Statistica Sinica Preprint No: SS-2017-0341	
Title	Statistical Inference for Structurally Changed Threshold Autoregressive Models
Manuscript ID	SS-2017-0341
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202017.0341
Complete List of Authors	Zhaoxing Gao and Shiqing Ling
Corresponding Author	Zhaoxing Gao
E-mail	mazxgao@gmail.com
Notice: Accepted version subject to English editing.	

Statistical Inference for Structurally Changed Threshold Autoregressive Models

Zhaoxing Gao[1] and Shiqing Ling[2]

[1]*Booth School of Business, University of Chicago*

[2]*Hong Kong University of Science and Technology*

Abstract: In this paper, we study the theory and methodology for statistical inferences of threshold and change-point in the threshold autoregressive models. It is shown that the least squares estimators (LSE) of the threshold and change-point are n-consistent, and converge weakly to the minimizer of a compound Poisson process and the location of minima of a two-sided random walk, respectively. When the magnitude of change in the parameters of the state regimes or the time horizon is small, it is further shown that these limiting distributions can be approximated by a class of known distributions. The LSE of slope parameters are $\sqrt{n}$-consistent and asymptotically normal. Furthermore, a likelihood-ratio based confidence set is given for the threshold and change-point, respectively. Simulation study is carried out to assess the performance of our procedure. The proposed theory and methodology are further illustrated by a tree-ring dataset.

Key words and phrases: Threshold, Change-point, Least squares estimation, Compound Poisson process, Brownian motion.

1 Introduction

Structural changes have been important problems in econometrics, engineering and statistics for a long time. They are ubiquitous in economic and financial time series and were widely recognized as early as the 1940s, see Haavelmo (1944). Many approaches have been developed to detect whether or not structural change exists in a statistical model, see recent articles by Ling (2007), Aue et al. (2009) and Hidalgo and Seo (2013), and references therein. Hinkley (1970) is the first to investigate the maximum likelihood estimator (MLE) of the change-point in a sequence of i.i.d. random variables and showed that the MLE converges in distribution to the location of the maxima of a double-sided random walk. Except for the i.i.d. normal or binomial random variables, its limiting distribution does not have a closed form and it is very difficult to use in practice. When the magnitude of change is small, Yao (1987) showed that Hinkley's limiting distribution can be approximated by a very nice distribution. Picard (1985) studied the MLE of change-points in autoregressive (AR) models and obtained the same limiting distribution as that in Yao (1987) when the magnitude of changed parameters approaches zero with the sample size tending to ∞. Along this direction, Bai (1994, 1995) and Bai et al. (1998) studied the estimated change-point in the multivariate AR model and the co-integrated time series model, see also Chong (2001). Saikkonen et al. (2006) and Kejriwal and Perron (2008) used a similar method to estimate the change-point in vector AR models and co-integrated regression models, respectively. Davis et al. (2006) proposed a minimum description length principle to locate the change-points in AR models with multiple structural changes. In an autoregressive setting, Chan et al. (2014) used the group Lasso to estimate clusters of parameters with identical values over time and Qian and Su (2014) considered the problems of estimation in time series with endogenous regressors and an unknown number of breaks using the group fused Lasso. Ling (2016) developed an asymptotic theory for the change-point in linear and nonlinear time series models. Other contributions include the monograph of Csörgö and Horváth (1997) and Shao and Zhang (2012), among others.

The previous change-point problems often refer to statistical models with a structural change in the time horizon. In the dynamic system, however, the structural change may occur over the state regimes. The threshold autoregressive (TAR) model proposed by Tong (1978) is to capture these phenomena. It has been extensively applied in many areas, including economics, finance, biological and environmental sciences, among others, see Chan and Kutoyants (2010) and Tong (2011) for nice reviews. The asymptotic theory of the least squares estimator (LSE) of two-regime TAR models was established by Chan (1993) and Chan and Tsay (1998), and was further developed by Li and Ling (2012) and Li et al. (2013) for the multiple-regime TAR model and the TMA model, respectively. Hansen (2000) studied the LSE for two-regime threshold AR/regression models when the threshold effect is vanishingly small. Seo and Linton (2007) proposed a smoothed LSE for the TAR/regression model and showed that the estimated threshold is asymptotically normal with a lower rate of convergence. Liu et al. (2011) and Gao et al. (2013) studied the LSE for the non-stationary first-order TAR model and Chan et al. (2015) adopted the LASSO method to estimate TAR models with multiple thresholds. Gao et al. (2017) proposed a non-nested test for TAR models vs. Smooth TAR models. However, the research on threshold models is limited to the cases without change-point over the time horizon. The main difficulty with a change-point lies in the non-smooth and nonlinear functions in the threshold models. The smoothness of the objective function in terms of parameters is required in Ling (2016) to establish the asymptotic properties, and hence his results cannot be applied to the threshold models. Recently, Yau et al. (2015) constructed a genetic algorithm to estimate multiple-regime TAR models with structural breaks. However, they only established the consistency results of the parameters under their setting without further studying the limiting distributions of the slope parameters, thresholds and change-point. As far as we know, no limiting distributions have been obtained when the structural change occurs over both the time horizon and the state regimes and no methodology is available for statistical inferences of the thresholds and change-points in this case.

On the other hand, empirical time series often exhibit complex patterns, which may include nonlinearity and nonstationarity. For example, Tong and Lim (1980) analyzed the Canadian lynx data using TAR models and found obvious limit cycles, and the one-step-ahead prediction is better than using pure AR models in terms of the root-mean-square error. In this modern information age, as the availability of large collections of time series, many long time series sequences often possess nonstationarities and it is often not adequate to characterize the data by a single stationary model; see, for example, Shao and Zhang (2012) for some applications. Therefore, it is important and interesting to combine the threshold and change-point together, which can be used to characterize the nonlinearity and nonstationarity, respectively.

This paper is to establish the theory and methodology for statistical inferences of thresholds and change-point in TAR models. We first study the LSE of the TAR model with a structural change. It is shown that both the estimated threshold and the change-point are n-consistent, and they converge weakly to the minimizer of a compound Poisson process and the location of the minima of a two-sided random walk, respectively. When the magnitude of change in parameters of the state regimes or the time horizon is small, it is further shown that these limiting distributions can be approximated by a class of known distributions. Furthermore, a likelihood-ratio based confidence set is given for the threshold and change-point, respectively. Similar to Bai (1994) and Bai (1997), other estimated slope parameters are $\sqrt{n}$-consistent and asymptotically normal. This paper also applies our procedure to study the growth of tree rings in China, in which we found an evidence of possible climate change during the period of 1641-1663.

This paper proceeds as follows. Section 2 presents the model and the estimation procedure. The asymptotic properties are presented in Section 3. Section 4 presents the approximating distribution of these limiting distributions. A likelihood-ratio based inference method is studied in Section 5. Section 6 reports simulation results. Section 7 analyzes data on the growth of tree rings. Owing to space constraint, some tables and figures of the numeric studies and all the proofs of the theorems in this paper are given in the Supplementary material.

Throughout this paper, $\|\cdot\|$ denotes the Euclidean norm of a matrix or vector, $O_p(1)$ (or $o_p(1)$) denotes a series of random variables that are bounded (or converge to zero) in probability, $\triangleq$ means "is defined as" and $\longrightarrow_{\mathcal{L}}$ denotes convergence in distribution.

2 Model and Least Squares Estimation

We consider the following TAR model with order p [TAR(p)]:

$$y_t = \Phi' Z_{t-1} I(q_{t-1} > r) + \Psi' Z_{t-1} I(q_{t-1} \leq r) + \varepsilon_t, \tag{2.1}$$

where p is some known positive integer, $\Phi = (\mu, \phi_1, \cdots, \phi_p)'$, $\Psi = (\nu, \psi_1, \cdots, \psi_p)'$, $Z_t = (1, y_t, \cdots, y_{t-p+1})'$, $I(\cdot)$ is an indicator function with a univariate threshold variable $q_t = q(y_t, y_{t-1}, ..., y_{t-p+1})$ and $\{\varepsilon_t\}$ are the errors, where $q(x)$ is a known function from R^p to R. It is often to take $q_{t-1} = y_{t-d}$ for some integer $1 \leq d \leq p$ in the literature. We only consider the simple case as Section 3 of Chan (1993) that there is no threshold effect on ε_t since the general case is similar. Denote $\theta = (\Phi', \Psi')'$. We assume that $\theta \in \Theta \subset R^{2p+2}$ and $r \in \Gamma = [\underline{r}, \bar{r}] \subset R$, where Θ is a compact set. We denote model (2.1) by $Y(\theta, r)$ with $\theta \in \Theta$ and $r \in \Gamma$, i.e. we will write $\{y_1, ..., y_m\} \in Y(\theta, r)$ for some $m > 0$ in the sense that they are generated from model (2.1) with parameters θ and r.

Let $\{y_1, y_2, ..., y_n\}$ be the random sample. We assume the data before the time k are generated from $Y(\theta_1, r_1)$ and those after the time k are generated from $Y(\theta_2, r_2)$, i.e.,

$$\{y_1, ..., y_k\} \in Y(\theta_1, r_1) \quad \text{and} \quad \{y_{k+1}, ..., y_n\} \in Y(\theta_2, r_2)$$

with $(\theta_1', r_1)' \neq (\theta_2', r_2)'$, and $k \in \{1, 2, \cdots, n-1\}$. k is called the unknown change-point and its true value is k_0. The true parameters of $(\theta_1', r_1)'$ and $(\theta_2', r_2)'$ are $(\theta_{10}', r_{10})'$ and $(\theta_{20}', r_{20})'$, respectively. We parameterize the unknown change-point k as $k = [n\tau]$ with $\tau \in (0,1)$ and $k_0 = [n\tau_0]$, where $[x]$ is the integer part of x. For each k, we use the pre-sample to estimate $(\theta_{10}', r_{10})'$ and the post-sample to estimate $(\theta_{20}', r_{20})'$ by least squares estimation. Let

$$\ell_t(\theta, r) = [y_t - \Phi' Z_{t-1} I(q_{t-1} > r) - \Psi' Z_{t-1} I(q_{t-1} \leq r)]^2. \tag{2.2}$$

Then the corresponding objective functions are as follows.

$$S_{1n}(\theta_1, r_1, k) = \sum_{t=1}^{k} \ell_t(\theta_1, r_1) \quad \text{and} \quad S_{2n}(\theta_2, r_2, k) = \sum_{t=k+1}^{n} \ell_t(\theta_2, r_2). \tag{2.3}$$

The objective function based on the whole sample is

$$S_n(\theta_1, \theta_2, r_1, r_2, k) = S_{1n}(\theta_1, r_1, k) + S_{2n}(\theta_2, r_2, k). \tag{2.4}$$

The minimizer $(\hat{\theta}_{1n}, \hat{\theta}_{2n}, \hat{r}_{1n}, \hat{r}_{2n}, \hat{k}_n)$ of $S_n(\theta_1, \theta_2, r_1, r_2, k)$ is called the LSE of the true ones, that is,

$$(\hat{\theta}_{1n}, \hat{\theta}_{2n}, \hat{r}_{1n}, \hat{r}_{2n}, \hat{k}_n) = \mathop{\arg\min}_{\substack{\theta_1, \theta_2 \in \Theta, r_1, r_2 \in \Gamma \\ 1 \leq k < n}} S_n(\theta_1, \theta_2, r_1, r_2, k).$$

The way to obtain $(\hat{\theta}_{1n}, \hat{\theta}_{2n}, \hat{r}_{1n}, \hat{r}_{2n}, \hat{k}_n)$ is as follows. Using a similar method to that in Chan (1993), for each $1 \leq k < n$ and fixed r_i, we first obtain

$$\hat{\theta}_{in}(r_i, k) = \mathop{\arg\min}_{\theta_i \in \Theta} S_{in}(\theta_i, r_i, k), \quad i = 1,\ 2,$$

where $\hat{\theta}_{in}(r,k)$ depends on r and k. Specifically, we have

$$\hat{\Phi}_{1n}(r_1,k)=\left[\sum_{t=1}^{k} Z_{t-1}Z_{t-1}' I(q_{t-1}>r_1)\right]^{-1}\sum_{t=1}^{k} Z_{t-1} I(q_{t-1}>r_1)y_t, \tag{2.5}$$

and similarly for others. Then we can obtain the minimizers $\hat{r}_{in}(k)$ for the fixed k as follows,

$$\hat{r}_{in}(k)=\arg\min_{r_i\in\Gamma} S_{in}\left(\hat{\theta}_{in}(r_i,k),r_i,k\right),\quad i=1,\ 2.$$

Finally, putting $\hat{\theta}_{in}(\hat{r}_{in}(k),k)$ and $\hat{r}_{in}(k)$ into (2.4) gives

$$\begin{aligned}\hat{k}_n &= \arg\min_{1\le k<n} S_n\left(\hat{\theta}_{1n}(\hat{r}_{1n}(k),k),\hat{\theta}_{2n}(\hat{r}_{2n}(k),k),\hat{r}_{1n}(k),\hat{r}_{2n}(k),k\right),\\ \hat{r}_{in} &= \hat{r}_{in}(\hat{k}_n) \text{ and } \hat{\theta}_{in}=\hat{\theta}_{in}(\hat{r}_{in},\hat{k}_n),\ i=1,2.\end{aligned}$$

The range $1\le k<n$ can be replaced by $\tilde{p}\le k\le n-\tilde{p}$ for some integer $\tilde{p}$ in practice. Let $\{q_{(1)},q_{(2)},...,q_{(k)}\}$ and $\{q_{(k+1)},...,q_{(n)}\}$ be the corresponding order statistics of the subsamples $\{q_1,q_2,...,q_k\}$ and $\{q_{k+1},...,q_n\}$. $S_{1n}(\hat{\theta}_{1n}(r_1,k),r_1,k)$ and $S_{2n}(\hat{\theta}_{2n}(r_2,k),r_2,k)$ are constants when $r_1\in[q_{(i)},q_{(i+1)})$ and $r_2\in[q_{(j)},q_{(j+1)})$. Thus, for each k, there exist two intervals, say $[q_{(i)},q_{(i+1)})$ and $[q_{(j)},q_{(j+1)})$, on which $S_{1n}(\hat{\theta}_{1n}(r_1,k),r_1,k)$ and $S_{2n}(\hat{\theta}_{2n}(r_2,k),r_2,k)$ achieve their global minima, respectively. We take $\hat{r}_{1n}(k)=q_{(i)}$ and $\hat{r}_{2n}(k)=q_{(j)}$. It is not difficult to show that $(\hat{\theta}_{1n},\hat{\theta}_{2n},\hat{r}_{1n},\hat{r}_{2n},\hat{k}_n)$ is the LSE of $(\theta_{10},\theta_{20},r_{10},r_{20},k_0)$.

The current paper only focuses on model (2.1) with one change-point. If model (2.1) has m unknown change-points $(k_2,k_3,...,k_{m+1})$ such that $1<k_2<k_3<...<k_{m+1}<n$ and the

corresponding changed parameters are $\{(\theta_j, r_j) : j = 1, 2, ..., m+1\}$, then the profile objective function becomes

$$S_n(k_2, ..., k_{m+1}) = \sum_{j=1}^{m+1} \min_{\theta_j, r_j} \sum_{t=k_j+1}^{k_{j+1}} \ell_t(\theta_j, r_j), \tag{2.6}$$

where $k_1 = 0$ and $k_{m+2} = n$. The estimators of the true change-points are

$$(\hat{k}_2, ..., \hat{k}_{m+1}) = \operatorname*{arg\,min}_{(k_2,...,k_{m+1}) \in \Lambda} S_n(k_2, ..., k_{m+1}), \tag{2.7}$$

where Λ is some appropriate partition set. The dynamic programming algorithm of Bai and Perron (2003) should be able to apply for the searching thresholds and change-points. However, how to determine the number of change-points m and how to establish its asymptotic results are still challenging, and this may be done by following the approach of Bai and Perron (1998). Yau et al. (2015) proposed an efficient algorithm to estimate the number of change-points together with all other parameters by a different approach, but their asymptotic distributions are still unclear.

3 Asymptotic Properties

We first study the rates of convergence of the estimated threshold and change-point. It is an essential part in the threshold and change-point problems. We need several assumptions as follows.

Assumption 3.1. $\varepsilon_t = \sigma_{10} u_t$ for $t \le k_0$ and $\varepsilon_t = \sigma_{20} u_t$ for $t > k_0$, where u_t is a sequence of i.i.d. random variables with $Eu_1 = 0$, $Eu_1^2 = 1$ and has an absolutely continuous distribution with a uniformly continuous and positive density $f_u(x)$ on R. Furthermore, $E|u_t|^{2+\iota} < \infty$ for some $\iota \in (0, 1)$.

Assumption 3.2. q_t has an absolutely continuous distribution with a uniformly continuous and positive density function $\pi_i(r)$ on Γ, where $i = 1$ when $t \leq k_0$ and $i = 2$ when $t > k_0$.

Assumption 3.3. $(\theta_{10}', r_{10})'$ and $(\theta_{20}', r_{20})'$ are interior points in $\Theta \times \Gamma$, and $\Phi_{10} \neq \Psi_{10}$, $\Phi_{20} \neq \Psi_{20}$ and $(\theta_{10}', r_{10})' \neq (\theta_{20}', r_{20})'$.

Assumption 3.1 allows a variance change on the errors across the change-point and the conditions of the density function is the same as Condition 2 in Chan (1993). It implies Assumption 3.2 and entails that $\pi(\cdot)$ is bounded, continuous and positive on Γ if we take $q_{t-1} = y_{t-d}$ as a special case, see (ii) in Remark B of Chan (1993). Assumption 3.3 guarantees the identification of r_{10}, r_{20} and k_0. In order to obtain the consistency of $\hat{\theta}_{in}$ and $\hat{r}_{in}$ for $i = 1, 2$, we need an assumption similar to Condition 1 in Chan (1993). Let $\mathcal{Z}_{it} = (y_t, \cdots, y_{t-p+1}) \in Y(\theta_{i0}, r_{i0})$ for $i = 1, 2$, then $\mathcal{Z}_{it}$ is a Markov chain. Denote its l-step transition probability by $P^l(x, A)$, where $x \in R$ and A is a Borel set of R.

Assumption 3.4. $\mathcal{Z}_{it}$ admits a unique invariant measure $\Pi_i(\cdot)$ such that there exist a $K > 0$ and $\rho \in [0, 1)$, and for any $x \in R$ and any positive integer l, $\|P^l(x, \cdot) - \Pi_i(\cdot)\|_v \leq K(1 + \|x\|)\rho^l$, where $\|\cdot\|_v$ and $\|\cdot\|$ denote the total variation norm and the Euclidean norm, respectively.

When $t \leq k_0$, $\ell_t(\theta, r) = \ell(\theta, r, y_t, ..., y_1, \mathcal{Z}_{10})$ and when $t > k_0$, $\ell_t(\theta, r) = \ell(\theta, r, y_t, ..., y_{k_0+1}, \mathcal{Z}_{2k_0})$, where $\ell(\cdot)$ is a measurable function of $\{y_t\}$ with parameters θ and r. That is, there are two processes $\{y_{1t}\} \in Y(\theta_{10}, r_{10})$ and $\{y_{2t}\} \in Y(\theta_{20}, r_{20})$ and we observe $y_t = y_{1t}$ when $t \leq k_0$ and $y_t = y_{2t}$ when $t > k_0$. A set of sufficient conditions for Assumption 3.4 is

$$\max_{i=1,2}\{\sum_{j=1}^{p}|\phi_{ij}|, \sum_{j=1}^{p}|\psi_{ij}|\} < 1$$

and Assumption 3.1, see Chan and Tong (1985) and Chan (1989). When $p = 1$, the above coefficient condition can be weakened to $\phi_{i1} < 1$, $\psi_{i1} < 1$ and $\phi_{i1}\psi_{i1} < 1$ for $i = 1, 2$. Under Assumption 3.4, $\{\mathcal{Z}_{it}\}$ is V-uniformly ergodic with $V(x) = K(1 + \|x\|)$, see Meyn and Tweedie (2009). This condition is stronger than geometric ergodicity and the strong mixing ($\alpha-$mixing) condition, see the definition in Rosenblatt (1956) and the discussion in Hansen (2000). If the initial values $\mathcal{Z}_{10}$ and $\mathcal{Z}_{2k_0}$ are from the distributions $\Pi_1(\cdot)$ and $\Pi_2(\cdot)$, respectively, then Assumption 3.4 implies that $\{y_t\}_{t=1}^{k_0}$ and $\{y_t\}_{k_0+1}^{n}$ are two strictly stationary and ergodic sequences. In practice, the initial values $\mathcal{Z}_{10}$ and $\mathcal{Z}_{2k_0}$ are replaced by some chosen constants and we will discuss their effect on the estimators in Section 5.

Our first result is stated as follows.

Theorem 3.1. *If Assumptions 3.1-3.4 hold, then*

$$(a) \quad \hat{\theta}_{in} = \theta_{i0} + o_p(1) \quad \textit{and} \quad \hat{r}_{in} = r_{i0} + o_p(1), \quad i = 1, 2,$$

$$(b) \quad \hat{k}_n = k_0 + O_p(1).$$

From Theorem 3.1, we see that the thresholds and the slop parameters are all consistent.

Similar results are also obtained in Theorem 2 of Yau et al. (2015). We can write $\hat{k}_n = [n\hat{\tau}_n]$, then $\hat{\tau}_n$ is an estimator of τ_0. Theorem 3.1 implies that the rate of convergence of $\hat{\tau}_n$ is n, which is the same with that in Bai and Perron (1998) and Yau et al. (2015), while it is faster than those in Picard (1985) and Bai et al. (1998) for AR models since their convergence rate is essentially nd_n^2, where d_n is the magnitude of the change with $d_n \to 0$. To obtain the rate of convergence of $\hat{r}_{in}$, we need two more assumptions. Define $M_i(r) = E(Z_t Z_t'|q_t = r)$, where $i = 1$ when $t \leq k_0$ and $i = 2$ when $t > k_0$.

Assumption 3.5. (i) $M_i(r)$ is continuous at $r = r_{i0}$; (ii) $(\Psi_{i0} - \Phi_{i0})' M_i(r_{i0})(\Psi_{i0} - \Phi_{i0}) > 0$; $i = 1,\, 2$.

We say that model (2.1) has a discontinuous autoregressive function if Assumption 3.5 is satisfied. For Assumption 3.5 to hold, we require the threshold variable to have a continuous distribution, see Assumption 1.5 of Hansen (2000) or Assumption 2.2 of Gao et al. (2017) for details. Assumption 3.5 (ii) is natural for a positive difinite matrix $M_i(r_{i0})$. We note that if we choose $q_{t-1} = y_{t-d}$ as that in Chan (1993), Assumption 3.5(ii) implies Condition 4 in Chan (1993) and the threshold r_{i0} becomes the jump point of the autoregressive function. For simplicity, let

$$M_i(r^+) = EZ_t Z_t'(r^+),\, M_i(r^-) = EZ_t Z_t'(r^-) \text{ and } M_i(r_1, r_2) = EZ_t Z_t'(r_1, r_2),$$

where $i = 1$ when $1 \leq t \leq k_0$ and $i = 2$ when $k_0 + 1 \leq t \leq n$, $Z_t(r^+) = Z_t I(q_t > r)$, $Z_t(r^-) = Z_t I(q_t \leq r)$ and $Z_t(r_1, r_2) = Z_t I(r_1 < q_t \leq r_2)$. Under Assumptions 3.1-3.3, it is not difficult to show that $M_i(r^+) > 0$, $M_i(r^-) > 0$ and $M_i(r_1, r_2) > 0$ for all $r, r_1, r_2 \in \Gamma$ with $r_1 < r_2$, $i = 1, 2$. By adopting the techniques in Chan (1993) to each segment of $\{y_t\}$, we have the convergence rate of $\hat{r}_{in}$ and the asymptotic normality of $\hat{\theta}_{in}$ as follows.

Theorem 3.2. *If Assumptions 3.1-3.5 hold, then*

$$(a) \quad n(\hat{r}_{in} - r_{i0}) = O_p(1),$$

$$(b) \quad \sqrt{n}(\hat{\theta}_{in} - \theta_{i0}) \longrightarrow_{\mathcal{L}} N(0, \sigma_{i0}^2 \Sigma_i^{-1}),$$

$i = 1,\ 2$, *as* $n \to \infty$, *where* $\Sigma_1 = \tau_0\, diag\{M_1(r_{10}^+), M_1(r_{10}^-)\}$ *and* $\Sigma_2 = (1-\tau_0) diag\{M_2(r_{20}^+),$ $M_2(r_{20}^-)\}$. *Furthermore,* $n(\hat{r}_{in} - r_{i0})$ *is asymptotically independent of* $\sqrt{n}(\hat{\theta}_{in} - \theta_{i0})$ *which is always asymptotically normal and* $\hat{\theta}_{1n}$ *and* $\hat{\theta}_{2n}$ *are also asymptotically independent of each other.*

This theorem shows that the rate of convergence of $\hat{r}_{in}$ is the same as that in Chan (1993) and Yau et al. (2015), but it is faster than those in Hansen (2000) and Seo and Linton (2007) since their convergence rate is essentially $n^{1-2\tilde{\alpha}}$ for some $\tilde{\alpha} \in (0, 1/2)$ by assuming a vanishingly small thershold effect. Yau et al. (2015) only obtained the $\sqrt{n}$-consistency of the slope parameters, while here we show that those parameters have the same asymptotic properties as those in Picard (1985) and Bai (1997), among others, that is, they are not affected by the threshold parameters.

In Theorem 3.2, τ_0, σ_{i0}^2 and Σ_i are not known in practice, but they can be replaced by consistent estimates from $(\hat{\theta}_{1n}, \hat{\theta}_{2n}, \hat{r}_{1n}, \hat{r}_{2n}, \hat{k}_n)$. For example, σ_{i0}^2 can be estimated by

$$\hat{\sigma}_{1n}^2 = \frac{1}{\hat{k}_n}\sum_{t=1}^{\hat{k}_n} \ell_t(\hat{\theta}_{1n}, \hat{r}_{1n}) \quad , \quad \hat{\sigma}_{2n}^2 = \sum_{t=\hat{k}_n+1}^{n} \ell_t(\hat{\theta}_{2n}, \hat{r}_{2n}), \tag{3.1}$$

and $M_i(r_{i0}^{\pm})$ can be estimated by

$$\hat{M}_1(\hat{r}_{1n}^{\pm}) = \frac{1}{\hat{k}_n}\sum_{t=1}^{\hat{k}_n} Z_t Z_t'(\hat{r}_{1n}^{\pm}) \quad , \quad \hat{M}_2(\hat{r}_{2n}^{\pm}) = \frac{1}{n-\hat{k}_n}\sum_{t=\hat{k}_n+1}^{n} Z_t Z_t'(\hat{r}_{2n}^{\pm}), \tag{3.2}$$

respectively. Hence, Σ_i can be estimated by

$$\hat{\Sigma}_{1n} = \hat{\tau}_n diag\{\hat{M}_1(\hat{r}_{1n}^{+}), \hat{M}_1(\hat{r}_{1n}^{-})\} \quad \text{and} \quad \hat{\Sigma}_{2n} = (1-\hat{\tau}_n) diag\{\hat{M}_2(\hat{r}_{2n}^{+}), \hat{M}_2(\hat{r}_{2n}^{-})\}. \tag{3.3}$$

By Theorems 3.1-3.2, it is not hard to show that the above sample estimators are all consistent to their corresponding true ones and hence can be used for statistical inferences.

To study the limiting distributions of $\hat{r}_{in}$, we consider the profile objective function

$$\tilde{S}_{in}(z, \hat{k}_n) \equiv S_{in}(\hat{\theta}_{in}(r_{i0} + \frac{z}{n}, \hat{k}_n), r_{i0} + \frac{z}{n}, \hat{k}_n) - S_{in}(\hat{\theta}_{in}(r_{i0}, \hat{k}_n), r_{i0}, \hat{k}_n), \tag{3.4}$$

where $z \in R$, $i = 1, 2$. According to the procedure for $\hat{r}_{in}$, we can see that

$$n(\hat{r}_{in} - r_{i0}) = \arg\min_{z \in R} \tilde{S}_{in}(z, \hat{k}_n). \tag{3.5}$$

Now, we define two jump processes as follows:

$$\tilde{\mathcal{S}}_{1n}(k_0,\theta_{10},r_{10},r_1)=\begin{cases}\sum_{t=1}^{k_0}\xi_{1t}I(r_1<q_{t-1}\le r_{10}), & r_1<r_{10},\\ 0, & r_1=r_{10},\\ \sum_{t=1}^{k_0}\eta_{1t}I(r_{10}<q_{t-1}\le r_1), & r_1>r_{10},\end{cases} \tag{3.6}$$

$$\tilde{\mathcal{S}}_{2n}(k_0,\theta_{20},r_{20},r_2)=\begin{cases}\sum_{t=k_0+1}^{n}\xi_{2t}I(r_2<q_{t-1}\le r_{20}), & r_2<r_{20},\\ 0, & r_2=r_{20},\\ \sum_{t=k_0+1}^{n}\eta_{2t}I(r_{20}<q_{t-1}\le r_2), & r_2>r_{20},\end{cases} \tag{3.7}$$

where $\xi_{it}=(\Phi_{i0}-\Psi_{i0})'Z_{t-1}Z_{t-1}'(\Phi_{i0}-\Psi_{i0})-2(\Phi_{i0}-\Psi_{i0})'Z_{t-1}\varepsilon_t$ and $\eta_{it}=(\Psi_{i0}-\Phi_{i0})'Z_{t-1}Z_{t-1}'\,(\Psi_{i0}-\Phi_{i0})-2(\Psi_{i0}-\Phi_{i0})'Z_{t-1}\varepsilon_t$ for $i=1,\,2$. We reparameterize r_i as $r_{i0}+\frac{z}{n}$ in (3.6) and (3.7). On the event $\{|z|\le B,|\hat{k}_n-k_0|\le M\}$ for any fixed $B,M\in(0,\infty)$, $\tilde{S}_{in}(z,\hat{k}_n)$ can be approximated by $\tilde{\mathcal{S}}_{in}(k_0,\theta_{i0},r_{i0},r_{i0}+z/n)$ in $D(R)$, the space of all càdlàg functions on R being equipped with the Skorokhod metric, i.e.,

$$\tilde{S}_{in}(z,\hat{k}_n)=\tilde{\mathcal{S}}_{in}(k_0,\theta_{i0},r_{i0},r_{i0}+\frac{z}{n})+o_p(1),\quad i=1,2, \tag{3.8}$$

see the proof of Lemma 4 in the supplementary material. We will show that $\tilde{\mathcal{S}}_{in}(k_0,\theta_{i0},r_{i0},$ $r_{i0}+z/n)$ weakly converges to a two-sided compound Poisson process $\mathcal{P}_1(\tau_0 z)$ for $i=1$ and $\mathcal{P}_2((1-\tau_0)z)$ for $i=2$, which are defined as follows:

$$\mathcal{P}_i(z)=I(z<0)\sum_{t=1}^{N_1^i(-z)}Y_{it}+I(z\ge 0)\sum_{t=1}^{N_2^i(z)}Z_{it},\quad z\in R,i=1,\,2, \tag{3.9}$$

where $\{N_1^i(z), z \geq 0\}$ and $\{N_2^i(z), z \geq 0\}$ are two independent poisson processes with $N_1^i(0) = 0$ and $N_2^i(0) = 0$ a.s. and with the same jump rate $\pi_i(r_{i0})$, where $\pi_i(\cdot)$ is the density function of $\{q_t\}$, see Assumption 3.2. $\{Y_{it}, t \geq 1\}$ are i.i.d. random variables with the distribution function $F_{1i}(\cdot|r_{i0})$ and $\{Z_{it}, t \geq 1\}$ are i.i.d. random variables with the distribution function $F_{2i}(\cdot|r_{i0})$, where $F_{1i}(\cdot|x)$ and $F_{2i}(\cdot|x)$ are the conditional distribution functions of ξ_{it} and η_{it} given $q_{t-1} = x$, respectively, and $\{Y_{it}, t \geq 1\}$ and $\{Z_{it}, t \geq 1\}$ are mutually independent. Clearly $\mathcal{P}_i(z) \to +\infty$ a.s. when $|z| \to \infty$ since $EY_{it} = EZ_{it} > 0$ by Assumptions 3.3-3.4. Therefore, there exists a unique random interval $[M_-^{(i)}, M_+^{(i)})$ on which the process $\mathcal{P}_i(z)$ attains its global minimum a.s.. That is,

$$[M_-^{(i)}, M_+^{(i)}) = \arg\min_{z \in R} \mathcal{P}_i(z), \quad i = 1, 2. \tag{3.10}$$

To obtain the limiting distribution of $\hat{k}_n$, we define a two-sided random walk as follows:

$$W(k, \theta_{10}, \theta_{20}, r_{10}, r_{20}) = \begin{cases} \sum_{t=k}^{-1} [\ell_t(\theta_{20}, r_{20}) - \ell_t(\theta_{10}, r_{10})], & k < 0, \\ 0, & k = 0, \\ \sum_{t=1}^{k} [\ell_t(\theta_{10}, r_{10}) - \ell_t(\theta_{20}, r_{20})], & k > 0, \end{cases} \tag{3.11}$$

where $\ell_t(\theta, r)$ is defined in (2.2). $y_t \in Y(\theta_{10}, r_{10})$ when $t < 0$ and $y_t \in Y(\theta_{20}, r_{20})$ when $t > 0$. Now we can state our theorem as follows.

Theorem 3.3. *If Assumptions 3.1-3.5 hold, then $\hat{k}_n$, $\hat{r}_{1n}$ and $\hat{r}_{2n}$ are asymptotically independent* *with each other, and*

$$(a)\quad n(\hat{r}_{1n} - r_{10}) \longrightarrow_{\mathcal{L}} \frac{1}{\tau_0} M_-^{(1)}$$
$$and\ n(\hat{r}_{2n} - r_{20}) \longrightarrow_{\mathcal{L}} \frac{1}{1-\tau_0} M_-^{(2)},$$
$$(b)\quad \hat{k}_n - k_0 \longrightarrow_{\mathcal{L}} \arg\min_k W(k, \theta_{10}, \theta_{20}, r_{10}, r_{20}),$$

as $n \to \infty$.

From this theorem, we can see that the limiting distribution of $\hat{r}_{in}$ is the same as that in Chan (1993) subject to a scale τ_0 or $1 - \tau_0$. The limiting distribution of the estimated change-point correspondes to that of the MLE in Ling (2016) is also related to a two-sided random walk even if the objective function is not continuous. On the other hand, this theorem can be treated as a complement to the results in Yau et al. (2015) since they only obtained the consistency results without investigating the limiting distributions of either the thresholds or the change-point.

4 Approximating the Limiting Distributions

Except for the MLE of the threshold in a simple regression model with i.i.d. data in Yu (2012), we generally do not have a closed form solution for the estimated threshold. The distributions in Theorem 3.3 are difficult to use directly for statistical inference. Li and Ling (2012) proposed a numerical method to simulate the limiting distribution of the estimated thresholds, when the threshold effect is fixed, while Hansen (2000) adopted a differenct approach and obtained a closed form for the distribution by assuming the threshold effect is vanishingly small as the sample size $n \to \infty$. In this section, we use Hansen (2000)'s method to obtain the approximating distributions of r_{10} and r_{20}. By borrowing a similar idea to that in Bai (1997) and Hansen (2000), we also obtain the approximating distribution of $W(k, \theta_{10}, \theta_{20}, r_{10}, r_{20})$. Our discussions are separated into two subsections.

4.1 Approximating Distributions of the Estimated r_{10} and r_{20}

Let $\bar{\xi}_{it} = \delta'_{in} Z_{t-1} Z'_{t-1} \delta_{in} + 2\delta'_{in} Z_{t-1}\varepsilon_t$ and $\bar{\eta}_{it} = \delta'_{in} Z_{t-1} Z'_{t-1} \delta_{in} - 2\delta'_{in} Z_{t-1}\varepsilon_t$ for $i = 1, 2$. Along (3.6)-(3.7), we first define the following two processes:

$$\bar{\mathcal{S}}_{1n}(k_0, \theta_{10}, r_{10}, r_1) = \begin{cases} \sum_{t=1}^{k_0} \bar{\xi}_{1t} I(r_1 < q_{t-1} \le r_{10}), & r_1 < r_{10}, \\ 0, & r_1 = r_{10}, \\ \sum_{t=1}^{k_0} \bar{\eta}_{1t} I(r_{10} < q_{t-1} \le r_1), & r_1 > r_{10}, \end{cases} \tag{4.1}$$

and

$$\bar{\mathcal{S}}_{2n}(k_0, \theta_{20}, r_{20}, r_2) = \begin{cases} \sum_{t=k_0+1}^{n} \bar{\xi}_{2t} I(r_2 < q_{t-1} \le r_{20}), & r_2 < r_{20}, \\ 0, & r_2 = r_{20}, \\ \sum_{t=k_0+1}^{n} \bar{\eta}_{2t} I(r_{20} < q_{t-1} \le r_2), & r_2 > r_{20}. \end{cases} \tag{4.2}$$

The difference between the processes in (3.6)-(3.7) and those in (4.1)-(4.2) lie in (ξ_{it}, η_{it}) and $(\bar{\xi}_{it}, \bar{\eta}_{it})$. y_t in (ξ_{it}, η_{it}) depends on Φ_{i0} and Ψ_{i0} and hence y_t is changing if $\Psi_{i0} - \Phi_{i0} \to 0$. But y_t in $(\bar{\xi}_{it}, \bar{\eta}_{it})$ is irrelevant to δ_{in} and hence y_t is still stationary and ergodic when $\delta_{in} \to 0$. This is the reason we introduce the two new processes. We make one assumption.

Assumption 4.1. $\delta_{in} = c_i n^{-\beta}$, where $\beta \in (0, 1/2)$, and c_1 and c_2 are nonzero constant vectors.

Define

$$\hat{z}_{in} = \arg\min_{z \in R} \bar{\mathcal{S}}_{in}(k_0, \theta_{i0}, r_{i0}, r_{i0} + \frac{z}{n}), \quad i = 1, 2, \tag{4.3}$$

and $B(r)$ is a standard Brownian motion on $(-\infty, \infty)$.

By Assumptions 3.1-3.5, 4.1, and the proof of Theorem 1 in Hansen (2000), we can easily obtain the following theorem.

Theorem 4.1. *If Assumptions 3.1-3.5 and 4.1 hold and $Ey_t^4 < \infty$, then*

$$n^{-2\beta} \hat{z}_{in} \longrightarrow_{\mathcal{L}} w_i \arg\min_{-\infty < r < \infty} [\frac{|r|}{2} + B(r)], \quad i = 1, 2, \tag{4.4}$$

where $w_1 = \sigma_{10}^2 / [\tau_0 c_1' M_1(r_{10}) c_1 \pi_1(r_{10})]$, $w_2 = \sigma_{20}^2 / [(1 - \tau_0) c_2' M_2(r_{20}) c_2 \pi_2(r_{20})]$, $M_1(r) = E(Z_t Z_t' | q_t = r)$ with $t \le k_0$ and $M_2(r) = E(Z_t Z_t' | q_t = r)$ with $t > k_0$.

Remark 4.1. The finite fouth moment in Theorem 4.1 is inherited from Hansen (2000). When there is a threshold effect in variance in model (2.1) as Chan (1993), Li and Ling (2012) discussed the limiting distribution of the estimated threshold. In this case, the approximation in Theorem 4.1 should be asymmetric, see a further discussion in Yu (2012) and Yu (2015).

Let $T = \arg\min_{-\infty < r < \infty} [\frac{|r|}{2} + B(r)]$ and $\Phi(x)$ denote the cumulative standard normal distribution function. Then for $x \ge 0$,

$$P(T \le x) = 1 + \sqrt{\frac{x}{2\pi}} \exp(-\frac{x}{8}) + \frac{3}{2} \exp(x) \Phi(-\frac{3\sqrt{x}}{2}) - (\frac{x + 5}{2}) \Phi(-\frac{\sqrt{x}}{2}).$$

and for $x < 0$, $P(T \geq x) = 1 - P(T \geq -x)$, see Yao (1987). By Theorem 4.1, if $\Psi_{i0} - \Phi_{i0} \approx \delta_{in}$, then we can use the following approximation:

$$n^{1-2\beta}(\hat{r}_{in} - r_{i0})/w_i \approx_d \mathop{\arg\min}_{-\infty<r<\infty}[\frac{|r|}{2} + B(r)], \quad i = 1,\ 2, \tag{4.5}$$

where $n^{1-2\beta}/w_i$ can be replaced by $\sigma_{10}^{-2}[(\Psi_{10} - \Phi_{10})' M_1(r_{10})(\Psi_{10} - \Phi_{10})\pi_1(r_{10})]k_0$ when $i = 1$ and by $\sigma_{20}^{-2}[(\Psi_{20} - \Phi_{20})' M_2(r_{20})(\Psi_{20} - \Phi_{20})\pi_2(r_{20})](n - k_0)$ when $i = 2$. In practice, k_0 can be replaced by $[n\hat{\tau}_n]$, other true values can be consistently estimated in a similar way to (3.1)-(3.2), $\pi_i(r_{i0})$ can be estimated by its kernel density estimator and $M_i(r_{i0})$ can be estimated by a polynomial regression as that in Hansen (2000). Thus, we can use the distribution of T to make statistical inferences for the thresholds r_{10} and r_{20}.

4.2 Approximating Distribution of the Estimated k_0

In this subsection, we investigate the limiting distribution of $W(k, \theta_{10}, \theta_{20}, r_{10}, r_{20})$ in Theorem 3.3. From (S2.24) in Section S2 of the supplementary material, we have

$$\begin{aligned}\hat{k}_n = \mathop{\arg\min}_{1\leq k<n} \Bigg\{ & I(k < k_0) \sum_{t=k+1}^{k_0} (A_{1t}^2 + 2A_{1t}\varepsilon_t) \\ & + I(k \geq k_0) \sum_{t=k_0+1}^{k} (A_{2t}^2 + 2A_{2t}\varepsilon_t) + o_p(1) \Bigg\},\end{aligned} \tag{4.6}$$

where

$$\begin{aligned} A_{1t} =&(\Phi_{10}-\Phi_{20})'Z_{t-1}I(q_{t-1}>r_{20})+(\Psi_{10}-\Psi_{20})'Z_{t-1}I(q_{t-1}\le r_{10}) \\ &+(\Phi_{10}-\Psi_{20})'Z_{t-1}I(r_{10}<q_{t-1}\le r_{20}) \end{aligned} \tag{4.7}$$

and $A_{2t}=-A_{1t}$ if $r_{20}>r_{10}$, and

$$\begin{aligned} A_{1t} =&(\Phi_{10}-\Phi_{20})'Z_{t-1}I(q_{t-1}>r_{10})+(\Psi_{10}-\Psi_{20})'Z_{t-1}I(q_{t-1}\le r_{20}) \\ &+(\Psi_{10}-\Phi_{20})'Z_{t-1}I(r_{20}<q_{t-1}\le r_{10}) \end{aligned} \tag{4.8}$$

and $A_{2t}=-A_{1t}$ if $r_{10}>r_{20}$. We only consider the case when $r_{20}>r_{10}$ in (4.6) since it is similar for the case when $r_{10}>r_{20}$. Now, we define the following process:

$$\bar{W}_n(k)=I(k<k_0)\sum_{t=k+1}^{k_0}(\bar{A}_{1t}^2+2\bar{A}_{1t}\varepsilon_t)+I(k\ge k_0)\sum_{t=k_0+1}^{k}(\bar{A}_{2t}^2+2\bar{A}_{2t}\varepsilon_t), \tag{4.9}$$

where

$$\bar{A}_{1t} =\kappa_{1n}'Z_{t-1}I(q_{t-1}>r_{20})+\kappa_{2n}'Z_{t-1}I(q_{t-1}\le r_{10})+\kappa_{3n}'Z_{t-1}I(r_{10}<q_{t-1}\le r_{20}) \tag{4.10}$$

and $\bar{A}_{2t}=-\bar{A}_{1t}$. y_t in A_{it} depends on Φ_{i0} and Ψ_{i0} and hence y_t is changing if $\Phi_{20}-\Phi_{10}\to 0$, $\Psi_{20}-\Psi_{10}\to 0$ and $\Psi_{20}-\Phi_{10}\to 0$. But y_t in $\bar{A}_{it}$ is irrelevant to κ_{in} and is stationary and ergodic when $\kappa_{in} \to 0$. We make one assumption as follows.

Assumption 4.2. $\kappa_{1n} = c_3 n^{-\beta}$, $\kappa_{2n} = c_4 n^{-\beta}$ and $\kappa_{3n} = c_5 n^{-\beta}$, where $\beta \in (0, 1/2)$, c_3, c_4 and c_5 are nonzero constant vectors.

We can parameterize k as $k = k_0 + [n^{2\beta} s]$ in (4.9). By Assumptions 3.1-3.5 and Theorem 1 in the supplementary material, we can show that the sum of $\bar{A}_{it}^2$ goes to $\varpi_i |s|$ in probability and uniformly on any compact set of s, $i = 1, 2$, respectively, where

$$\varpi_i = c_3' M_i(r_{20}^+) c_3 + c_4' M_i(r_{10}^-) c_4 + c_5' M_i(r_{10}, r_{20}) c_5.$$

By Theorem A.1 in Li et al. (2016), the sum of $\bar{A}_{it}\varepsilon_t$ weakly converges to a Gaussian process $G(s)$ with covariance kernel $\varpi_i \sigma^2(|s| \wedge |t|)$, $i = 1, 2$, respectively. Then we can state our theorem as follows.

Theorem 4.2. *If Assumptions 3.1-3.5 and 4.2 hold, then*

$$\sigma_{10}^{-2} \varpi_1 \mathop{\arg\min}_{-\infty < s < \infty} \bar{W}_n(k_0 + [n^{2\beta} s]) \longrightarrow_{\mathcal{L}} T_{\phi,\xi},$$

where $\phi = \varpi_2 \sigma_{20}^2 / (\varpi_1 \sigma_{10}^2)$, $\xi = \varpi_2 / \varpi_1$ *and*

$$T_{\phi,\xi} = \mathop{\arg\min}_{s \in R} \left\{ \left[\frac{\xi}{2}|s| + \sqrt{\phi} B(s) \right] I(s \geq 0) + \left[\frac{1}{2}|s| + B(s) \right] I(s < 0) \right\},$$

$M_i(r_{i0}^+)$ *and* $M_i(r_{i0}^-)$ *are defined as in Theorem 3.2.*

Remark 4.2. Under Assumption 4.2, we can see that the third term in (4.10) and in ϖ_i will disappear if $r_{20} - r_{10} \to 0$, and the assumption on κ_{3n} will be redundant if $r_{20} = r_{10}$. Alternatively, if κ_{3n} is fixed and $r_{20} - r_{10}$ is vanishingly small such that it matches the convergence rate of s in Theorem 4.2, it is possible to obtain a similar result.

The distribution of $T_{\phi,\xi}$ can be found in Bai (1997) and its density is asymmetric unless $\phi = \xi = 1$. We can use the cumulative distribution functions in Appendix B of Bai (1997) to construct confidence intervals once we know the values of ϕ and ξ.

If $\Phi_{20} - \Phi_{10} \approx \kappa_{1n}$, $\Psi_{20} - \Psi_{10} \approx \kappa_{2n}$ and $\Phi_{10} - \Psi_{20} \approx \kappa_{3n}$, by Theorem 4.2, (4.6) and (4.9) with parameterizing $k = k_0 + [n^{2\beta}s]$, then we can see that

$$\frac{\varpi_1}{\sigma_{10}^2 n^{2\beta}}(\hat{k}_n - k_0) \approx \frac{\varpi_1}{\sigma_{10}^2} \mathop{\arg\min}_{-\infty<s<\infty} \bar{W}_n(k_0 + [n^{2\beta}s]) \approx_d T_{\phi,\xi}. \tag{4.11}$$

Here, we can approximate ϕ, ξ and $\varpi_1/(\sigma^2 n^{2\beta})$ as follows:

$$\phi \approx \frac{d_2\sigma_{20}^2}{d_1\sigma_{10}^2}, \xi \approx \frac{d_2}{d_1} \text{ and } \varpi_1/(\sigma_{10}^2 n^{2\beta}) \approx d_1/\sigma_{10}^2, \tag{4.12}$$

where $d_i = (\Phi_{20} - \Phi_{10})' M_i(r_{20}^+)(\Phi_{20} - \Phi_{10}) + (\Psi_{20} - \Psi_{10})' M_i(r_{10}^-)(\Psi_{20} - \Psi_{10}) + (\Phi_{10} - \Psi_{20})' M_i(r_{10}, r_{20})(\Phi_{10} - \Psi_{20})$, $i = 1, 2$. In practice, d_i can be estimated by

$$\begin{aligned}\hat{d}_{in} =& (\hat{\Phi}_{2n} - \hat{\Phi}_{1n})' \hat{M}_i(\hat{r}_{2n}^+)(\hat{\Phi}_{2n} - \hat{\Phi}_{1n}) + (\hat{\Psi}_{2n} - \hat{\Psi}_{1n})' \hat{M}_i(\hat{r}_{1n}^-)(\hat{\Psi}_{2n} - \hat{\Psi}_{1n}) \\ &+ (\hat{\Phi}_{1n} - \hat{\Psi}_{2n})' \hat{M}_i(\hat{r}_{1n}, \hat{r}_{2n})(\hat{\Phi}_{1n} - \hat{\Psi}_{2n}), \quad i = 1, 2,\end{aligned} \tag{4.13}$$

where $\hat{M}_i(\hat{r}_{1n}, \hat{r}_{2n})$ is defined in a similar way as $\hat{M}_i(\hat{r}_{in}^{\pm})$ in (3.2). Let $\hat{\phi}_n = \hat{d}_{2n}\hat{\sigma}_{2n}^2/(\hat{d}_{1n}\hat{\sigma}_{1n}^2)$ and $\hat{\xi}_n = \hat{d}_{2n}/\hat{d}_{1n}$, it is not hard to show that

$$\hat{\phi}_n \to_p \phi, \quad \hat{\xi}_n \to_p \xi \quad \text{and} \quad \frac{\hat{d}_{1n}}{\hat{\sigma}_{1n}^2} \to_p \frac{d_1}{\sigma_{10}^2}. \tag{4.14}$$

In practice, we do not know whether $r_{20} > r_{10}$ or not. If $\hat{r}_{2n} > \hat{r}_{1n}$, we use (4.13) to obtain the consistent estimates for (4.12). Otherwise, based on the expression of A_{1t} in (4.8) when $r_{10} > r_{20}$, we replace $\hat{\Phi}_{1n} - \hat{\Psi}_{2n}$ by $\hat{\Psi}_{1n} - \hat{\Phi}_{2n}$ and interchange the positions of $\hat{r}_{2n}$ and $\hat{r}_{1n}$ in (5.13). Thus, we can use $T_{\hat{\phi}_n,\hat{\xi}_n}$ to make statistical inferences for the change-point k_0. Note that $T_{\hat{\phi}_n,\hat{\xi}_n}$ is asymmetric and this is different from the symmetric T which is commonly used in approximating the distribution of the estimated change-point in the literature. See Bai (1994), Chong (2001), and Ling (2016), among others.

5 Likelihood-ratio Based Inference for Threshold and Change-point

The simulation method of Li and Ling (2012) for the inferences of the thresholds may not be accurate when the threshold effect is small. At the same time, the confidence interval based on the approximating method in Section 4 has a coverage rate below the nominal level when the threshold/structural change effect is large, see the discussion in Hansen (2000) for threshold scenario, and Elliott and Müller (2007), Eo and Morley (2015) and Elliott et al. (2015) for change-point scenario. Furthermore, they commented that the likelihood-ratio test is asymptotically pivotal when the threshold/structural change effect is small and the confidence region based on the inverted likelihood-ratio test is asymptotically valid, even if the threshold/structural change effect is relatively large. In the regression model for i.i.d. data, the nonparametric approach of Yu (2015) seems work well when the threshold effect is relatively strong but undercovers when the threshold effect is weak. However, it is not clear whether his method can be applied to threshold model with time series dependence. In this section, we consider the likelihood-ratio based confidence sets for the thresholds and the change-change, respectively.

We first investigate the likelihood-ratio test statistics for the thresholds. Following Hansen (2000), we consider the likelihood-ratio statistic $LR_{in}(r)$ for $H_{i0} : r = r_{i0}$ as

$$LR_{in}(r) = \frac{1}{\hat{\sigma}_{in}^2}[S_{in}(\hat{\theta}_{in}(r, \hat{k}_n), r, \hat{k}_n) - S_{in}(\hat{\theta}_{in}, \hat{r}_{in}, \hat{k}_n)], \quad i = 1,\, 2, \tag{5.1}$$

where $\hat{\sigma}_{in}^2$ and S_{in} are defined in (3.1) and (2.3), respectively. Under H_{i0}, by (3.4), (3.5) and (5.1), we have

$$LR_{in}(r_{i0}) = \frac{1}{\hat{\sigma}_{in}^2} \max_{z \in R} \left[-\tilde{\mathcal{S}}_{in}(k_0, \theta_{i0}, r_{i0}, r_{i0} + \frac{z}{n})\right] + o_p(1), \tag{5.2}$$

where $\tilde{\mathcal{S}}_{in}(k_0, \theta_{i0}, r_{i0}, r)$ are defined in (3.6) and (3.7) for $i = 1, 2$, respectively. By a similar argument as that in Section 4.1, we use $\bar{\mathcal{S}}_{in}$ in (4.1)-(4.2) to approximate $\tilde{\mathcal{S}}_{in}$ for $i = 1,\, 2$, respectively. By Theorem 4.1, it is not hard to show that

$$\frac{1}{\hat{\sigma}_{in}^2}\max_{z\in R}\left[-\bar{\mathcal{S}}_{in}(k_0,\theta_{i0},r_{i0},r_{i0}+\frac{z}{n})\right]=\frac{1}{\hat{\sigma}_{in}^2}\max_{v\in R}\left[-\bar{\mathcal{S}}_{in}(k_0,\theta_{i0},r_{i0},r_{i0}+\frac{v}{n^{1-2\beta}})\right]$$
$$\Rightarrow \max_{v\in R}[2B(v)-|v|]. \tag{5.3}$$

Then, if $\Psi_{i0}-\Phi_{i0}\approx\delta_{in}$, then we can use the following approximation:

$$LR_{in}(r_{i0})\approx_d \Delta := \max_{v\in R}[2B(v)-|v|], \tag{5.4}$$

where the distribution of Δ is $P(\Delta\le x)=(1-e^{-x/2})^2$, see Hansen (2000). From (5.4), we can see that the asymptotic distribution of $LR_{in}(r_{i0})$ is free of nuisance parameters since the errors $\{\varepsilon_t\}$ are homoskedastic on each segment by Assumption 3.1, see Theorem 2 in Hansen (2000) and the discussion thereafter. We use the distribution of Δ to solve for the critical value $c_{1-\alpha}$ (e.g., $\alpha=0.05$) and a $1-\alpha$ likelihood-ratio based confidence set for r_{i0} is given by

$$\Gamma_{1-\alpha}^i=\{r: LR_{in}(r)\le c_{1-\alpha}\}. \tag{5.5}$$

Next, we study the likelihood-ratio based confidence set for the change-point. Following Eo and Morley (2015), we define the likelihood-ratio test statistic for $H_0: k=k_0$ as follows:

$$LR_n(k)=S_n\left(\hat{\theta}_{1n}(\hat{r}_{1n},k),\hat{\theta}_{2n}(\hat{r}_{2n},k),\hat{r}_{1n},\hat{r}_{2n},k\right)-S_n\left(\hat{\theta}_{1n},\hat{\theta}_{2n},\hat{r}_{1n},\hat{r}_{2n},\hat{k}_n\right), \tag{5.6}$$

where S_n is defined in (2.4). By our estimation procedure, the second term of (5.6) is $S_n(\hat\theta_{1n}, \hat\theta_{2n}, \hat r_{1n}, \hat r_{2n}, \hat k_n) = \min_{1\le k<n} S_n(\hat\theta_{1n}, \hat\theta_{2n}, \hat r_{1n}, \hat r_{2n}, k)$. Then, under H_0, by the results in Sections 3-4 and (S3.4) in the supplementary material, it is not hard to show that

$$\begin{aligned}
LR_n(k_0) &= \max_{1\le k<n}\left[S_n\left(\hat\theta_{1n}(\hat r_{1n}, k_0), \hat\theta_{2n}(\hat r_{2n}, k_0), \hat r_{1n}, \hat r_{2n}, k_0\right) - S_n\left(\hat\theta_{1n}, \hat\theta_{2n}, \hat r_{1n}, \hat r_{2n}, k\right)\right] \\
&= \max_{1\le k<n}\left\{ I(k<k_0)\sum_{t=k+1}^{k_0}[\ell_t(\theta_{10}, r_{10}) - \ell_t(\theta_{20}, r_{20})] \right. \\
&\qquad \left. + I(k\ge k_0)\sum_{t=k_0+1}^{k}[\ell_t(\theta_{20}, r_{20}) - \ell_t(\theta_{10}, r_{10})]\right\} + o_p(1) \\
&\triangleq \max_{1\le k<n}\{-W_n(k)\} + o_p(1). \qquad (5.7)
\end{aligned}$$

Without loss of generality, we assume $r_{20} > r_{10}$ and we will see that this assumption will not affect our limiting theory later. By a similar argument as that in Section 4.2, we use $\bar W_n(k)$ in (4.9) to approximate $W_n(k)$ in (5.7). We only consider the case when $k < k_0$ in (4.9) since the other case is similar. By (4.10) and Assumption 4.2, we first define

$$\bar A_t = c_3' Z_{t-1} I(q_{t-1} > r_{20}) + c_4' Z_{t-1} I(q_{t-1} \le r_{10}) + c_5' Z_{t-1} I(r_{10} < q_{t-1} \le r_{20}) \qquad (5.8)$$

and hence $\bar{A}_{1t} = n^{-\beta}\bar{A}_t$. Under Assumption 4.2, we have

$$\begin{aligned}
\max_{1\le k<n} \sum_{t=k+1}^{k_0} (-\bar{A}_{1t}^2 - 2\bar{A}_{1t}\varepsilon_t) &= \max_{s\in R} \sum_{t=k_0+[n^{2\beta}s]+1}^{k_0} (-\bar{A}_{1t}^2 - 2\bar{A}_{1t}\varepsilon_t) \\
&\Rightarrow \max_{s\in R}[2\sqrt{E\bar{A}_t^2\sigma_{10}^2}B(s) - E\bar{A}_t^2|s|] \\
&=_d \max_{v\in R}\{\sigma_{10}^2[2B(v) - |v|]\},
\end{aligned} \tag{5.9}$$

where we make a change of variables $s = v\sigma_{10}^2/E\bar{A}_t^2$ in the last step of (5.9). Then, we conclude that

$$\max_{1\le k<n}\{-\bar{W}_n(k)\} \Rightarrow \tilde{\Delta} = \max_v \begin{cases} \sigma_{10}^2[2B(v) - |v|] \text{ for } v \in (-\infty, 0), \\ \sigma_{20}^2[2B(v) - |v|] \text{ for } v \in [0, \infty), \end{cases} \tag{5.10}$$

and

$$LR_n(k_0) \approx_d \tilde{\Delta}, \tag{5.11}$$

where the distribution function of $\tilde{\Delta}$ is

$$P(\tilde{\Delta} \le x) = \left(1 - \exp(-\frac{x}{2\sigma_{10}^2})\right)\left(1 - \exp(-\frac{x}{2\sigma_{20}^2})\right), \tag{5.12}$$

see, for example, Eo and Morley (2015). Then, we use (5.12) to solve for the critical value $\tilde{c}_{1-\alpha}$

and a $1-\alpha$ likelihood-ratio based confidence set for k_0 is given by

$$C_{1-\alpha} = \{k : LR_n(k) \leq \tilde{c}_{1-\alpha}\}. \tag{5.13}$$

The different scaling factors σ_{10}^2 and σ_{20}^2 generally make the distribution of (5.12) to be asymmetric. In practice, σ_{10}^2 and σ_{20}^2 are replaced by consistent estimates from (3.1), and the calculation of a critical value using (5.12) is straightforward.

6 Simulation Study

This section examines the performance of our asymptotic results in finite samples via Monte Carlo experiments and all the tables are placed in the supplement. We use sample sizes $n = 400$, 800 and 1200 with true change-points $k_0 = 200, 400$ and 600, respectively. The data is generated from the following TAR(1) model with a change-point k:

$$y_t = \begin{cases} (\mu_1 + \phi_1 y_{t-1}) I(y_{t-1} > r_1) \\ +(\nu_1 + \psi_1 y_{t-1}) I(y_{t-1} \leq r_1) + \varepsilon_t, \text{ if } t \leq k, \\ (\mu_2 + \phi_2 y_{t-1}) I(y_{t-1} > r_2) \\ +(\nu_2 + \psi_2 y_{t-1}) I(y_{t-1} \leq r_2) + \varepsilon_t, \text{ if } t > k, \end{cases} \tag{6.1}$$

with the true values given as follows:

$$(\theta_{10}', r_{10}) = (-1, -0.6, 1, 0.4, 0.8) - \gamma(-1, -1, 1, 1, 0),$$

$$(\theta_{20}', r_{20}) = (-0.8, -0.9, 0.7, 0.6, 0.5) - \gamma(-1, -1, 1, 1, 0),$$

and $\gamma = 0$, 0.2 and 0.4, respectively, where $\varepsilon_t \sim N(0,1)$, i.e., $\sigma_{10} = \sigma_{20} = 1$. Clearly, the autoregressive functions are not continuous over two thresholds $\{0.8, 0.5\}$ in all cases. We use 1000 replications for each case. Table S1 summarizes the averages of the bias, the empirical standard deviation (ESD), the asymptotic standard deviation (ASD) and the estimated asymptotic standard deviation (EASD) when $\gamma = 0$. The results are similar for the other cases and hence are not reported. Here, the ASDs of $\hat{\theta}_{in}$ are computed using the true $\sigma_{i0}^2 \Sigma_i$ in Theorem 3.2, where Σ_i is calculated by (3.2) and (3.3) using true thresholds r_{i0} and change-point k_0, the ASDs of $\hat{r}_{in}$ are obtained by the simulation method in section 4 of Li and Ling (2012) with true τ_0 in Theorem 3.3, the EASDs of of $\hat{\theta}_{in}$ are computed using the estimtors in (3.1)-(3.3) and the EASDs of $\hat{r}_{in}$ are obtained by the simulation method of Li and Ling (2012) with replacing τ_0 by $\hat{\tau}_n$ in Theorem 3.3. From Table S1, we can see that the bias is very close to 0 for large n while it is not strictly decreasing as n increases since the average of the empirical bias also depends the variance of the estimator. In addition, the larger the sample size is, the closer the ESDs, ASDs and EASDs are on the whole. We also see that the convergence rate of the thresholds is n, for example, the ESDs of $\hat{r}_{in}$ in sample of size 800 are half of those in sample of size 400, the ESDs of $\hat{r}_{in}$ in sample of size 1200 are one third of those in sample of size 400, $i = 1$, 2. Similarly, we can see that the convergence rates of other parameters are lower than those of the thresholds. These findings are similar to those reported in Li and Ling (2012). Furthermore, Table S1 shows that all the estimated thresholds have a negative bias. This is because we took the left end-point of the interval on which (2.4) achieves its minimum. This negative bias issue can be overcome by using its middle-point, see Yu (2012) and Yu (2015). We do not pursue this issue here since the bias is negligible for large sample size and we also use the left end-point $M_{-}^{(i)}$ to make statistical inferences in Theorem 3.3 and the simulation studies below.

We now study the coverage probabilities of r_{i0} and the performance of the approximating distributions and the likelihood-ratio based confidence sets in Section 4.1 and Section 5, respectively. To do this, we first simulate the quantiles of $M_{-}^{(1)}$ and $M_{-}^{(2)}$ with 10000 replications. Based on these quantiles and the corresponding ones in Table 1 of Hansen (1997), the coverage probabilities of r_{i0} are reported in Table S2 when $\gamma = 0$, 0.2 and 0.4, respectively, where $\|\delta_{in}\| = \|\Psi_{i0} - \Phi_{i0}\|$, $i = 1, 2$. It can be seen that the coverage probabilities based on $M_{-}^{(1)}$ and $M_{-}^{(2)}$ are rather accurate in all cases on the whole and it tends to undercover when the threshold effect $\|\delta_{in}\|$ is small. The coverage probabilities based on the quantiles of T are often relatively worse than the other two methods, especially when $\gamma = 0$ (i.e. $\|\delta_{in}\|$ is very large), but their accuracy improves when $\|\delta_{in}\|$ becomes smaller. The coverage probabilities based on LR_{in} tend to exceed the nominal levels for large threshold effects $\|\delta_{in}\|$ and they will decrease as $\|\delta_{in}\|$ becomes smaller; similar results can be found in Hansen (2000), in which he also found that the likelihood-ratio based method may undercover for very small threshold effects and small sample sizes. Overall, when the threshold effect is large, the simulation method of Li and Ling (2012) is rather accurate, the approximating method in Section 4.1 will undercover and the likelihood-ratio based method will overcover and is quite conservative. When the threshold effect is small, the simulation method of Li and Ling (2012) will undercover and the methods based on the approximation and the likelihood-ratio may be more accurate. In practice, we do not know the exact magnitude of the threshold effect, we recommend to use the likelihood-ratio method and the simulation method to make statistical inferences since they are generally more accurate.

We next study the performance of the approximating distribution in Theorem 4.2 and the likelihood-ratio based confidence sets in Section 5 for the estimated change-points in finite samples. We will cover both large and small structural change effects in our experiments. To see the finite sample performance when the structural change effect is relatively large, we first fix the true parameter $(\theta_{10}', r_{10}) = (-1, -0.6, 1, 0.4, 0.8)$ and let the true parameter

$$(\theta_{20}', r_{20}) = (-\theta_{10}', r_{10}) - \tilde{\gamma}_1(1, 1, -1, -1, 1), \tag{6.2}$$

with $\tilde{\gamma}_1 = 0,\ 0.1, 0.2$ and 0.4, respectively. It is easy to see that the structural change effects are relatively large for the choices of $\tilde{\gamma}_1$ since θ_{10} and θ_{20} are two vectors of different signs. When $\tilde{\gamma}_1$ increases, the structural change effect will decrease. The number of replications is 1000 for each case in this experiment. Table S3 summarizes the mean, ESD, ASD and the estimators of d_{i0}, ϕ and ξ for (6.2). Here, $\kappa_n = \|\Phi_{20} - \Phi_{10}\| + \|\Psi_{20} - \Phi_{10}\| + \|\Psi_{10} - \Phi_{20}\|$ which represents the structural change effect, ESD is calculated based on the 1000 estimators, ASD is computed using the approximating distribution in (4.11) and $\hat{d}_{in}$ is based on (4.13). From the expressions of $\hat{\phi}_n$ and $\hat{\xi}_n$ in (4.14), our simulation results show that $\hat{\phi}_n$ and $\hat{\xi}_n$ are almost the same since we assume $\sigma_{10} = \sigma_{20} = 1$ in our experiment. We can also use the true parameters and (4.12) to obtain d_{i0}, ϕ and ξ, the results are similar to those in Table S3 and hence are not reported here. For the purpose of reference, we only report the values of $(\hat{d}_{10}, \hat{d}_{20})$ corresponding to different $\tilde{\gamma}_1$, which are calculated using (4.12) with the true parameters and the sample estimates of $M_i(r_{i0}^{\pm})$. From Table S3, we can see that the means of the estimated change-points are close to the true change-points in all cases. The ASDs are smaller then the ESDs in all cases and they tend to be closer as the structural change effect decreases. This is reasonable since our approximating distribution is based on a small change effect. We can also see that the estimated d_{i0}, ϕ and ξ are almost the same for different sample sizes with fixed $\tilde{\gamma}_1$.

Based on the results in Table S3 and the density function of $T_{\phi,\xi}$ in Bai (1997), we study the coverage probabilities of the estimated change-points. The results are reported in Table S4. From Table S4, we can see that the likelihood-ratio based confidence sets of LR_n overcover at all the three nominal levels, while the approximating distribution $T_{\phi,\xi}$ considerably undercovers for relatively large structural change effects in (6.2). These findings are similar to those in Eo and Moley (2015), where they found that the coverage rate based on likelihood-ratio approach is more precise than those of Bai (1997) and Elliott and Müller (2007).

Now, we study the finite sample performance when the structural change effect is relatively small. We set (θ_{10}', r_{10}) as before and let the true parameter

$$(\theta_{20}', r_{20}) = (\theta_{10}', r_{10}) - \tilde{\gamma}_2(-1, -1, 1, 1, 1), \tag{6.3}$$

with $\tilde{\gamma}_2 = 0.1$, 0.2, 0.3 and 0.5, respectively. It is easy to see that the structural change effects are relatively small for the choices of $\tilde{\gamma}_2$. When $\tilde{\gamma}_2$ increases, the structural change effect will increase. Table S5 summarizes the results as those in Table S3. From Table S5, we have similar findings as Table S3 and the ESDs and ASDs are larger than those reported in Table S3. This is reasonable since it is not easy to locate the change-point when the structural change effect is small. Based on the estimators in Table S5 and the density function of $T_{\phi,\xi}$ in (1997), we study the coverage probabilities of the estimated change-points. The results are reported in Table S6. From Table S6, we can see that the likelihood-ratio based confidence sets of LR_n undercover at all the three nominal levels, while the approximating distribution $T_{\phi,\xi}$ performs better than the likelihood-ratio based method for relatively small structural change effects in (6.3).

Overall, when the structural change effect is large, the likelihood-ratio based confidence sets of LR_n will overcover and are somewhat conservative. When the structural change effect is very small, the approximating method in Section 5.2 may be more accurate. In practice, it is not easy to evaluate the magnitude of the change especially when the two thresholds are different. According to our limited simulation experience, we recommend to use the likelihood-ratio based method since the performance is better in general.

7 A Real Data Example

Yau et al. (2015) applied the TAR model with structural breaks to an U.S. GNP data and found some breaks associated with substantial changes in the U.S. economy. This section uses model

(2.1) with a structural change to study a long time series of annual tree ring width (Figure S1(a)), with the measurements taken from a Qilian Juniper tree in the northeastern Tibetan Plateau of China. The time series spans the period from 1079 to 2009 and it was obtained from the NOAA paleoclimatology database at `https://www.ncdc.noaa.gov/paleo/study/16645`. Tree rings provide one of the most important records of the past climates and hence they become an important sources to study climate change, see Cook (1985). In the past decades, TAR models have been recognized as an important nonlinear time series model to study climate changes, see Ellis and Post (2004) and Tong (2011) for demonstrations of the merits of using TAR models rather than linear AR ones.

Let x_t denote the original data and y_t denote its continuously annualized average growth rate, i.e., $y_t = \log(x_t/x_{t-1})$. Figure S1(b) is the time plot of y_t. We can see large fluctuations in Figure S1(a)-(b). Figure S2 is the autocorrelation function (ACF) and partial ACF (PACF) of $\{y_t\}$, which indicate that $\{y_t\}$ is a sequence of dependent time series. We now illustrate the procedure of building a TAR model with a structural change to $\{y_t\}$ as follows.

Step 1. We first perform the threshold nonlinearity test. There are many ways to do this in the literature, see Tsay (1989) and Chan (1991), among others. Here we adopt the likelihood-ratio approach of Chan (1991) and use the corresponding package `TSA` in `R` , see also Cryer and Chan (2008) for details. For each null model AR(p), we choose the threshold variable $q_{t-1} = y_{t-d}$ with $1 \leq d \leq p$. Table S7 reports the $p-$values when performing the threshold nonlinearity test under some possible linear AR models. The choice of the AR order p is based on the PACF in Figure S2(b), as suggested by Tsay (1989) when testing for threshold nonlinearity. From Table

S7, we can see that most of the $p-$values are close to 0 except for some cases when $d = 2$, in which the $p-$values are only slightly larger than 5%, we conclude that most likely there is a threshold effect in the data $\{y_t\}$. In other words, it is better to use a threshold model to fit the data than a pure AR model.

Step 2. We now fit a TAR(p) model to $\{y_t\}$ with a threshold y_{t-d}, where $1 \leq d \leq p$. Define

$$AIC(p,d) = n\log(\hat{\sigma}_n^2) + 2(p+1) \quad \text{and} \quad BIC(p,d) = n\log(\hat{\sigma}_n^2) + (p+1)\log(n), \quad (7.1)$$

where $\hat{\sigma}_n^2$ is similarly defined as (3.1) using the whole sample. (7.1) is slightly different from that in Li and Ling (2012) since we do not allow a threshold effect in the error term of (2.1). To make the model simple, according to the PACF in Figure S2(b), we set $1 \leq p \leq 12$ and $1 \leq d \leq p$ and the results of the AICs and BICs are summarized in Table S8. From Table S8, we can see that the AIC selects the model TAR(12) with $d = 10$ and BIC selects TAR(8) with $d = 8$. For ease of exposition, we choose the simpler model TAR(8) by BIC and the fitted model is as follows:

$$\begin{aligned} y_t &= (\mu + \sum_{i=1}^{8} \phi_i y_{t-i}) I(y_{t-8} > 0.1252) \\ &\quad + (\nu + \sum_{i=1}^{8} \psi_i y_{t-i}) I(y_{t-8} \leq 0.1252) + 0.247 u_t, \end{aligned} \quad (7.2)$$

where the standard deviaton 0.247 is calculated by the residual sum of squares and the other parameters are summarized in Table S9. We now use the portmanteau test in (15.8.3) of Cryer and Chan (2008) (pp. 412) to check if model (7.2) is adequate. Figure S3 displays the $p-$values of the test and it shows that the $p-$values are all larger than 5%, which means that model (7.2) is adequate for data $\{y_t\}$.

Step 3. We then use the Sup-likelihood-ratio test statistic in Andrews (1993) (see also Davis et al. (1995)), $\sup_{\tau\in[0.05,0.95]} LR(\tau)$, to test whether or not there exists a structural change in model (7.2). Since the estimate of the threshold is super-efficient with convergence rate n under the null hypothesis of no change-point, the limiting distribution in Andrews (1993) or Davis et al. (1995) is still applicable to TAR(p) model with degrees of freedom $2(p+1)$. It turns out that $\sup_{\tau\in[0.1,0.9]} LR(\tau) = 89.77$ which exceeds the critical value 46.69 at the 0.01 significance level, see Table 1 in Andrews (1993). Hence, model (7.2) most likely has a structural change during this period. This doesn't contradict the finding in Step 2 that model (7.2) is adequate for the data since likelihood-ratio test is a different criterion for model selection in this step.

Step 4. Based on the estimation procedure in Section 2, a TAR(8) model with $d = 8$ and a structural change is used to fit the data. The result is as follows:

$$y_t = \begin{cases} (\mu_1 + \sum\limits_{i=1}^{8} \phi_{1i} y_{t-i}) I(y_{t-8} > -0.1744) \\ +(\nu_1 + \sum\limits_{i=1}^{8} \psi_{1i} y_{t-i}) I(y_{t-8} \le -0.1744) + 0.241 u_t, \quad t \le 578 \\ (\mu_2 + \sum\limits_{i=1}^{8} \phi_{2i} y_{t-i}) I(y_{t-8} > 0.0972) \\ +(\nu_2 + \sum\limits_{i=1}^{8} \psi_{2i} y_{t-i}) I(y_{t-8} \le 0.0972) + 0.232 u_t, \quad t > 578, \end{cases} \tag{7.3}$$

where the two standard deviations 0.241 and 0.232 are calculated by (3.1) and the other coefficients are reported in Table S10. The standard deviations are given in the parentheses and some parameters are not significant at 5% level. We use the method of Cryer and Chan (2008) (pp. 412) again for model diagnostic checking. Figure S4 displays the $p-$values of the test in two segments of (7.3) and shows that model (7.3) is also adequate for $\{y_t\}$. We note that there appears to be at least two change-points in the data from Figure S1(b), but our method finds the most visually obvious one. For multiple change-point cases, interested readers may consult the approach provided in Section 2 on page 9.

From the model, we can see that almost all the AR coefficients are negative. This is reasonable because a higher growth rate of the current year will result in a lower growth rate in the next year, and vice versa. Furthermore, the AR coefficients before and after the change-point $\hat{k}_n = 578$ change drastically and almost all their absolute values after the change-point are larger than the counterparts before the change-point, which means the dependence of the growth rates grows stronger after the change-point. Based on the methods of Li and Ling (2012), the approximation in Section 4.1 and the likelihood-ratio one in Section 5, the 95% confidence intervals for r_{10} are $[-0.189, -0.166]$, $[-0.213, -0.135]$ and $[-0.174, -0.014]$, respectively, and those of r_{20} are $[0.084, 0.119]$, $[-0.179, 0.374]$ and $[-0.028, 0.301]$, respectively. We can see that the simulation method provides rather tight confidence intervals for the thresholds, while other two methods tend to give wider confidence intervals. The likelihood-ratio based method gives similar intervals as the simulation one for r_{10} while it provides a much wider interval for r_{20}. According to our simulation experience, we suggest to use the intervals produced by the simulation based method since they are pretty tight in this case. The 95% confidence interval of k_0 based on the approximation method is $[563, 585]$. This tight interval indicates the estimator $\hat{k}_n$ is very accurate.

After checking with the data, it is found that $\hat{k}_n$ represents the year 1656, and the 95%-confidence intervals show that there was most likely a great climate change around $1641 - 1663$. From Figure S1(a), we can see that the growth of tree rings declined from the 1600s onwards, which indicates that the temperatures might have been changing rapidly around this period. Based on the historical records, there were a lot of disasters such as dry weather and crop failures in this period and the Ming dynasty collapsed in the year 1644. There was not a climate record during this period in China. A lot of historians suspect that the bad weather is because of the climate change. Our findings provide an evidence to support this view and may be useful for future study of Chinese history.

In the supplementary material, we further examined some steps in this section and demonstrated the merits of model (7.3) with a change-point through the forecasting errors. See Section S5 in the supplement for details.

Supplementary Material

Owing to space constraint, a new SLLN, the proofs of the theorems above, the effect of the initial values and some tables and figures are given in the supplementary material.

Acknowledgments

We are grateful to the Editors and the anonymous referees for their insightful comments and suggestions that have substantially improved the presentation and the content of this paper. This project is supported by Hong Kong Research Grants Commission Grants GRF 16500915, GRF 16307516, GRF 16500117, and the fifth batch of excellent talent support program for Chongqing Colleges and Universities.

References

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 821–856.

Aue, A., Hörmann, S., Horváth, L. and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics 37*(6B), 4046–4087.

Bai, J. (1994). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis 15*(5), 453–472.

Bai, J. (1995). Least absolute deviation estimation of a shift. *Econometric Theory 11*(03), 403–436.

Bai, J. (1997). Estimation of a change point in multiple regression models. *Review of Economics and Statistics 79*(4), 551–563.

Bai, J., Lumsdaine, R. L. and Stock J. H.(1998). Testing for and dating common breaks in multivariate time series. *The Review of Economic Studies 65*(3), 395–432.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 47–78.

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

Chan, K.-S. (1989). A note on the geometric ergodicity of a Markov chain. *Advances in Applied Probability 21*, 702–704.

Chan, K. S. (1991). Percentage points of likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society. Series B 53*, 691–696.

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

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

Chan, K.-S. and Tsay, R. S. (1998). Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika 85*(2), 413–426.

Chan, N. H., and Kutoyants, Y. A. (2010). Recent developments of threshold estimation for nonlinear time series. *Journal of the Japan Statistical Society 40*(2), 277–303.

Chan, N. H., Yau, C. Y. and Zhang, R.-M. (2014). Group lasso for structural break time series. *Journal of the American Statistical Association 109*(506), 590–599.

Chan, N. H., Yau, C. Y. and Zhang, R.-M. (2015). LASSO estimation of threshold autoregressive models. *Journal of Econometrics 189*(2), 285–296.

Chong, T. T.-L. (2001). Structural change in AR(1) models. *Econometric Theory 17*(01), 87–155.

Cook, E. (1985). *A time series analysis approach to tree ring standardization*. Thesis. The University of Arizona.

Cryer, D. and Chan,K. S. (2008). *Time Series Analysis With Applications in R*. Springer.

Csörgö, M. and Horváth, L. (1997). *Limit theorems in change-point analysis*, Volume 18. John Wiley & Sons Inc.

Davis, R. A., Huang, D. and Yao, Y.-C. (1995). Testing for a change in the parameter values and order of an autoregressive model. *The Annals of Statistics 23*(1), 282–304.

Davis, R. A., Lee, T. C. M. and Rodriguez-Yam, G. A. (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association 101*(473), 223–239.

Elliott, G. and Müller, U. K. (2007). Confidence sets for the date of a single break in linear time series regressions. *Journal of Econometrics 141*(2), 1196–1218.

Elliott, G. , Müller, U. K. and Watson, M. W. (2015). Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica 83*(2), 771–811.

Ellis, A. M. and Post, E. (2004). Population response to climate change: linear vs. non-linear modeling approaches. *BMC ecology 4*(1), 2.

Eo, Y. and Morley, J. (2015). Likelihood-ratio-based confidence sets for the timing of structural breaks. *Quantitative Economics 6*(2), 463–497.

Gao, J., Tjøstheim, D. and Yin, J. (2013). Estimation in threshold autoregressive models with a stationary and a unit root regime. *Journal of Econometrics 172*(1), 1–13.

Gao, Z., Ling, S. and Tong, H. (2017). Tests for TAR models vs. STAR models–a separate family of hypotheses approach. *To appear in Statistica Sinica*.

Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica*, 1–118.

Hansen, B. E. (1997). Inference in tar models. *Studies in Nonlinear Dynamics & Econometrics 2*(1), 119–131.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica 68*(3), 575–603.

Hidalgo, J. and Seo, M. H. (2013). Testing for structural stability in the whole sample. *Journal of Econometrics 175*(2), 84–93.

Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika 57*(1), 1–17.

Kejriwal, M. and Perron. P. (2008). The limit distribution of the estimates in cointegrated regression models with multiple structural changes. *Journal of Econometrics 146*(1), 59–73.

Li, D. and Ling, S. (2012). On the least squares estimation of multiple-regime threshold autoregressive models. *Journal of Econometrics 167*(1), 240–253.

Li, D., Ling, S. and Li, W. K. (2013). Asymptotic theory on the least squares estimation of threshold moving-average models. *Econometric Theory 29*(03), 482–516.

Li, D., Ling, S. and Zhang R. M. (2016). On a threshold double autoregressive model. *Journal of Business & Economic Statistics 34*(1), 68–80.

Ling, S. (2007). Testing for change points in time series models and limiting theorems for ned sequences. *The Annals of Statistics 35*(3), 1213–1237.

Ling, S. (2016). Estimation of change-points in linear and nonlinear time series models. *Econometric Theory 32*, 402–430.

Liu, W., Ling, S. and Shao, Q.-M. (2011). On non-stationary threshold autoregressive models. *Bernoulli 17*(3), 969–986.

Meyn, S. P. and Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability*. Cambridge Univ Press.

Picard, D. (1985). Testing and estimating change-points in time series. *Advances in Applied Probability*, 841–867.

Qian, J. and Su, L. (2014). Structural change estimation in time series regressions with endogenous variables. *Economics Letters 125*(3), 415–421.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America 42*(1), 43.

Saikkonen, P., Lütkepohl, H. and Trenkler, C. (2006). Break date estimation for var processes with level shift with an application to cointegration testing. *Econometric Theory 22*(01), 15–68.

Seo, M. H. and Linton, O. (2007). A smoothed least squares estimator for threshold regression models. *Journal of Econometrics 141*(2), 704–735.

Shao, X. and Zhang, X. (2012). Testing for change points in time series. *Journal of the American Statistical Association 105*, 1228–1240.

Tong, H. (1978). *On a threshold model. In Pattern Recognition and Signal Processing (C. H. Chen, ed.) 101–141*. Sijthoff & Noordhoff.

Tong, H. and Lim, K.S. (1980). Threshold autoregression, limit cycles and cyclical data-with discussion. *Journal of the Royal Statistical Society, Series B,* 42(3), 245–292.

Tong, H. (2011). Threshold models in time series analysis–30 years on. *Statistics and its Interface 4*(2), 107–118.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association 84*(405), 231–240.

Yao, Y.-C. (1987). Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *The Annals of Statistics*, 1321–1328.

Yau, C. Y., Tang, C. M. and Lee, T. C. (2015). Estimation of multiple-regime threshold autoregressive models with structural breaks. *Journal of the American Statistical Association 110*(511), 1175–1186.

Yu, P. (2012). Likelihood estimation and inference in threshold regression. *Journal of Econometrics 167*(1), 274–294.

Yu, P. (2015). Adaptive estimation of the threshold point in threshold regression. *Journal of Econometrics 189*(1), 83–100.

Booth School of Business, University of Chicago

E-mail: mazxgao@gmail.com

Department of Mathematics, Hong Kong University of Science and Technology

E-mail: maling@ust.hk