (1978). On a threshold model. In *Pattern Recognition and Signal Processing*. (Chen, C.H., ed.). Sijthoff and noordhoff, Amsterdam. 575–586.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford University. Press, New York.

Tong, H. (2010). Threshold models in time series analysis — 30 years on. *Manuscript.*

Tsay, R.S. (1989). Testing and modeling threshold autoregressive processes. *J. Amer. Statist. Assoc.* **84** 231–240.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *J. Amer. Statist. Assoc.* **93** 1188–1202.

List of Figures

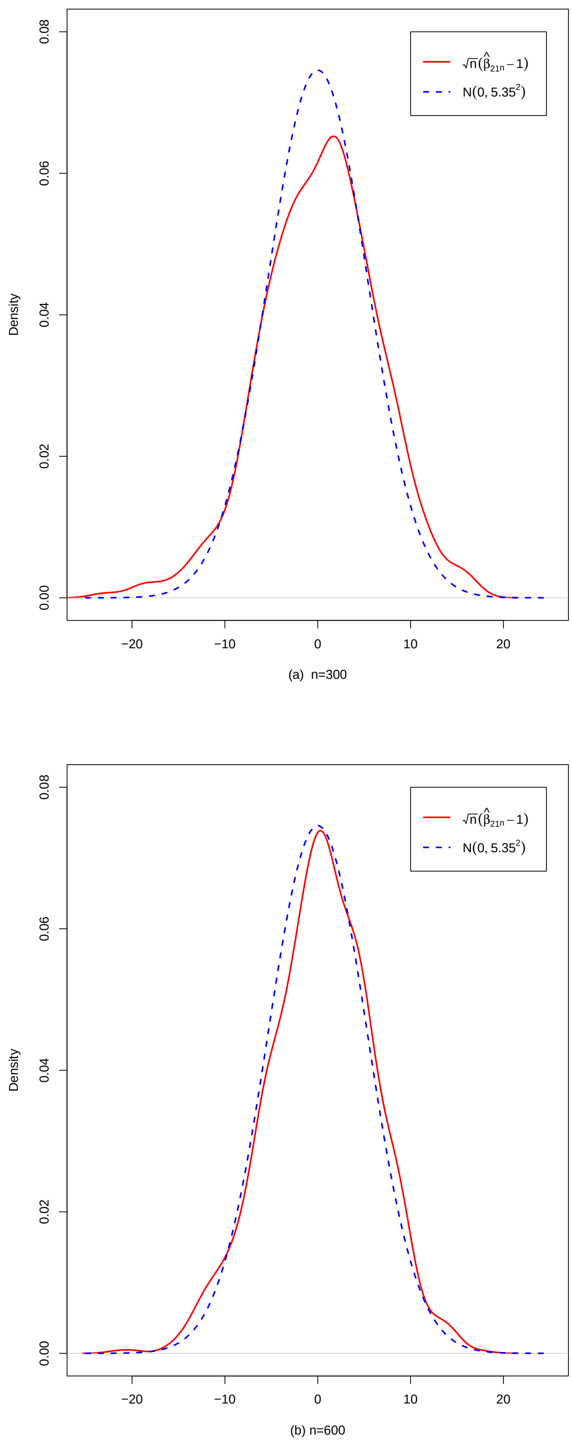

Figure 1: The density functions of $\sqrt{n}(\widehat{\beta}_{21,n} - 1)$ and $\mathcal{N}(0, 5.35^2)$. (a) The sample size is 300 and (b) The sample size is 600.

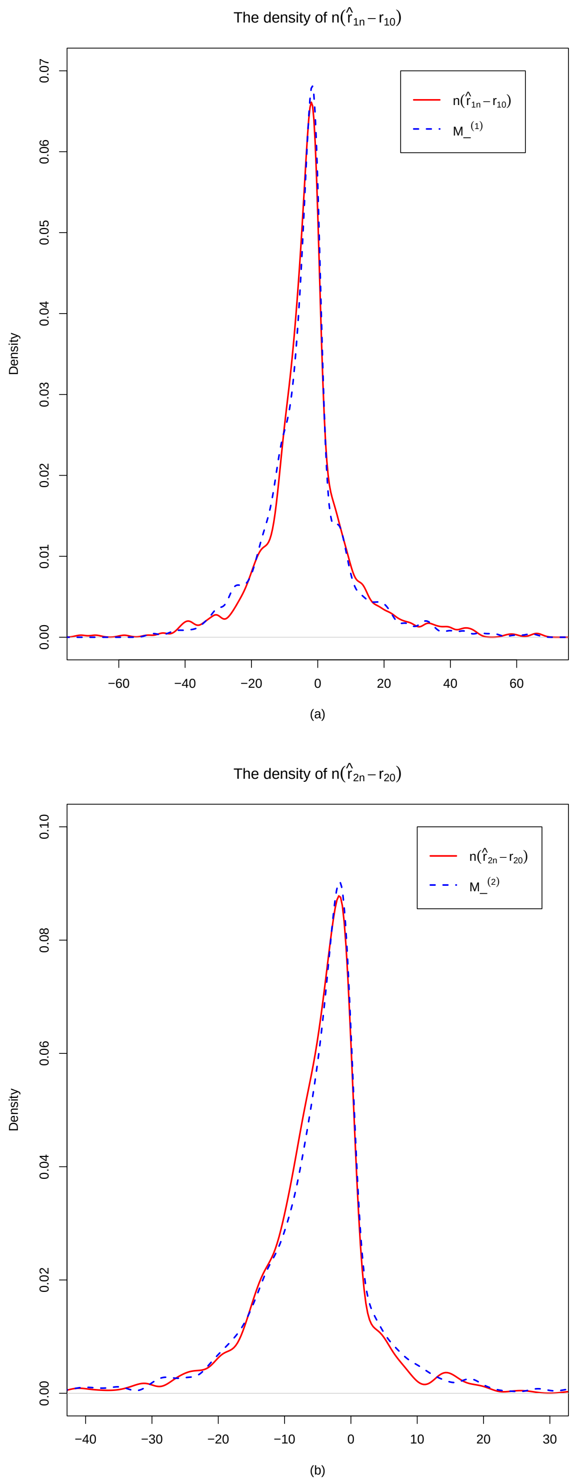

Figure 2: (a) The density functions of $n(\widehat{r}_{1n} - r_{10})$ and $M_{-}^{(1)}$. (b) The density functions of $n(\widehat{r}_{2n} - r_{20})$ and $M_{-}^{(2)}$. The sample size is 900.

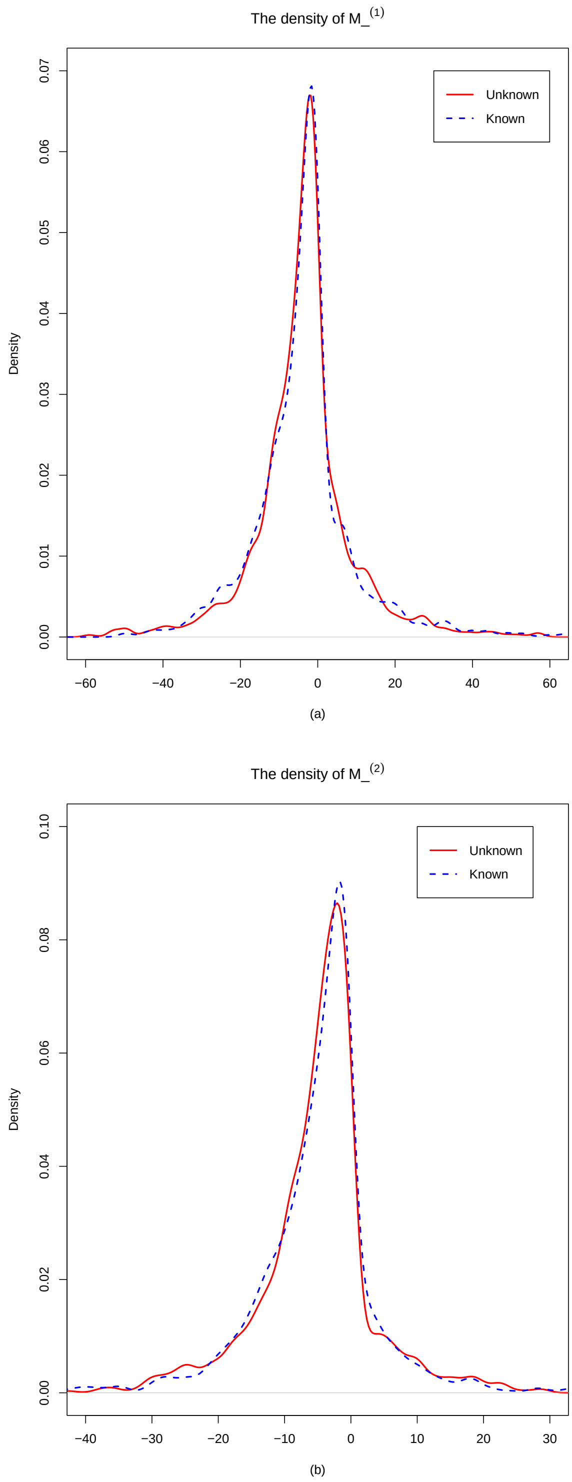

Figure 3: The densities of $M_{-}^{(i)}$, $i = 1, 2$, when the parameters used in the algorithm are known and unknown, respectively.

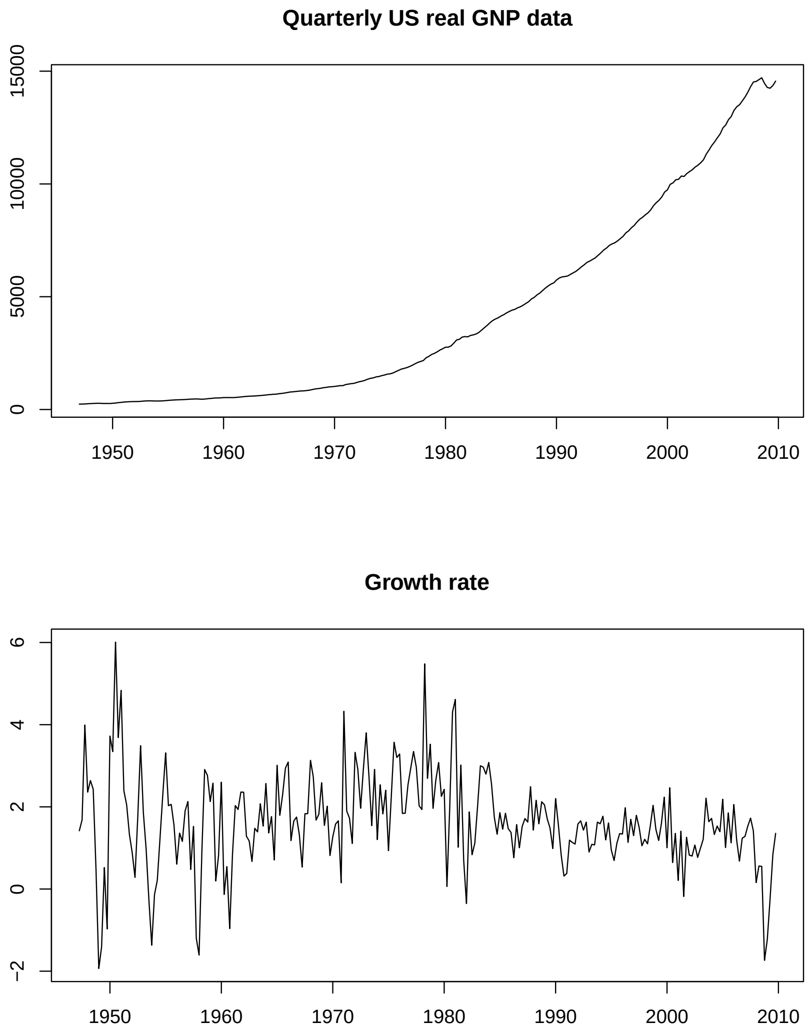

Figure 4: The original data and the growth rate.

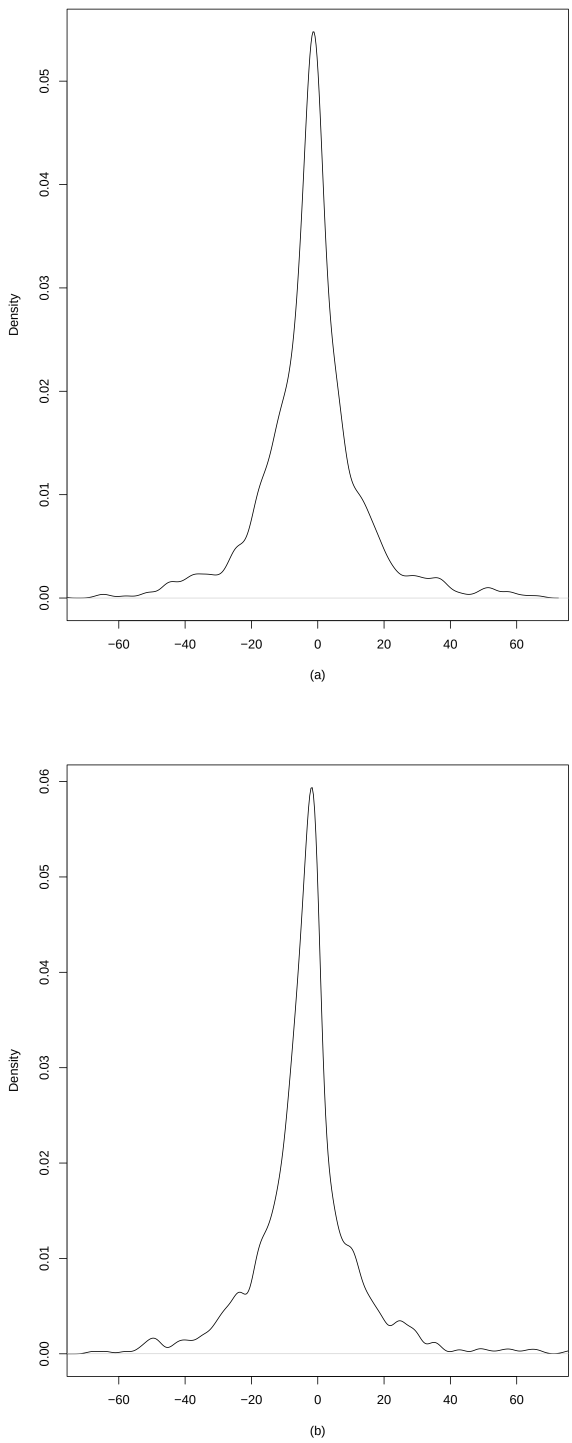

Figure 5: The densities of the normalized estimated thresholds. (a) for the estimated threshold 1.2029 and (b) for 2.4266.

List of Tables

Table 1: Simulation studies for model (5.1) with the true value $\boldsymbol{\theta}_0 = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, \beta_{30}, \beta_{31}, r_1, r_2)' = (1, -0.4, 0.6, 1, -1, -0.2, -0.8, 0.5)'$.

n		β_{10}	β_{11}	β_{20}	β_{21}	β_{30}	β_{31}	r_1	r_2
	Bias	0.0315	0.0119	0.0003	0.0153	-0.0183	0.0091	-0.0178	-0.0125
300	ESD	0.2924	0.1508	0.1503	0.3609	0.2083	0.1042	0.0760	0.0533
	ASD	0.2716	0.1409	0.1261	0.3087	0.2024	0.1041	0.0468	0.0312
	Bias	0.0071	0.0023	0.0035	0.0115	-0.0022	0.0013	-0.0070	-0.0073
600	ESD	0.1972	0.1014	0.0864	0.2304	0.1428	0.0735	0.0267	0.0153
	ASD	0.1922	0.0996	0.0892	0.2186	0.1430	0.0737	0.0234	0.0156
	Bias	0.0018	0.0003	0.0013	0.0187	0.0011	-0.0002	-0.0034	-0.0059
900	ESD	0.1561	0.0837	0.0756	0.1919	0.1162	0.0609	0.0164	0.0100
	ASD	0.1569	0.0814	0.0730	0.1785	0.1168	0.0601	0.0156	0.0104
	Bias	0.0090	0.0033	0.0015	0.0085	-0.0034	0.0021	-0.0030	-0.0041
1200	ESD	0.1400	0.0730	0.0656	0.1599	0.0987	0.0501	0.0135	0.0078
	ASD	0.1358	0.0704	0.0631	0.1547	0.1011	0.0521	0.0117	0.0078

Table 2: Empirical quantiles for $M_-^{(1)}$ and $M_-^{(2)}$ under model (5.1).

α	0.5%	1%	2.5%	5%	95%	97.5%	99%	99.5%
$M_-^{(1)}$	-58.5	-47.4	-33.8	-25.7	17.8	27.0	39.4	48.6
$M_-^{(2)}$	-37.5	-31.4	-24.3	-19.4	9.2	14.7	22.8	27.9

Table 3: Coverage probabilities of r_{10} and r_{20}.

	α	300	600	900	1200
	0.01	0.972	0.989	0.984	0.990
r_{10}	0.05	0.928	0.928	0.936	0.943
	0.10	0.864	0.878	0.874	0.895
	0.01	0.987	0.988	0.992	0.989
r_{20}	0.05	0.941	0.959	0.939	0.952
	0.10	0.890	0.916	0.885	0.917

Table 4: p-values for the multivariate independence test.

n	300	600	900	1200
p-value	0.394	0.827	0.608	0.715

Table 5: The coefficients for model (6.1).

Regime	$\boldsymbol{\beta}$										
	β_{i0}	β_{i1}	β_{i2}	β_{i3}	β_{i4}	β_{i5}	β_{i6}	β_{i7}	β_{i8}	β_{i9}	$\beta_{i,10}$
1	0.7566	0.6585	0.3831	-0.4261	-0.0278†	-0.2160	0.1542				
	(0.2146)	(0.0849)	(0.0903)	(0.0942)	(0.1001)	(0.0788)	(0.1253)				
2	-0.4939	0.3094	0.0190†	0.1196	0.1633	-0.2680	0.6696	0.2105			
	(0.4447)	(0.0935)	(0.0935)	(0.1034)	(0.0897)	(0.0975)	(0.2593)	(0.0949)			
3	0.9404	0.0707†	0.3516	-0.1627	0.1802	0.1333	-0.2506	-0.0215†	-0.2449	0.2391	0.4087
	(0.6764)	(0.1344)	(0.1285)	(0.1186)	(0.1726)	(0.1497)	(0.1971)	(0.1228)	(0.1177)	(0.1360)	(0.1090)

Note: Standard errors are in parentheses and † denotes that the coefficient is not significant at the 5% level.