Statistica Sinica: Supplement

Supplementary Material for "Statistical Inference for Structurally Changed Threshold Autoregressive Models"

Zhaoxing Gao[1] and Shiqing Ling[2]

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

Supplementary Material

In this supplementary material, we provide all the proofs of the main article in Sections S1-S3. The effect of the initial values is discussed in Section S4. A further analysis of the real data example is given in Section S5. Some tables and figures of the simulation study and the real data analysis of the main article are also presented in this supplement.

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, $\Longrightarrow$ denotes the weak convergence for sequences of (measurable) random elements of a space of bounded Euclidean-valued càdlàg functions on a compact set, $\triangleq$ means " is defined as", and $\longrightarrow_{\mathcal{L}}$ denotes convergence in distribution.

S1 A Strong Law of Large Numbers

The SLLN plays an important role in statistics. The existing theorems for the forward sums of dependent random sequences, such as those in Stout (1974) (

e.g. Theorem 3.7.6) and Hall and Heyde (2014) (e.g. Theorem 2.20), require at least a finite second moment or do not have a rate of convergence. Shao (1995) established a maximal inequality of partial sums of $\rho-$mixing sequences under a finite second moment, which can be applied for the SLLN. Wu (2007) established the strong invariance principles for stationary and ergodic Markov chains, in which the SLLN requires a finite $(1+\epsilon)$th moment with other conditions. These conditions essentially are the NED in Ling (2007). It is not clear if the process from the TAR model is NED yet and additional conditions may be needed even if it is. In the proof of Theorem 3.1, it involves the convergence of the following partial sum:

$$\frac{1}{n}\sum_{t=-n}^{-1}[\ell_t(\theta,r)-E\ell_t(\theta,r)],$$

But the usual ergodic theorem cannot apply to this, since very few time series have been known to be time-reversible, see Cheng (1999). On the other hand, the weakest moment condition is given in the ergodic theorem, but it does not have a rate of convergence. Furthermore, since the TAR model may not be the near-epoch dependence (NED) sequences without more restrictions, the strong law of large numbers (SLLN) for backward-sum in Ling (2007) cannot be applied either. It demands a new SLLN for the backward-sum of a strong mixing random sequence. In view of the moment condition in Assumption 3.1, we establish a SLLN for the forward and backward sums of $\alpha-$mixing sequences under a finite $(1+\epsilon)$th moment.

Theorem 1. *Let* $\{X_t : t = 0, \pm 1, \pm 2, \cdots\}$ *be a stationary process with zero mean and* $E|X_t|^p < \infty$ *for some* $1 < p < 2$*. If it is a strong mixing sequence with* $\alpha(n) \le K a^n$ *for some* $a \in (0,1)$ *and a constant* $K > 0$*, then there exists a constant* $\delta \in (0,1)$ *such that*

$$(a) \quad \frac{1}{n^{1-\delta}} \sum_{t=1}^{n} X_t \longrightarrow 0 \quad a.s.,$$

$$(b) \quad \frac{1}{n^{1-\delta}} \sum_{t=-n}^{-1} X_t \longrightarrow 0 \quad a.s.,$$

as $n \to \infty$.

PROOF. We only prove it for (a) since (b) is similar. We choose $\tilde{p}$ with $1 < \tilde{p} < p$ and a constant $c > 0$ which will be determined later. Denote $\xi_t = X_t I(|X_t| \le t^c)$ and $\eta_t = X_t I(|X_t| > t^c)$. By Minkowski's inequality, it follows that

$$\begin{aligned} E|\sum_{t=1}^{n} X_t|^{\tilde{p}} &= E|\sum_{t=1}^{n}(\xi_t - E\xi_t) + \sum_{t=1}^{n}(\eta_t - E\eta_t)|^{\tilde{p}} \\ &\le O(1)E|\sum_{t=1}^{n}(\xi_t - E\xi_t)|^{\tilde{p}} + O(1)E|\sum_{t=1}^{n}(\eta_t - E\eta_t)|^{\tilde{p}} \\ &:= I + II, \end{aligned}$$

where $O(1)$ holds uniformly in n.

We note that $|\xi_i| \le i^c$ and $|\xi_j| \le j^c$, $1 \le i < j \le n$. By Lemma 1.2 in Ibragimov (1962) and the strong mixing condition, we have

$$|E\xi_i\xi_j - E\xi_i E\xi_j| \le 4 i^c j^c \alpha(j-i). \qquad \text{(S1.1)}$$

By Hölder's inequality, Minkowski's inequality and (S1.1), we have

$$
\begin{aligned}
I &\leq O(1)\left[E|\sum_{t=1}^{n}(\xi_t - E\xi_t)|^2\right]^{\frac{\tilde{p}}{2}} \\
&\leq O(1)\left[\sum_{t=1}^{n}E(\xi_t - E\xi_t)^2 + 2\sum_{i<j}(E\xi_i\xi_j - E\xi_i E\xi_j)\right]^{\frac{\tilde{p}}{2}} \\
&\leq O(1)\left[\sum_{t=1}^{n}E\xi_t^2 + 2\sum_{i<j}|E\xi_i\xi_j - E\xi_i E\xi_j|\right]^{\frac{\tilde{p}}{2}} \\
&\leq O(1)\left[\sum_{t=1}^{n}t^{2c} + 8\sum_{i<j}i^c j^c \alpha(j-i)\right]^{\frac{\tilde{p}}{2}} \\
&\leq O(n^{\frac{\tilde{p}}{2}+\tilde{p}c}). \qquad \text{(S1.2)}
\end{aligned}
$$

By Hölder's inequality and Markov's inequality, we have

$$
E|\eta_t|^{\tilde{p}} \leq (E|X_t|^p)^{\tilde{p}/p}[P(|X_t| > t^c)]^{(p-\tilde{p})/p} \leq \frac{E|X_t|^p}{t^{c(p-\tilde{p})}}. \qquad \text{(S1.3)}
$$

By Minkowski's inequality, Hölder's inequality and (S1.3), we have

$$
\begin{aligned}
II &= O(1)E|\sum_{t=1}^{n}(\eta_t - E\eta_t)|^{\tilde{p}} \\
&\leq O(1)\left[\sum_{t=1}^{n}(E|\eta_t|^{\tilde{p}})^{1/\tilde{p}} + \sum_{t=1}^{n}|E\eta_t|\right]^{\tilde{p}} \\
&\leq O(1)\left[\sum_{t=1}^{n}(E|\eta_t|^{\tilde{p}})^{1/\tilde{p}}\right]^{\tilde{p}} \\
&\leq O(1)\left[\sum_{t=1}^{n}(\frac{E|X_t|^p}{t^{c(p-\tilde{p})}})^{1/\tilde{p}}\right]^{\tilde{p}} \\
&\leq O(1)\left[\sum_{t=1}^{n}\frac{t^{1+\epsilon-c(p-\tilde{p})/\tilde{p}}}{t^{1+\epsilon}}\right]^{\tilde{p}} \\
&\leq O(n^{\tilde{p}+\tilde{p}(\epsilon-c(p-\tilde{p})/\tilde{p})}), \qquad \text{(S1.4)}
\end{aligned}
$$

where we choose ϵ such that $0 < \varepsilon < c(p - \tilde{p})/\tilde{p}$.

Let $0 < c < 1/2$ and $0 < \varepsilon < c(p - \tilde{p})/\tilde{p}$. By (S1.2) and (S1.4), there exists a constant $\rho \in (0, 1)$ such that

$$E|\sum_{t=1}^{n} X_t|^{\tilde{p}} \leq O(n^{\tilde{p}\rho}). \tag{S1.5}$$

Let $S_n = \sum_{t=1}^{n} X_t$. By Proposition 1 in Wu (2007) and (S1.5), we have

$$\begin{aligned}
\left[E\left|\max_{1\leq t\leq 2^k} |S_t|\right|^{\tilde{p}}\right]^{1/\tilde{p}} &\leq \sum_{r=0}^{k}\left[\sum_{i=1}^{2^{k-r}} E|S_{2^r i} - S_{2^r(i-1)}|^{\tilde{p}}\right]^{1/\tilde{p}} \\
&\leq \sum_{r=0}^{k}\left[\sum_{i=1}^{2^{k-r}} E|\sum_{t=1}^{2^r} X_t|^{\tilde{p}}\right]^{1/\tilde{p}} \qquad \text{(by stationarity)} \\
&\leq O(1)\sum_{r=0}^{k}\left[\sum_{i=1}^{2^{k-r}} 2^{r\tilde{p}\rho}\right]^{1/\tilde{p}} \\
&\leq O(2^{k\rho}).
\end{aligned} \tag{S1.6}$$

Let $T_n = \frac{S_n}{n^{1-\delta}}$. Using (S1.6), Borel-Cantelli Lemma and Lemma 2.3.1 in Stout (1974), it follows that $T_n \longrightarrow 0$ a.s. as $n \to \infty$, which proves (a). This completes the proof of Theorem 1. □

S2 Proof of Theorems 3.1-3.2 and 4.1

We first give several useful lemmas. The proofs are given in Section S3.

Lemma 1. *If Assumptions 1-4 hold, then* $E\sup_{\theta_i\in\Theta, r_i\in\Gamma} |\ell_t(\theta_i, r_i)| < \infty$ *and*

$E\ell_t(\theta_i, r_i)$ has a unique minimizer at $\theta_i = \theta_{i0}$ and $r_i = r_{i0}$, where $i = 1$ when $1 \le t \le k_0$ and $i = 2$ when $k_0 + 1 \le t \le n$.

Lemma 2. *If Assumptions 1-4 hold, we have*

$$\begin{aligned} &(a) \quad \frac{1}{n}\max_{\Theta\times\Gamma}\left|\sum_{t=1}^{n}[\ell_t(\theta, r) - E\ell_t(\theta, r)]\right| \longrightarrow 0 \quad a.s., \\ &(b) \quad \frac{1}{n}\max_{\Theta\times\Gamma}\left|\sum_{t=-n}^{-1}[\ell_t(\theta, r) - E\ell_t(\theta, r)]\right| \longrightarrow 0 \quad a.s., \end{aligned}$$

as $n \to \infty$.

The following Lemma is from Li et al. (2013).

Lemma 3. *(i). If the density $f_\varepsilon(x)$ of ε_t is continuous and bounded, then the density $f_i(x)$ of $\{y_t\} \in Y(\theta_{i0}, r_{i0})$ is continuous and bounded for $i = 1$, 2. (ii). Under Assumption 2, there exist constants $0 < m_0 < M_0 < \infty$ such that $m_0 u \le P(r < q_t \le r + u) \le M_0 u$ for fixed $r \in R$ and any $u \in [0, 1]$, $i = 1$, 2.*

Proof of Theorem 3.1. (a). By Lemmas 1-2, and a similar proof to that of Lemma 9.1 in Ling (2016), we can show that $\hat{\tau}_n = \tau_0 + o_p(1)$. So we can assume that $\hat{k}_n, k \in [k_L, n - k_L]$, where $k_L = [n\tilde{\tau}]$, $\tilde{\tau} \in (0, 1/2)$ and $\tau_0 \in (\tilde{\tau}, 1 - \tilde{\tau})$. We prove only the case when $k \le k_0$, $r_1 \ge r_{10}$ and $r_2 \ge r_{20}$, as the other cases are similar. Denote

$$\Delta_n(\theta_1, \theta_2, r_1, r_2, k) = S_n(\theta_1, \theta_2, r_1, r_2, k) - S_n(\theta_{10}, \theta_{20}, r_{10}, r_{20}, k_0).$$

We use the convention $\sum_{k_0+1}^{k_0} X_t = 0$ for any series X_t. When $k \le k_0$, we have

$$\begin{aligned}\Delta_n(\theta_1,\theta_2,r_1,r_2,k) =& \sum_{t=1}^{k}\{\ell_t(\theta_1,r_1)-\ell_t(\theta_{10},r_{10})\} + \sum_{k_0+1}^{n}\{\ell_t(\theta_2,r_2)-\ell_t(\theta_{20},r_{20})\}\\ &+\sum_{k+1}^{k_0}\{\ell_t(\theta_2,r_2)-\ell_t(\theta_{10},r_{10})\}. \qquad \text{(S2.1)}\end{aligned}$$

Let $\Theta_{1\delta} = \{(\theta_1', r_1)', \|\theta_1-\theta_{10}\| + |r_1 - r_{10}| \ge \delta\}$. By Lemma 1, $C = \min_{\Theta_{1\delta}}[E\ell_t(\theta_1,r_1) - E\ell_t(\theta_{10},r_{10})] > 0$ when $t \le k_0$. Thus, by Lemma 2 and Lemma 1 in Chow (1978) (pp. 66), we have

$$\begin{aligned}&\frac{1}{n}\min_{k_L\le k\le k_0}\min_{\Theta_{1\delta}}\left[\sum_{t=1}^{k}\{\ell_t(\theta_1,r_1)-\ell_t(\theta_{10},r_{10})\}\right]\\ \ge& -\frac{2}{n}\max_{k_L\le k\le k_0}\max_{\Theta_{1\delta}}\left|\sum_{t=1}^{k}[\ell_t(\theta_1,r_1)-E\ell_t(\theta_1,r_1)]\right|\\ &+\tilde{\tau}\min_{\Theta_{1\delta}}[E\ell_t(\theta_1,r_1)-E\ell_t(\theta_{10},r_{10})]\\ =&\tilde{\tau}C + o_p(1). \qquad \text{(S2.2)}\end{aligned}$$

Since $\min_{\theta_2\in\Theta, r_2\in\Gamma}[E\ell_t(\theta_2,r_2)-E\ell_t(\theta_{20},r_{20})] = 0$ when $t > k_0$, by Lemma 2, it follows that

$$\begin{aligned}&\frac{1}{n}\min_{\theta_2\in\Theta,r_2\in\Gamma}\left[\sum_{t=k_0+1}^{n}\{\ell_t(\theta_2,r_2)-\ell_t(\theta_{20},r_{20})\}\right]\\ \ge& -\frac{2}{n}\max_{k_L\le k\le k_0}\max_{\theta_2\in\Theta,r_2\in\Gamma}\left|\sum_{t=k_0+1}^{n}[\ell_t(\theta_2,r_2)-E\ell_t(\theta_2,r_2)]\right|\\ &+\frac{n-k_0}{n}\min_{\theta_2\in\Theta,r_2\in\Gamma}[E\ell_t(\theta_2,r_2)-E\ell_t(\theta_{20},r_{20})]\\ =&o_p(1). \qquad \text{(S2.3)}\end{aligned}$$

Note that $\min_{\theta\in\Theta,r\in\Gamma}[E\ell_t(\theta,r)-E\ell_t(\theta_{10},r_{10})]=0$ when $t\leq k_0$. We have

$$\begin{aligned}
&\frac{1}{n}\min_{k-k_0\leq -M}\inf_{\theta_2\in\Theta,r_2\in\Gamma}\left[\sum_{t=k+1}^{k_0}\{\ell_t(\theta_2,r_2)-\ell_t(\theta_{10},r_{10})\}\right]\\
\geq&-\frac{2}{n}\max_{k-k_0\leq -M}\sup_{\theta\in\Theta,r\in\Gamma}\left|\sum_{t=k+1}^{k_0}\{\ell_t(\theta,r)-E\ell_t(\theta,r)\}\right|\\
&+\frac{k_0-k}{n}\min_{\theta\in\Theta,r\in\Gamma}[E\ell_t(\theta,r)-E\ell_t(\theta_{10},r_{10})]\\
&\geq -2\max_{k-k_0\leq -M}\sup_{\theta\in\Theta,r\in\Gamma}\frac{1}{k_0-k}\left|\sum_{t=k+1}^{k_0}\{\ell_t(\theta,r)-E\ell_t(\theta,r)\}\right|\\
=_d&-2\max_{u\geq M}\sup_{\theta\in\Theta,r\in\Gamma}\frac{1}{u}\left|\sum_{t=-u}^{-1}\{\ell_t(\theta,r)-E\ell_t(\theta,r)\}\right|\\
=&o_{pM}(1), \qquad \text{(S2.4)}
\end{aligned}$$

where $o_{pM}(1)\to 0$ in probability when $M\to\infty$ and it holds uniformly in n, where the last step holds by Lemma 2, and Lemma 1 in Chow (1978) (pp. 66), and " $=_d$ " denotes " $=$ " in distribution. On the event $\{\|\hat\theta_{1n}-\theta_{10}\|+|\hat r_{1n}-r_{10}|\geq\delta,|\hat k_n-k_0|>M\}$, by (S2.2)-(S2.4), we have

$$\begin{aligned}
&P(\|\hat\theta_{1n}-\theta_{10}\|+|\hat r_{1n}-r_{10}|\geq\delta,|\hat k_n-k_0|>M)\\
&\leq P\Big(\frac{1}{n}\min_{|k_0-k|>M}\min_{\substack{(\theta_1',r_1)'\in\Theta_{1\delta}\\(\theta_2',r_2)'\in\Theta\times\Gamma}}\Delta_n(\theta_1,\theta_2,r_1,r_2,k)\leq 0\Big)\\
&\leq P(\tilde\tau C+o_{pM}(1)+o_p(1)\leq 0)\to 0, \qquad \text{(S2.5)}
\end{aligned}$$

as $M,n\to\infty$. When $|k_0-k|\leq M$, the third term of (S2.1) is larger than

$$-2\sum_{k+1}^{k_0}\max_{\theta_1\in\Theta,r_1\in\Gamma}|\ell_t(\theta_1,r_1)|=o_p(n). \qquad \text{(S2.6)}$$

On the event $\{\|\hat{\theta}_{1n}-\theta_{10}\|+|\hat{r}_{1n}-r_{10}|\geq\delta, |\hat{k}_n-k_0|\leq M\}$, by (S2.2)-(S2.3) and (S2.6), we have

$$\begin{aligned}
&P(\|\hat{\theta}_{1n}-\theta_{10}\|+|\hat{r}_{1n}-r_{10}|\geq\delta, |\hat{k}_n-k_0|\leq M)\\
&\leq P(\frac{1}{n}\min_{|k_0-k|\leq M}\min_{\substack{(\theta_1',r_1)'\in\Theta_{1\delta}\\(\theta_2',r_2)'\in\Theta\times\Gamma}}\Delta_n(\theta_1,\theta_2,r_1,r_2,k)\leq 0)\\
&\leq P(\tilde{\tau}C+o_p(1)\leq 0)\to 0, \qquad\qquad \text{(S2.7)}
\end{aligned}$$

as $n\to\infty$ for any given M.

By (S2.5) and (S2.7), we can see that $P(\|\hat{\theta}_{1n}-\theta_{10}\|+|\hat{r}_{1n}-r_{10}|\geq\delta)\to 0$ as $n\to\infty$, which implies $\hat{\theta}_{1n}-\theta_{10}=o_p(1)$ and $\hat{r}_{1n}-r_{10}=o_p(1)$. Similarly, we can show that $\hat{\theta}_{2n}-\theta_{20}=o_p(1)$ and $\hat{r}_{2n}-r_{20}=o_p(1)$. Thus, (a) holds.

(b). We only consider the case with $\hat{k}_n\leq k_0$. We note that

$$\sum_{t=1}^{\hat{k}_n}\ell_t(\hat{\theta}_{1n},\hat{r}_{1n})+\sum_{t=\hat{k}_n}^{n}\ell_t(\hat{\theta}_{2n},\hat{r}_{2n})\leq\sum_{t=1}^{k_0}\ell_t(\hat{\theta}_{1n},\hat{r}_{1n})+\sum_{t=k_0}^{n}\ell_t(\hat{\theta}_{2n},\hat{r}_{2n}). \quad \text{(S2.8)}$$

Thus,

$$-\sum_{t=\hat{k}_n+1}^{k_0}\ell_t(\hat{\theta}_{1n},\hat{r}_{1n})+\sum_{t=\hat{k}_n+1}^{k_0}\ell_t(\hat{\theta}_{2n},\hat{r}_{2n})\leq 0. \qquad \text{(S2.9)}$$

Let $d_{\theta_i,r_i}=\|\theta_i-\theta_{i0}\|+|r_i-r_{i0}|$. By (a) of Theorem 3.1, Assumptions 3.1-3.4 and dominated convergence theorem, we have

$$\begin{aligned}
&\lim_{n\to\infty}|E\ell_t(\hat{\theta}_{in},\hat{r}_{in})-E\ell_t(\theta_{i0},r_{i0})|\\
&\leq\lim_{\delta\to 0}[E\sup_{d_{\theta_i,r_i}<\delta}|\ell_t(\theta_i,r_i)-\ell_t(\theta_{i0},r_{i0})|+\lim_{n\to\infty}E\xi_tI(d_{\theta_i,r_i}\geq\delta)]=0, \quad i=1,\,2,
\end{aligned}$$

where $\xi_t = 2\max_{\theta\in\Theta,r\in\Gamma}|\ell_t(\theta,r)|$. Then, there exists a constant $C>0$ such that, for $\hat{k}_n+1\le t\le k_0$,

$$\lim_{n\to\infty}\{E\ell_t(\hat{\theta}_{2n},\hat{r}_{1n})-E\ell_t(\hat{\theta}_{1n},\hat{r}_{2n})\} =E\ell_t(\theta_{20},r_{20})-E\ell_t(\theta_{10},r_{10}) \tag{S2.10}$$
$$=C>0,$$

since (θ_{10},r_{10}) is the only minimizer of $E\ell_t(\theta,r)$ if $t\le k_0$. When $|\hat{k}_n-k_0|>M$, by (S2.9)-(S2.10),

$$\begin{aligned}
&\frac{2}{k_0-\hat{k}_n}\sup_{\theta\in\Theta,r\in\Gamma}\left|\sum_{t=\hat{k}_n+1}^{k_0}[\ell_t(\theta,r)-E\ell_t(\theta,r)]\right|\\
\ge&\frac{1}{k_0-\hat{k}_n}\left\{\sum_{t=\hat{k}_n+1}^{k_0}[\ell_t(\hat{\theta}_{1n},\hat{r}_{1n})-E\ell_t(\hat{\theta}_{1n},\hat{r}_{1n})]\right.\\
&\left.-\sum_{t=\hat{k}_n+1}^{k_0}[\ell_t(\hat{\theta}_{2n},\hat{r}_{2n})-E\ell_t(\hat{\theta}_{2n},\hat{r}_{2n})]\right\}\\
\ge&\frac{1}{k_0-\hat{k}_n}\sum_{t=\hat{k}_n+1}^{k_0}[E\ell_t(\hat{\theta}_{2n},\hat{r}_{2n})-E\ell_t(\hat{\theta}_{1n},\hat{r}_{1n})]\\
=&C+o(1),
\end{aligned}$$

as $n\to\infty$. By the previous inequality, the stationarity and Lemma 2, for any

$\epsilon > 0$, we have

$$
\begin{aligned}
& P(k_0 - \hat{k}_n > M) \\
& = P(k_0 - \hat{k}_n > M, \frac{2}{k_0 - \hat{k}_n} \sup_{\theta \in \Theta, r \in \Gamma} | \sum_{t=\hat{k}_n+1}^{k_0} [\ell_t(\theta, r) - E\ell_t(\theta, r)]| \geq C + o(1)) \\
& \leq P(\max_{u > M} \frac{2}{u} \sup_{\theta \in \Theta, r \in \Gamma} | \sum_{t=-u}^{-1} [\ell_t(\theta, r) - E\ell_t(\theta, r)]| \geq C + o(1)) \\
& = P(o_p(1) \geq \frac{C}{2} + o(1)) \\
& < \epsilon,
\end{aligned}
$$

as $M > 0$ is large enough. Thus, $k_0 - \hat{k}_n = O_p(1)$. This completes the proof. □

Proof of Theorem 3.2. (a). We only prove the case when $i = 1$. Since $\hat{\theta}_{1n}$ and $\hat{r}_{1n}$ are strongly consistent and $\hat{k}_n - k_0 = O_p(1)$, we restrict the parameter space to a neighborhood of θ_{10} :

$$
V_{1\delta} = \{\theta_1 \in \Theta : \|\theta_1 - \theta_{10}\| < \delta, \ |r_1 - r_{10}| < \delta, |k - k_0| \leq M\},
$$

for some $\delta \in (0, 1)$ and $M > 0$. By definition of $(\hat{\theta}_{1n}, \hat{r}_{1n}, \hat{k}_n)$, we have

$$
S_{1n}(\hat{\theta}_{1n}, \hat{r}_{1n}, \hat{k}_n) \leq S_{1n}(\hat{\theta}_{1n}, r_{10}, \hat{k}_n). \tag{S2.11}
$$

By Theorem 3.1 and (S2.11), we have

$$\begin{aligned}
&P(k_0|\hat{r}_{1n}-r_{10}|>B)\\
&\le P(\inf_{\substack{B/k_0<|r_1-r_{10}|<\delta\\ \theta_1\in V_{1\delta}}}\frac{S_{1n}(\theta_1,r_1,k)-S_{1n}(\theta_1,r_{10},k)}{k_0G(|r_1-r_{10}|)}\le 0)\\
&\le 1-P(\inf_{\substack{B/k_0<|r_1-r_{10}|<\delta\\ \theta_1\in V_{1\delta}}}\frac{S_{1n}(\theta_1,r_1,k)-S_{1n}(\theta_1,r_{10},k)}{k_0G(|r_1-r_{10}|)}>\gamma),
\end{aligned}$$

where $\gamma>0$ is a constant and $G(x)=EI(r_{10}<q_t\le r_{10}+x)$ for $1\le t\le k_0$. Now, it suffices to show that, for any $\epsilon>0$, there exists a constant $\gamma>0$, a small δ and a large B such that, as n is large enough,

$$P(\inf_{\substack{B/k_0<|r_1-r_{10}|<\delta\\ \theta_1\in V_{1\delta}}}\frac{S_{1n}(\theta_1,r_1,k)-S_{1n}(\theta_1,r_{10},k)}{k_0G(|r_1-r_{10}|)}>\gamma)>1-\epsilon. \tag{S2.12}$$

Here, we only treat the case when $r_1>r_{10}$ and $k<k_0$. The other cases are similar. For simplicity, write $r_1=r_{10}+v$ for some $v<\delta$. Then

$$\begin{aligned}
&\frac{1}{k_0}[S_{1n}(\theta_1,r_{10}+v,k)-S_{1n}(\theta_1,r_{10},k)]\\
&\quad=-\frac{1}{k_0}\{[S_{1n}(\theta_1,r_{10}+v,k_0)-S_{1n}(\theta_1,r_{10}+v,k)]-\\
&\qquad[S_{1n}(\theta_1,r_{10},k_0)-S_{1n}(\theta_1,r_{10},k)]\}\\
&\qquad+\frac{1}{k_0}[S_{1n}(\theta_1,r_{10}+v,k_0)-S_{1n}(\theta_1,r_{10},k_0)]\\
&\quad:=A_{1n}+A_{2n}.
\end{aligned}$$

By Lemma 3, we have

$$m_0v\le G(v)=P(r_{10}<q_t\le r_{10}+v)\le M_0v. \tag{S2.13}$$

By Lemma 1, on the event $\{|\hat{k}_n - k_0| \leq M\}$, there exists a constant $C > 0$ such that

$$\begin{aligned} E \sup_{\substack{B/k_0<v<\delta \\ V_{1\delta}}} |A_{1n}| =& E \sup_{\substack{B/k_0<v<\delta \\ V_{1\delta}}} \frac{1}{k_0} \left| \sum_{t=k+1}^{k_0} [\ell_t(\theta_1, r_{10}+v) - \ell_t(\theta_1, r_{10})] \right| \\ \leq& 2E(\frac{M}{k_0} \sup_{\substack{B/k_0<v<\delta \\ V_{1\delta}}} |\ell_t(\theta_1, r_{10}+v)|) \\ \leq& \frac{CM}{k_0}. \end{aligned} \tag{S2.14}$$

Then by (S2.13)-(S2.14), for any $\epsilon > 0$ and $\eta > 0$, we have

$$\begin{aligned} P(\sup_{\substack{B/k_0<v<\delta \\ V_{1\delta}}} \frac{|A_{1n}|}{G(v)} < \eta) &\geq 1 - P(\sup_{\substack{B/k_0<v<\delta \\ V_{1\delta}}} \frac{|A_{1n}|}{m_0 B/k_0} \geq \eta) \\ &\geq 1 - \frac{CM/k_0}{\eta B m_0/k_0} \\ &\geq 1 - \epsilon, \end{aligned} \tag{S2.15}$$

where we choose $B > \frac{CM}{\eta m_0 \epsilon}$.

By the proof of Proposition 1 in Chan (1993) or Theorem 2 in Qian (1998) with y_{t-d} replaced by q_{t-1}, we can show that there is a constant $\gamma > 0$ such that

$$P(\inf_{\substack{B/k_0<v<\delta \\ \theta_1 \in V_{1\delta}}} \frac{A_{2n}}{G(v)} > 2\gamma) > 1 - \epsilon. \tag{S2.16}$$

By (S2.15) and (S2.16), we have

$$
\begin{aligned}
&P(\inf_{\substack{B/k_0<v<\delta\\ \theta_1\in V_{1\delta}}}\frac{S_{1n}(\theta_1,r_1,k)-S_{1n}(\theta_1,r_{10},k)}{k_0G(v)}>\gamma)\\
&\qquad=P(\inf_{\substack{B/k_0<v<\delta\\ \theta_1\in V_{1\delta}}}\frac{A_{1n}+A_{2n}}{G(v)}>\gamma)\\
&\qquad\geq 1-P(\inf_{B/n<v<\delta}\frac{A_{2n}(v)}{G(v)}\leq 2\gamma)-P(\sup_{B/n<v<\delta}|\frac{A_{1n}(v)}{G(v)}|\geq\gamma)\\
&\qquad\geq 1-2\epsilon.
\end{aligned}
$$

This proves (S2.12) and hence $n(\hat{r}_{1n}-r_{10})=O_p(1)$. Similarly, we can prove $n(\hat{r}_{2n}-r_{20})=O_p(1)$. Thus, (a) holds.

(b). We consider the case with $\hat{k}_n\leq k_0$ and $\hat{r}_{1n}\geq r_{10}$. By the definition of the LSE, we can show that

$$
\begin{aligned}
\hat{\Phi}_{1n}=&[\sum_{t=1}^{\hat{k}_n}Z_{t-1}Z_{t-1}'(\hat{r}_{1n}^+)]^{-1}\sum_{t=1}^{\hat{k}_n}Z_{t-1}(\hat{r}_{1n}^+)y_t\\
=&\Phi_{10}+[\frac{1}{n}\sum_{t=1}^{\hat{k}_n}Z_{t-1}Z_{t-1}'(\hat{r}_{1n}^+)]^{-1}\frac{1}{n}\sum_{t=1}^{\hat{k}_n}Z_{t-1}(\hat{r}_{1n}^+)\varepsilon_t, \qquad \text{(S2.17)}
\end{aligned}
$$

By Theorem 1, we have $\hat{k}_n-k_0=O_p(1)$ and $\hat{r}_{1n}-r_{10}=O_p(\frac{1}{n})$. Thus, we can show that

$$
\begin{aligned}
\sqrt{n}(\hat{\Phi}_{1n}-\Phi_{10})=&[\frac{1}{n}\sum_{t=1}^{k_0}Z_{t-1}Z_{t-1}'(r_{10}^+)]^{-1}[\frac{1}{\sqrt{n}}\sum_{t=1}^{k_0}Z_{t-1}(r_{10}^+)\varepsilon_t]+o_p(1)\\
&\longrightarrow_{\mathcal{L}} N(0,\frac{\sigma^2}{\tau_0}M_1^{-1}(r_{10}^+)), \qquad \text{(S2.18)}
\end{aligned}
$$

as $n\to\infty$. The proof is similar for $\hat{\Psi}_{1n}$ and the other cases. This completes the proof of Theorem 3.2. $\square$

Lemma 4. *If Assumptions 3.1-3.5 hold, then, for any B, M* $\in (0,\infty)$, *it follows that*

$$\sup_{\substack{|z|\leq B \\ |\hat{k}_n-k_0|\leq M}} |\tilde{S}_{in}(z,\hat{k}_n) - \tilde{\mathcal{S}}_{in}(k_0,\theta_{i0},r_{i0},r_{i0}+\frac{z}{n})| = o_p(1), \quad i=1,\,2.$$

where $\tilde{S}_{in}(z,\hat{k}_n)$ *and* $\tilde{\mathcal{S}}_{in}(k_0,\theta_{i0},r_{i0},r_{i0}+\frac{z}{n})$ *are defined in (4.1) and (4.3)-(4.4).*

Proof of Theorem 4.1. By Theorem 3.2, we can reparameterize $\hat{r}_{1n}$ and $\hat{r}_{2n}$ as $r_{10}+z_1/n$ and $r_{20}+z_2/n$, respectively, for some $z_1, z_2 \in R$. Thus,

$$\begin{aligned} &S_n\left(\hat{\theta}_{1n}(r_{10}+\frac{z_1}{n},\hat{k}_n),\hat{\theta}_{2n}(r_{20}+\frac{z_2}{n},\hat{k}_n),r_{10}+\frac{z_1}{n},r_{20}+\frac{z_2}{n},\hat{k}_n\right) \\ &\quad - S_n\left(\hat{\theta}_{1n}(r_{10},k_0),\hat{\theta}_{2n}(r_{20},k_0),r_{10},r_{20},k_0\right) \\ &=\tilde{S}_{1n}(z_1,\hat{k}_n)+\tilde{S}_{2n}(z_2,\hat{k}_n)+S_{3n}(\hat{k}_n,k_0), \end{aligned} \tag{S2.19}$$

where $\tilde{S}_{1n}(z_1,\hat{k}_n)$, $\tilde{S}_{2n}(z_2,\hat{k}_n)$ are defined in (4.1) and

$$\begin{aligned} S_{3n}(\hat{k}_n,k_0) =&\{I(\hat{k}_n\leq k_0)\sum_{\hat{k}_n+1}^{k_0}[\ell_t(\hat{\theta}_{2n}(r_{20},k_0),r_{20})-\ell_t(\hat{\theta}_{1n}(r_{10},k_0),r_{10})] \\ &+I(\hat{k}_n>k_0)\sum_{k_0+1}^{\hat{k}_n}[\ell_t(\hat{\theta}_{1n}(r_{10},k_0),r_{10})-\ell_t(\hat{\theta}_{2n}(r_{20},k_0),r_{20})]\}. \end{aligned}$$

By Lemma 4, it follows that

$$\begin{aligned} &S_n\left(\hat{\theta}_{1n}(r_{10}+\frac{z_1}{n},\hat{k}_n),\hat{\theta}_{2n}(r_{20}+\frac{z_2}{n},\hat{k}_n),r_{10}+\frac{z_1}{n},r_{20}+\frac{z_2}{n},\hat{k}_n\right) \\ &\quad - S_n\left(\hat{\theta}_{1n}(r_{10},k_0),\hat{\theta}_{2n}(r_{20},k_0),r_{10},r_{20},k_0\right) \\ &=\tilde{\mathcal{S}}_{1n}(k_0,\theta_{10},r_{10},r_{10}+\frac{z_1}{n})+\tilde{\mathcal{S}}_{2n}(k_0,\theta_{20},r_{20},r_{20}+\frac{z_2}{n}) \\ &\quad +S_{3n}(\hat{k}_n,k_0)+o_p(1), \end{aligned} \tag{S2.20}$$

where $o_p(1)$ holds uniformly on $\{|z_1| \leq B, |z_2| \leq B, |\hat{k}_n - k_0| \leq M\}$. On the event $\{|\hat{k}_n - k_0| \leq M\}$, by (S2.17) and (S2.18), we can show that

$$\left\{\sum_{\hat{k}_n+1}^{k_0}[\ell_t(\hat{\theta}_{in}(r_{i0}, k_0), r_{i0}) - \ell_t(\theta_{i0}, r_{i0})]\right\} = o_p(1), \quad i = 1,\, 2.$$

It follows that

$$\begin{aligned} S_{3n}(\hat{k}_n, k_0) = \Bigg\{ & I(\hat{k}_n \leq k_0) \sum_{\hat{k}_n+1}^{k_0}[\ell_t(\theta_{20}, r_{20}) - \ell_t(\theta_{10}, r_{10})] \\ & + I(\hat{k}_n > k_0) \sum_{k_0+1}^{\hat{k}_n}[\ell_t(\theta_{10}, r_{10}) - \ell_t(\theta_{20}, r_{20})]\Bigg\} + o_p(1). \end{aligned} \quad \text{(S2.21)}$$

By Lemma 4 and (4.2), we have

$$n(\hat{r}_{in} - r_{i0}) = \arg\min_{z\in R}\left[\tilde{\mathcal{S}}_{in}(k_0, \theta_{i0}, r_{i0}, r_{i0} + \frac{z}{n}) + o_p(1)\right], \; i = 1, 2. \quad \text{(S2.22)}$$

Thus, for $i = 1$, by a similar proof to Theorem 3.3 in Li and Ling (2012) with $\frac{z}{n}$ in (S2.22) rewritten as $\frac{k_0 z/n}{k_0}$ and noting that $k_0 = [n\tau_0]$, we have

$$\begin{aligned} n(\hat{r}_{1n} - r_{10}) &= \arg\min_{z\in R}\left[\tilde{\mathcal{S}}_{1n}(k_0, \theta_{i0}, r_{i0}, r_{i0} + \frac{z}{n}) + o_p(1)\right] \\ &\Longrightarrow \arg\min_{z\in R} \mathcal{P}_1(\tau_0 z) \\ &= \frac{1}{\tau_0}\arg\min_{z\in R} \mathcal{P}_1(z). \end{aligned} \quad \text{(S2.23)}$$

The proof is similar for the case of $i = 2$. Thus, (a) holds.

Since $\tilde{\mathcal{S}}_{1n}$ and $\tilde{\mathcal{S}}_{2n}$ do not depend on $\hat{k}_n$, we replace $\hat{k}_n$ with k in (S2.20). By

(S2.21), it follows that

$$\begin{aligned}\hat{k}_n =& \arg\min_{1\le k<n} S_{3n}(k,k_0)\\ =& \arg\min_{1\le k<n}\Bigg\{ I(k<k_0)\sum_{t=k+1}^{k_0}[\ell_t(\theta_{20},r_{20})-\ell_t(\theta_{10},r_{10})]\\ &+ I(k\ge k_0)\sum_{t=k_0+1}^{k}[\ell_t(\theta_{10},r_{10})-\ell_t(\theta_{20},r_{20})]+o_p(1)\Bigg\}, \qquad \text{(S2.24)}\end{aligned}$$

By the stationarity of $Y(\theta_{10}, r_{10})$ and $Y(\theta_{20}, r_{20})$,

$$\hat{k}_n - k_0 \longrightarrow_{\mathcal{L}} \arg\min_k W(k,\theta_{10},\theta_{20},r_{10},r_{20}). \qquad \text{(S2.25)}$$

This completes the proof of Theorem 4.1. □

S3 Proofs of Lemmas 1-2 and 4

Proof of Lemma 1. By Assumptions 3.1-3.4, $E\sup_{\theta_i\in\Theta,r_i\in\Gamma}|\ell_t(\theta_i,r_i)|<\infty$ for $i = 1, 2$. The rest of the proof is similar to that of Lemma 6.4 in Li et al. (2013). □

Proof of Lemma 2. We only prove it for (a) since (b) is similar. We first prove that, for each $(\theta', r)' \in \Theta\times\Gamma$ and any $\eta > 0$,

$$\lim_{\ell\to\infty} P(\max_{n\ge\ell}\frac{1}{n}|\sum_{t=1}^{n}[\ell_t(\theta,r)-E\ell_t(\theta,r)]|>\eta)=0. \qquad \text{(S3.1)}$$

Taking $X_t = \ell_t(\theta, r) - E\ell_t(\theta, r)$ and $p = 1 + \iota/2$, by Assumptions 3.1 and 3.4, $\{X_t\}$ is a strong mixing sequence with geometric rate and $E|X_t|^p < \infty$. By

Theorem 1 in section S1, it follows that, for each $(\theta', r)' \in \Theta \times \Gamma$,

$$\frac{1}{n^{1-\delta}}\sum_{t=1}^{n}[\ell_t(\theta,r)-E\ell_t(\theta,r)] \longrightarrow 0 \quad a.s.. \tag{S3.2}$$

By (S3.2) and Lemma 1 in Chow (1978) (pp. 66), we can see that (S3.1) holds. By the standard piece-wise argument, we can show that

$$\lim_{l\to\infty} P(\max_{n\geq l}\frac{1}{n}\max_{\Theta\times\Gamma}|\sum_{t=1}^{n}[\ell_t(\theta,r)-E\ell_t(\theta,r)]|>\eta)=0. \tag{S3.3}$$

Thus, by Lemma 1 in Chow (1978) (pp. 66), (a) holds. This completes the proof. $\square$

Proof of Lemma 4. We only prove it for $i=1$ and $\hat{k}_n \leq k_0$. On the event $\{|z|\leq B, |\hat{k}_n-k_0|\leq M\}$, by (S2.17)-(S2.18), it follows that

$$\sqrt{n}\sup_{\substack{|z|\leq B\\ |k-k_0|\leq M}}\|\hat{\theta}_{in}(r_{10}+\frac{z}{n},k)-\hat{\theta}_{in}(r_{10},k_0)\|=o_p(1),\ i=1,\,2, \tag{S3.4}$$

where $o_p(1)$ uniformly goes to zero in probability as $n\to\infty$ for any fixed $B, M\in(0,\infty)$. By (S3.4), it is not hard to show that

$$S_{in}(\hat{\theta}_{in}(r_{i0},\hat{k}_n),r_{i0},\hat{k}_n)=S_{in}(\hat{\theta}_{in}(r_{i0},k_0),r_{i0},\hat{k}_n)+o_p(1).$$

Thus, (4.1) becomes

$$\begin{aligned}
\tilde{S}_{1n}(z,\hat{k}_n)=&S_{in}(\hat{\theta}_{in}(r_{i0}+\frac{z}{n},\hat{k}_n),r_{i0}+\frac{z}{n},\hat{k}_n)-S_{in}(\hat{\theta}_{in}(r_{i0},k_0),r_{i0},\hat{k}_n)+o_p(1)\\
=&\left\{\sum_{t=1}^{k_0}\ell_t(\hat{\theta}_{1n}(r_{10}+\frac{z}{n},\hat{k}_n),r_{10}+\frac{z}{n})-\sum_{t=1}^{k_0}\ell_t(\hat{\theta}_{1n}(r_{10},k_0),r_{10})\right\}\\
&-\left\{\sum_{t=\hat{k}_n+1}^{k_0}[\ell_t(\hat{\theta}_{1n}(r_{10}+\frac{z}{n},\hat{k}_n),r_{10}+\frac{z}{n})-\ell_t(\hat{\theta}_{1n}(r_{10},k_0),r_{10})]\right\}\\
:=&R_{1n}(z,\hat{k}_n)-R_{2n}(z,\hat{k}_n).
\end{aligned}$$

Then it is easy to obtain

$$\sup_{\substack{|z|\le B \\ |\hat{k}_n - k_0|\le M}} |R_{2n}(z,\hat{k}_n)| = o_p(1). \tag{S3.5}$$

For notational simplicity, by Theorem 3.2, we can replace $\hat{\theta}_{1n}(r_{10}+\frac{z}{n},\hat{k}_n)$ and $r_{10}+\frac{z}{n}$ with $\hat{\theta}_{1n}$ and $\hat{r}_{1n}$, respectively, without affecting the asymptotic properties of $R_{1n}(z,\hat{k}_n)$. We only consider the case when $\hat{r}_{1n}\ge r_{10}$. Then

$$\begin{aligned}
\sum_{t=1}^{k_0}\ell_t(\hat{\theta}_{1n},\hat{r}_{1n}) =& \sum_{t=1}^{k_0}[\varepsilon_t-(\hat{\Phi}_{1n}-\Phi_{10})'Z_{t-1}]^2 I(q_{t-1}>\hat{r}_{1n}) \\
&+\sum_{t=1}^{k_0}[\varepsilon_t-(\hat{\Psi}_{1n}-\Phi_{10})'Z_{t-1}]^2 I(r_{10}<q_{t-1}\le\hat{r}_{1n}) \\
&+\sum_{t=1}^{k_0}[\varepsilon_t-(\hat{\Psi}_{1n}-\Psi_{10})'Z_{t-1}]^2 I(q_{t-1}\le r_{10}) \\
:=& B_{1n}+B_{2n}+B_{3n}.
\end{aligned} \tag{S3.6}$$

On the event $\{k_0-\hat{k}_n\le M\}\cap\{r_{10}<\hat{r}_{1n}\le r_{10}+\frac{B}{n}\}$, by Theorem 3.2 and (S2.17)-(S2.18), we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{k_0} Z_{t-1}\varepsilon_t I(q_{t-1}>\hat{r}_{1n}) = \sqrt{n}(\hat{\Phi}_{1n}-\Phi_{10})'\tau_0 M_1(r_{10}^{+})+o_p(1). \tag{S3.7}$$

Thus,

$$\begin{aligned}
B_{1n} =& \sum_{t=1}^{k_0}\varepsilon_t^2 I(q_{t-1}>\hat{r}_{1n}) - 2(\hat{\Phi}_{1n}-\Phi_{10})'\sum_{t=1}^{k_0} Z_{t-1}\varepsilon_t I(q_{t-1}>\hat{r}_{1n}) \\
&+(\hat{\Phi}_{1n}-\Phi_{10})'\sum_{t=1}^{k_0} Z_{t-1}Z_{t-1}'(\hat{r}_{1n}^{+})(\hat{\Phi}_{1n}-\Phi_{10}) \\
=& \sum_{t=1}^{k_0}\varepsilon_t^2 I(q_{t-1}>\hat{r}_{1n}) - \sqrt{n}(\hat{\Phi}_{1n}-\phi_{10})'\tau_0 M_1(r_{10}^{+})\sqrt{n}(\hat{\Phi}_{1n}-\Phi_{10}) \\
&+o_p(1).
\end{aligned} \tag{S3.8}$$

Similarly we have

$$B_{3n} = \sum_{t=1}^{k_0} \varepsilon_t^2 I(q_{t-1} \le r_{10}) \qquad \text{(S3.9)}$$
$$- \sqrt{n}(\hat{\Psi}_{1n} - \Psi_{10})' \tau_0 M_1(r_{10}^-) \sqrt{n}(\hat{\Psi}_{1n} - \Psi_{10}) + o_p(1).$$

We note that

$$B_{2n} = \sum_{t=1}^{k_0} \varepsilon_t^2 I(r_{10} < q_{t-1} \le \hat{r}_{1n}) - 2(\hat{\Psi}_{1n} - \Phi_{10})' \sum_{t=1}^{k_0} Z_{t-1}(r_{10}, \hat{r}_{1n})\varepsilon_t$$
$$+ (\hat{\Psi}_{1n} - \Phi_{10})' \sum_{t=1}^{k_0} Z_{t-1}Z_{t-1}'(r_{10}, \hat{r}_{1n})(\hat{\Psi}_{1n} - \Phi_{10})$$
$$:= C_{1n} + C_{2n} + C_{3n}. \qquad \text{(S3.10)}$$

We continue to calculate each term of B_{2n}. On the event $\{k_0 - \hat{k}_n \le M\} \cap \{r_{10} < \hat{r}_{1n} \le r_{10} + \frac{B}{n}\}$,

$$C_{2n} = -2(\hat{\Psi}_{1n} - \Psi_{10} + \Psi_{10} - \Phi_{10})' \sum_{t=1}^{k_0} Z_{t-1}(r_{10}, \hat{r}_{1n})\varepsilon_t$$
$$= -2(\Psi_{10} - \Phi_{10})' \sum_{t=1}^{k_0} Z_{t-1}(r_{10}, \hat{r}_{1n})\varepsilon_t + o_p(1), \qquad \text{(S3.11)}$$

where the last equality holds by Markov's inequality and Theorem 3.2.

$$C_{3n} = (\Psi_{10} - \Phi_{10})' \sum_{t=1}^{k_0} Z_{t-1}'Z_{t-1}(r_{10}, \hat{r}_{1n})(\Psi_{10} - \Phi_{10})$$
$$+ 2(\hat{\Psi}_{1n} - \Psi_{10})' \sum_{t=1}^{k_0} Z_{t-1}'Z_{t-1}(r_{10}, \hat{r}_{1n})(\Psi_{10} - \Phi_{10})$$
$$+ (\hat{\Psi}_{1n} - \Psi_{10})' \sum_{t=1}^{k_0} Z_{t-1}'Z_{t-1}(r_{10}, \hat{r}_{1n})(\hat{\Psi}_{1n} - \Psi_{10})$$
$$= (\Psi_{10} - \Phi_{10})' \sum_{t=1}^{k_0} Z_{t-1}'Z_{t-1}(r_{10}, \hat{r}_{1n})(\Psi_{10} - \Phi_{10}) + o_p(1), \qquad \text{(S3.12)}$$

where the last equality holds by Markov's inequality and Theorem 3.2.

Then by (S3.6) and (S3.8)-(S3.12),

$$\begin{aligned}\sum_{t=1}^{k_0}\ell_t(\hat{\theta}_{1n},\hat{r}_{1n}) =& \sum_{t=1}^{k_0}\varepsilon_t^2 - \sqrt{n}(\hat{\Phi}_{1n}-\Phi_{10})'\tau_0 M_1(r_{10}^{+})\sqrt{n}(\hat{\Phi}_{1n}-\Phi_{10}) \\ &- \sqrt{n}(\hat{\Psi}_{1n}-\Psi_{10})'\tau_0 M_1(r_{10}^{-})\sqrt{n}(\hat{\Psi}_{1n}-\Psi_{10}) \\ &+ (\Psi_{10}-\Phi_{10})'\sum_{t=1}^{k_0} Z_{t-1}'Z_{t-1}(r_{10},\hat{r}_{1n})(\Psi_{10}-\Phi_{10}) \\ &- 2(\Psi_{10}-\Phi_{10})'\sum_{t=1}^{k_0} Z_{t-1}(r_{10},\hat{r}_{1n})\varepsilon_t + o_p(1). \qquad \text{(S3.13)}\end{aligned}$$

Now, we replace $\hat{\theta}_{in}$ and $\hat{r}_{1n}$ in (S3.13) with $\hat{\theta}_{1n}(r_{10}+\frac{z}{n},\hat{k}_n)$ and $r_{10}+\frac{z}{n}$, respectively. Then we have

$$\begin{aligned}&\sum_{t=1}^{k_0}\ell_t\left(\hat{\theta}_{1n}(r_{10}+\frac{z}{n},\hat{k}_n),r_{10}+\frac{z}{n}\right) \\ &= -\sqrt{n}\left(\hat{\Phi}_{1n}(r_{10}+\frac{z}{n},\hat{k}_n)-\Phi_{10}\right)'\tau_0 M_1(r_{10}^{+}) \\ &\quad \times\sqrt{n}\left(\hat{\Phi}_{1n}(r_{10}+\frac{z}{n},\hat{k}_n)-\Phi_{10}\right) \\ &\quad -\sqrt{n}\left(\hat{\Psi}_{1n}(r_{10}+\frac{z}{n},\hat{k}_n)-\Psi_{10}\right)'\tau_0 M_1(r_{10}^{-}) \\ &\quad \times\sqrt{n}\left(\hat{\Psi}_{1n}(r_{10}+\frac{z}{n},\hat{k}_n)-\Psi_{10}\right) \\ &\quad + I(z\geq 0)\tilde{\mathcal{S}}_{1n}(k_0,\theta_{10},r_{10},r_{10}+\frac{z}{n}) + \sum_{t=1}^{k_0}\varepsilon_t^2 + o_p(1). \qquad \text{(S3.14)}\end{aligned}$$

Similarly we can show that

$$\begin{aligned}
&\sum_{t=1}^{k_0}\ell_t(\hat{\theta}_{1n}(r_{10},k_0),r_{10}) \\
&\quad= -\sqrt{n}(\hat{\Phi}_{1n}(r_{10},k_0)-\Phi_{10})'\tau_0 M_1(r_{10}^{+})\sqrt{n}(\hat{\Phi}_{1n}(r_{10},k_0)-\Phi_{10}) \\
&\qquad -\sqrt{n}(\hat{\Psi}_{1n}(r_{10},k_0)-\Psi_{10})'\tau_0 M_1(r_{10}^{-})\sqrt{n}(\hat{\Psi}_{1n}(r_{10},k_0)-\Psi_{10}) \qquad \text{(S3.15)} \\
&\qquad +\sum_{t=1}^{k_0}\varepsilon_t^2+o_p(1).
\end{aligned}$$

By (S3.4) and (S3.14)-(S3.15), for any $|z|<B$, we have

$$\begin{aligned}
R_{1n}(z,\hat{k}_n) &= \sum_{t=1}^{k_0}\ell_t\left(\hat{\theta}_{1n}(r_{10}+\frac{z}{n},\hat{k}_n),r_{10}+\frac{z}{n}\right)-\sum_{t=1}^{k_0}\ell_t\left(\hat{\theta}_{1n}(r_{10},k_0),r_{10}\right) \\
&= \tilde{\mathcal{S}}_{1n}(k_0,\theta_{10},r_{10},r_{10}+\frac{z}{n})+o_p(1). \qquad \text{(S3.16)}
\end{aligned}$$

By (S3.5) and (S3.16), it follows that

$$\sup_{\substack{|z|\le B\\ |\hat{k}_n-k_0|\le M}}|\tilde{S}_{1n}(z,\hat{k}_n)-\tilde{\mathcal{S}}_{1n}(k_0,\theta_{10},r_{10},r_{10}+\frac{z}{n})|=o_p(1). \qquad \text{(S3.17)}$$

This completes the proof. □

S4 The Effect of the Initial Values

Now, we discuss the effect of the initial values $\mathcal{Z}_{10}$ and $\mathcal{Z}_{2k_0}$ on the results obtained in Sections 3-5. Since we only have one data set $\{y_1,...,y_n\}$, we use this and replace $\mathcal{Z}_{10}$ by some constant $\tilde{\mathcal{Z}}_{10}$ to calculate $\ell_t(\theta,r)$. Although we do not know k_0, this calculation has implied that we replace $\mathcal{Z}_{2k_0}$ by

$\tilde{\mathcal{Z}}_{2k_0} = \{y_{k_0}, ..., y_1, \mathcal{Z}_{10}\}$ when $t > k_0$. That is, we first choose some constant $\mathcal{Z}_{10}$ to generate the data and the initial value of the second period are from the data of the first period, as implied by the model presentation. With these initial values, the results of Ling and Tong (2005) and Li et al. (2011) show that they do not affect the asymptotic properties of $\hat{\theta}_{in}$ and $\hat{r}_{in}$. To see their effect on the estimated change-point k_0, we denote $\tilde{\ell}_t(\theta_1, r_1) = \ell(\theta_1, r_1, y_t, ...y_1, \tilde{\mathcal{Z}}_{10})$ when $t \leq k_0$ and $\tilde{\ell}_t(\theta_2, r_2) = \ell(\theta_2, r_2, y_t, ...y_{k_0+1}, \tilde{\mathcal{Z}}_{2k_0})$ when $t > k_0$. From the proof of Theorem 3.3, we can see that

$$\tilde{k}_n - k_0 \longrightarrow_{\mathcal{L}} \arg\min_k \tilde{W}(k, \theta_{10}, \theta_{20}, r_{10}, r_{20}),$$

where $\tilde{k}_n$ is the estimator given these initial values and $\tilde{W}(\cdot)$ is defined as (4.8) with replacing $\ell_t(\theta_{i0}, r_{i0})$ by $\tilde{\ell}_t(\theta_{i0}, r_{i0})$. Since the distribution of $W(\cdot)$ and $\tilde{W}(\cdot)$ are different, the initial values always affect the asymptotic distribution of the estimated k_0 in Theorem 3.3. However, under Assumption 4.2, if $\Phi_{20} - \Phi_{10} \approx \kappa_{1n}$, $\Psi_{20} - \Psi_{10} \approx \kappa_{2n}$ and $\Phi_{10} - \Psi_{20} \approx \kappa_{3n}$, it is not hard to show that

$$W(k, \theta_{10}, \theta_{20}, r_{10}, r_{20}) - \tilde{W}(k, \theta_{10}, \theta_{20}, r_{10}, r_{20}) = o_p(1),$$

as $n \to \infty$. Thus, we can see that the approximating distribution in Theorem 4.2 and the likelihood-ratio based distribution in (5.10) are still valid in this case.

S5 Further Analysis of the Real Data Example

In this section, we re-examine some steps in Section 7 of the main article. In Step 1, we have applied the threshold nonlinearity test to the data and found significant threshold effect on the data $\{y_t\}$. After we fitted a TAR(8) model with threshold variable y_{t-8} to the data, we further applied the Sup-likelihood-ratio test for the existence of a change-point in the TAR model and found an estimated change-point $\hat{k}_n = 578$.

Now, we change the order of the tests in Step 1 and Step 3, i.e. we first fit the data using an AR(8) model, where we adopt the order 8 of the TAR model in the main article for the purpose of comparison. Then we will perform the threshold nonlinearity test on the two segments, respectively. In this experiment, when we apply the Sup-likelihood-ratio test to the AR(8) model, we find that $\sup_{\tau\in[0.05,0.95]} LR(\tau) = 85.78$, which is larger than the critical value 31.61 at the 0.01 significance level, see Table 1 in Andrews (1993) with degrees of freedom $9(= p + 1)$. The estimated change-point $\hat{k}_n = 560$, which is different from the one ($\hat{k}_n = 578$) obtained in the main article. This is understandable since the structural change in AR model may be different from the TAR one. On the other hand, according to our Theorem 3.1(b), the $\hat{k}_n$ is not a consistent estimator and only $\hat{\tau}_n$ is consistent to τ_0. Therefore, we allow some discrepancy when it comes to the estimator for the change-point, as long as $\hat{k}_n/n$ is consistent to the true one. In this case, the estimator $\hat{k}_n = 560$ does not contradict with the one $\hat{k}_n = 578$ in the main article.

Next, we apply the threshold nonlinearity test to the two segments of the AR models, respectively. For the first segment, under the null hypothesis that there is no threshold effect, and the alternative is the threshold model with threshold variable y_{t-8}, the $p-value$ is $0.3088 > 0.05$, which suggests that there is no threshold effect in the direction of y_{t-8} as the threshold variable, and the $p-value$ of the second segment is 1.45×10^{-9} which suggests that there is threshold effect. But this does not mean the first segment does not have threshold effect since the $p-values$ for other alternative threshold variables may significantly small. For example, when the threshold variable is y_{t-4} in the alternative, the $p-value$ is $0.014 < 0.05$. However, when we set $\hat{k}_n = 578$ as the main article and apply the threshold nonlinearity test to the two segments, the $p-values$ in the direction of the threshold variable y_{t-8} are 0.00043 and 0.00024, respectively, which suggests that there is significant threshold effect along the two segments and the two fitted sub-models are more reasonable than pure AR models.

Finally, we examine the performance of the fitted model (7.3) in the main article by computing the average of the mean squared forecasting errors. We first choose the forecasting period from $t = 579$ to $t = 608$, which contains 30 data points after the estimated change-point. We denote the models in (7.3) for $t \leq 578$, $t > 578$ and model (7.2) by Models I, II, and III, respectively. The averages of mean squared errors of the forecasting values based on theses three models are reported in Table S11. From this table, we can see that the mean squared error produced by Model II is the smallest one and the performance of Model I is the worst one. Therefore, the performance of model (7.3) with a change-point is better than the model (7.2) which is fitted to the whole dataset, and the TAR model fitted to the second segment can characterize the second segment of the data in a better way. We then examine the performance of the out-of-sample forecastings by choosing the last 30 points (from $t = 901$ to $t = 930$) as the testing sample. We denote the fitted TAR models for the period $t = 1$ to $t = 930$ and the period $t = 579$ to $t = 900$ by Model IV and Model V, respectively. The mean squared errors based on Models IV and V are reported in Table S11. As we can see that with a change-point $\hat{k}_n = 578$, the fitted model using the period from $t = 579$ to $t = 900$ outperforms the model fitted to the whole series (from $t = 1$ to $t = 900$), which means that our model with a change-point in (7.3) improves the performance of the forecastings.

References

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

Table S1: Simulation studies for model (6.1) with true parameters $(\theta_{10}', r_{10}) = (-1, -0.6, 1, 0.4, 0.8)$ and $(\theta_{20}', r_{20}) = (-0.8, -0.9, 0.7, 0.6, 0.5)$, $\varepsilon_t \sim N(0,1)$, $k_0 = n/2$.

n		μ_{10}	ϕ_{10}	ν_{10}	ψ_{10}	r_{10}	μ_{20}	ϕ_{20}	ν_{20}	ψ_{20}	r_{20}
400	Bias	-.0070	.0026	.0100	.0029	-.0198	.0125	-.0133	.0123	.0014	-.0217
	ESD	.3753	.2191	.1142	.0717	.0289	.3439	.2433	.1241	.0748	.0366
	ASD	.3631	.2102	.1076	.0678	.0286	.3274	.2340	.1173	.0732	.0365
	EASD	.3559	.2068	.1053	.0699	.0282	.3205	.2281	.1156	.0722	.0373
800	Bias	-.0079	.0059	.0035	.0024	-.009	-.0066	.0022	.0048	-.0008	-.0083
	ESD	.2657	.1526	.0786	.0479	.0133	.2324	.1634	.0852	.0526	.0180
	ASD	.2537	.1469	.0758	.0479	.0145	.2289	.1630	.0826	.0516	.0180
	EASD	.2505	.1448	.0751	.0475	.0141	.2260	.1603	.0817	.0511	.0186
1200	Bias	.0011	.0005	.0056	.0026	-.0064	-.0033	.0002	.0045	.0017	-.0054
	ESD	.2119	.1213	.0617	.0394	.0089	.1894	.1361	.0650	.0427	.0112
	ASD	.2058	.1188	.0618	.0390	.0096	.1864	.1321	.0673	.0420	.0118
	EASD	.2048	.1181	.0614	.0388	.0094	.1841	.1306	.0670	.0418	.0124

* ESD: empirical standard deviation; ASD: asymptotic standard deviation; EASD: estimated asymptotic standard deviation.

Table S2: Coverage probabilities of r_{i0} by the simulation method of Li and Ling (2012), the approximation method in Section 4 and the likelihood-ratio method in Section 5.

γ	$\|\delta_{1n}\|$	$\|\delta_{2n}\|$	n	$\tilde{\alpha}$	CM_1	CB_1	CLR_1	CM_2	CB_2	CLR_2
				10%	0.901	0.598	0.969	0.874	0.692	0.946
			400	5%	0.953	0.736	0.980	0.933	0.809	0.966
				1%	0.988	0.906	0.997	0.983	0.935	0.988
				10%	0.906	0.600	0.967	0.902	0.710	0.959
0	2.236	2.121	800	5%	0.947	0.728	0.983	0.946	0.823	0.979
				1%	0.991	0.904	0.995	0.989	0.950	0.997
				10%	0.907	0.600	0.975	0.901	0.699	0.960
			1200	5%	0.955	0.735	0.984	0.952	0.817	0.982
				1%	0.991	0.909	0.996	0.989	0.939	0.998
				10%	0.887	0.746	0.947	0.892	0.817	0.941
			400	5%	0.935	0.848	0.969	0.929	0.896	0.962
				1%	0.983	0.948	0.988	0.975	0.961	0.991
				10%	0.901	0.736	0.955	0.889	0.810	0.944
0.2	1.709	1.56	800	5%	0.950	0.859	0.975	0.935	0.895	0.968
				1%	0.991	0.963	0.990	0.986	0.971	0.993
				10%	0.890	0.730	0.962	0.901	0.830	0.958
			1200	5%	0.944	0.841	0.981	0.953	0.911	0.979
				1%	0.988	0.952	0.994	0.989	0.981	0.993
				10%	0.852	0.807	0.917	0.810	0.801	0.888
			400	5%	0.915	0.878	0.955	0.848	0.847	0.944
				1%	0.959	0.952	0.986	0.881	0.908	0.974
				10%	0.885	0.832	0.946	0.841	0.833	0.909
0.4	1.217	0.99	800	5%	0.934	0.900	0.975	0.892	0.897	0.959
				1%	0.975	0.967	0.992	0.933	0.949	0.988
				10%	0.907	0.864	0.945	0.876	0.861	0.924
			1200	5%	0.959	0.932	0.976	0.925	0.927	0.960
				1%	0.993	0.990	0.997	0.963	0.983	0.993

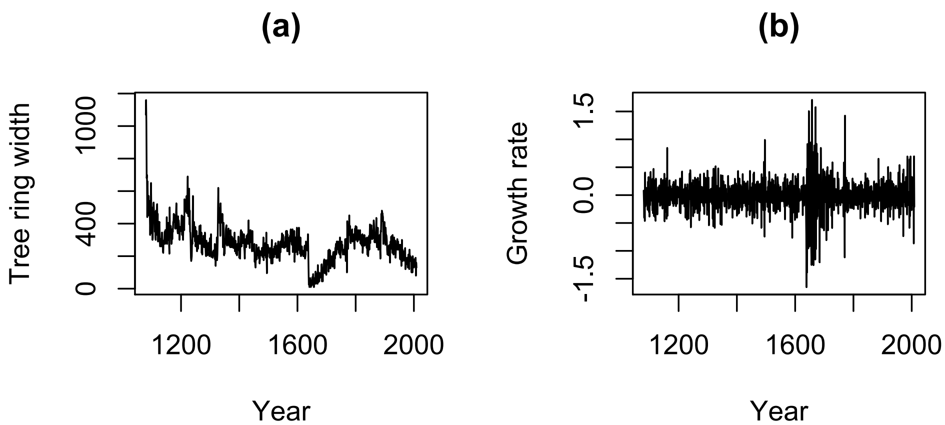


Figure S1: (a) The original tree ring width data; (b) the growth rate by taking the log difference.

Cheng, Q. (1999). Miscellanea. on time-reversibility of linear processes. *Biometrika 86*(2), 483–486.

Chow, Y. (1978). *Probability Theory: Independence, Interchangeability, Martingales.* Wiley Online Library.

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.

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

Hall, P. and C. C. Heyde (2014). *Martingale limit theory and its application.* Academic press.

Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory of Probability & Its*

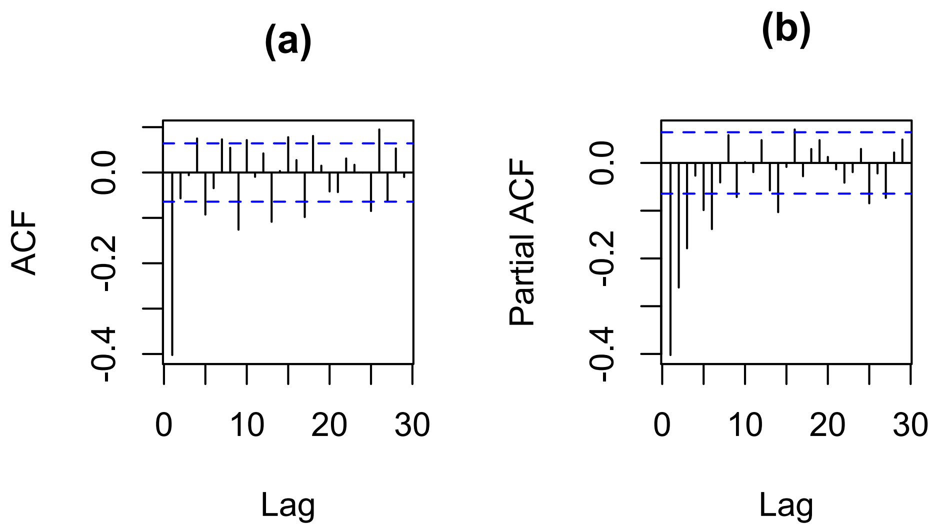


Figure S2: (a) Autocorrelation functions of y_t; (b) Partial autocorrelation functions of y_t.

Applications 7(4), 349–382.

Kuczmaszewska, A. (2011). On the strong law of large numbers for φ-mixing and ρ-mixing random variables. *Acta Mathematica Hungarica 132*(1), 174–189.

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

Li, D., Li, W. and Ling, S. (2011). On the least squares estimation of threshold autoregressive. *Statistics and its Interface 4*(1), 183–196.

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.

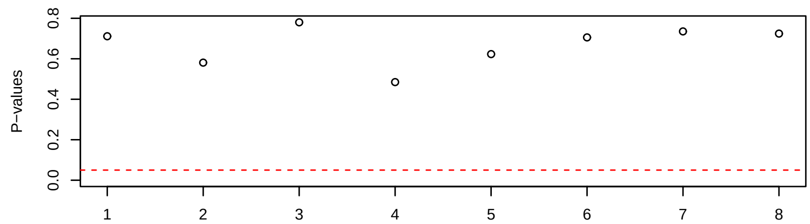


Figure S3: Model diagnostics of model (7.2).

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.

Ling, S. and Tong, H. (2005). Testing for a linear ma model against threshold ma models. *The Annals of Statistics 33*(6), 2529–2552.

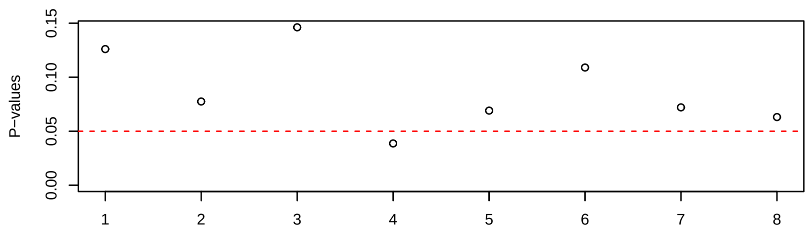
Qian, L. (1998). On maximum likelihood estimators for a threshold autoregression. *Journal of Statistical Planning and Inference 75*(1), 21–46.

Shao, Q.-M. (1995). Maximal inequalities for partial sums of ρ-mixing sequences. *The Annals of Probability*, 948–965.

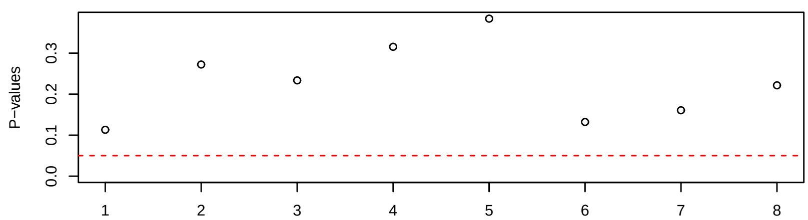
Stout, W. F. (1974). *Almost sure convergence*, Volume 154. Academic press New York.

Wu, W. B. (2007). Strong invariance principles for dependent random variables. *The Annals of Probability 35*(6), 2294–2320.

Yu, P. (2015). Adaptive estimation of the threshold point in threshold regression. *Journal of*



(a)



(b)

Figure S4: Model diagnostics of model (7.3)— (a) when $t \leq 578$ and (b) when $t > 578$.

Econometrics 189(1), 83–100.

Table S3: The mean, ESD, ASD and the estimators of d_{i0}, ϕ and ξ for (6.2).

$\tilde{\gamma}_1$	κ_n	n	k_0	Mean	ESD	ASD	$\hat{d}_{1n}$	$\hat{d}_{2n}$	$\hat{\phi}_n \approx \hat{\xi}_n$
0	4.69	400	200	199	1.464	0.559	6.684	17.196	2.573
		800	400	400	1.342	0.560	6.669	17.259	2.588
		1200	600	599	1.291	0.560	6.676	17.107	2.563
0.1	4.36	400	200	199	1.419	0.741	5.711	8.902	1.559
		800	400	399	1.534	0.743	5.712	8.853	1.550
		1200	600	599	1.446	0.743	5.714	8.818	1.543
0.2	4.15	400	200	199	1.780	0.980	4.882	5.528	1.132
		800	400	399	1.852	0.974	4.878	5.623	1.153
		1200	600	600	1.775	0.974	4.884	5.614	1.149
0.4	3.88	400	200	199	2.338	1.551	3.523	3.041	0.863
		800	400	399	2.393	1.553	3.524	3.034	0.861
		1200	600	599	2.288	1.550	3.523	3.043	0.864

* ESD: empirical standard deviation; ASD: asymptotic standard deviation; For one instance of 1000 replications, $(\hat{d}_{10}, \hat{d}_{20}) = (6.515, 17.447), (5.770, 8.905), (4.846, 5.618), (3.458, 2.996)$ for $\tilde{\gamma}_1 = 0, 0.1, 0.2, 0.4$, respectively, where $(\hat{d}_{10}, \hat{d}_{20})$ are obtained by (4.13) using the true parameters and sample estimates of $M_i(r_{i0}^{\pm})$ when the sample size $n = 1200$.

Table S4: The coverage probabilities of the estimated k_0 for (6.2).

$\tilde{\gamma}_1$	n	$T_{\hat{\phi}_n,\hat{\xi}_n}$			LR_n		
		10%	5%	1%	10%	5%	1%
0	400	0.690	0.690	0.876	0.975	0.985	0.992
	800	0.735	0.735	0.896	0.976	0.986	0.995
	1200	0.731	0.736	0.899	0.982	0.990	0.999
0.1	400	0.630	0.790	0.943	0.969	0.984	0.995
	800	0.633	0.812	0.925	0.966	0.987	0.995
	1200	0.637	0.797	0.947	0.972	0.984	0.995
0.2	400	0.733	0.799	0.945	0.957	0.976	0.993
	800	0.753	0.810	0.960	0.968	0.979	0.997
	1200	0.760	0.810	0.960	0.964	0.982	0.998
0.4	400	0.783	0.884	0.971	0.930	0.971	0.993
	800	0.792	0.885	0.971	0.942	0.963	0.991
	1200	0.799	0.880	0.977	0.937	0.970	0.992

* $T_{\hat{\phi}_n,\hat{\xi}_n}$ is the approximative distribution in Section 4 and LR_{in} represents the likelihood-ratio based distribution in Section 5.

Table S5: The mean, ESD, ASD and the estimators of d_{i0}, ϕ and ξ for (6.3).

$\tilde{\gamma}_2$	κ_n	n	k_0	Mean	ESD	ASD	$\hat{d}_{1n}$	$\hat{d}_{2n}$	$\hat{\phi}_n \approx \hat{\xi}_n$
0.1	2.385	400	200	202	41.770	25.186	0.198	0.205	1.036
		800	400	402	41.830	25.443	0.196	0.203	1.036
		1200	600	601	41.780	25.190	0.201	0.204	1.014
0.2	2.535	400	200	199	20.599	12.407	0.404	0.414	1.023
		800	400	398	20.600	12.454	0.405	0.410	1.014
		1200	600	599	19.398	12.521	0.403	0.411	1.007
0.3	2.687	400	200	199	12.733	8.144	0.632	0.613	0.971
		800	400	399	11.590	8.132	0.631	0.613	0.973
		1200	600	599	11.100	8.129	0.630	0.618	0.982
0.5	2.995	400	200	199	6.621	4.694	1.148	1.078	0.887
		800	400	399	6.205	4.693	1.147	1.109	0.889
		1200	600	599	6.380	4.670	1.147	1.020	0.889

* ESD: empirical standard deviation; ASD: asymptotic standard deviation; For one instance of 1000 replications, $(\hat{d}_{10}, \hat{d}_{20}) = (0.204, 0.201), (0.408, 0.415), (0.641, 0.623), (1.152, 1.050)$ for $\tilde{\gamma}_2 = 0.1, 0.2, 0.3, 0.5$, respectively, where $(\hat{d}_{10}, \hat{d}_{20})$ are obtained by (4.13) using the true parameters and sample estimates of $M_i(r_{i0}^{\pm})$ when the sample size $n = 1200$.

Table S6: The coverage probabilities of the estimated k_0 for (6.3).

$\tilde{\gamma}_2$	n	$T_{\hat{\phi}_n,\hat{\xi}_n}$			LR_n		
		10%	5%	1%	10%	5%	1%
0.1	400	0.710	0.820	0.941	0.550	0.670	0.835
	800	0.719	0.826	0.949	0.591	0.705	0.881
	1200	0.738	0.846	0.949	0.621	0.729	0.915
0.2	400	0.771	0.858	0.955	0.739	0.816	0.926
	800	0.781	0.868	0.960	0.735	0.828	0.958
	1200	0.799	0.881	0.962	0.762	0.848	0.957
0.3	400	0.805	0.879	0.960	0.777	0.851	0.945
	800	0.826	0.895	0.974	0.820	0.891	0.960
	1200	0.839	0.905	0.977	0.829	0.902	0.977
0.5	400	0.826	0.904	0.970	0.866	0.930	0.982
	800	0.836	0.907	0.981	0.878	0.931	0.981
	1200	0.837	0.905	0.976	0.874	0.935	0.979

* $T_{\hat{\phi}_n,\hat{\xi}_n}$ is the approximative distribution in Section 4 and LR_{in} represents the likelihood-ratio based distribution in Section 5.

Table S7: $p-$values when testing for threshold nonlinearity, the null model is AR(p) for $5 \leq p \leq 9$.

$d =$	1	2	3	4	5	6	7	8	9
AR(5)	0.013	0.093	0.015	0.000	0.003				
AR(6)	0.036	0.122	0.028	0.000	0.015	0.000			
AR(7)	0.001	0.056	0.013	0.000	0.019	0.000	0.000		
AR(8)	0.000	0.062	0.002	0.000	0.027	0.000	0.000	0.000	
AR(9)	0.000	0.007	0.003	0.000	0.004	0.000	0.000	0.000	0.000

* d denotes the delay lag of the threshold variable y_{t-d} in the alternative hypothesis.

Table S8: The AICs and BICs of TAR(p) models based on (7.1) for $1 \leq p \leq 12$.

	d	1	2	3	4	5	6	7	8	9	10	11	12
TAR(1)	AIC	−2392.27											
	BIC	−2382.6											
TAR(2)	AIC	−2456.36	−2450.52										
	BIC	−2441.84	−2436.02										
TAR(3)	AIC	−2491.12	−2480.9	−2485.91									
	BIC	−2471.78	−2461.56	−2466.57									
TAR(4)	AIC	−2493.51	−2488.4	−2496.56	−2527.82								
	BIC	−2469.33	−2464.22	−2472.39	−2503.64								
TAR(5)	AIC	−2511.57	−2503.17	−2511.45	−2539.21	−2516.78							
	BIC	−2482.56	−2474.16	−2482.44	−2510.2	−2487.77							
TAR(6)	AIC	−2526.35	−2520.19	−2529.66	−2555.3	−2527	−2544.93						
	BIC	−2492.5	−2486.34	−2495.81	−2521.46	−2493.15	−2511.08						
TAR(7)	AIC	−2537.59	−2528.99	−2537	−2556.02	−2528.55	−2552.95	−2541.77					
	BIC	−2498.91	−2490.31	−2498.32	−2517.34	−2489.87	−2514.27	−2503.09					
TAR(8)	AIC	−2545.8	−2533.07	−2548.22	−2561.1	−2532.29	−2558.35	−2549.36	−2584.92				
	BIC	−2502.28	−2489.56	−2504.7	−2517.58	−2488.77	−2514.84	−2505.84	**-2541.4**				
TAR(9)	AIC	−2552.35	−2540.81	−2551.7	−2571.71	−2549.79	−2570.42	−2552.2	−2586.79	−2555.68			
	BIC	−2504	−2492.46	−2503.35	−2523.36	−2501.44	−2522.07	−2503.85	−2538.44	−2507.32			
TAR(10)	AIC	−2550.97	−2559.02	−2550.33	−2573.27	−2552.75	−2573.15	−2552	−2585.35	−2555.94	−2590.06		
	BIC	−2497.78	−2505.83	−2497.14	−2520.09	−2499.56	−2519.97	−2498.81	−2532.16	−2502.75	−2536.87		
TAR(11)	AIC	−2554.45	−2558.34	−2551.88	−2574.83	−2554.98	−2573.36	−2550.31	−2586.24	−2558.59	−2590.91	−2587.25	
	BIC	−2496.43	−2500.31	−2493.86	−2516.81	−2496.96	−2515.34	−2492.28	−2528.22	−2500.57	−2532.88	−2529.23	
TAR(12)	AIC	−2566.71	−2568.93	−2562.71	−2580.2	−2561.69	−2579.22	−2559.51	−2586.48	−2564.03	**-2595.3**	−2590.1	−2591.1
	BIC	−2503.85	−2506.08	−2499.85	−2517.34	−2498.83	−2516.36	−2496.66	−2523.62	−2501.18	−2532.44	−2527.24	−2528.24

⋆ For each $1 \leq p \leq 12$, the AICs and BICs are calculated for TAR(p) models with the threshold variable y_{t-d}, where $1 \leq d \leq p$.

Table S9: Estimated coefficients of model (7.2).

μ	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8
−0.0920	−0.8766	−0.7905	−0.6514	−0.4799	−0.4184	−0.3661	−0.3173	0.2004
(0.0267)	(0.0638)	(0.0702)	(0.0676)	(0.064)	(0.0634)	(0.0683)	(0696)	(0.0701)
ν	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7	ψ_8
−0.0075	−0.5176	−0.244	−0.0893	−0.0022	−0.0695	−0.0718	0.0919	0.0605
(0.0115)	(0.0375)	(0.0446)	(0.0494)	(0.0525)	(0.0516)	(0.0489)	(0.0445)	(0.0509)

* Standard deviations are given in the parentheses.

Table S10: Estimated coefficients of model (7.3). (* not significant at 5% level
)

μ_1^*	ϕ_{11}	ϕ_{12}	ϕ_{13}^*	ϕ_{14}	ϕ_{15}	ϕ_{16}	ϕ_{17}^*	ϕ_{18}^*
−0.0089	−0.3828	−0.2246	−0.0934	−0.2101	−0.2518	−0.2296	−0.067	−0.0529
(0.0131)	(0.048)	(0.053)	(0.0521)	(0.0557)	(0.0545)	(0.0623)	(0.0618)	(0.0661)
ν_1^*	ψ_{11}	ψ_{12}	ψ_{13}^*	ψ_{14}^*	ψ_{15}^*	ψ_{16}^*	ψ_{17}	ψ_{18}^*
−0.0468	−0.2874	−0.4187	−0.1766	0.1702	0.0656	−0.0528	0.2138	−0.0253
(0.0452)	(0.0988)	(0.1038)	(0.1391)	(0.1113)	(0.1195)	(0.0836)	(0.0825)	(0.1151)
μ_2	ϕ_{21}	ϕ_{22}	ϕ_{23}	ϕ_{24}	ϕ_{25}	ϕ_{26}	ϕ_{27}	ϕ_{28}^*
−0.099	−1.1219	−1.0419	−0.9784	−0.7289	−0.6346	−0.698	−0.6104	0.1481
(0.0399)	(0.0854)	(0.0997)	(0.1063)	(0.1236)	(0.1469)	(0.1329)	(0.1172)	(0.0961)
ν_2^*	ψ_{21}	ψ_{22}	ψ_{23}	ψ_{24}^*	ψ_{25}	ψ_{26}	ψ_{27}^*	ψ_{28}^*
0.0085	−0.6389	−0.3462	−0.2101	−0.1731	−0.186	−0.1033	−0.0519	0.0355
(0.018)	(0.0589)	(0.0811)	(0.0939)	(0.0968)	(0.0847	(0.0807)	(0.073)	(0.0806)

* Standard deviations are given in the parentheses.

Table S11: The forecasting errors for the period 579-608 using the Models I, II and III, and 901-930 using Models IV and V. MSE denotes the average of the mean squared errors of the 30 forecasting values.

	I	II	III	IV	V
MSE	0.4961	0.1238	0.1868	0.0747	0.0601