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

ARTICLE INFO

Article history:
Received 25 June 2010
Received in revised form 14 November 2011
Accepted 16 November 2011
Available online 25 November 2011

JEL classification:
C13
C22

Keywords:
Asymptotic distribution
Compound Poisson process
Least squares estimation
Multiple-regime TAR model

ABSTRACT

This paper studies the least squares estimator (LSE) of the multiple-regime threshold autoregressive (TAR) model and establishes its asymptotic theory. It is shown that the LSE is strongly consistent. When the autoregressive function is discontinuous over each threshold, 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. The theory of Chan (1993) is revisited and a numerical approach is proposed to simulate the limiting distribution of the estimated threshold via simulating a related compound Poisson process. Based on the numerical result, one can construct a confidence interval for the unknown threshold. This issue is not straightforward and has remained as an open problem since the publication of Chan (1993). This paper provides not only a solution to this long-standing open problem, but also provides methodological contributions to threshold models. 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.

1. Introduction

The threshold autoregressive (TAR) model, proposed by Tong (1978), has received great attention in the nonlinear time series literature and has been widely used in econometrics, finance and statistics, among others. Chan (1993) considers the LSE and establishes an asymptotic theory for the threshold model when the autoregressive function is discontinuous. He shows that the LSE of the threshold has an extra-fast convergence rate and obtains its limiting distribution, which is the smallest minimizer of a one-dimensional two-sided compound Poisson process. It is the first time to obtain the limiting distribution of the threshold estimator and to provide a theoretical foundation for statistical inferences about the threshold. This result is very important and significant in the development of threshold models. However, it seems that the theory of Chan (1993) has been used only to justify the extra-fast consistency of the threshold estimator, and it is unclear whether or not this theory does yield a practical approach to construct a confidence interval for the unknown threshold. There are two reasons. One is the dependence on a host of nuisance parameters, including the distribution of the innovation and all autoregressive coefficients in each regime. The other is from difficulties in sampling from the jump distributions contained in the related compound Poisson process. There is no related algorithms provided in the literature up to now.

To make statistical inferences doable about the unknown thresholds in TAR models, many econometricians and statisticians have proposed other alternative methods in the literature. It is worth noting that Hansen (1997, 2000) develops a manageable limiting distribution for the threshold estimator at the cost of the convergence rate under the diminishing threshold effect assumption that the threshold model reduces to the linear model as the sample size goes to infinity. Although the limiting distribution is readily available, its validity may be limited to the “small effect” case, as he calls it. Seo and Linton (2007) proposed a smoothed LSE for a threshold regression model. In the model, they allow the threshold variable to be a linear combination of the regressors or other variables. They also show that the estimated threshold is asymptotically normal at the expense of the convergence rate. Their result also allows for time series data, a special case being TAR models. However, its convergence rate to ensure the normality is less than that obtained in Chan (1993) and relates to the choice of the bandwidth. In the econometric literature, there are other methods to conduct inference for TAR models, for example, the sequential estimation procedure in Gonzalo and Pitarakis (2002), the subsampling in Gonzalo and Wolf (2005), etc.

More earlier related results on the LSE for TAR models can be found in Petruccelli (1986), Chan and Tsay (1998), Qian (1998),
Tsai (1998), Caner and Hansen (2001), etc. At the same time, probabilistic structures of TAR models were studied intensively by Chan et al. (1985), Chan and Tong (1985), Chen and Tsay (1991), Brockwell et al. (1992), Liu and Susko (1992), An and Huang (1996), Ling (1999), Cline and Pu (2004) and so on.

The numerous applied econometrics literature has witnessed a growing interest in TAR models. The popularity of TAR models lies in that it 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. 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. A very all-sided survey on TAR models is available in Tong (1990) and a selective review of the history of threshold models is given by Tong (2011).

In most of the papers listed above, they mainly focus on two-regime TAR models, and there is only a small literature investigating the asymptotic theory of the estimation associated with TAR models. However, as an extension of two-regime TAR models, multiple-regime TAR models have many far-ranging applications in practice as well. 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. Unfortunately, multiple-regime TAR models do not have a satisfactory basis for inference even in the case of the LSE. The large sample theory of the estimator, particularly the limiting distributions of the estimated thresholds, is still an open problem up to now.

This paper follows the framework of Chan (1993). We will study the LSE of multiple-regime TAR models and establish 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 √n-consistent and asymptotically normal.

Furthermore, we revisit the theory of Chan (1993) and propose a numerical approach to simulate the limiting distribution of the estimated threshold via simulating a related compound Poisson process. Based on the numerical result, we can construct a confidence interval for the unknown threshold. This issue is not straightforward and has remained as an open problem since the publication of Chan (1993). The most difficulty is how to simulate the jump distribution associated with the compound Poisson process since it is the conditional one of $\varepsilon_i^{j,(i)}$ (defined in (3.2) below) given $y_1 = r$. We provide not only a solution to this problem, but also methodological contributions to general threshold models. 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 1947–1998 is given.

The rest of the paper is organized as follows. Section 2 presents the model and its estimation procedure. The asymptotic properties of the LSE is established in Section 3. Section 4 proposes a numerical method to implement the limiting distribution of the estimated thresholds. Section 5 reports simulation results. Section 6 analyzes an empirical example. Proofs of Theorems are collected 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

$$y_t = \sum_{j=1}^{m} (Y_{t-1}^{j} \beta_j + \sigma_j \varepsilon_t) I(r_{j-1} < y_{t-d} \leq r_j),$$

(2.1)

where $Y_{t-1}^{j} = (y_{t-1}, \ldots, y_{t-p})'$, $\beta_j = (\beta_{j0}, \beta_{j1}, \ldots, \beta_{jp})'$ $\in \mathbb{R}^{p+1}$, $\sigma_j > 0$, $j = 1, \ldots, m$, $-\infty = r_0 < r_1 < \cdots < r_m < \infty$ and $I(\cdot)$ is an indicator function. The number $m$ of regimes and the order $p$ of model (2.1) are positive integers. $d$ is a positive integer called the delay lag. $(r_1, \ldots, r_{m-1})$ are threshold parameters. The errors $\{\varepsilon_t\}$ are independent and identically distributed (i.i.d.) random variables with zero mean and unit variance, and $\varepsilon_t$ is independent of the past information $\{y_{t-j} : j \geq 1\}$.

Let $r = (r_1, \ldots, r_{m-1})'$ $\in \mathbb{R}^{m-1}$ and $\theta = (\beta', \rho', d)' = (\beta_1, \ldots, \beta_m, \rho, d)' \in \mathbb{R}^{m(p+1)+(m-1)} \times \{1, \ldots, D_0\}$, where $D_0$ is a known positive integer. Suppose that a sample $(y_1, \ldots, y_n)$ is from model (2.1) with true value $\theta_0 = (\beta_{10}, \ldots, \beta_{m0}, \rho_0, d_0)$. Given the initial values $(y_0, \ldots, y_1)$, the sum of square errors function $L_n(\theta)$ is defined as

$$L_n(\theta) = \sum_{t=1}^{n} [y_t - \varepsilon(y_t | F_{t-1})]^2,$$

where $F_t$ is the $\sigma$-algebra generated by $\{y_1, \ldots, y_t]\}$ and $\varepsilon(\cdot | \cdot)$ denotes conditional expectation assuming $\theta$ to be the true parameter. The minimizer $\hat{\theta}_n$ of $L_n(\theta)$ is called a LSE of $\theta_0$, that is,

$$\hat{\theta}_n = \arg \min \{L_n(\theta)\}.$$

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

- Fix $r \in \mathbb{R}^{m-1}$ and $d \in \{1, \ldots, D_0\}$, then minimize $L_n(\theta)$ and get its minimizer $\hat{\theta}_n(r, d)$ and minimum $L^*_n(r, d) = L_n(\hat{\theta}_n(r, d))$.
- Since $L^*_n(r, d)$ only takes finite possible values, one can get the minimizer $(\hat{r}_n, \hat{d}_n)$ of $L^*_n(r, d)$ by the enumeration approach.
- Use a plug-in method, one can finally get $\hat{\theta}_n = \hat{\theta}_n(\hat{r}_n, \hat{d}_n)$ and $\hat{\theta}_n$.

Generally, $\hat{r}_n$ is taken as the form $(y_{i_1}, \ldots, y_{i_{(m-1)}})'$, where $1 < \cdots < i_{(m-1)}$ and $(y_{i_1}, \ldots, y_{i_m})$ is the order statistics of the sample $(y_1, \ldots, y_n)$. If $(y_{i_1}, \ldots, y_{i_{(m-1)}})'$ is an estimator of $r_0$, then $L_n^*(r, \hat{d}_n)$ is a constant over the $(m-1)$-dimensional cube $A$, where

$$A = \{(r = (r_1, \ldots, r_{m-1})' : n_l \in \{y_{i_l}, y_{i_{l+1}}\}, l = 1, \ldots, m-1\}.$$

Thus, there exist infinitely many $r$ such that $L_n(r)$ can achieve its global minimum and each $r \in A$ can be considered as an estimator of $r_0$. In this case, we choose $(y_{i_1}, \ldots, y_{i_{(m-1)}})'$ as a representative of $A$ and denote it as the estimator of $r_0$. According to the procedure for obtaining $\hat{\theta}_n$, it is not hard to show that $\hat{\theta}_n$ is the LSE of $\theta_0$.

Let $\sigma^2_{\theta}$ be the true value of $\sigma_j$ for $j = 1, \ldots, m$. Once $\hat{\theta}_n$ is obtained, we then can estimate $\sigma^2_{\theta}$ by

$$\hat{\sigma}^2_{\theta} = \frac{1}{n \hat{d}_n} \sum_{t=1}^{n} (y_t - Y_{t-1}^{j} \hat{\beta}_j)^2 I(\hat{r}_{j-1,n}^* < y_{t-\hat{d}_n} \leq \hat{r}_j),$$

(2.2)

where $n_j = \sum_{t=1}^{n} I(\hat{r}_{j-1,n} < y_{t-\hat{d}_n} \leq \hat{r}_j)$.

---

1 Generally, since $d$ is small and $1 \leq d \leq \max(p, 1)$ is assumed in practice, we can set $D_0 = \max(p, 1)$ when $p$ is known and $D_0 = 12$ when $p$ is unknown, for example.
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 $n$ diverges. This issue is similar to the computational one 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 the same rearranged technique, Coakley et al. (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 $L_n(\cdot)$.

In practice, however, $m$ and $p$ are also unknown and needed to be specified. When $m$ is known, we can use the AIC below to determine the order in each regime.

$$
\text{AIC}(p_j) = \sum_{j=1}^{m} \left[ n_j \log(\sigma^2_{m,j}) + 2(p_j + 1) \right],
$$

where $p_j$ is the order of the $j$th regime. See Tsay (1998). More information criteria as model selection tools for nonlinear threshold models, see Kapetanios (2001), in which the author established the consistency of lag selection and compared the small sample performance among different criteria. For the choice of $m$, Gonzalo and Pitarakis (2002) proposed a sequential model selection approach and considered its weak consistency under some conditions for model (2.1) with all $\sigma_i$’s being equal. For general threshold models, it seems that the literature does not offer any formal methodology for selecting the number of regimes. More work should be needed in the future. Throughout our paper, we assume that both $m$ and $p$ are known.

3. Main results

Let $\Theta = \{i, \ldots, D_0\}$ be the parameter space, where $\Theta = \Theta_m \times \Theta_f \times \Theta_a$ is a compact subset of $\mathbb{R}^{|\Theta_m|+1} \times \mathbb{R}^{m-1}$ and $\mathbb{R}^{m-1} = \{(r_1, \ldots, r_{m-1}) : -\infty < r_1 < \ldots < r_{m-1} < \infty\}$. The following theorem states the strong consistency of $\hat{\theta}_n$. The proof is similar to that of Theorem 1 in Chan (1993) and hence it is omitted.

**Theorem 3.1.** Suppose that (i) $\{y_t\}$ satisfying (2.1) is strictly stationary and ergodic, having finite second moments, (ii) $\beta_0 \neq \beta_{j+1,0}$ for $j = 1, \ldots, m-1$, and (iii) $\varepsilon_t$ admits a bounded, continuous and positive density $f(x)$ on $\mathbb{R}$ with zero mean and unit variance. Then, $\hat{\theta}_n \to \theta_0$ a.s. as $n \to \infty$ and so are $\hat{\alpha}_n$’s in (2.2).

The condition (ii) in Theorem 3.1 is required to guarantee the identification of $\theta_0$. The strong consistency of $\hat{\theta}_n$ holds regardless if the autoregressive function is continuous over all thresholds or not. From Theorem 3.1, we know that $\hat{\alpha}_n$ equals $d_0$ eventually. Thus, without loss of generality, we assume that the delay lag $d_0$ is known for the remainder of this paper and it is deleted from $\theta_0$, i.e., $\theta_0 = (\beta_0, \sigma_f, \varepsilon_1)$, and so is $\hat{\alpha}_n$ from $\hat{\theta}_n$. The parameter space becomes $\Theta_0$, accordingly, and we write $d$ for $d_0$ in what follows.

To obtain the convergence rate of $\hat{\theta}_n$, the asymptotic normality of $\hat{\beta}_n$ and the limiting distribution of $n(\hat{\theta}_n - \theta_0)$, we first give four assumptions as follows.

**Assumption 3.1.** $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with $E\varepsilon_1 = 0$, $E\varepsilon_1^2 = 1$ and $E\varepsilon_1^4 < \infty$, $\varepsilon_1$ has a bounded, continuous and positive density $f(x)$ on $\mathbb{R}$.

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

Let $Z_i = (Y_1, \ldots, Y_{(p+d+1)})^T$, where $p \geq d = \max\{p, d\}$. Then $\{Z_i\}$ is a Markov chain. Denote its $l$-step transition probability by $P^l(z, A)$, where $z \in \mathbb{R}^{pd}$ and $A$ is a Borel set of $\mathbb{R}^{qd}$.

**Assumption 3.3.** $\{Z_i\}$ admits a unique invariant measure $\pi(\cdot)$ such that there exist $K > 0$ and $\rho \in [0, 1)$, for any $z \in \mathbb{R}^{pd}$ and any $n$, $\|P^n(z) - \pi(\cdot)\|_\rho \leq K(1 + \|z\|)^n$, where $\| \cdot \|_\rho$ and $\| \cdot \|$ denote the total variation norm and the Euclidean norm, respectively.

Under Assumption 3.3, $\{Z_i\}$ is $V$-uniformly ergodic with $V(z) = K(1 + \|z\|)$, which is stronger than geometric ergodicity. For the concept of $V$-uniform ergodicity, see Meyn and Tweedie (1993). If Assumption 3.1 holds and $\max_{1 \leq i \leq m} |\beta_i| < 1$, then Assumption 3.3 holds and $E\varepsilon_t^4 < \infty$, see Chan (1989) andChan and Tong (1985). If the initial value $Z_0$ is from the distribution $\pi(\cdot)$, then Assumption 3.3 implies that $\{y_t\}$ is strictly stationary.

**Assumption 3.4.** There exist nonrandom vectors $w_i^* = (1, w_{i1}, \ldots, w_{id})$ with $w_{id} = r_0$ such that $(\beta_0 - \beta_{i+1,0}) w_i^* \neq 0$ for $i = 1, \ldots, m - 1$.

In Assumption 3.4, $w_i$ may not be a component of $w_i^*$ if $d > p$. In this case, Assumption 3.4 is equivalent to the conditions $\|\beta_0 - \beta_{i+1,0}\| > 0$ for $i = 1, \ldots, m-1$. The latter is necessary and sufficient for the identification of all thresholds. When $p = d = 1$, Assumption 3.4 implies that the autoregressive mean function is discontinuous at all thresholds $\{r_1, \ldots, r_{m-1}\}$. Assumption 3.4 in the general case implies that $\|\beta_{i+1,0} - \beta_{i+1,0}\|$ is bigger than a positive constant with a positive probability and plays a key role in obtaining the $n$-convergence rate of $\hat{\beta}_n$ and its limiting distribution.

**Theorem 3.2.** If Assumptions 3.1–3.4 hold, then

(i) $n(\hat{\beta}_n - \beta_0) \to \text{N}(0, \Sigma)$

(ii) $\sqrt{n} \sup_{|r_0| < \epsilon B/n} \|\hat{\beta}_n(r) - \beta_0(r_0)\| = o_p(1)$ for any fixed $B \in (0, \infty)$.

Furthermore,

$$
\sqrt{n}(\hat{\sigma}_n - \sigma_0) = \sqrt{n}(\hat{\sigma}_n(r_0) - \sigma_0) + o_p(1) \to \text{N}(0, \Sigma)
$$

as $n \to \infty$,

where ‘$\to$’ denotes convergence in distribution and $\Sigma = \text{diag}(\sigma_{10}^2, \ldots, \sigma_{m0}^2, \Sigma_m)$ with

$$
\Sigma_m^{-1} = \mathbb{E}[\text{Y}_{t-1}]^T (\text{Y}_{t-1} < y_{t-1} \leq y_t), \quad j = 1, \ldots, m.
$$

The proof of Theorem 3.2 is similar to that of Proposition 1 in Chan (1993) and that of Theorem 4 in Qian (1998) and hence it is omitted. From Theorem 3.2(i), we know that the convergence rate of $\hat{\beta}_n$ is $n$. To study the limiting distribution of $n(\hat{\beta}_n - \beta_0)$, we consider the following profile sum of squares errors function:

$$
\tilde{L}_n(s) = L_n\left(\hat{\beta}_n \left(\rho_0 + \frac{s}{n}\right), \rho_0 + \frac{s}{n}\right) - L_n\left(\beta_0, \rho_0\right),
$$

$s \in \mathbb{R}^{m-1}$. (3.1)

Using Theorem 3.2 and Taylor’s expansion, we can show that $\tilde{L}_n(s)$ can be approximated in the function space $\mathbb{D}(\mathbb{R}^{m-1})$ (defined in the proof of Theorem 3.3) by

$$
g_0(s) = L_n\left(\beta_0, \rho_0 + \frac{s}{n}\right) - L_n(\beta_0, \rho_0) \approx \sum_{i=1}^{m-1} \sum_{t=1}^{n} \left[\left(\text{Y}_{t-1}(\beta_0 - \beta_{i+1,0}) \right)^2 - 2\sigma_{00}^2 \varepsilon_t \text{Y}_{t-1}(\beta_0 - \beta_{i+1,0})^2\right]
$$
for $s_i < 0$) \\
+ 2\sigma_1 \varepsilon_1 Y_{i-1} (\beta_{1+1} - \beta_0) \\
\times I (r_0 < y_{i-d} \leq r_0 + \frac{s_i}{n}) I (s_i \geq 0) \\
\} \\
= \sum_{i=1}^{m-1} \sum_{l=1}^{n} \varepsilon_{t-l}^{(i+1)} I (r_0 + \frac{s_i}{n} < y_{i-d} \leq r_0) \\
\times I (s_i < 0) + \varepsilon_{t-l}^{(i+1)} I (r_0 < y_{i-d} \leq r_0 + \frac{s_i}{n}) I (s_i \geq 0),
\end{align*}

where

$$
\varepsilon_{t-l}^{(i)} = \left[ Y_{i-1} (\beta_{1+1} - \beta_0) \right]^2 + 2\sigma_0 \varepsilon_t Y_{i-1} (\beta_{1+1} - \beta_0),
$$

$i, j, \ldots, m$.

(3.2)

Let $F_{i,j} (\cdot | r)$ be the conditional distribution function of $\varepsilon_{t-l}^{(i)}$ given $y_i = r$. We first define $(m-1)$ independent one-dimensional two-sided compound Poisson processes $\{ \varepsilon_i (z), z \in \mathbb{R} \}$ as

$\varepsilon_j (z) = I (z < 0) \sum_{k=1}^{n} Y_{k+1}^{(j+1)} + I (z \geq 0) \sum_{k=1}^{n} Z_{k}^{(j+1)}$.

(3.3)

for $j = 1, \ldots, m-1$, where $\{N_0 (z), z \geq 0\}$ and $\{N_2 (z), z \geq 0\}$ are two independent Poisson processes with $N_0 (0) = N_2 (0) = 0$ a.s. and with the same jump rate $\pi (r_0)$, where $\pi (\cdot)$ is the density function of $y_i$, $Y_{k+1}^{(j+1)} : k \geq 1$ are i.i.d. random variables with the distribution $F_{i,j+1} (\cdot | r_0)$, and $Z_k^{(j+1)} : k \geq 1$ are i.i.d. random variables with the distribution $F_{i,j+1} (\cdot | r_0)$. $Y_{k}^{(j+1)} : k \geq 1$ and $Z_k^{(j+1)} : k \geq 1$ are mutually independent. Here, we work with the left continuous version for $N_0 (\cdot)$ and the right continuous version for $N_2 (\cdot)$ for $j = 1, \ldots, m-1$.

We further define a spatial compound Poisson process $\phi (s)$ as follows,

$$
\phi_j (s) = \sum_{i=1}^{m-1} \varepsilon_{t-l}^{(i)} ,
$$

(3.4)

Clearly, $\phi (s)$ goes to $+\infty$ a.s. when $|s| \to \infty$ since $\varepsilon_0 Y_{k+1}^{(j+1)} = \varepsilon_0 Z_k^{(j+1)} > 0$ by Assumption 3.4 for $i = 1, \ldots, m-1$. Therefore, there exists a unique $(m-1)$-dimensional cube $[\mathbf{M}_0, \mathbf{M}_0] := [M_0^{(1)}, M_0^{(1)}] \times \cdots \times [M_0^{(m-1)}, M_0^{(m-1)}]$ on which the process $\phi (s)$ attains its global minimum a.s. That is,

$$
\mathbf{M}_0 := \arg \min_{s \in \mathbb{R}^{m-1}} \phi (s).
$$

(3.4)

From (3.4), the minimization above is equivalent to

$$
\mathbf{M}_0 := \arg \min_{z \in \mathbb{R}^{m-1}} \varepsilon_{t-l}^{(i)} (z),
\quad j = 1, \ldots, m-1,
$$

Note that the processes $\{ \varepsilon_j (z) : j = 1, \ldots, m-1 \}$ are independent, so are $\{M_0^{(j)} : j = 1, \ldots, m-1 \}$.

**Theorem 3.3.** If Assumptions 3.1–3.4 hold, then $n(\mathbf{r}_n - \mathbf{r}_0)$ converges weakly to $\mathbf{M}_0$ and its components are asymptotically independent as $n \to \infty$. Furthermore, $n(\mathbf{r}_n - \mathbf{r}_0)$ is asymptotically independent of $\sqrt{n}(\beta_{1+1} - \beta_0)$ which is always asymptotically normal.

Intuitively, it is easy to understand the asymptotic independence among the components of $n(\mathbf{r}_n - \mathbf{r}_0)$. For the asymptotic behavior of the component $n(\mathbf{r}_m - \mathbf{r}_0)$, it is determined by $Y_{i-1}$'s with $y_{i-d}$'s lying in the $n^{-1}$-neighborhood of $r_0$ and no information is provided by other $Y_{i-1}$'s with $y_{i-d}$'s out of $n^{-1}$-neighborhood of $r_0$. When $m = 1$, Theorem 3.3 reduces to Theorem 2 of Chan (1993). The limit distribution of $\mathbf{M}_0$ does not have a closed form and depends on the nuisance parameters and the distribution of $e_t$. In next section, we will describe how to simulate $\mathbf{M}_0$ via a numerical approach.

**4. Numerical implementation of $\mathbf{M}_0$.**

From Theorem 3.3, we know that obtaining $\mathbf{M}_0$ is equivalent to obtaining each component $M_0^{(j)}$ separately for $j = 1, \ldots, m-1$. From (3.3), we know that two factors determine $M_0^{(j)}$, that is, the jump rate $\pi (r_0)$ and the jump distributions $F_{i,j+1} (\cdot | r_0)$ and $F_{i,j+1} (\cdot | r_0)$. We can simulate $M_0^{(j)}$ via simulating the two-sided compound Poisson process (3.3) on the interval $[-T, T]$ for any given $T \geq 0$ large enough. Modifying Algorithm 6.2 in Cont and Tankov (2004, page 174) for a one-sided compound Poisson process, we have an algorithm for a two-sided compound Poisson process as follows.

**Algorithm A.**

**Step A.1.** Simulate two independent Poisson random variables $N_0^{(1)}$ and $N_0^{(2)}$ from Poisson distribution with the same parameter $\pi (r_0)$T which are the total number of jumps on the intervals $[-T, 0]$ and $[0, T]$, respectively.

**Step A.2.** Simulate two independent jump time sequences: $\{U_1, \ldots, U_{N_0^{(1)}}\}$ and $\{V_1, \ldots, V_{N_0^{(2)}}\}$, where $U_i$’s and $V_i$’s are independently and uniformly distributed on $[-T, 0]$ and $[0, T]$, respectively.

**Step A.3.** Simulate two independent jump-size sequences: $\{Y_{N_1}, \ldots, Y_{N_1}\}$ and $\{Z_1, \ldots, Z_{N_2}\}$ from $F_{1,j+1} (\cdot | r_0)$ and $F_{1,j+1} (\cdot | r_0)$, respectively.

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

$$
\phi_j (z) = I (z < 0) \sum_{i=1}^{n} I (U_i \leq z) Y_{i+1}^{(j)} + I (z \geq 0) \sum_{i=1}^{n} I (V_i < z) Z_{i+1}^{(j)},
$$

Then, we take the smallest minimizer of $\phi_j (z)$ on $[-T, T]$ as one observed value of $M_0^{(j)}$. Repeating above algorithm, we can get a sequence of observations of $M_0^{(j)}$ and then do some statistical inferences for $n(\mathbf{r}_n - \mathbf{r}_0)$.

In the above algorithm, the key is how to sample the jump-size sequences from $F_{1,j+1} (\cdot | r_0)$ and $F_{1,j+1} (\cdot | r_0)$ in Step A.3. When $p = d = 1$, the sampling is easy because the conditional distributions $F_{1,j+1} (\cdot | r_0)$ and $F_{1,j+1} (\cdot | r_0)$ become the unconditional ones. For general cases, the sampling is more complicated.

By the property of conditional expectation, the strong law of large numbers and $\mathbb{E} [\pi (r_0) Z_0] = \pi (r_0)$, we have

$$
F_{1,j+1} (\cdot | r_0) = \int_{\mathbb{R}^d} F_{1,j+1} (\cdot | r_0; Z_0 = z) \pi (r_0 | Z_0) \pi (r_0) d(z),
$$

$$
= \frac{1}{n} \sum_{i=1}^{n} F_{1,j+1} (\cdot | r_0; Z_0 = z_i) \pi (r_0 | Z_0) + o(1)
\quad \pi (r_0 | Z_0) \pi (r_0) d(z).
$$

(4.1)

a.s. as $n \to \infty$, uniformly in $x \in \mathbb{R}$ by Theorem 2 in Pollard (1984, page 8), where $Z_0 = (y_0, \ldots, y_{1-p:d})$, $z_i \in \mathbb{R}^d$, $G (\cdot)$ is
the distribution of \( Z_0 \), and \( \pi(r_0|z) \) is the conditional density of \( y_1 \) given \( Z_0 = z \). Let

\[
h(V, \theta) = \sum_{j=1}^{m} V_j \hat{\beta}_j |(r_{j-1} < u_d \leq r_j) \quad \text{and} \quad \sigma(V, \theta) = \sum_{j=1}^{m} \sigma_j |(r_{j-1} < u_d \leq r_j),
\]

(4.2)

where \( V = (1, v_1, \ldots, v_p)^T \in \mathbb{R}^{p+1} \). Then,

\[
\pi(r_0|z) = \{ \sigma(\theta, \theta_0)^{-1} f_1((\sigma(\theta, \theta_0)^{-1} |(r_0 - h(y_j, \theta_0))) \}
\]

(4.3)

where \( y_j = (1, y_{1}, \ldots, y_{p+1}) \) if \( z_j = (y_{1}, \ldots, y_{p+1}, \ldots, y_{1-(p+d)+1}, \ldots) \).

Let \( F_j(x) \) be the cumulative distribution function of \( e_t \). Suppose that the sample \( X \equiv \{y_1, \ldots, y_\n\} \) is from model (2.1). Then \( \theta_0, \sigma, \pi(r_0), \) and \( F_j(x) \) are known, the following algorithm describes how to sample \( Y_1 \) from \( F_{(j+1)}(r_0|z). \)

**Algorithm B.**

**Step B.1.** Set \( z_j = (y_1, \ldots, y_{(p+d)+1}) \) for \( i = 1, \ldots, n \).
**Step B.2.** For each \( i \in \{1, \ldots, n\} \), sample independently \( \{e_{2i}, \ldots, e_{(d+1)i}\} \) from \( F_j(x) \) and generate \((y_2, \ldots, y_{d+1})\) by iterating model (2.1) with the initial values \( y_1 = r_0 \) and \( Z_0 = z \). Then calculate \( \xi^{(j+1)}(i) \) in (3.2), denoted by \( \xi^{(j+1)}_i \).
**Step B.3.** Calculate \( \pi(r_0|z) \) in (4.3) and sample \( A \) from the conditional discrete density:

\[
P(U = i|X) = \pi(r_0|z)/\sum_{j=1}^{\n} \pi(r_0|z)
\]

for \( i = 1, \ldots, n \), independent of all \( \{e_{2i}, \ldots, e_{(d+1)i}\} \).
**Step B.4.** Obtain \( Y_1 = \xi^{(j+1)}_i \).

Clearly, by steps B.2 and B.3, all \( \{y_2, \ldots, y_{d+1}\} \) are independent of \( U \) given \( X \), and so is \( \xi^{(j+1)}_i \). Denote the conditional measure \( P_X(\cdot|A) \equiv P(\cdot|A, X) \). Thus,

\[
P_X(Y_1 \leq x) = \sum_{i=1}^{\n} P_X(\xi^{(j+1)}_i \leq x, U = i) = \sum_{i=1}^{\n} P_X(\xi^{(j+1)}_i \leq x) P_X(U = i) = \sum_{i=1}^{\n} P(\xi^{(j+1)}_i \leq x) \frac{\pi(r_0|z)}{\sum_{j=1}^{\n} \pi(r_0|z)} = F^{(n)}_{(j+1)}(x|r_0, z).
\]

(4.4)

Since \( F^{(n)}_{(j+1)}(x|r_0, z) \) is determined by the jump rate \( \pi(r_0) \) and jumps distribution \( F_{(j+1)}(x|\xi^{(j+1)}_i, z) \) and \( F_{(j+1)}(x|\xi^{(j+1)}_i, z) \), note that every compound Poisson process is a stationary independent increment process. By Theorem 16 in Pollard (1984, p. 134), we have a.s. that \( \psi^{(n)}(z) \rightarrow \psi(\xi^{(j+1)}_i, z) \) weakly converges to \( \xi^{(j+1)}(z) \) conditionally on \( X \) in \( \mathcal{D}([0, \infty)) \).

**Theorem 4.1.** Suppose \( f_s(x) \) is bounded and continuous positive on \( \mathbb{R} \) and \( \{y_1\} \) is strictly stationary and ergodic. Then, a.s., for \( j = 1, \ldots, m - 1 \),

\[
\lim_{n \rightarrow \infty} |P_X(M_n^{(j)} \leq x) - P(M_1^{(j)} \leq x)| = 0
\]

at each \( x \) for which \( P(M_1^{(j)} = x) = 0 \).

Since \( M_n^{(j)} \) is only relevant to model (2.1) and independent of the estimation of the parameter, without loss of generality, in what follows, we regard \( M_n^{(j)} \) as \( M^{(j)} \).

However, in practice, since all parameters \( \theta_0, \sigma, \pi(r_0), \) and \( F_j(x) \) are unknown, we first use the sample \( X \) to consistently estimate \( \theta_0, \sigma, \) and \( \pi(r_0) \) by the least squares method and the nonparametric kernel method, respectively, denoting the estimators by \( \hat{\theta}_0, \hat{\sigma}, \) and \( \hat{\pi}(r_0), \) where \( \hat{\pi}(r_0) \) is the kernel density estimator of \( \pi(r_0) \). Then we calculate the residuals \( \{\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n\}. \) Based on the residuals, we can get the estimator for \( \sigma_0 \) and construct the estimators \( \hat{\theta}_0, \) \( \hat{\sigma}, \) and \( \hat{\pi}(r_0) \) as follows

**Algorithm C.**

**Step C.1.** Set \( z_j = (y_1, \ldots, y_{(p+d)+1}) \) for \( i = 1, \ldots, n \).
**Step C.2.** For each \( i \in \{1, \ldots, n\} \), sample independently \( \{\hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_{d+1}\} \) from \( \hat{F}_j(x) \) and generate \( \{y_2, \ldots, y_d\} \) by iterating model (2.1) with the initial values \( y_1 = \hat{r}_n \) and \( Z_0 = z \) and \( \theta_0 \) and \( \sigma_0 \) being replaced by \( \hat{\theta}_0, \) and \( \hat{\sigma}_0, \) respectively. Then calculate \( \hat{\xi}^{(d+1)}(i) \), denoted by \( \hat{\xi}^{(d+1)}_i \), where

\[
\hat{\xi}^{(d+1)}_i = Y_{d+1} \left( \hat{\beta}_n - \hat{\beta}_{1+n} \right) + 2\hat{\sigma}_n \hat{\xi}^{(d+1)}_i \left( Y_{d} \left( \hat{\beta}_n - \hat{\beta}_{1+n} \right) \right)
\]

and

\[
y_{d+1} = \begin{cases} 
  y_{k}, & \text{if } k \geq 2, \\
  \hat{r}_m, & \text{if } k = 1, \\
  y_{i+k}, & \text{if } k \leq 0.
\end{cases}
\]

**Step C.3.** Calculate \( \hat{\pi}^{(j+1)}(z_j) \) in (4.3) with \( f_s(x), \theta_0, \) and \( \sigma_0 \) being replaced by their estimators, respectively, and sample a \( U \) from the conditional discrete density:

\[
P(U = i|X) = \pi^{(d+1)}(r_0|z_j) \sum_{j=1}^{\n} \hat{\pi}(r_0|z_j)
\]

for \( i = 1, \ldots, n \), conditionally independent of all \( \{\hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_{d+1}\} \) given \( X \).

**Step C.4.** Let \( \hat{Y}_1 = \hat{\xi}^{(d+1)}_i. \)

---

2 When \( d = 1 \), the iteration is not necessary since \( y_2 = y_1 = r_0 \) is given.
Denote the counterpart of $\hat{M}_n^{(1)}$ by $\bar{M}_n^{(1)}$ obtained by using Algorithm C. Then, we have

**Theorem 4.2.** If Assumptions 3.1–3.4 hold and $f_i(x)$ is uniformly continuous on $\mathcal{X}$, then, in probability,

$$\lim_{n \to \infty} \mathbb{P}(\hat{M}_n^{(j)} \leq x) - \mathbb{P}(M^{(j)} \leq x) = 0, \quad j = 1, \ldots, m - 1,$$

at each $x$ for which $\mathbb{P}(M^{(j)} = x) = 0$. That is, $\hat{M}_n^{(j)}$ converges weakly to $M^{(j)}$ conditionally on $\mathcal{X}$, in probability.

Simulation studies show that Algorithm C does work well for simulating the distribution of $\mathcal{M}_n$. See the example in (5.2) in Section 5.

5. Simulation studies

To assess the performance of the LSE of $\theta_0$ in finite samples, we use sample sizes $n = 300, 600, 900$ and 1200, each with replications 1000 for the following three-regime TAR model:

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

(5.1)

with the true value $\theta_0 = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, \beta_{30}, \beta_{31}, r_1, r_2)' = (1, -0.4, 0.6, 1, -1, -0.2, -0.8, 0.5)'$ and $\epsilon_t \sim i.i.d. \mathcal{N}(0, 1)$. Clearly, the autoregressive function is not continuous at 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 $\hat{\beta}_{10}$ are computed by using $\Sigma$ in Theorem 3.2 and the ASD of $\hat{\theta}_n$ 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 values of the ESDS for $\hat{\theta}_n$ are about halved when the value of $n$ is doubled, for example, from 600 to 1200. This illustrates the $n$-consistency of the estimated thresholds, under which they would approach the true thresholds much faster than other estimated parameters.

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), Chen and Tsay (1991), etc. 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. (2011). $\sqrt{n}(\hat{\beta}_{21, n} - \beta_{21})$ is still asymptotically normal. Fig. 1 displays the densities of $\sqrt{n}(\hat{\beta}_{21, n} - \beta_{21})$ and $\mathcal{N}(0, 5.35^2)$ when $n = 300$ and 600, respectively. The number 5.35 is the estimator of the asymptotic variance of $\hat{\beta}_{21, n}$ in Theorem 3.2(ii). From Fig. 1, we see that they are very close each other when $n = 600$.

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 $M_n^{(1)}$ and $M_n^{(2)}$ by 20,000 replication. When the significance level $\alpha$ equals 0.5%, 1%, 2.5%, 5%, 95%, 97.5%, 99% and 99.5%, the values are given in Table 2.

Table 2

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>0.5%</th>
<th>1%</th>
<th>2.5%</th>
<th>5%</th>
<th>95%</th>
<th>97.5%</th>
<th>99%</th>
<th>99.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_n^{(1)}$</td>
<td>-58.5</td>
<td>-47.4</td>
<td>-33.8</td>
<td>-25.7</td>
<td>17.8</td>
<td>27.0</td>
<td>39.4</td>
<td>48.6</td>
</tr>
<tr>
<td>$M_n^{(2)}$</td>
<td>-37.5</td>
<td>-31.4</td>
<td>-24.3</td>
<td>-19.4</td>
<td>9.2</td>
<td>14.7</td>
<td>22.8</td>
<td>27.9</td>
</tr>
</tbody>
</table>

Table 3

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>300</th>
<th>600</th>
<th>900</th>
<th>1200</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{10}$</td>
<td>0.01</td>
<td>0.05</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>$r_{20}$</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>

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.

To see the overall approximation of the estimated thresholds, Fig. 2(a) and (d) shows the density functions of $n(f_{n,p} - r_{10})$, $j = 1, 2$, respectively, when $n = 600$. Fig. 2(b) and (e) show the densities of $M_n^{(1)}$ and $M_n^{(2)}$, respectively. From the figure, we see that the densities of both $n(f_{n,p} - r_{10})$ and $M_n^{(1)}$ are very close. We also note that the density of $M_n^{(2)}$ is leptokurtic and asymmetric, skewing toward the left hand side of the origin. In fact, the skewness is $-0.04$ and the kurtosis is 12.5 for $M_n^{(1)}$, and the skewness is $-0.18$ and the kurtosis is 9.37 for $M_n^{(2)}$. Owing to the skewness, a caution should be taken in constructing confidence intervals of thresholds in practice.

To see whether the simulation method in Section 4 works well, we generate a sample $\{y_1, \ldots, y_{600}\}$ from model (5.1) and use it to obtain $\hat{M}_n^{(1)}$ by Algorithm C. Fig. 2(c) and (f) exhibit the densities of $\hat{M}_n^{(1)}$ and $\hat{M}_n^{(2)}$, respectively. From the figure, we can see that the
The sample size is 300. (b) The sample size is 600.

Fig. 1. The densities of $\sqrt{n}(\hat{\beta}_{21, n} - \beta_{21, 0})$ and $\mathcal{N}(0, 5.35^2)$.

(a) The sample size is 300. (b) The sample size is 600.

Fig. 2. (a) The density of $n(r_{1n} - r_{10})$; (b) the density of $M^{(1)}$; (c) the density of $\hat{M}^{(1)}$; (d) the density of $n(r_{2n} - r_{20})$; (e) the density of $M^{(2)}$; (f) the density of $\hat{M}^{(2)}$. The sample size is 600.

densities of $\hat{M}^{(j)}_n$ and $M^{(j)}$ match well, $j = 1, 2$. This illustrates that Algorithm C is a useful approach to implement the limit distribution of the estimated threshold.

**Theorem 3.3** claims that $n(r_{1n} - r_{10})$ and $n(r_{2n} - r_{20})$ 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 $n(r_{1n} - r_{10})$ and $n(r_{2n} - r_{20})$ are indeed independent at 5% significant level. This is a further evidence of **Theorem 3.3** to a certain extent.

To further evaluate the performance of the simulation method in Section 4 for higher-order TAR models, we consider another...
example:
\[
y_t = \begin{cases} 
1 + 0.2y_{t-1} - 0.4y_{t-2} + 0.3y_{t-3} + \varepsilon_t, & y_{t-3} \leq 0, \\
1 + 0.6y_{t-1} + 0.1y_{t-2} - 0.2y_{t-3} + \varepsilon_t, & y_{t-3} > 0,
\end{cases}
\]
where \( \varepsilon_t \sim i.i.d. N(0, 1) \) and the sample size 600 is used. Clearly, model \( (5.2) \) satisfies Assumptions 3.1–3.4. Here, we do not report the empirical results about the estimator \( \hat{\beta} \), since they are similar to those in Table 1. To obtain more precise densities of \( n(n_0 - n_0) \) and \( M_0 \), we increase the number of replications to 10,000. Fig. 3(a) and (b) display their densities, respectively. When a sample \( y_1, \ldots, y_{252} \) is fixed and given, we can obtain the density of \( M_0 \), by Algorithm C in Section 4. Fig. 3(c) and (d) show the density of \( M_0 \), where 1000 replications are used for (c) and 10,000 replications for (d). From Fig. 3(a), (b) and (d), we can see that the densities are almost identical. Compared (c) with (d) in Fig. 3, we see that the more the number of replications, the more precise the density of \( M_0 \).

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 growth rate of the quarterly U.S. real GNP data over the period 1947–2009 with a total of 252 observations.

Let \( y_1, \ldots, y_{252} \) denote the original data. We define the growth rate series as
\[
x_t = 100(\log y_t - \log y_{t-1}), \quad t = 2, \ldots, 252.
\]
The data \( (y_t) \) and the growth rate series \( (x_t) \) are plotted in Fig. 4. Setting \( m_0 = \max(p_1, p_2, p_3) \leq 12 \) and \( 1 \leq d \leq \max(m_0, 1) \), we use the AIC in (2.3) to select the model as follows:
\[
\begin{aligned}
\beta_{10} + \sum_{i=1}^{6} \beta_{1i}x_{t-i} + \sigma_1 \varepsilon_t, & \quad x_{t-d} \leq 1.20, \\
\beta_{20} + \sum_{i=1}^{10} \beta_{2i}x_{t-i} + \sigma_2 \varepsilon_t, & \quad 1.20 < x_{t-d} \leq 2.43, \\
\beta_{30} + \sum_{i=1}^{10} \beta_{3i}x_{t-i} + \sigma_3 \varepsilon_t, & \quad x_{t-d} > 2.43.
\end{aligned}
\]

The coefficients \( \{\beta_{ij}\} \) and their standard deviations are summarized in Table 5. \( \beta_{11}, \beta_{12}, \beta_{13} \) and \( \beta_{23} \) are not significant at the 5% level. The estimators of \( \sigma_i \)'s are \( \hat{\sigma}_1 = 0.66, \hat{\sigma}_2 = 0.81 \) and \( \hat{\sigma}_3 = 0.90 \), respectively. The Ljung–Box test statistics \( Q(6) \) has a value of 1.45 which has a p-value of 0.96 and the McLeod–Li test statistics \( Q^2(6) \) has a value of 5.50 which has a p-value of 0.48 (see Li and Li (1996), McLeod and Li (1983)), which suggest that model \( (6.1) \) is adequate for \( (x_t) \). The number of data \( (x_t) \) in three regimes are 80, 119 and 51, respectively. The 95% confidence intervals of \( r_{10} \) and \( r_{20} \) are \((1.10, 1.34)\) and \((2.30, 2.57)\), respectively. Based on the simulation approach in Section 4, the densities of the estimated thresholds are given in Fig. 5.

According to model \( (6.1) \), the normal growth rate of the U.S. GNP is in the interval \((1.20, 2.43)\). The growth rate is considered as the high one if it is bigger than 2.43. Otherwise, it is regarded as the low one if it is smaller than 1.20. In each regime, the growth rate can be fitted by different AR models, respectively.

7. Proofs of theorems

7.1. Proof of Theorem 3.3

We first define the space \( D(\mathcal{F}) \) with its topology. Recall that \( D(\mathcal{F}) \) is the function space consisting of uniform limits of sequences of simple functions defined on an arbitrary space \( X \). Consider a closed interval \( I \subset \mathbb{R} \) and the class \( A_1 \) of one to one and strictly monotone increasing functions from \( I \) into itself. Define the function \( \|\| : A_1 \to \mathbb{R} \) by \( \|\| = \sup_{x \in I} \left\{ \frac{\log (1 + \frac{1}{u})}{u} \right\} \). Let \( A_k^{(m-1)} = \{A_{[k-1]} \times \cdots \times A_{[k-1]}\} \) and \( \|\| \equiv \max_{1 \leq i \leq m-1} \|\| \|\|_k \).

Now, we give the proof of Theorem 3.3. For simplicity, we only deal with the case when \( m = 3 \) and the proof is similar when \( m > 3 \). By Theorem 3.2 and Taylor's expansion, it follows that \( d(L_0(s), \varphi_0(s)) \to 0 \) in probability as \( n \to \infty \). To prove weak convergence of \( L_m(s) \), it suffices to prove that \( \varphi_0(s) \) converges weakly as \( n \to \infty \) by Theorem 3.1 in Billingsley (1999, p. 27). Note that
\[
\varphi_0(s) = \sum_{i=1}^{n} \left[ \xi_i^{(1,2)} I \left( r_{10} + \frac{s_i}{n} < y_{t-d} \leq r_{10} \right) I(s_1 < 0) + \xi_i^{(1,2)} I \left( r_{10} < y_{t-d} \leq r_{10} + \frac{s_i}{n} \right) I(s_1 \geq 0) \right]
\]
with \( \xi_i^{(1,2)} I \left( r_{10} < y_{t-d} \leq r_{10} + \frac{s_i}{n} \right) I(s_2 < 0) + \xi_i^{(1,2)} I \left( r_{20} < y_{t-d} \leq r_{20} + \frac{s_i}{n} \right) I(s_2 \geq 0) \equiv \xi_{1n}(s_1) + \xi_{2n}(s_2), \)

where \( s = (s_1, s_2)^\top \in \mathbb{R}^2 \) and \( \xi_i^{(1,2)} \) is defined in \( (3.2) \). By Theorem A.1 in Li et al. (2010), we can prove that \( \xi_{1n}(s_1) \) converges weakly as \( n \to \infty \) for \( i = 1, 2, 3 \). Thus, \( \varphi_0(s) \) is tight by Theorem 5.5 in Straf (1972).

Next, we characterize convergence of finite dimensional distributions of \( \varphi_0(s) \). Without loss of generality, we assume \( \xi_i^{(1,3)} \) is bounded. Otherwise, use the truncating technique in Li et al. (2010) to truncate \( \xi_i^{(1,3)} \) and then consider a truncated process. To this end, for any \( s_i = (s_{1i}, s_{2i})^\top \in \{0, \infty\} \times \{0, \infty\} \), satisfying \( s_{1i} \leq s_{2i} < s_{1i} \leq s_{2i}, i = 1, \ldots, 4, j = 1, 2 \), and for any constants \( c_1 \) and \( c_2 \), the linear combination of the increments of \( \varphi_0(s) \) is
\[
S_n \equiv c_1 \{ \varphi_0(s_2) - \varphi_0(s_1) \} + c_2 \{ \varphi_0(s_4) - \varphi_0(s_3) \}
\]
\[
= \sum_{i=1}^{n} \left[ c_1^{(1,2)} \left\{ c_1^{(1,1)} + c_2^{(1,2)} \right\} + c_2^{(1,3)} \right]
\]

\[
+ \sum_{i=1}^{n} \left[ c_1^{(3,2)} \left\{ c_1^{(2,1)} + c_2^{(2,3)} \right\} + c_2^{(2,3)} \right].
\]
Fig. 3. (a) The density of $n(\hat{r}_n - r_0)$; (b) the density of $M_n$; (c) the density of $\hat{M}_n$ when 1000 replications are used; (d) the density of $\hat{M}_n$ when 10,000 replications are used. The sample size is 600.

Quarterly US real GNP data

Growth rate

Fig. 4. The original data and the growth rate.

where

$J_t^{(i,j)} = I \left( r_0 + \frac{s_j}{n} < y_{t-d} \leq r_0 + \frac{s_j+1}{n} \right), \quad i = 1, 2, \; j = 1, 3.  \quad J_t^{(i,j)} = \xi_t^{(2,1)} \left[ c_{i1-t}^{(1,1)} + c_{i2}^{(1,3)} \right] + \xi_t^{(3,2)} \left[ c_{i1-t}^{(2,1)} + c_{i2}^{(2,3)} \right].$

Let $\varepsilon = 1/n$ and
We verify Assumptions A.1–A.3 in Li et al. (2010) for $f'_i$. By Assumption 3, we have

$$
\lambda = \lim_{\varepsilon \to 0 \; m \to \infty} e^{-\varepsilon} \mathbb{P}(f'_i \neq 0) = \pi(r_{10})(s_{21} - s_{11}) + \pi(r_{20})(s_{42} - s_{22}),
$$

(7.1)

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

$$
Q^*(B) = \sum_{i=1}^{4} w_i Q_i^*(B),
$$

(7.2)

where

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

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

$$
\lim_{\varepsilon \to 0 \; m \to \infty} \mathbb{E}_n \left[ f(x + f'_i m) - f(x) | f'_i m \neq 0 \right]
$$

$$
= \lim_{\varepsilon \to 0} \mathbb{E}_n \left[ f(x + f'_i m) - f(x) | f'_i m \neq 0 \right]
$$

$$
= \int [f(x + u) - f(x)] Q^*(du).
$$

(7.3)

By (7.1)-(7.3), Assumptions A.1-A.3 in Li et al. (2010) hold. Furthermore, by their Theorem A.1, we claim that $\mathcal{S}_c$ weakly converges to a compound Poisson random variable $J$ with jump rate $\lambda$ and the jump distribution $Q^*$. The characteristic function $f_j(t)$ of $J$ can be written as

$$
f_j(t) = \exp \left\{ -\lambda \left[ 1 - \int_{\mathbb{R}} e^{\lambda Q^*(dx)} \right] \right\}
$$

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

which is equal to that of the linear combination $c_1 \{ \mathcal{P}(s_2) - \mathcal{P}(s_1) \} + c_2 \{ \mathcal{P}(s_3) - \mathcal{P}(s_4) \}$ of the independent increments of a spatial compound Poisson process

$$
\mathcal{P}(s) = \sum_{i=1}^{n_{s_1}} \xi_{i_1}^{(2)} + \sum_{i=1}^{n_{s_2}} \xi_{i_2}^{(3)}, \quad s_1 \geq 0, s_2 \geq 0,
$$

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

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

$$
\varphi_n(s) \Rightarrow \left\{ \sum_{i=1}^{n_{s_1}} \xi_{i_1}^{(2)} + \sum_{i=1}^{n_{s_2}} \xi_{i_2}^{(3)} \right\} | s_1 \geq 0, s_2 \geq 0
$$

Note: Standard errors are in parentheses.

* The coefficient is not significant at the 5% level.
\[
\begin{align*}
&+ \left\{ \sum_{i=1}^{N_{2,1}(s_1)} \xi_{2,1}^{(2,1)} + \sum_{i=1}^{N_{2,2}(s_2)} \xi_{2,2}^{(2,3)} \right\} I(s_1 \geq 0, s_2 < 0) \\
&+ \left\{ \sum_{i=1}^{N_{1,1}(s_1)} \xi_{1,1}^{(1,2)} + \sum_{i=1}^{N_{2,1}(s_2)} \xi_{2,2}^{(2,1)} \right\} I(s_1 < 0, s_2 \geq 0) \\
&\times I(s_1 < 0, s_2 < 0),
\end{align*}
\]

where \( N_{k+1,i}(z) \) and \( N_{k,(i+1)}(z) \) are independent Poisson processes with jump rate \( \pi(r_0) \) for \( k, i = 1, 2, \{ \xi_{k,i}^{(i)} : i \geq 1 \} \) are i.i.d random variables having the same distribution \( F_{k,i}(|r_{ij}|) \). All random variables and processes defined above are mutually independent. Since \( N_{1,1}(z) \) and \( N_{2,1}(z) \) have the same jump rate \( \pi(r_0) \) and \( \xi_{1,1}^{(1,2)} \) and \( \xi_{2,2}^{(2,1)} \) have the same distribution, there exist the process \( V_k^{(2,1)} \) and the i.i.d. random variables \( \{ V_k^{(2,1)} \} \), defined by (3.3), such that

\[
I(s_1 \geq 0, s_2 \geq 0) \geq \sum_{i=1}^{N_{1,1}(s_1)} \xi_{1,1}^{(1,2)} + \sum_{i=1}^{N_{2,1}(s_2)} \xi_{2,2}^{(2,1)} I(s_1 \geq 0, s_2 \geq 0) \geq \sum_{i=1}^{N_{1,1}(s_1)} \xi_{1,1}^{(1,2)} + \sum_{i=1}^{N_{2,1}(s_2)} \xi_{2,2}^{(2,1)}
\]

where “\( X \Rightarrow Y \)” indicates the random elements \( X \) and \( Y \) have the same distribution. Similarly, using the same way to combine the other terms, we can get

\[
g_{\theta^n}(\mathbf{s}) \Rightarrow g_{\theta}(\mathbf{s}), \quad \text{as } n \rightarrow \infty,
\]

where \( g_{\theta}(\mathbf{s}) \) is defined in (3.4). Thus, \( \bar{L}_n(s) \) defined in (3.1) converges weakly to \( g_{\theta}(\mathbf{s}) \) as \( n \rightarrow \infty \). By Theorem 3.1 in Seijo and Sen (2011), it is readily seen that \( n(\bar{r}_n - r_0) \) converges weakly to \( \mathbf{M}_n \) where \( \{ \mathbf{M}_n \} \) is the unique \((m-1)\)-dimensional random cube over which \( g_{\theta}(\mathbf{s}) \) attains its global minimum. The remainder of the proof is similar to that of Theorem 2 in Chan (1993).

\[\Box\]

7.2. Proof of Theorem 4.2

Before the proof, we first give a technical lemma.

**Lemma 7.1.** Suppose that the conditions in Theorem 4.2 hold, then, in probability,

\[
\sup_{x \in \mathbb{R}} |P_X(\bar{Y}_1 \leq x) - F_{\mathcal{Y}_{j+1}}(x|r_0)| \rightarrow 0.
\]

**Proof.** From Algorithm C, it follows that

\[
P_X(\bar{Y}_1 \leq x) = \sum_{i=1}^{n} P_X(\xi_{j+1}^{(j+1)} \leq x, U = i) = \sum_{i=1}^{n} P_X(\xi_{j+1}^{(j+1)} \leq x) P_X(U = i) = \sum_{i=1}^{n} P_X(\xi_{j+1}^{(j+1)} \leq x | y_1 = \hat{r}_m, Z_0 = z) \times \frac{\hat{f}(\hat{r}_m|Z_1)}{\sum_{i=1}^{n} \hat{f}(\hat{r}_m|Z_1)}.
\]

By a simple calculation, using (4.4), we have the equation given in Box 1. Note that

\[
\frac{1}{n} \sum_{i=1}^{n} |\hat{f}(\hat{r}_m|Z_i) - \pi(r_0|Z_i)| \leq \delta_0 \left[ \|f_r - f_h\|_{\infty} + \frac{1}{n} \sum_{i=1}^{n} f_r \left( \frac{\hat{r}_m - h(y_i, \hat{\theta}_n)}{\sigma(y_i, \hat{\theta}_n)} \right) - f_r \left( r_0 - h(y_i, \theta_0) \right) \right] + \delta \left[ 2 \|f_r - f_h\|_{\infty} + \frac{1}{n} \sum_{i=1}^{n} |\hat{\sigma}(y_i, \hat{\theta}_n) - \sigma(y_i, \theta_0)| \right].
\]

where \( \hat{\sigma}(V, \theta) = \sum_{j=1}^{m} \hat{\sigma}_j l(r_{j-1} < \nu_d \leq r_j) \), similar to the definition of \( \sigma(V, \theta) \) in (4.2), and \( \delta = \max(\hat{\sigma}_{m_1}^{-1}, \ldots, \hat{\sigma}_{m_1}^{-1}) \). Since \( f_r(x) \) is uniformly continuous on \( \mathbb{R} \), using the similar method in Silverman (1978), we have

\[
\|f_r - f_h\|_{\infty} = o_P(1).
\]

Using the following inequalities:

\[
|\hat{\sigma}(y_i, \hat{\theta}_n) - \sigma(y_i, \theta_0)| \leq \sum_{j=1}^{m} |\hat{\sigma}_j - \sigma_j| \leq \sum_{j=1}^{m} (\hat{\sigma}_j + \sigma_j) \left| |(y_i - d + 1) - r_0| \right| I(y_i - d + 1 - r_0) \leq \sum_{j=1}^{m} \|\hat{\beta}_j - \beta_j\|_{\infty} \|y_i\|,
\]

we can obtain that \( \hat{\sigma}(y_i, \hat{\theta}_n) \rightarrow \sigma(y_i, \theta_0) \) a.s. for each \( i \) by Theorem 3.1. Then,

\[
\frac{1}{n} \sum_{i=1}^{n} |\hat{\sigma}(y_i, \hat{\theta}_n) - \sigma(y_i, \theta_0)| = o(1) \quad \text{a.s.}
\]

and

\[
\lim_{n \rightarrow \infty} \mathbb{E} \left| f_r \left( \frac{\hat{r}_m - h(y_i, \hat{\theta}_n)}{\hat{\sigma}(y_i, \hat{\theta}_n)} \right) - f_r \left( r_0 - h(y_i, \theta_0) \right) \right| = 0
\]

by the boundedness and continuity of \( f_r(x) \) and the dominated convergence theorem, which yields that

\[
\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left| f_r \left( \frac{\hat{r}_m - h(y_i, \hat{\theta}_n)}{\hat{\sigma}(y_i, \hat{\theta}_n)} \right) - f_r \left( \frac{r_0 - h(y_i, \theta_0)}{\sigma(y_i, \theta_0)} \right) \right| = 0.
\]

Thus, from (7.4)–(7.6) and \( \delta \rightarrow \delta_0 \) a.s., it follows that

\[
\frac{1}{n} \sum_{i=1}^{n} |\hat{f}(\hat{r}_m|Z_i) - \pi(r_0|Z_i)| = o_P(1),
\]

which implies that

\[
\frac{1}{n} \sum_{i=1}^{n} \hat{f}(\hat{r}_m|Z_i) = \frac{1}{n} \sum_{i=1}^{n} \pi(r_0|Z_i).
\]
by the ergodic theorem. Therefore, from (7.7) and (7.8), we have
\[ \sup_{x \in \mathbb{R}} |\mathbb{P}_X(\hat{Y}_1 \leq x) - \mathbb{P}_X(Y_1 \leq x)| 
\leq \mathcal{O}_p(1) \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}_X(\hat{\xi}_{d+1}^{(1)} \leq x | y_1 = \hat{r}_{jm}, Z_0 = z_j) \right\} = \mathcal{O}_p(1), \] (7.10)

In particular, when \( k = 2 \), after a simple calculation, we can get
\[ \left| \hat{H}_2 - H_2 \right| \leq \sup_{x \in \mathbb{R}} |\mathbb{P}_X(\hat{Y}_2 \leq x | y_1 = \hat{r}_{jm}, Z_0 = z_j) - \mathbb{P}_X(Y_2 \leq x | Z_0 = z_j)| \] (7.11)

Proof of Theorem 4.2. Let \( \hat{F}_{ij}(x) = \mathbb{P}_X(\hat{Y}_{ij} \leq x) \) and \( \hat{F}_{ij}(x) = \mathbb{P}_X(\hat{Z}_{ij} \leq x) \). We define a two-sided CPP by \( \hat{P}_j(z) \) which is determined by the jump rate \( \hat{\pi}(r_m) \) and jump distributions \( \hat{F}_{ij}(x) \) and \( \hat{F}_{ij}(x) \). More specifically,
\[ \hat{P}_j(z) = I(z < 0) \sum_{k=1}^{N_{ij}(z)} + I(z \geq 0) \sum_{k=1}^{N_{ij}(z)} \] (7.12)

where both \( \hat{N}_{ij}(z) : z \in [0, \infty) \) and \( \hat{N}_{ij}(z) : z \in [0, \infty) \) are conditional independent compound Poisson processes given \( \mathcal{X} \) and have the same jump rate \( \hat{\pi}(r_m) \). \( \hat{Y}_{ij}^{(k+1)} \) is i.i.d. from \( \hat{F}_{ij}(x) \) and \( \hat{Z}_{ij}^{(k+1)} \) is i.i.d. from \( \hat{F}_{ij}(x) \) with \( i, j, k = 1, 2, \ldots \) mutually conditional independent given \( \mathcal{X} \). Then, we claim that the probability \( \hat{P}_j(z) \to P_j(z) \) defined in (3.3) conditionally on \( \mathcal{X} \) in
Indeed, since every compound Poisson process is a stationary independent increment process, we have, in probability,
\[
(\hat{P}_1(\rho_0 + \delta_n) - \hat{P}_1(\rho_0), \hat{P}_2(\rho_0 + \delta_n) - \hat{P}_2(\rho_0)) \to 0 \quad \text{a.s.}
\]
by \(N_1(0) = N_2(0) = 0\) a.s., wherever \(\delta_n\) is a sequence of positive numbers converging to zero and \(\rho_0\) is a sequence of stopping times taking values in \([0, T]\) for each fixed \(T > 0\). On the other hand, for any \(0 < s_1 < \ldots < s_k\), the characteristic function of \((\hat{P}_1(s_1), \hat{P}_2(s_1), \hat{P}_1(s_2), \hat{P}_2(s_2), \ldots, \hat{P}_1(s_k), \hat{P}_2(s_k))\) conditionally on \(X\) is
\[
\hat{\psi}_k(U_1, \ldots, U_k; v_1, \ldots, v_k) = E_X \left\{ \exp \left[ i \left( \sum_{i=1}^{k} v_i \hat{P}_1(s_i) + \sum_{i=1}^{k} v_i \hat{P}_2(s_i) \right) \right] \right\}
\]
\[
= E_X \left\{ \exp \left[ i \sum_{i=1}^{k} v_i \hat{P}_1(s_i) \right] \right\} E_X \left\{ \exp \left[ i \sum_{i=1}^{k} v_i \hat{P}_2(s_i) \right] \right\}
\]
\[
= \prod_{i=1}^{k} \exp \left\{ -\hat{\pi}(f_{jm}) (s_i - s_{i-1}) \right\} \times \int [1 - \exp(iax)] dF_{(j+1)}(x|f_{jm}) \times \prod_{i=1}^{k} \exp \left\{ -\hat{\pi}(f_{jm}) (s_i - s_{i-1}) \right\} \times \int [1 - \exp(ibx)] dF_{(j+1)}(x|f_{jm})
\]
where \(a_i = u_i + \cdots + u_1\) and \(b_i = v_i + \cdots + v_1\). By Lemma 7.1, \(\hat{\pi}(f_{jm}) \to \pi(f_{jm})\) in probability and the dominated convergence theorem, we have, in probability,
\[
\hat{\psi}_k(U_1, \ldots, U_k; v_1, \ldots, v_k) \to \prod_{i=1}^{k} \exp \left\{ -\pi(f_{jm}) (s_i - s_{i-1}) \right\} \times \int [1 - \exp(iax)] dF_{(j+1)}(x|f_{jm}) \times \prod_{i=1}^{k} \exp \left\{ -\pi(f_{jm}) (s_i - s_{i-1}) \right\} \times \int [1 - \exp(ibx)] dF_{(j+1)}(x|f_{jm})
\]
which is the characteristic function of \((\hat{P}_1(s_1), \hat{P}_2(s_1), \hat{P}_1(s_2), \hat{P}_2(s_2), \ldots, \hat{P}_1(s_k), \hat{P}_2(s_k))\), where
\[
\mathcal{P}(z) = \sum_{j=0}^{N_j(z)} \Gamma_{j+1}^{(j)} \quad \text{and} \quad \mathcal{P}(z) = \sum_{j=0}^{N_j(z)} Z_{j+1}^{(j)}
\]
By Theorem 16 in Pollard (1984, page 134), it follows that, in probability,
\[
(\hat{P}_1(z), \hat{P}_2(z)) \to (P_1(z), P_2(z)) \quad \text{in } D^2[0, \infty).
\]
Thus, \(\hat{P}(z)\) converges weakly to \(P(z)\) conditionally on \(X\) in \(D(\mathbb{R})\), in probability. By Theorem 3.1 in Seijo and Sen (2011), \(\hat{M}_n^{(j)}\) converges weakly to \(M^{(j)}\) conditionally on \(X\), in probability, i.e., the result holds.

**Acknowledgments**

The authors thank two anonymous referees, the Associate Editor and the Editor for their helpful comments that improved the presentation. The research was partially supported by Hong Kong Research Grants Commission Grants HKUST601607 and HKUST602609.

**References**


