On the Least Squares Estimation of Multiple-Threshold-Variable Autoregressive Models

Xinyu Zhang[*], Dong Li[†] and Howell Tong[‡]

Abstract

Most threshold models to-date contain a single threshold variable. However, in many empirical applications, models with multiple threshold variables may be needed and are the focus of this paper. For the sake of readability, we start with the two-threshold-variable autoregressive (2-TAR) model and study its least squares estimation (LSE). Among others, we show that the respective estimated thresholds are asymptotically independent. We propose a new method, namely the weighted Nadaraya-Watson method, to construct confidence intervals for the threshold parameters, that turns out to be, as far as we know, the only method to-date that enjoys good probability coverage, regardless of whether the threshold variables are endogenous or exogenous. Finally, we describe in some detail how our results can be extended to the K-threshold-variable autoregressive (K-TAR) model, $K > 2$. We assess the finite-sample performance of the LSE by simulation and present two real examples to illustrate the efficacy of our modelling.

Keywords: Compound Poisson process, degeneracy of a spatial process, multiple threshold variables, TAR model, weighted Nadaraya-Watson method.

[*]Xinyu Zhang, Department of Statistics and Actuarial Science, The University of Iowa, Iowa city, US (Email: xinyu-zhang@uiowa.edu).

[†]Dong Li, Center for Statistical Science, Department of Industrial Engineering, Tsinghua University, Beijing, China (Email: malidong@tsinghua.edu.cn). Corresponding author.

[‡]Howell Tong, University of Electronic Science and Technology of China, Chengdu, China; Tsinghua University, Beijing, China; London School of Economics and Political Science, London, UK (Email: howell.tong@gmail.com).

1 Introduction

The threshold autoregressive (TAR) model holds a prominent place in the nonlinear time series literature, since its introduction by Tong (1978). See also Tong and Lim (1980) and Tong (1990). It is capable of modeling many nonlinear phenomena, as well as often offering interpretable results in substantive fields, thus leading to its popularity. Indeed, it has extensive applications in many fields, such as economics (Hansen; 2011), finance (Chen et al.; 2011), ecology (Stenseth; 2009), epidemiology (Samia et al.; 2007), population dynamics (Chan et al.; 2009; Stenseth et al.; 1999) and actuarial science (Chan et al.; 2004), among others.

Most studies to-date of the TAR models are concerned with a *single* threshold variable. However, in many applications, we may need multiple threshold variables. Among several others, we cite two examples. Tong and Lim (1980) showed that improvements could be made by using both the "level" and "slope" as threshold variables. Using similar ideas, Tiao and Tsay (1994) separated the U.S. quarterly real Gross National Product(GNP) growth rate into four regimes according to the sign of a past growth rate and that of the first difference of two consecutive past growth rates as follows:

$$y_t = \begin{cases} 0.004 + 0.443 y_{t-1} + 0.0082\epsilon_t, & \text{if } y_{t-1} - y_{t-2} > 0,\ y_{t-2} > 0, \\ 0.006 + 0.438 y_{t-1} + 0.0094\epsilon_t, & \text{if } y_{t-1} - y_{t-2} \le 0,\ y_{t-2} > 0, \\ -0.015 - 1.076 y_{t-1} + 0.0062\epsilon_t, & \text{if } y_{t-1} - y_{t-2} \le 0,\ y_{t-2} \le 0, \\ -0.006 + 0.630 y_{t-1} - 0.756 y_{t-2} + 0.0132\epsilon_t, & \text{if } y_{t-1} - y_{t-2} > 0,\ y_{t-2} \le 0, \end{cases} \tag{1.1}$$

where the threshold parameters were *set* at zero. As we shall show later, it is more appropriate to estimate the latter from data.

This paper addresses multiple threshold variables that can be exogenous variables besides endogenous ones. For readability, we start by detailing the large sample theory for LSE of two threshold variables, the 2-TAR model, in the fixed-threshold-effect framework of Chan (1993), before generalizing it to the K-TAR model, $K > 2$. We show that Chan's large sample results carry over to the present context. However, his methodology deals with only one threshold variable. In the case of multiple threshold variables, it is important to study the independence or otherwise of the estimated threshold parameters. For this, we adopt a fundamentally different methodology and succeed in proving the asymptotic independence of the threshold parameters under some mild conditions. We stress that the asymptotic independence reduces substantially computation time for large samples, as well as facilitates the construction of their large-sample confidence intervals, because we can do so individually instead of jointly.

To simulate the limiting distribution of the estimated thresholds and derive confidence intervals for the threshold parameters had been open problems until Li and Ling (2012) proposed a resampling method to simulate the jump distribution. Unfortunately, their method is only for the self-exciting TAR model with an endogenous threshold variable at a single lag for reasons given in Section 4. We modify their method to allow for linear combinations of different lags of the endogenous threshold variable. However, to cope with both endogenous and exogenous threshold variables, we need a completely new method. This we achieve by developing the weighted Nadaraya-Watson (WNW) method. Both the WNW method and the modified resampling method are evaluated. We conduct simulations to assess the performance of the LSE in finite samples and showcase the efficacy of our approach with two real data sets.

Major contributions of our paper are as follows.

First, we introduce a framework for the multiple-threshold-variable TAR model that can cover two or more threshold variables and is without any need for a Gaussian assumption. It should not be confused with the single-threshold-variable-multiple-regime TAR model in Li and Ling (2012). Specifically, based on the classic TAR model with one indicator function, the generalization by Li and Ling (2012) has a number of shortcomings. Let us name just a couple of them here. First it is through a summation of indicator functions, implying the absence of interactions, while ours is through a multiplication of indicator functions involving interactions among the threshold variables. Interactions are clearly important and pose nontrivial and theoretically significant challenges. Next, Li and Ling (2012) can only cope with endogenous threshold variables, while ours can cope with both endogenous and exogenous threshold variables. Regarding Chen et al. (2012), we note that it is limited to only two threshold variables and, more significantly, to data driven by Gaussian errors. In contrast, our method enjoys much wider applicability beyond two threshold variables and beyond Gaussian errors. Later we discuss the poor performance of confidence intervals based on Chen et al. (2012).

Second, we prove the asymptotic independence of the estimated thresholds. It is a new and important result. To prove it, we discover that the methodology initiated by Chan (1993) is not appropriate. Instead, we utilize a fundamentally new methodology based on the degeneracy of a spatial process. It affords us substantial practical convenience for large samples in that instead of a time consuming two-dimensional joint search, two one-dimensional searches will suffice, thus reducing computational time by an order of magnitude.

Third, we develop a new method, the WNW method, to construct confidence intervals (CIs) for the threshold parameters. As far as we know, this is the first time that a valid method for exogenous threshold variables is developed. We stress that it overcomes the difficulties encountered in constructing CIs for the threshold parameters of such variables. To compare, under the diminishing-threshold-effect framework in Chen et al. (2012), CIs are constructed jointly for all thresholds instead of individually, which might be hard, if not impossible, to implement, especially for large samples or when K is large. Another important advantage of the new method is that it produces CIs with coverage probabilities that are much closer to the nominal levels, in sharp contrast to those produced by the methods of Chen et al. (2012); the latter tend to give misleading, and often conservative, coverage probabilities. For more detail, see Supplementary S.3. The favourable comparative results regarding CIs enjoyed by our method relative to the diminishing-threshold-effect framework of Hansen (2000) stems from the super n-consistency of the threshold estimates. We refer to Yu and Phillips (2018) and Li et al. (2019) for more discussion of the connection and difference between Chan's framework and Hansen's.

The remainder of the paper is organized as follows. Section 2 presents the 2-TAR model and its estimation. The asymptotic properties of the estimates are established in Section 3. Section 4 gives numerical methods to obtain the limiting distribution of the estimated thresholds. Section 5 reports simulation results. Section 6 analyzes two empirical examples. Section 7 addresses the general K-TAR case, $K \geq 2$. We conclude in Section 8. The Supplementary Material discusses issues of information criteria and regime structure specification, gives extended numerical algorithms, compares confidence intervals constructed under two frameworks with theory and simulation, and includes all technical proofs.

2 Model and Least Squares Estimation

A time series $\{y_t\}$ is said to follow a p-th order 2-TAR model if it satisfies

$$y_t = \sum_{j=1}^{4} (\boldsymbol{\beta}_j' \mathbf{y}_{t-1} + \sigma_j \epsilon_t) I_{jt}(r, s), \tag{2.1}$$

where $\mathbf{y}_{t-1} = (1, y_{t-1}, ..., y_{t-p})'$, $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, ..., \beta_{jp})' \in \mathbb{R}^{p+1}$, $\sigma_j > 0$, and

$$I_{1t}(r, s) = I(z_{t-1} > r, w_{t-1} > s), \quad I_{2t}(r, s) = I(z_{t-1} \le r, w_{t-1} > s),$$
$$I_{3t}(r, s) = I(z_{t-1} \le r, w_{t-1} \le s), \quad I_{4t}(r, s) = I(z_{t-1} > r, w_{t-1} \le s),$$

in which z_{t-1} and w_{t-1} are the threshold variables that classify $\{y_t\}$ into four regimes. Here, z_{t-1} and w_{t-1} are given real-valued random variables, measurable with respect to the natural filtration generated by $\{(y_{t-i}, v_{t-i}) : i \ge 1\}$, where v_t is an exogenous time series. Different choices of (z_{t-1}, w_{t-1}) can be used in applications. For example, they each can be exogenous or endogenous. Here $\boldsymbol{\tau} = (r, s)'$ is the threshold parameter; $\{\epsilon_t\}$ is independent and identically distributed (i.i.d.) with zero mean and unit variance, and is independent of the past information $\mathcal{F}_{t-1} = \sigma\{(y_{t-i}, z_{t-i}, w_{t-i}) : i \ge 1\}$. Let $e_t = \epsilon_t \sum_{j=1}^{4} \sigma_j I_{jt}(r, s)$.

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\tau}')' = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\beta}_3', \boldsymbol{\beta}_4', r, s)' \in \mathbb{R}^{4(p+1)+2}$. Let $\{y_1, ..., y_n\}$ denote the observations from model (2.1) with true parameter $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_{10}', \boldsymbol{\beta}_{20}', \boldsymbol{\beta}_{30}', \boldsymbol{\beta}_{40}', r_0, s_0)'$. Given initial values $\{y_{1-p}, ..., y_0\}$, the sum-of-squared-error function $L_n(\boldsymbol{\theta})$ is defined as

$$L_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} [y_t - \mathbb{E}_{\boldsymbol{\theta}}(y_t | \mathcal{F}_{t-1})]^2 = \sum_{t=1}^{n} \left[y_t - \sum_{j=1}^{4} \boldsymbol{\beta}_j' \mathbf{y}_{t-1} I_{jt}(\boldsymbol{\tau}) \right]^2.$$

Henceforth, $I_{jt}(\boldsymbol{\tau})$ and $I_{jt}(r, s)$ are used interchangeably. The minimizer of $L_n(\boldsymbol{\theta})$ is called the LSE of $\boldsymbol{\theta}_0$, i.e.,

$$\widehat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta} \in \Theta}{\arg\min}\, L_n(\boldsymbol{\theta}).$$

Since $L_n(\boldsymbol{\theta})$ is discontinuous in $\boldsymbol{\tau}$, a multi-parameter grid-search algorithm is needed. We can obtain $\widehat{\boldsymbol{\theta}}_n$ as follows.

- Fix $\boldsymbol{\tau}$, then minimize $L_n(\boldsymbol{\theta})$ and get its minimizer $\widehat{\boldsymbol{\beta}}_n(\boldsymbol{\tau})$ and minimum $L_n^*(\boldsymbol{\tau}) = L_n(\boldsymbol{\theta})|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_n(\boldsymbol{\tau})}$.

- As $L_n^*(\boldsymbol{\tau})$ only takes finitely many values, we can obtain the minimizer $\widehat{\boldsymbol{\tau}}_n$ of $L_n^*(\boldsymbol{\tau})$ by an enumeration approach.

- Using a plug-in method, we can finally obtain $\widehat{\boldsymbol{\beta}}_n \equiv \widehat{\boldsymbol{\beta}}_n(\widehat{\boldsymbol{\tau}}_n)$. Thus, $\widehat{\boldsymbol{\theta}}_n = (\widehat{\boldsymbol{\beta}}_n', \widehat{\boldsymbol{\tau}}_n')'$.

Generally, $\widehat{\boldsymbol{\tau}}_n \equiv (\widehat{r}_n, \widehat{s}_n)'$ takes the form of $(z_{(i)}, w_{(j)})'$, where $\{z_{(1)}, z_{(2)}, ..., z_{(n)}\}$ and $\{w_{(1)}, w_{(2)}, ..., w_{(n)}\}$ are, respectively, the order statistics of the observations $\{z_1, z_2, ..., z_n\}$ and $\{w_1, w_2, ..., w_n\}$. If $(z_{(i_0)}, w_{(j_0)})'$ is an estimator of $\boldsymbol{\tau}_0$ for some subscript (i_0, j_0), then $L_n^*(\boldsymbol{\tau})$ is a constant over the rectangular region $\{(r, s) : r \in [z_{(i_0)}, z_{(i_0+1)}), s \in [w_{(j_0)}, w_{(j_0+1)})\}$. Thus, there exist infinitely many $\boldsymbol{\tau}$s such that $L_n(\cdot)$ can achieve its global minimum and each can be considered an estimator of $\boldsymbol{\tau}_0$. In this case, we choose the lower left vertex of the rectangular region, $(z_{(i_0)}, w_{(j_0)})'$, as the estimator of $\boldsymbol{\tau}_0$.

For large data sets, Theorem 3.3 below justifies separate grid searches, one for each threshold parameter, and we can use the fast algorithm of Li and Tong (2016). For not so large data sets, a joint grid-search may be needed and time consuming since it involves two threshold parameters. Our experience suggests that we may relax the one-by-one enumeration by searching say every other point, so as to save substantial computation time with slight loss of precision.

Let σ_{j0} be the true value of σ_j for $j = 1, ..., 4$. Once $\widehat{\boldsymbol{\theta}}_n$ is obtained, we can estimate σ_{j0}^2 by

$$\widehat{\sigma}_{jn}^2 = \frac{1}{n_j} \sum_{t=1}^{n} (y_t - \widehat{\boldsymbol{\beta}}_{jn}' \mathbf{y}_{t-1})^2 I_{jt}(\widehat{\boldsymbol{\tau}}_n),$$

where $n_j = \sum_{t=1}^{n} I_{jt}(\widehat{\boldsymbol{\tau}}_n)$.

In practical applications, the unknown order of the j-th regime, say p_j, for each j needs to be specified. This challenging and open problem involves parameters that may not be all independently adjusted. For the self-exciting threshold autoregressive (SETAR) model with one threshold variable, there is a moderate amount of literature proposing different information criteria and assessing their performance. See, e.g., Wong and Li (1998), De Gooijer (2001), Peña and Rodriguez (2005), among others. A good choice is the AICu, which was proposed by McQuarrie et al. (1997) as an approximate unbiased estimator of

Kullback-Leibler information. It is shown, in the literature, that it enjoys competitive small sample performance and computational convenience for threshold models. However, further research is needed, especially for the case of multiple or exogenous threshold variables. As an interim suggestion, we adopt the following AICu

$$\mathrm{AICu}(\{p_j\}) = \sum_{j=1}^{4} \left[n_j \log(\widehat{\sigma}_{jn}^2) + \frac{2n_j(p_j+2)}{n_j - p_j - 3} + n_j \log\left(\frac{n_j}{n_j - p_j - 2} \right) \right]. \tag{2.2}$$

Alternatively, we can consider other model selection criteria such as the BIC (appropriately modified) and others. Supplementary S.1.1 discusses the issues of information criteria for threshold models. Throughout the paper, we assume that p_j's are known.

3 Asymptotic Properties

We first introduce several assumptions.

Assumption 3.1. *$\{\epsilon_t\}$ is i.i.d. with zero mean and unit variance. Its density $f_\epsilon(\cdot)$ is bounded, continuous and positive on $\mathbb{R}$.*

Assumption 3.2. *The parameter space Θ is a compact subset of $\mathbb{R}^{4(p+1)+2}$.*

Assumption 3.3. *Let $\{(y_t, z_t, w_t)\}$ be strictly stationary and ergodic, where $\{(z_t, w_t)\}$ are random vectors with a bounded, continuous and positive density $\pi(\cdot,\cdot)$ on $\mathbb{R}^2$. Denote the marginal density of z_t and w_t as $\pi_1(\cdot)$ and $\pi_2(\cdot)$, respectively. If $\{(z_t, w_t)\}$ are exogenous, we further assume they are Markovian.*

Assumptions 3.1 and 3.2 are standard. See Chan (1993) and Li and Ling (2012). If the model is endogenous with multiple different lagged variables as threshold variables and with homoscedastic errors, a sufficient condition for the strictly stationarity and ergodicity of y_t in Assumption 3.3 to hold is $\sum_i \max_j |\beta_{ji,0}| < 1$, as illustrated by Brachner et al. (2009). Now, we have the following theorem.

Theorem 3.1. *Suppose Assumptions 3.1-3.3 hold. We further assume (i) $\mathbb{E}y_t^2 < \infty$; (ii) there exists some $j_0 \in \{1, 2, 3, 4\}$ for which $\boldsymbol{\beta}_{j_0,0} \neq \boldsymbol{\beta}_{j_0+1,0}$ and $\boldsymbol{\beta}_{j_0,0} \neq \boldsymbol{\beta}_{j_0-1,0}$, with the convention $\boldsymbol{\beta}_{00} = \boldsymbol{\beta}_{40}$ and $\boldsymbol{\beta}_{50} = \boldsymbol{\beta}_{10}$. Then, $\widehat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ a.s. as $n \to \infty$.*

Condition (ii) in Theorem 3.1 guarantees the identification of $\boldsymbol{\tau}_0$, so that the threshold parameter can be identified not only under the four-regime case, but also when some regimes coalesce, resulting in a three-regime case or even a two-regime case, as illustrated in Figure 1. For example, Figure 1(b) means $\boldsymbol{\beta}_{10} \neq \boldsymbol{\beta}_{20}, \boldsymbol{\beta}_{20} \neq \boldsymbol{\beta}_{30}, \boldsymbol{\beta}_{30} = \boldsymbol{\beta}_{40}, \boldsymbol{\beta}_{40} \neq \boldsymbol{\beta}_{10}$,

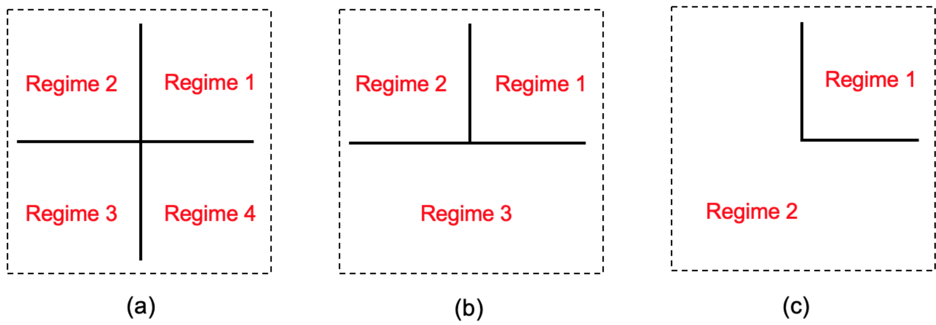

Figure 1: Cases where $\boldsymbol{\tau}_0 = (r_0, s_0)'$ can be identified: (a) four-regime case; (b) three-regime case;(c) two-regime case.

in which case the model has three regimes and condition (ii) is satisfied for $j_0 = 1$ and $j_0 = 2$. Figure 1(c) means $\boldsymbol{\beta}_{10} \neq \boldsymbol{\beta}_{20} = \boldsymbol{\beta}_{30} = \boldsymbol{\beta}_{40}$, resulting in two regimes and condition (ii) being satisfied for only $j_0 = 1$. However, not all two-regime cases are admissible; only those that satisfy condition (ii) are. For example, if $\boldsymbol{\beta}_{10} = \boldsymbol{\beta}_{20} \neq \boldsymbol{\beta}_{30} = \boldsymbol{\beta}_{40}$, we still have two regimes. However, condition (ii) is not satisfied and only one threshold, s_0, can be identified, so this case is not admissible. In the three- or two-regime cases, the model (2.1) and the parameters should be modified accordingly. In Supplementary S.1.2, we discuss specifying the regime structure using information criterion and we give a real data example in Section 6. Within this paper, we assume the 4-regime structure.

For the convergence rate and the limiting distribution of the estimated parameters, we introduce several assumptions, as follows. Let

$$
\begin{aligned}
&f_1(r)=\mathbb{E}\big(|w_t|\big|z_t=r\big), \quad g_1(s)=\mathbb{E}\big(|z_t|\big|w_t=s\big),\\
&f_2(r)=\mathbb{E}\big(\|\mathbf{y}_t\|^3\big|z_t=r\big), \quad g_2(s)=\mathbb{E}\big(\|\mathbf{y}_t\|^3\big|w_t=s\big),\\
&f_3(r)=\mathbb{E}\big(\|\mathbf{y}_t\mathbf{y}_t'\|\big|z_t=r\big), \quad g_3(s)=\mathbb{E}\big(\|\mathbf{y}_t\mathbf{y}_t'\|\big|w_t=s\big),\\
&f_4(r)=\mathbb{E}\big(\|\mathbf{y}_t\|^2|w_t|\big|z_t=r\big), \quad g_4(s)=\mathbb{E}\big(\|\mathbf{y}_t\|^2|z_t|\big|w_t=s\big),\\
&f_5(r)=\mathbb{E}\big(\|\mathbf{y}_te_{t+1}z_{t+1}\|\big|z_t=r\big), \quad g_5(s)=\mathbb{E}\big(\|\mathbf{y}_te_{t+1}z_{t+1}\|\big|w_t=s\big),\\
&f_6(r)=\mathbb{E}\big(\|\mathbf{y}_te_{t+1}w_{t+1}\|\big|z_t=r\big), \quad g_6(s)=\mathbb{E}\big(\|\mathbf{y}_te_{t+1}w_{t+1}\|\big|w_t=s\big).
\end{aligned}
\tag{3.1}
$$

Assumption 3.4. *Suppose $\mathbb{E}(\epsilon_t^4+y_t^4)<\infty$, and all $f_i(r)$'s in (3.1) are continuous at r_0, and $g_i(s)$'s at s_0, for $i=1,...,6$.*

Let $\mathbf{x}_t=(y_t,y_{t-1},...,y_{t-p+1},z_t,w_t)'$. Then $\{\mathbf{x}_t: t\geq 0\}$ is automatically a Markov chain with respect to its natural filtration. Denote its k-step transition probability by $\mathbf{P}^k(\mathbf{x},A)$, where $\mathbf{x}\in\mathbb{R}^{p+2}$ and A is a Borel set of $\mathbb{R}^{p+2}$.

Assumption 3.5. *$\{\mathbf{x}_t: t=0,1,...\}$ admits a unique invariant measure $\Pi(\cdot)$ such that there exist $K>0$ and $0<\rho<1$, for any $\mathbf{x}\in\mathbb{R}^{p+2}$ and any integer $k\geq 1$, $\|\mathbf{P}^k(\mathbf{x},\cdot)-\Pi(\cdot)\|_{\mathrm{v}}\leq K\rho^k(1+\|\mathbf{x}\|)$, where $\|\cdot\|_{\mathrm{v}}$ and $\|\cdot\|$ denote the total variation norm and the Euclidean norm, respectively.*

Assumption 3.6. *There exists at least one regime $j\in\{1,2,3,4\}$, in which $\Gamma'\boldsymbol{\beta}_{j0}$ is different from that of its two neighboring regimes. Here $\Gamma=\mathbb{E}(\mathbf{y}_t|z_t=r_0,w_t=s_0)$. Specifically, there is at least one $j\in\{1,2,3,4\}$ such that $\Gamma'(\boldsymbol{\beta}_{j0}-\boldsymbol{\beta}_{j-1,0})\neq 0$ and $\Gamma'(\boldsymbol{\beta}_{j0}-\boldsymbol{\beta}_{j+1,0})\neq 0$.*

The first part of Assumption 3.4 is standard, and the second part is mainly for cases with exogenous threshold variables. Under Assumption 3.5, $\{\mathbf{x}_t\}$ is V-uniformly ergodic with $V(\cdot)=K(1+\|\cdot\|)$, which is stronger than geometric ergodicity. For the concept of V-uniform ergodicity, see Chapter 16 in Meyn and Tweedie (2009). In the special case that all the threshold variables are lags of y_t and errors are homoskedastic across all regimes, a sufficient condition for Assumption 3.5 is Assumption 3.1 together with $\sum_i \max_j|\beta_{ji,0}|<1$. See Chan and Tong (1985) and Chan (1989). Assumption 3.6 implies that the autoregressive function is discontinuous at the threshold (r_0,s_0). The process y_t is generated from a fixed dynamic mechanism. It is different from Hansen (2000), who assumes that the parameters change with the sample size, n, and the magnitude of change goes to zero as $n \to \infty$. In contrast, Assumption 3.6 implies that for all $\mathbf{y}_t$ near Γ, $|\mathbf{y}_t'(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{j0})|$ exceeds a positive constant. The identifiability condition (ii) in Theorem 3.1 is a necessary condition of Assumption 3.6.

Theorem 3.2. *If Assumptions 3.1-3.6 hold and $\boldsymbol{\theta}_0$ is an interior point of Θ, then*

(i) $n\|\widehat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0\| = O_p(1)$;

(ii) $\sqrt{n}\sup_{\|\boldsymbol{\tau}-\boldsymbol{\tau}_0\|<B/n} \|\widehat{\boldsymbol{\beta}}_n(\boldsymbol{\tau}) - \widehat{\boldsymbol{\beta}}_n(\boldsymbol{\tau}_0)\| = o_p(1)$ *for any fixed* $B \in (0, \infty)$. *Further, for the four-regime case, as* $n \to \infty$,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = \sqrt{n}(\widehat{\boldsymbol{\beta}}_n(\boldsymbol{\tau}_0) - \boldsymbol{\beta}_0) + o_p(1) \to_d \mathcal{N}(0, \Sigma^{-1}),$$

where $\Sigma = \mathrm{diag}(\Sigma_1/\sigma_{10}^2, \Sigma_2/\sigma_{20}^2, \Sigma_3/\sigma_{30}^2, \Sigma_4/\sigma_{40}^2)$, *and* $\Sigma_j = \mathbb{E}[\mathbf{y}_{t-1}\mathbf{y}_{t-1}' I_{jt}(\boldsymbol{\tau}_0)]$ *for* $j = 1, ..., 4$.

For the three- or two-regime case, Σ should be modified accordingly.

Theorem 3.2(i) implies that the convergence rate of $\widehat{\boldsymbol{\tau}}_n$ is n, i.e. super-efficient. In order to establish the limiting distribution of $n(\widehat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0)$, we consider the following profile sum-of-squared-error function:

$$\bar{L}_n(u, v) = L_n\big(\widehat{\boldsymbol{\beta}}_n(r_0 + u/n, s_0 + v/n), r_0 + u/n, s_0 + v/n\big) - L_n\big(\widehat{\boldsymbol{\beta}}_n(\boldsymbol{\tau}_0), \boldsymbol{\tau}_0\big).$$

Using Theorem 3.2 and Taylor's expansion, similar to Li and Ling (2012), we can approximate $\bar{L}_n(u, v)$ in the function space $\mathbb{D}(\mathbb{R}^2)$ by

$$\tilde{L}_n(u, v) = L_n(\boldsymbol{\beta}_0, r_0 + u/n, s_0 + v/n) - L_n(\boldsymbol{\beta}_0, r_0, s_0),$$

where the definition of $\mathbb{D}(\mathbb{R}^2)$ is given in subsection 7.1 of Li and Ling (2012). Define

$$\begin{aligned}
\tilde{R}_n(u) =& L_n\big(\boldsymbol{\beta}_0, r_0 + u/n, s_0\big) - L_n\big(\boldsymbol{\beta}_0, r_0, s_0\big) \\
=& \sum_{t=1}^{n} \big[\gamma_t^{(1)} I\big(r_0 < z_{t-1} \le r_0 + u/n\big) I(u > 0) + \gamma_t^{(2)} I\big(r_0 + u/n < z_{t-1} \le r_0\big) I(u \le 0)\big], \\
\tilde{Q}_n(v) =& L_n\big(\boldsymbol{\beta}_0, r_0, s_0 + v/n\big) - L_n\big(\boldsymbol{\beta}_0, r_0, s_0\big) \\
=& \sum_{t=1}^{n} \big[\gamma_t^{(3)} I\big(s_0 < w_{t-1} \le s_0 + v/n\big) I(v > 0) + \gamma_t^{(4)} I\big(s_0 + v/n < w_{t-1} \le s_0\big) I(v \le 0)\big].
\end{aligned}$$

Here

$$
\begin{aligned}
\gamma_t^{(1)} =& \xi_t^{(1,2)} I(w_{t-1} > s_0) + \xi_t^{(4,3)} I(w_{t-1} \le s_0), \\
\gamma_t^{(2)} =& \xi_t^{(2,1)} I(w_{t-1} > s_0) + \xi_t^{(3,4)} I(w_{t-1} \le s_0), \\
\gamma_t^{(3)} =& \xi_t^{(1,4)} I(z_{t-1} > r_0) + \xi_t^{(2,3)} I(z_{t-1} \le r_0), \\
\gamma_t^{(4)} =& \xi_t^{(4,1)} I(z_{t-1} > r_0) + \xi_t^{(3,2)} I(z_{t-1} \le r_0),
\end{aligned} \tag{3.2}
$$

with

$$
\xi_t^{(i,j)} = [(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{j0})'\mathbf{y}_{t-1}]^2 + 2\sigma_{i0}\epsilon_t(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{j0})'\mathbf{y}_{t-1}, \quad i, j = 1, ..., 4.
$$

A significant result is given in Proposition 3.1, the proof of which can be found in the Supplementary, and some intuitive discussion can be found in Remark 3.1.

Proposition 3.1. *If Assumptions 3.1–3.6 hold, then* $\sup_{|u|,|v|\le B} \left|\tilde{L}_n(u,v) - \tilde{R}_n(u) - \tilde{Q}_n(v)\right| = o_p(1)$ *for any fixed* $B \in (0, \infty)$*.*

By Proposition 3.1, the process $\tilde{L}_n(u,v)$ of (u,v) can be written as the summation of a process of u and a process of v, while the cross term of (u,v) will degenerate, which is the key to the asymptotic independence in the following Theorem 3.3.

We define two independent one-dimensional two-sided compound Poisson processes $\{\mathcal{P}_1(u), u \in \mathbb{R}\}$ and $\{\mathcal{P}_2(v), v \in \mathbb{R}\}$ as

$$
\mathcal{P}_1(u) = I(u > 0) \sum_{k=1}^{N_1(u)} \zeta_k^{(1)} + I(u \le 0) \sum_{k=1}^{N_2(-u)} \zeta_k^{(2)}, \tag{3.3}
$$

$$
\mathcal{P}_2(v) = I(v > 0) \sum_{k=1}^{N_3(v)} \zeta_k^{(3)} + I(v \le 0) \sum_{k=1}^{N_4(-v)} \zeta_k^{(4)}, \tag{3.4}
$$

where $\{N_1(u), u \ge 0\}$ and $\{N_2(u), u \ge 0\}$ are two independent Poisson processes with $N_1(0) = N_2(0) = 0$ a.s. and the same jump rate $\pi_1(r_0)$. Here $\{\zeta_k^{(1)} : k \ge 1\}$ are i.i.d. from $F_1(\cdot|r_0)$ and $\{\zeta_k^{(2)} : k \ge 1\}$ from $F_2(\cdot|r_0)$, and they are mutually independent, where $F_1(\cdot|r_0)$ is the conditional distribution of $\gamma_2^{(1)}$ given $z_1 = r_0$, and $F_2(\cdot|r_0)$ that of $\gamma_2^{(2)}$.

Similarly, $\{N_3(v), v \ge 0\}$ and $\{N_4(v), v \ge 0\}$ are two independent Poisson processes with $N_3(0) = N_4(0) = 0$ a.s. and the same jump rate $\pi_2(s_0)$. Also $\{\zeta_k^{(3)} : k \ge 1\}$ are i.i.d. from $G_1(\cdot|s_0)$ and $\{\zeta_k^{(4)} : k \ge 1\}$ from $G_2(\cdot|s_0)$, and they are mutually independent, where $G_1(\cdot|s_0)$ is the conditional distribution of $\gamma_2^{(3)}$ given $w_1 = s_0$, and $G_2(\cdot|s_0)$ that of

$\gamma_2^{(4)}$. Here, we work with the right continuous version for $N_1(u)$ and $N_3(v)$, and the left continuous version for $N_2(u)$ and $N_4(v)$.

We further define a spatial compound Poisson process

$$\psi(u, v) = \mathcal{P}_1(u) + \mathcal{P}_2(v).$$

Clearly, $\psi(u, v)$ goes to $+\infty$ a.s. when $|u|$, $|v| \to \infty$ since all jump distributions have positive means by Assumption 3.6. Therefore, there exists a unique 2-dimensional cube $[\boldsymbol{M}_{-}, \boldsymbol{M}_{+}) \equiv [M_{-}^{(1)}, M_{+}^{(1)}) \times [M_{-}^{(2)}, M_{+}^{(2)})$ on which the process $\psi(u, v)$ attains its global minimum a.s., namely,

$$[\boldsymbol{M}_{-}, \boldsymbol{M}_{+}) = \arg \min_{(u,v)\in\mathbb{R}^2} \psi(u, v).$$

Actually, since $\mathcal{P}_1(u)$ and $\mathcal{P}_2(v)$ are independent, the minimization above is equivalent to

$$[M_{-}^{(1)}, M_{+}^{(1)}) = \arg \min_{u\in\mathbb{R}} \mathcal{P}_1(u), \quad [M_{-}^{(2)}, M_{+}^{(2)}) = \arg \min_{v\in\mathbb{R}} \mathcal{P}_2(v).$$

Accordingly, $M_{-}^{(1)}$ and $M_{-}^{(2)}$ are independent. Now, we can state our result.

Theorem 3.3. *If Assumptions 3.1–3.6 hold, then $n(\widehat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0)$ converges weakly to $\boldsymbol{M}_{-}$ and its components are asymptotically independent as $n \to \infty$. Furthermore, $n(\widehat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0)$ is asymptotically independent of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{\mathbf{0}})$.*

Remark 3.1. *At first sight, the asymptotic independence might seem counter-intuitive since the two threshold parameters are not separable in the modeling procedure. However, it is a direct result of Proposition 3.1, which proves the degeneracy of the cross term of (u, v). Figure 2 gives an intuitive graphical representation of $\tilde{L}_n(u, v)$ when $u, v > 0$, which is nonzero only in the grey "road" zones. Intuitively, $\tilde{R}_n(u)$ and $\tilde{Q}_n(v)$ are related to $I\big(r_0 < z_{t-1} \leq r_0 + u/n\big)$ (the vertical "road") and $I\big(s_0 < w_{t-1} \leq s_0 + v/n\big)$ (the horizontal "road"), respectively, while the cross term is related to $I\big(r_0 < z_{t-1} \leq r_0 + u/n, s_0 < w_{t-1} \leq s_0 + v/n\big)$ (the "crossing" zone). When $\boldsymbol{\tau}_n$ approaches $\boldsymbol{\tau}_0$ at the rate of n, by Assumption 3.3 on the density $\pi(\cdot, \cdot)$, $\pi_1(\cdot)$ and $\pi_2(\cdot)$, the area of $\tilde{R}_n(u)$ and $\tilde{Q}_n(v)$ will shrink at the same rate of n. However, the "crossing" area will shrink at a faster rate of n^2 and is thus negligible, resulting in the asymptotic independence between the two estimated thresholds.*

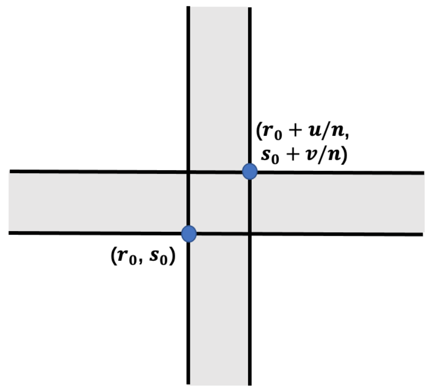

Figure 2: An intuitive "crossroad" illustration of $\tilde{L}_n(u, v)$ and the asymptotic independence between the two estimated thresholds.

In practice, z_t and w_t may be highly correlated, such as $z_t = y_{t-1} - y_{t-2}$ and $w_t = y_{t-2}$ in the motivating example (1.1). The correlation between z_t and w_t will not affect the asymptotic independence, as long as Assumption 3.3 on the density $\pi(\cdot, \cdot)$ is satisfied. A violation of Assumption 3.3 occurs when z_t and w_t are exactly the same, in which case $\pi(\cdot, \cdot)$ is not well defined. In fact, this case reduces to the single-threshold-multiple-regime TAR model in Li and Ling (2012), and asymptotic independence still holds as proved by them.

4 Numerical Implementation of $\boldsymbol{M}_{-}$

The distribution of $\boldsymbol{M}_{-}$ does not have a closed form. From Theorem 3.3, $\boldsymbol{M}_{-}$ is equivalent to $M_{-}^{(1)}$ and $M_{-}^{(2)}$ separately. Modifying Algorithm 6.2 in Cont and Tankov (2004), we obtain an algorithm to output a sequence of $M_{-}^{(1)}$, by simulating the two-sided compound Poisson process on the interval $[-T, T]$, for any given $T > 0$ large enough. We can deal with $M_{-}^{(2)}$ similarly and thus omit the details.

Algorithm A:

Step A.1 Simulate two independent Poisson random variables N_1 and N_2 with the same parameter $\pi_1(r_0)T$, which are the total number of jumps on the intervals $[0, T]$ and $[-T, 0]$, respectively.

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

Step A.3 Simulate two independent jump-size sequences: $\{\zeta_1^{(1)}, ..., \zeta_{N_1}^{(1)}\}$ from $F_1(\cdot|r_0)$, and $\{\zeta_1^{(2)}, ..., \zeta_{N_2}^{(2)}\}$ from $F_2(\cdot|r_0)$.

Step A.4 For $u \in [-T, T]$, the trajectory of $\mathcal{P}_1(u)$ in (3.3) is given by

$$\mathcal{P}_1(u) = I(u > 0) \sum_{k=1}^{N_1} I(U_k < u)\zeta_k^{(1)} + I(u \leq 0) \sum_{k=1}^{N_2} I(V_k > u)\zeta_k^{(2)}.$$

Take the smallest minimizer of $\mathcal{P}_1(u)$ on $[-T, T]$ as an observed value of $M_-^{(1)}$.

Step A.5 Repeating Step A.1-A.4, we can obtain a sequence of observations of $M_-^{(1)}$.

In practice, Step A.1 can be implemented by substituting $\pi_1(r_0)$ by its nonparametric kernel estimate. However, the difficulty lies in Step A.3, which involves sampling from conditional distributions $F_1(\cdot|r_0)$ and $F_2(\cdot|r_0)$. It remained a problem until Li and Ling (2012) developed a resampling method for self-exciting TAR models, whose key idea is to estimate the jump distribution by simulating the time series y_t following the TAR mechanism conditional on $z_{t-1} = r_0$. As such, this procedure is invalid for exogenous threshold variables. Thus, to implement Step A.3 in the general setting, we need to take a new route. In subsection 4.1, we develop the weighted Nadaraya-Watson (WNW) method. The WNW method estimates the conditional distribution from the viewpoint of nonparametric regression, so it is valid whether the threshold variable is endogenous or exogenous. As far as we know, this is the first valid method to handle the exogenous case. Besides, in subsection 4.2, we give a modified resampling method when the threshold variables are linear functions of lags of y_t.

Conditional on $\mathcal{X} \equiv \{y_{1-p}, ..., y_n; z_0, \ldots, z_{n-1}; w_0, \ldots, w_{n-1}\}$ from model (2.1), we give two estimators of $F_1(\cdot|r_0)$, namely $\widehat{F}_1^{\mathrm{WNW}}(\gamma|r_0, \mathcal{X})$ and $\widehat{F}_1^{\mathrm{RS}}(\gamma|r_0, \mathcal{X})$ with the WNW and modified resampling methods respectively, and illustrate how to sample from them. For simplicity, in subsections 4.1 and 4.2, we assume the ideal case that the true model is known, i.e., $\boldsymbol{\theta}_0$, σ_{j0}, $\pi_1(r_0)$, $f_\epsilon(\cdot)$ and $F_\epsilon(\cdot)$, which are respectively the density and distribution function of ϵ_t, are all known. We give the algorithms with estimated parameters in Supplementary S.2. The procedures are similar for $F_2(\cdot|r_0)$ and thus omitted.

4.1 The Weighted Nadaraya-Watson Method

Needs for estimating a conditional distribution arise in settings such as quantile regression and prediction interval of time series. Some nonparametric methods have been proposed, including the local linear estimators in Yu and Jones (1998), the local logistic estimator in Hall et al. (1999), the weighted Nadaraya-Watson estimator in Hall et al. (1999) and Cai (2002), among others. The WNW estimator enjoys two nice properties. It reproduces superior properties of local linear estimators in respect of bias and automatic boundary behavior, and it preserves the property of always being a distribution function. Given $\mathcal{X}$, $\boldsymbol{\theta}_0$, and σ_{j0}, we can obtain $\gamma_t^{(1)}$ in (3.2) for $t = 1, \ldots, n$. We introduce the WNW estimator for $F_1(\gamma|z)$, the conditional distribution of $\gamma_t^{(1)}$ given $z_{t-1} = z$.

Let $p_t(z)$ for $1 \le t \le n$, denote the weight functions of the data z_{t-1} and the point z with the property

$$p_t(z) \ge 0, \quad \sum_{t=1}^{n} p_t(z) = 1, \quad \text{and} \quad \sum_{t=1}^{n} (z_{t-1} - z) p_t(z) K_h(z - z_{t-1}) = 0, \tag{4.1}$$

where $K(\cdot)$ is a kernel function, $K_h(\cdot) = K(\cdot/h)/h$, and $h = h_n$ is the bandwidth. Then, the WNW estimator of the conditional distribution $F_1(\gamma|z)$ is defined as

$$\widehat{F}_1^{\mathrm{WNW}}(\gamma|z, \mathcal{X}) = \frac{\sum_{t=1}^{n} p_t(z) K_h(z - z_{t-1}) I(\gamma_t^{(1)} \le \gamma)}{\sum_{t=1}^{n} p_t(z) K_h(z - z_{t-1})}.$$

We specify $p_t(z)$ by maximizing $\sum_{t=1}^{n} \log\{p_t(z)\}$ subject to the constraints in (4.1) through the Lagrange multiplier, which gives the solution: $p_t(z) = n^{-1}\{1 + \lambda(z_{t-1} - z) K_h(z - z_{t-1})\}^{-1}$, where λ is uniquely defined by (4.1) and can be found with the Newton-Raphson scheme in implementation. As for the selection of the bandwidth h_n, we adopt the nonparametric Akaike information criterion proposed by Cai and Tiwari (2000).

Thus, using the following algorithm, we derive the estimator $\widehat{F}_1^{\mathrm{WNW}}(\gamma|r_0)$ conditional on $\mathcal{X}$ and sample from it.

Algorithm B:

Step B.1 For a fixed γ,

$$\widehat{F}_1^{\mathrm{WNW}}(\gamma|r_0, \mathcal{X}) = \frac{\sum_{t=1}^{n} p_t(r_0) K_h(r_0 - z_{t-1}) I(\gamma_t^{(1)} \le \gamma)}{\sum_{t=1}^{n} p_t(r_0) K_h(r_0 - z_{t-1})}.$$

Step B.2 $\widehat{F}_1^{\mathrm{WNW}}(\cdot|r_0, \mathcal{X})$ is defined as the cumulative distribution function of a discrete distribution taking values at $\{\gamma_t^{(1)} : t = 1, \ldots, n\}$ with the accorded cumulative probability $\widehat{F}_1^{\mathrm{WNW}}(\gamma_t^{(1)}|r_0, \mathcal{X})$. Draw a random sample from this discrete distribution, and denote it as $\zeta_1^{(1),\mathrm{WNW}}$.

By Theorem 1 in Cai (2002), $\widehat{F}_1^{\mathrm{WNW}}(\cdot|r_0, \mathcal{X})$ is a consistent estimator of $F_1(\cdot|r_0)$, which is summarized as follows.

Proposition 4.1. *Besides Assumptions 3.1–3.6, further assume the following: (i) $F_1(\cdot|z)$ has continuous second-order derivative with respect to z. (ii) The kernel function $K(\cdot)$ is a symmetric, bounded and compactly supported density. (iii) As $n \to \infty$, $h \to 0$ and $nh \to \infty$. (iv) Let $\pi_1^{(1,t)}(\cdot, \cdot)$ be the joint density of z_1 and z_t for $t \geq 2$. Assume $|\pi_1^{(1,t)}(u, v) - \pi_1(u)\pi_1(v)| \leq C < \infty$ for all u and v and some C. Then, as $n \to \infty$,*

$$\left|\widehat{F}_1^{\mathrm{WNW}}(\gamma|r_0, \mathcal{X}) - F_1(\gamma|r_0)\right| \to 0, \quad \textit{in probability.}$$

4.2 The Modified Resampling Method

When the threshold variables are both linear functions of the lags of y_t, we assume $z_{t-1} = l(y_{t-d_1}, ..., y_{t-d_m})$, where $l : \mathbb{R}^m \to \mathbb{R}$ is a known linear function and $1 \leq d_1 \leq ... \leq d_m \leq p$. In order to estimate $F_1(\gamma|r_0) = \mathbb{P}(\gamma_2^{(1)} \leq \gamma|z_1 = r_0)$, the key idea of the resampling method is based on conditional arguments of $\boldsymbol{Q} = (y_{1-d_1}, ..., y_{2-d_1-p})'$. Dilatate the function l to $l^* : \mathbb{R}^{p+1} \to \mathbb{R}$ by appending the linear coefficients of l with $(p + 1 - m)$ 0s. Specifically,

$$z_1 = l(y_{2-d_1}, ..., y_{2-d_m}) = l^*(y_{2-d_1}, \ldots, y_{2-d_m}, \ldots, y_{2-d_1-p}) = l^*(y_{2-d_1}, \boldsymbol{Q}).$$

Then, conditional on $\boldsymbol{Q} = \boldsymbol{q} \in \mathbb{R}^p$, z_1 and y_{2-d_1} have the mapping relationship that $z_1 = l_{\boldsymbol{q}}(y_{2-d_1}) \equiv l^*(y_{2-d_1}, \boldsymbol{Q} = \boldsymbol{q})$. Denote $\pi_1(\cdot|\boldsymbol{q})$ and $\phi(\cdot|\boldsymbol{q})$ as the conditional densities of z_1 and y_{2-d_1} given $\boldsymbol{Q} = \boldsymbol{q}$, respectively. Rewrite $I_{jt}(\boldsymbol{\tau})$ as $I_{jt}(\mathbf{y}_{t-1}, \boldsymbol{\tau})$ just to emphasize the fact that z_{t-1} and w_{t-1} are both linear functions of $\mathbf{y}_{t-1}$. Let $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3, \sigma_4)$, and let $\boldsymbol{\sigma}_0$ be the true value. Define

$$h(\mathbf{y}_{t-1}, \boldsymbol{\theta}) = \sum_{j=1}^{4} \boldsymbol{\beta}_j' \mathbf{y}_{t-1} I_{jt}(\mathbf{y}_{t-1}, \boldsymbol{\tau}), \quad \nu(\mathbf{y}_{t-1}, \boldsymbol{\theta}, \boldsymbol{\sigma}) = \sum_{j=1}^{4} \sigma_j I_{jt}(\mathbf{y}_{t-1}, \boldsymbol{\tau}).$$

Let $\boldsymbol{q}_i = (y_i, ..., y_{i-p+1})'$, and $\tilde{\boldsymbol{q}}_i = (1, \boldsymbol{q}_i)$. Then,

$$\begin{aligned}
\phi(x|\boldsymbol{q}_i) &= [\nu(\tilde{\boldsymbol{q}}_i, \boldsymbol{\theta}_0, \boldsymbol{\sigma}_0)]^{-1} f_\epsilon\big([\nu(\tilde{\boldsymbol{q}}_i, \boldsymbol{\theta}_0, \boldsymbol{\sigma}_0)]^{-1}[x - h(\tilde{\boldsymbol{q}}_i, \boldsymbol{\theta}_0)]\big) \\
\pi_1(r_0|\boldsymbol{q}_i) &= \phi(l^{-1}_{\boldsymbol{q}_i}(r_0)|\boldsymbol{q}_i) \left|(l^{-1}_{\boldsymbol{q}_i})'(r_0)\right|.
\end{aligned} \tag{4.2}$$

where $(l^{-1}_{\boldsymbol{q}_i})'(r_0)$ is the derivative at r_0, and is a constant among all q_i.

By the property of conditional expectation and the strong law of large numbers, it follows that

$$\begin{aligned}
F_1(\gamma|r_0) =& \mathbb{P}(\gamma_2^{(1)} \leq \gamma | z_1 = r_0) \\
=& \int_{\mathbb{R}^p} \mathbb{P}(\gamma_2^{(1)} \leq \gamma | z_1 = r_0, \boldsymbol{Q} = \boldsymbol{q}) \frac{\pi_1(r_0|\boldsymbol{q})}{\pi_1(r_0)} H(d\boldsymbol{q}) \\
=& \sum_{i=1}^n \mathbb{P}(\gamma_2^{(1)} \leq \gamma | z_1 = r_0, \boldsymbol{Q} = \boldsymbol{q}_i) \frac{\pi_1(r_0|\boldsymbol{q}_i)}{\sum_{j=1}^n \pi_1(r_0|\boldsymbol{q}_j)} + o(1) \\
\equiv& \widehat{F}_1^{\mathrm{RS}}(\gamma|r_0, \mathcal{X}) + o(1),
\end{aligned} \tag{4.3}$$

a.s. as $n \to \infty$, uniformly in $\gamma \in \mathbb{R}$ by Theorem 2 in Pollard (1984), where $H(\cdot)$ is the distribution of $\boldsymbol{Q}$. It means that $\widehat{F}_1^{\mathrm{RS}}(\gamma|r_0, \mathcal{X})$ is a uniformly consistent estimator of $F_1(\gamma|r_0)$. Then, we can sample a $\zeta_1^{(1),\mathrm{RS}}$ from $\widehat{F}_1^{\mathrm{RS}}(\gamma|r_0, \mathcal{X})$ following **Algorithm C**.

Algorithm C:

Step C.1 For each $i = 1, ..., n$, set $\boldsymbol{q}_i = (y_i, ..., y_{i-p+1})'$ and generate a sample $\{y_1^{(i)}, \ldots, y_{2-d_1-p}^{(i)}\}$ in the following way: first let $(y_{1-d_1}^{(i)}, ..., y_{2-d_1-p}^{(i)})' = \boldsymbol{q}_i$ and $y_{2-d_1}^{(i)} = l^{-1}_{\boldsymbol{q}_i}(r_0)$; draw $\{\epsilon_{3-d_1}^{(i)}, ..., \epsilon_1^{(i)}\}$ independently from $F_\epsilon(\cdot)$ and generate $y_{3-d_1}^{(i)}, ..., y_1^{(i)}$ by iterating model (2.1) based on the initial values $\{y_{2-d_1}^{(i)}, y_{1-d_1}^{(i)}, ..., y_{2-d_1-p}^{(i)}\}$. Then, calculate $\gamma_2^{(1)}$ in (3.2) based on the sample $\{y_1^{(i)}, \ldots, y_{2-d_1-p}^{(i)}\}$ and denote it as ζ_i.

Step C.2 Calculate $\pi_1(r_0|\boldsymbol{q}_i)$'s in (4.2) for $i = 1, ..., n$ and draw a U from the discrete distribution: $\mathbb{P}(U = i|\mathcal{X}) = \pi_1(r_0|\boldsymbol{q}_i)/\{\sum_{j=1}^n \pi_1(r_0|\boldsymbol{q}_j)\}$, independent of all $\{\epsilon_{3-d_1}^{(i)}, ..., \epsilon_1^{(i)}\}$.

Step C.3 Obtain $\zeta_1^{(1),\mathrm{RS}} = \zeta_U$.

Compared to the original resampling algorithm for self-exciting threshold models in Li and Ling (2012), our algorithm differs in the design of $\boldsymbol{Q}$ in order to adapt to threshold variables in the form of linear functions of lags of y_t.

4.3 Theoretical Justification of the Algorithms

Now, for the ideal case that the true model is known, **Algorithm A** can be implemented. Define a two-sided compound Poisson process by $\mathcal{P}_1^{\mathrm{WNW}}(u)$ which is determined by the jump rate $\pi_1(r_0)$ and jump distributions $\widehat{F}_1^{\mathrm{WNW}}(\gamma|r_0,\mathcal{X})$ and $\widehat{F}_2^{\mathrm{WNW}}(\gamma|r_0,\mathcal{X})$. Similarly define $\mathcal{P}_1^{\mathrm{RS}}(u)$ by $\pi_1(r_0)$, $\widehat{F}_1^{\mathrm{RS}}(\gamma|r_0,\mathcal{X})$ and $\widehat{F}_2^{\mathrm{RS}}(\gamma|r_0,\mathcal{X})$. Note that every compound Poisson process is a stationary independent increment process. By Proposition 4.1, (4.3) and Theorem 16 in Pollard (1984), we have in probability

$$\mathcal{P}_1^{\mathrm{WNW}}(u) \Rightarrow \mathcal{P}_1(u), \quad \mathcal{P}_1^{\mathrm{RS}}(u) \Rightarrow \mathcal{P}_1(u), \quad \text{in } D(\mathbb{R})$$

conditionally on $\mathcal{X}$ as $n \to \infty$, where $\Rightarrow$ denotes weak convergence. Minimizing the processes $\mathcal{P}_1^{\mathrm{WNW}}(u)$ and $\mathcal{P}_1^{\mathrm{RS}}(u)$, we obtain the smallest minimizer M_1^{WNW} and M_1^{RS}, respectively. By Theorem 3.1 in Seijo and Sen (2011), we have the following result.

Theorem 4.1. *(i) Under the conditions in Proposition 4.1, we have in probability that* $\lim_{n\to\infty} |\mathbb{P}(M_1^{\mathrm{WNW}} \le x|\mathcal{X}) - \mathbb{P}(M_-^{(1)} \le x)| = 0$ *at each x for which* $\mathbb{P}(M_-^{(1)} = x) = 0$.

(ii) Suppose Assumptions 3.1–3.6 hold, and further assume the threshold variables z_{t-1} and w_{t-1} are linear functions of $\mathbf{y}_{t-1}$. Then, we have in probability that $\lim_{n\to\infty} |\mathbb{P}(M_1^{\mathrm{RS}} \le x|\mathcal{X}) - \mathbb{P}(M_-^{(1)} \le x)| = 0$ *at each x for which* $\mathbb{P}(M_-^{(1)} = x) = 0$.

Since all parameters $\boldsymbol{\theta}_0$, σ_{j0}, $\pi_1(r_0)$, $f_\epsilon(\cdot)$ and $F_\epsilon(\cdot)$ are unknown in practice, we first use the sample $\mathcal{X}$ to estimate them consistently and then extend **Algorithm B** and **C** by substituting all parameters by their estimate. We follow Li and Ling (2012) for the details of estimation and the extended algorithm, and thus relegate them to the Supplementary S.2. For the extended algorithm, denote the counterparts of M_1^{WNW} and M_1^{RS} by $\hat{M}_1^{\mathrm{WNW}}$ and $\hat{M}_1^{\mathrm{RS}}$. Then, both $\hat{M}_1^{\mathrm{WNW}}$ and $\hat{M}_1^{\mathrm{RS}}$ converge weakly to $M_-^{(1)}$ conditionally on $\mathcal{X}$, in probability, which is summarized in the following theorem the proof of which can be found in Supplementary S.3.

Theorem 4.2. *(i) Under the conditions in Proposition 4.1, we have in probability that* $\lim_{n\to\infty} |\mathbb{P}(\hat{M}_1^{\mathrm{WNW}} \le x|\mathcal{X}) - \mathbb{P}(M_-^{(1)} \le x)| = 0$ *at each x for which* $\mathbb{P}(M_-^{(1)} = x) = 0$.

(ii) Suppose Assumptions 3.1–3.6 hold, and further assume the threshold variables z_{t-1} and w_{t-1} are linear functions of $\mathbf{y}_{t-1}$. Assume $f_\epsilon(\cdot)$ is uniformly continuous on $\mathbb{R}$. Then,

we have in probability that $\lim_{n\to\infty} |\mathbb{P}(\hat{M}_1^{\mathrm{RS}} \leq x|\mathcal{X}) - \mathbb{P}(M_-^{(1)} \leq x)| = 0$ *at each* x *for which* $\mathbb{P}(M_-^{(1)} = x) = 0$.

5 Simulation Studies

In order to assess the performance of the LSE of $\boldsymbol{\theta}_0$ in finite samples, we consider the following 2-TAR model:

$$y_t = \begin{cases} 1 + 0.4y_{t-1} + 0.8\epsilon_t, & \text{if } y_{t-1} - y_{t-2} > 0,\ y_{t-2} > 0, \\ 0.3y_{t-1} + 0.5y_{t-2} + 1.3\epsilon_t, & \text{if } y_{t-1} - y_{t-2} \leq 0,\ y_{t-2} > 0, \\ 0.5 + 0.7y_{t-1} + 0.1y_{t-2} + 0.9\epsilon_t, & \text{if } y_{t-1} - y_{t-2} \leq 0,\ y_{t-2} \leq 0, \\ -0.8y_{t-1} + 0.4y_{t-2} + 0.8\epsilon_t, & \text{if } y_{t-1} - y_{t-2} > 0,\ y_{t-2} \leq 0, \end{cases} \tag{5.1}$$

where ϵ_t is i.i.d from $\mathcal{N}(0,1)$. Let the sample sizes be $n = 300$, 600, 900, and 1200, each with 1000 replications. For this model, the two thresholds are respectively the 48% and 45% quantile of the according threshold variables. The average proportions of observations in the four regimes are 16.8% : 37.7% : 10.4% : 35.1%.

Table 1 summarizes the bias, the empirical standard deviation (ESD) and the asymptotic standard deviation (ASD). The ASDs of $\widehat{\boldsymbol{\beta}}_n$ are provided by Theorem 3.2 in closed form, however they can not be directly computed based on the model due to the difficulty in computing Σ_j. Thus, to obtain Σ_j, we first draw a large number of long series following (5.1), randomly and independently. Take the averaged empirical Σ_j as the true Σ_j, based on which we obtain the ASDs of $\widehat{\boldsymbol{\beta}}_n$. The ASDs of $\widehat{r}_n$ and $\widehat{s}_n$ are estimated by the WNW and resampling method, with the superscript ^{W} and ^{R} respectively. Table 1 shows that the ESDs and ASDs become closer with increasing sample size. In line with the n-consistency, the ESDs for $\widehat{r}_n$ and $\widehat{s}_n$ are about halved with n doubled.

To show the asymptotic normality of $\widehat{\boldsymbol{\beta}}_n$, we present the densities of each of its elements. Take β_{30} as a typical example. Figure 3 displays the densities of $\sqrt{n}(\widehat{\beta}_{30,n} - \beta_{30,0})$ and $\mathcal{N}(0, 5.68^2)$ when $n = 600$ and $n = 1200$, respectively. The number 5.68 is the asymptotic standard deviation of $\widehat{\beta}_{30,n}$ multiplied by $\sqrt{n}$. From Figure 3, we can see that they are very close when $n = 1200$.

Table 1: Simulation studies for model (5.1).

n	300			600			900			1200		
	Bias	ESD	ASD	Bias	ESD	ASD	Bias	ESD	ASD	Bias	ESD	ASD
β_{10}	0.060	0.357	0.253	0.042	0.192	0.179	0.032	0.154	0.146	0.023	0.136	0.126
β_{11}	-0.027	0.149	0.102	-0.018	0.078	0.072	-0.014	0.061	0.059	-0.009	0.054	0.051
β_{20}	-0.022	0.299	0.243	-0.008	0.185	0.172	-0.008	0.148	0.140	-0.005	0.129	0.122
β_{21}	-0.009	0.101	0.092	-0.005	0.070	0.065	-0.003	0.054	0.053	-0.003	0.046	0.046
β_{22}	-0.010	0.153	0.133	-0.008	0.101	0.094	-0.004	0.080	0.077	-0.003	0.067	0.067
β_{30}	-0.029	0.565	0.327	0.000	0.293	0.231	0.023	0.208	0.189	0.002	0.178	0.163
β_{31}	-0.015	0.446	0.283	0.002	0.242	0.200	0.011	0.182	0.163	-0.001	0.152	0.141
β_{32}	-0.006	0.585	0.369	-0.005	0.306	0.261	0.001	0.239	0.213	0.003	0.196	0.184
β_{40}	0.002	0.203	0.151	-0.003	0.113	0.107	-0.003	0.094	0.087	0.000	0.073	0.075
β_{41}	0.004	0.093	0.080	0.003	0.061	0.056	0.002	0.048	0.046	0.000	0.042	0.040
β_{42}	-0.001	0.120	0.097	-0.003	0.074	0.068	-0.003	0.060	0.056	-0.001	0.048	0.048
r	-0.047	0.207	$^{W}0.150$ $^{R}0.150$	-0.029	0.097	$^{W}0.075$ $^{R}0.075$	-0.015	0.058	$^{W}0.050$ $^{R}0.050$	-0.013	0.042	$^{W}0.037$ $^{R}0.037$
s	0.006	0.071	$^{W}0.055$ $^{R}0.060$	-0.002	0.030	$^{W}0.027$ $^{R}0.030$	-0.002	0.018	$^{W}0.018$ $^{R}0.020$	-0.001	0.014	$^{W}0.014$ $^{R}0.015$

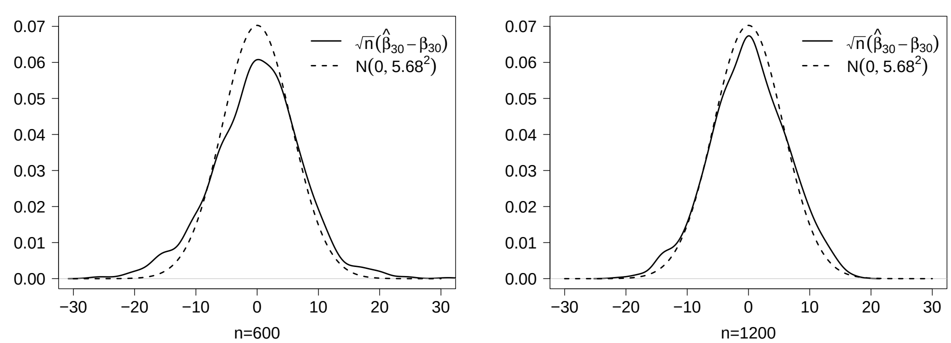

Figure 3: The densities of $\sqrt{n}(\widehat{\beta}_{30,n} - \beta_{30,0})$ when $n = 600$ and $n = 1200$.

To study the coverage probabilities of r_0 and s_0, we use the WNW and modified resampling methods respectively, and obtain the empirical quantiles of $M_-^{(1)}$ and $M_-^{(2)}$ with 10,000 replications. Table 2 shows the coverage probabilities of r_0 and s_0 by the two methods, which are quite accurate. In Supplementary S.3, we also provide the coverage probabilities with our method and method under the diminishing-threshold-effect framework. It turns out that our methods enjoys the accuracy without being conservative.

Table 2: Coverage probabilities using the WNW and the modified resampling methods.

		WNW				RS			
	$1-\alpha$	300	600	900	1200	300	600	900	1200
	0.99	0.959	0.976	0.981	0.988	0.957	0.971	0.978	0.986
r_0	0.95	0.897	0.929	0.938	0.951	0.898	0.933	0.937	0.951
	0.90	0.832	0.881	0.888	0.901	0.845	0.885	0.894	0.905
	0.99	0.979	0.985	0.993	0.991	0.962	0.971	0.985	0.987
s_0	0.95	0.930	0.940	0.946	0.944	0.911	0.932	0.942	0.945
	0.90	0.866	0.883	0.886	0.906	0.830	0.887	0.876	0.880

For an overall assessment of the estimated thresholds, Figure 4(a) and (d) respectively show the histograms of $n(\widehat{r}_n - r_0)$ and $n(\widehat{s}_n - s_0)$ when $n = 600$. Figure 4(b) and (c) show the histograms of M_1^{WNW} and M_1^{RS} respectively. Figure 4(e) and (f) show those of M_2^{WNW} and M_2^{RS} respectively, which are similarly defined as in Theorem 4.1. It shows that both methods behave well, and the WNW is better.

To check the asymptotic independence between the two estimated thresholds, we use the multivariate independence test that is based on the empirical copula process as proposed by Genest and Rémillard (2004). By implementing the functions "`indepTestSim`" and "`indepTest`" in the package `copula` in the software R, we got the p-values 0.08, 0.74, 0.90, and 0.80 for $n = 300, 600, 900,$ and 1200, respectively. None is rejected at the 5% significance level. This result is in line with Theorem 3.3.

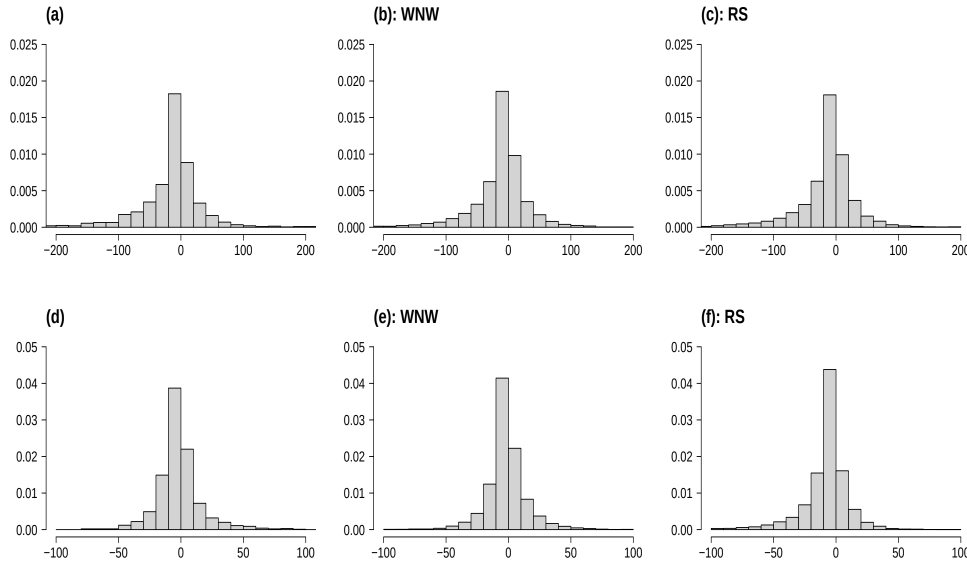

Figure 4: The histogram of (a) $n(\widehat{r}_n - r_0)$; (b) M_1^{WNW}; (c) M_1^{RS}; (d) $n(\widehat{s}_n - s_0)$; (e) M_2^{WNW}; (f) M_2^{RS}.

6 Empirical Examples

6.1 Analysis of U.S. Real GNP

The quarterly U.S. real GNP data have been analyzed by many statisticians and econometricians. Regimes have been suggested to reflect various economic scenarios, e.g. expansion, recession, depression, and recovery. See, e.g., Schumpeter (1939), Tiao and Tsay (1994) and Li and Ling (2012). To model the 4 scenarios, we build a 2-TAR model to fit the growth rate of the quarterly data.

Let x_t denote the original data from the first quarter of 1947 to the second quarter of 2019, totaling 290 observations. We define the growth rate series as $y_t = 100(\log x_t - \log x_{t-1})$. The GNP data $\{x_t\}$ and the growth rate $\{y_t\}$ are plotted in Figure 5. No trend is discernible in y_t, the growth rate of GNP data, and Tiao and Tsay (1994) fitted a stationary model with $y_{t-1} - y_{t-2}$ and y_{t-2} as the two threshold variables. It is well known that if $\{y_t\}$ is strictly stationary and ergodic, so is the vector sequence $\{(y_t, y_{t-1} - y_{t-2}, y_{t-2})', t \in Z\}$, since measurable transformations of strictly stationary and ergodic sequences are strictly

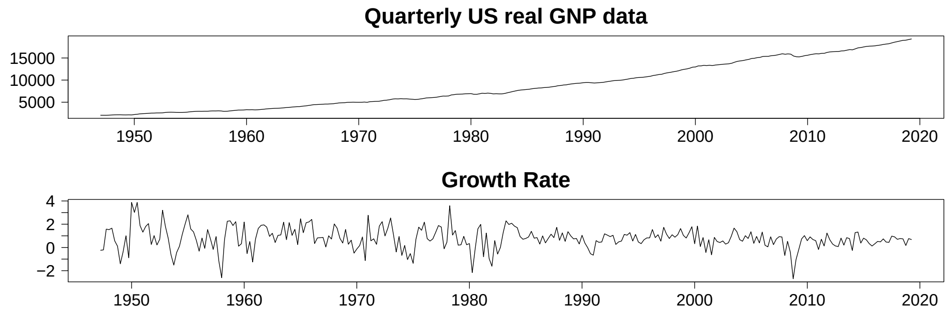

Figure 5: The original GNP data and growth rate.

stationary and ergodic; see Theorem A.1 in Francq and Zakoian (2019).

Thus, we follow the choice of $y_{t-1} - y_{t-2}$ and y_{t-2} as the two threshold variables. Instead of setting the threshold parameters arbitrarily at zero, we estimate them from data. Set $\max\{p_1, p_2, p_3, p_4\} = 7$. Using the AICu in (2.2), we select the following model for the growth :

$$y_t = \begin{cases} \beta_{10} + \sum_{i=1}^{2} \beta_{1i} y_{t-i} + \sigma_1 \epsilon_t, & \text{if } y_{t-1} - y_{t-2} > -0.95,\ y_{t-2} > 1.30, \\ \beta_{20} + \sum_{i=1}^{1} \beta_{2i} y_{t-i} + \sigma_2 \epsilon_t, & \text{if } y_{t-1} - y_{t-2} \leq -0.95,\ y_{t-2} > 1.30, \\ \beta_{30} + \sum_{i=1}^{1} \beta_{3i} y_{t-i} + \sigma_3 \epsilon_t, & \text{if } y_{t-1} - y_{t-2} \leq -0.95,\ y_{t-2} \leq 1.30, \\ \beta_{40} + \sum_{i=1}^{3} \beta_{4i} y_{t-i} + \sigma_4 \epsilon_t, & \text{if } y_{t-1} - y_{t-2} > -0.95,\ y_{t-2} \leq 1.30. \end{cases} \tag{6.1}$$

The coefficients with their standard errors are summarized in Table 3. The estimates of σ_i's are $\widehat{\sigma}_1 = 0.65$, $\widehat{\sigma}_2 = 0.72$, $\widehat{\sigma}_3 = 1.60$, $\widehat{\sigma}_4 = 0.78$, respectively. There are 37, 27, 19 and 197 observations for regimes 1, 2, 3 and 4, respectively. The 95% confidence intervals of r_0 and s_0 are $(-1.61, -0.35)$ and $(1.18, 1.53)$, respectively, computed by the WNW method.

To gain some insight, we refer to the National Bureau of Economic Research (NBER) Business Cycle Dating Committee, which has been classifying the state of the U.S. economy for the past 60 years. Members of the committee reach a subjective consensus about business cycle turning points, and their decision is generally accepted as the official dating of the U.S. business cycle. Their decision is based on a wide range of measurements of economic activities, including real personal income less transfers, nonfarm payroll employment, real personal consumption expenditures, industrial production, among others. They interpret expansion as the normal state of economy, while caution should be taken

Table 3: Coefficients and standard errors for model (6.1).

Regime	β_{i0}	β_{i1}	β_{i2}	β_{i3}
1	-1.44*	-0.10	1.44*	
	0.47	0.22	0.35	
2	0.85*	0.20		
	0.19	0.23		
3	0.28	0.16		
	0.67	0.54		
4	0.56*	0.31*	0.18*	-0.16*
	0.09	0.08	0.09	0.07

* denotes significance at 5%.

in recession. Note that there is a subtle difference in the terminology between their two-regime classification and our four-regime one, i.e., their recession actually corresponds to our Regime 3, namely depression epoch. Although their multi-information-source methodology is different from our model, it remains interesting to compare their recession epochs with those in our Regime 3. This we do in Figure 6, which also shows the Regime 3 for the case in which the thresholds are set at $(0, 0)$ as in Tiao and Tsay (1994). As we can see, the epochs of Regime 3 of our model match quite well the recession epochs declared by the NBER, except for two points (we miss the 1990 and add 2014). The matching is perhaps as good as could be reasonably expected, on bearing in mind the fact that the NBER uses multiple information sources while our model uses the sole information provided by the GNP. It is interesting to note that if the threshold values were arbitrarily fixed at $(0, 0)$ (-a practice we prefer not to adopt), the model would miss almost half of the depression epochs.

For the two thresholds, y_{t-2} reflects the speed of economic growth and $y_{t-1} - y_{t-2}$ the acceleration. Equivalently, $-(y_{t-1} - y_{t-2})$ reflects the deceleration. Thus, $y_{t-2} > 1.30$ (≤ 1.30) indicates high speed (normal speed); $-(y_{t-1} - y_{t-2}) < 0.95$ (≥ 0.95) indicates weak deceleration (strong deceleration) of, i.e. weak break (strong break) on, the economy. Based on all the estimated coefficients and thresholds, the four regimes can be interpreted as follows.

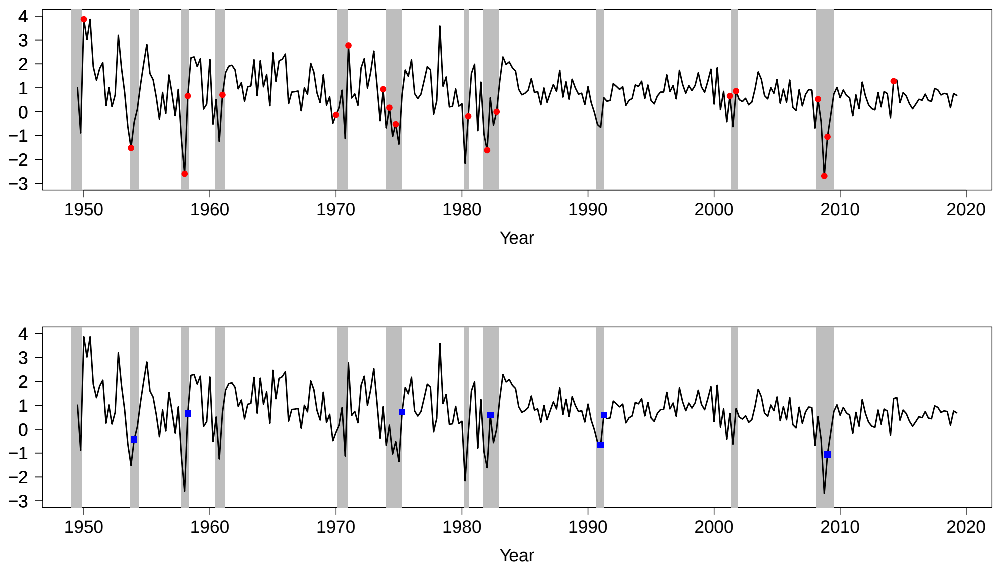

Figure 6: GNP growth rate with recession epoch indicated by NBER (grey band), Regime 3 by model (6.1) (red points), Regime 3 by 2-TAR model with (0,0) thresholds (blue squares).

- Regime 1 is an expansion epoch with high speed and weak deceleration. The fact that $y_{t-2} > 1.30$ and its significant positive coefficient indicates expansion.

- Regime 2 is a recession epoch with high speed and strong deceleration.

- Regime 3 is a depression epoch with normal speed and strong deceleration. Reassuringly it lasted the shortest. This regime has more fluctuations. During this epoch, $y_{t-1} \leq 0.35$, which tends to be negative with high probability. Thus, the positive coefficient at order 1 (albeit not significantly large) tends to suggest contraction.

- Regime 4 is a recovery epoch with normal speed and weak deceleration. This regime consists of most of the observations and can be seen as the normal state of the economy.

6.2 Analysis of Stock Return

The relationship between return autocorrelation and trading volume has attracted much attention in finance. See, e.g., Campbell et al. (1993). In Llorente et al. (2002), information asymmetry was also considered. For each individual stock, proxy time series such as quoted spread and market capitalization, *averaged* over the sample period, are used as information asymmetry. However, we suggest that *raw* data before averaging may also be useful in modeling the dynamics of the stock return.

Generally speaking, for large firms, the degree of informed trading is relatively low and the quoted spread is almost zero. So we pick the moderate-size stock Sinopec Shanghai Petrochemical Company Limited (SHI) over the period from January 2, 2015 to December 31, 2019, totalling 1256 observations. Multiply the daily return by 100 to remove decimal places and let it be y_t.

The two threshold variables are analogous to those of Llorente et al. (2002). We use daily turnover as a measure of trading volume, which is the total number of shares traded on the day divided by the total number of shares outstanding. Thus, the first threshold variable is defined as the detrended log turnover of the previous day,

$$z_{t-1} = \log(\text{turnover}_{t-1}) - \frac{1}{200}\sum_{j=1}^{200} \log(\text{turnover}_{t-1-j}).$$

We use the quoted spread to measure information asymmetry, which is the difference between the highest price favoured by a buyer and the lowest price favoured by a seller. Due to the inaccessibility of the transaction-to-transaction data, we use the daily (time weighted) quoted spread as a proxy of information asymmetry on that day. The second threshold variable, w_{t-1}, is defined as the quoted spread of the previous day. The return and turnover data are from CRSP (Center for Research of Security Prices), and the quoted spread data are from WRDS Intraday Indicator Dataset (IID). See Figure 7.

Let $\max\{p_1, p_2, p_3, p_4\} = 4$. Using the AICu in (2.2), we selected the following model:

$$y_t = \begin{cases} \beta_{10} + \sum_{i=1}^{3} \beta_{1i} y_{t-i} + \sigma_1 \epsilon_t, & \text{if } z_{t-1} > 0.722,\ w_{t-1} > 0.132, \\ \beta_{20} + \beta_{21} y_{t-1} + \sigma_2 \epsilon_t, & \text{if } z_{t-1} \le 0.722,\ w_{t-1} > 0.132, \\ \beta_{30} + \beta_{31} y_{t-1} + \sigma_3 \epsilon_t, & \text{if } z_{t-1} \le 0.722,\ w_{t-1} \le 0.132, \\ \beta_{40} + \beta_{41} y_{t-1} + \sigma_4 \epsilon_t, & \text{if } z_{t-1} > 0.722,\ w_{t-1} \le 0.132. \end{cases} \tag{6.2}$$

The estimated coefficients with their standard deviations are summarized in Table 4. The estimates of σ_i's are $\widehat{\sigma}_1 = 3.65$, $\widehat{\sigma}_2 = 1.90$, $\widehat{\sigma}_3 = 2.08$, $\widehat{\sigma}_4 = 3.52$, respectively. The numbers

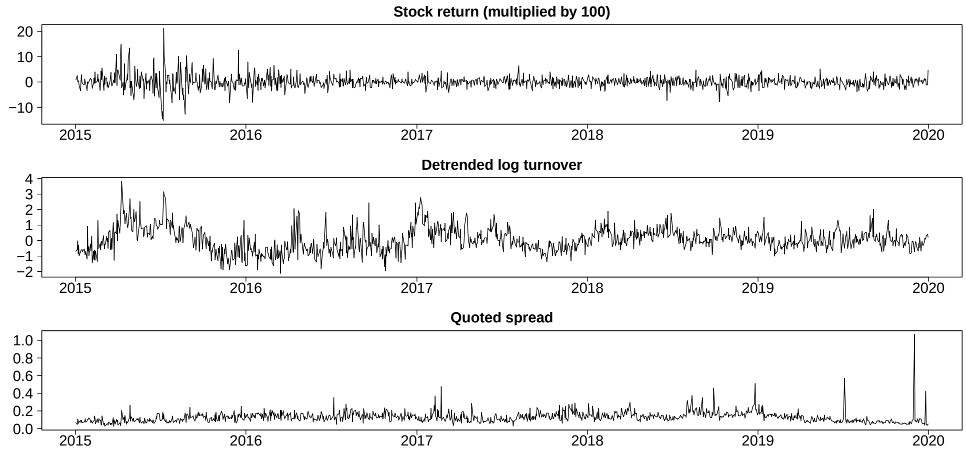

Figure 7: Stock return (multiplied by 100), detrended log turnover and quoted spread for the stock SHI.

of observations are 67, 466, 563 and 156. The 95% confidence intervals of r_0 and s_0 are $(0.515, 0.823)$ and $(0.110, 0.138)$, respectively, computed by the WNW method.

Table 4: Coefficients and standard deviations for model (6.2).

Regime	β_{i0}	β_{i1}	β_{i2}	β_{i3}
1	-0.23	0.12	-0.28	-0.62*
	0.46	0.12	0.15	0.14
2	0.11	-0.10*		
	0.09	0.05		
3	0.02	-0.10*		
	0.09	0.05		
4	-0.07	0.39*		
	0.28	0.07		

* denotes significance at 5%.

From the table, Regimes 2 and 3 have close coefficients, suggesting potential regime coalescence. Thus, we use the AICu to choose the most appropriate regime structure, and the detailed procedure can be found in Supplementary S.1.2. The AICu chose to coalesce

Regimes 2 and 3, and selected the following model with $\max\{p_1, p_2, p_3, p_4\} = 4$.

$$y_t = \begin{cases} \beta_{10} + \sum_{i=1}^{3} \beta_{1i} y_{t-i} + \sigma_1 \epsilon_t, & \text{if } z_{t-1} > 0.722,\ w_{t-1} > 0.129, \\ \beta_{20} + \beta_{21} y_{t-1} + \sigma_2 \epsilon_t, & \text{if } z_{t-1} \le 0.722, \\ \beta_{40} + \beta_{41} y_{t-1} + \sigma_4 \epsilon_t, & \text{if } z_{t-1} > 0.722,\ w_{t-1} \le 0.129. \end{cases} \tag{6.3}$$

The estimated coefficients with their standard deviations are summarized in Table 5. The estimates of σ_i's are $\widehat{\sigma}_1 = 3.55$, $\widehat{\sigma}_2 = 2.00$, $\widehat{\sigma}_4 = 3.56$, respectively. The numbers of observations are 74, 1029 and 149, respectively.

Table 5: Coefficients and standard deviations for model (6.3).

Regime	β_{i0}	β_{i1}	β_{i2}	β_{i3}
1	-0.29	0.10	-0.24	-0.61*
	0.43	0.12	0.14	0.13
2	0.06	-0.10*		
	0.06	0.03		
4	-0.01	0.40*		
	0.29	0.07		

* denotes significance at 5%.

Compared with model (6.2), the estimated thresholds are very close, and so are the estimated AR coefficients. Thus, this coalescence is reasonable since it reduces the AICu and simplifies the model. It seems that the return series on days with high trading volume are more volatile. Especially, Regime 1, which means days with high volume and potential information asymmetry, has the maximum volatility and the fewest observations. On high trading volume days, days with low information asymmetry (Regime 4) show a significant return continuation; while days with high information asymmetry (Regime 1) show less continuation or reversal. Most observations fall in the low trading volume regime (Regimes 2), with significant negative coefficients, suggesting reversal. Refer to Supplementary S.1.2 for more knowledge of regime coalescence.

7 K-TAR Cases

To generalize a 2-TAR model to a K-TAR model, let $\boldsymbol{z}_t = (z_{1t}, ..., z_{Kt})'$ be the K-dimensional threshold variables, which are observed and real-valued. Each such variable can be an exogenous time series or a linear function of the lags of y_t. The threshold parameter is $\boldsymbol{\tau} = (\tau_1, ..., \tau_K)'$. These K threshold variables divide the whole space $\mathbb{R}^K$ into 2^K regimes. Let $\boldsymbol{j} = (j_1, ..., j_K)$ be the regime index, where $j_k = I(z_{k,t-1} > \tau_k)$ for $1 \leq k \leq K$. It is similar to the binary encoding. Let $\mathcal{J} = \{\boldsymbol{j} : j_k = 0, 1,\ 1 \leq k \leq K\}$ be the set of all possible regime indexes arranged by the binary size. So, given a regime index $\boldsymbol{j}$, we can locate this regime by

$$I_t(\boldsymbol{j}, \boldsymbol{\tau}) = \prod_{k=1}^{K} I_t(j_k, \tau_k), \tag{7.1}$$

with $I_t(j_k, \tau_k) = I(z_{k,t-1} \leq \tau_k)I(j_k = 0) + I(z_{k,t-1} > \tau_k)I(j_k = 1)$. In the $\boldsymbol{j}$-indexed regime, the coefficient is $\boldsymbol{\beta}_{\boldsymbol{j}}$ and the standard deviation of the error term is $\sigma_{\boldsymbol{j}}$.

Thus, the p-th order K-TAR model is defined as

$$y_t = \sum_{\boldsymbol{j} \in \mathcal{J}} \left(\boldsymbol{\beta}_{\boldsymbol{j}}' \mathbf{y}_{t-1} + \sigma_{\boldsymbol{j}} \epsilon_t\right) I_t(\boldsymbol{j}, \boldsymbol{\tau}), \tag{7.2}$$

where $\{\epsilon_t\}$ is i.i.d. with zero mean and unit variance, and is independent of the past information. When $K = 2$, the K-TAR model reduces to the 2-TAR model in (2.1). The regime indexed by $\boldsymbol{j} = (1, 1)$ in (7.1) corresponds to Regime 1 in (2.1); $\boldsymbol{j} = (0, 1)$ to Regime 2; $\boldsymbol{j} = (0, 0)$ to Regime 3; $\boldsymbol{j} = (1, 0)$ to Regime 4.

Let $\boldsymbol{\beta}' = (\boldsymbol{\beta}_{\boldsymbol{j}}', \boldsymbol{j} \in \mathcal{J})$, and $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\tau}')'$. Assume that $\{y_1, ..., y_n\}$ is from model (7.2) with the true parameter $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0', \boldsymbol{\tau}_0')'$, where $\boldsymbol{\beta}_0' = (\boldsymbol{\beta}_{\boldsymbol{j},\mathbf{0}}', \boldsymbol{j} \in \mathcal{J})$ and $\boldsymbol{\tau}_0 = (\tau_{1,0}, ..., \tau_{K,0})'$. Then, using the LSE method and multi-parameter grid-search algorithm introduced in Section 2, we can estimate $\boldsymbol{\theta}_0$ and obtain $\widehat{\boldsymbol{\theta}}_n$.

Most assumptions in Section 3 remain valid here, with appropriate modifications. Further, the convergence rate given by Theorem 3.1 and 3.2 continues to hold. Here we will only restate the limiting distribution of the estimated threshold parameters and the required assumptions, since it is our main interest.

Let $\{(y_t, \boldsymbol{z}_t)\}$ be strictly stationary and ergodic, where $\{\boldsymbol{z}_t\}$ are random vectors with a bounded, continuous and positive density $\pi(\cdot)$ on $\mathbb{R}^K$. Denote the marginal density of $\boldsymbol{z}_t$ as $\pi_k(\cdot)$ for $k = 1, \ldots, K$. We still allow the less-than-2^K-regimes cases in Assumption 3.6.

To establish the limiting distribution of $n(\widehat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0)$, similar to the 2-TAR case, we consider $\tilde{L}_n(\boldsymbol{u}) = L_n(\boldsymbol{\beta}_0, \boldsymbol{\tau}_0 + \boldsymbol{u}/n) - L_n(\boldsymbol{\beta}_0, \boldsymbol{\tau}_0)$. Let $\boldsymbol{\tau}(u_k)$ be a vector that differs from $\boldsymbol{\tau}_0$ only for the k-th element by a magnitude u_k/n, specifically, $\boldsymbol{\tau}(u_k) = (\tau_{1,0}, ..., \tau_{k,0} + u_k/n, ..., \tau_{K,0})'$. Define

$$
\begin{aligned}
\tilde{R}_n(u_k) =& L_n\big(\boldsymbol{\beta}_0, \boldsymbol{\tau}(u_k)\big) - L_n\big(\boldsymbol{\beta}_0, \boldsymbol{\tau}_0\big) \\
=& \sum_{t=1}^{n} \Big[\gamma_{k,t}^{(1)} I\big(\tau_0 < z_{k,t-1} \le \tau_0 + u_k/n\big) I(u_k > 0) + \gamma_{k,t}^{(2)} I\big(\tau_0 + u_k/n < z_{k,t-1} \le \tau_0\big) I(u_k \le 0)\Big].
\end{aligned}
$$

In order to explicitly write the form of $\gamma_{k,t}^{(1)}$ and $\gamma_{k,t}^{(2)}$, let's consider the threshold variables other than the kth one. Specifically, let $\boldsymbol{z}_{-k,t} = (z_{1,t}, ..., z_{k-1,t}, z_{k+1,t}, ..., z_{K,t})'$ and $\boldsymbol{\tau}_{-k,0} = (\tau_{1,0}, ..., \tau_{k-1,0}, \tau_{k+1,0}, ..., \tau_{K,0})'$. Let $\boldsymbol{j}_{-k} = (j_1, ..., j_{k-1}, j_{k+1}, ..., j_K)$ denote the regime index $\boldsymbol{j}$ without the component j_k. Then, the region located by $\boldsymbol{j}_{-k}$ is $I_t(\boldsymbol{j}_{-k}, \boldsymbol{\tau}_{-k,0})$ by (7.1). Let $\mathcal{J}_{-K}$ be the set of all possible $\boldsymbol{j}_{-k}$, which includes 2^{K-1} elements. Let $\boldsymbol{j}_{k0} = (j_1, ..., j_{k-1}, 0, j_{k+1}, ..., j_K)$ and $\boldsymbol{j}_{k1} = (j_1, ..., j_{k-1}, 1, j_{k+1}, ..., j_K)$. Then,

$$
\begin{aligned}
\gamma_{k,t}^{(1)} &= \sum_{\boldsymbol{j}_{-k} \in \mathcal{J}_{-K}} I_t(\boldsymbol{j}_{-k}, \boldsymbol{\tau}_{-k,0}) \left\{ \left[(\boldsymbol{\beta}_{\boldsymbol{j}_{k1},0} - \boldsymbol{\beta}_{\boldsymbol{j}_{k0},0})' \mathbf{y}_{t-1} \right]^2 + 2\sigma_{\boldsymbol{j}_{k1},0} \epsilon_t \left(\boldsymbol{\beta}_{\boldsymbol{j}_{k1},0} - \boldsymbol{\beta}_{\boldsymbol{j}_{k0},0} \right)' \mathbf{y}_{t-1} \right\}, \\
\gamma_{k,t}^{(2)} &= \sum_{\boldsymbol{j}_{-k} \in \mathcal{J}_{-K}} I_t(\boldsymbol{j}_{-k}, \boldsymbol{\tau}_{-k,0}) \left\{ \left[(\boldsymbol{\beta}_{\boldsymbol{j}_{k0},0} - \boldsymbol{\beta}_{\boldsymbol{j}_{k1},0})' \mathbf{y}_{t-1} \right]^2 + 2\sigma_{\boldsymbol{j}_{k0},0} \epsilon_t \left(\boldsymbol{\beta}_{\boldsymbol{j}_{k0},0} - \boldsymbol{\beta}_{\boldsymbol{j}_{k1},0} \right)' \mathbf{y}_{t-1} \right\}.
\end{aligned}
$$

Let $F_k(\cdot|\tau_k)$ be the conditional distribution of $\gamma_{k,2}^{(1)}$ given $z_{k,1} = \tau_k$, and $G_k(\cdot|\tau_k)$ be that of $\gamma_{k,2}^{(2)}$ given $z_{k,1} = \tau_k$.

Then, we define K independent one-dimensional two-sided compound Poisson processes $\{\mathcal{P}_k(u_k), u_k \in \mathbb{R}\}$ for $1 \le k \le K$ as

$$
\mathcal{P}_k(u_k) = I(u_k > 0) \sum_{i=1}^{N_{k,1}(u_k)} \zeta_{k,i}^{(1)} + I(u_k \le 0) \sum_{i=1}^{N_{k,2}(-u_k)} \zeta_{k,i}^{(2)},
$$

where $\{N_{k,1}(u_k), u_k \ge 0\}$ and $\{N_{k,2}(u_k), u_k \ge 0\}$ are two independent Poisson processes with $N_{k,1}(0) = N_{k,2}(0) = 0$ a.s. and the same jump rate $\pi_k(\tau_{k,0})$. We work with the right continuous version for $N_{k,1}(u_k)$, and the left continuous version for $N_{k,2}(u_k)$. Here $\{\zeta_{k,i}^{(1)} : i \ge 1\}$ are i.i.d. from $F_k(\cdot|\tau_{k,0})$ and $\{\zeta_{k,i}^{(2)} : i \ge 1\}$ from $G_k(\cdot|\tau_{k,0})$, and they are mutually independent.

We further define a spatial compound Poisson process $\psi(\boldsymbol{u}) = \sum_{k=1}^{K} \mathcal{P}_k(u_k)$. Clearly, $\psi(\boldsymbol{u})$ goes to $+\infty$ a.s. when $|u_k| \to \infty, 1 \le k \le K$ since all jump distributions have positive means. Therefore, there exists a unique K-dimensional cube $[\boldsymbol{M}_-, \boldsymbol{M}_+) \equiv \prod_{k=1}^{K}[M_-^{(k)}, M_+^{(k)})$ on which the process $\psi(\boldsymbol{u})$ attains its global minimum a.s., namely,

$$[\boldsymbol{M}_-, \boldsymbol{M}_+) = \arg\min_{\boldsymbol{u}\in\mathbb{R}^K} \psi(\boldsymbol{u}).$$

Now we can state our result, which is a generalization of Theorem 3.3 from 2-TAR to K-TAR.

Theorem 7.1. *If Assumptions 3.1-3.6 with modifications for K-TAR hold, then $n(\widehat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0)$ converges weakly to $\boldsymbol{M}_-$ and its components are asymptotically independent as $n \to \infty$. Furthermore, $n(\widehat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0)$ is asymptotically independent of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$.*

8 Conclusion

In this paper, we have, first of all, generalized the classical 1-TAR model, with one threshold variable, to a 2-TAR model, with two threshold variables, in which we allow interaction terms and both exogenous and endogenous threshold variables. The driving noise terms are not restricted to be Gaussian. We have established the asymptotic theory and proved that the two estimated thresholds will each converge weakly to a smallest minimizer of a two-sided compound Poisson process and they are asymptotically independent. We have developed algorithms to obtain the limiting distribution of the estimated thresholds. A new method, the weighted Nadaraya-Watson method, has been introduced to construct confidence intervals of threshold parameters, which is a universal method not limited by the type of threshold variables. The associated computation is much lighter relative to existing methods. Above all, we have demonstrated the efficacy of our approach to 2-TAR modelling.

Finally, we have explained in some detail how to extend our results to a K-TAR model for $K \geq 2$. A number of interesting problems await future research: number of threshold variables, number of regimes, count data and others.

Supplemental Material

The supplementary contains five parts. Section S.1 discusses information criteria for order selection and regime specification, and provides an example of regime coalescence. Section S.2 gives the extended weighted Nadaraya-Watson and modified resampling algorithms with all the parameters estimated. In order to compare with the diminishing-threshold-effect framework in Chen et al. (2012), Section S.3 discusses on confidence interval construction for thresholds, provides theoretical support and gives simulation results for different distributions. Section S.4 and S.5 contain proofs of the main results and some auxiliary lemmas.

Acknowledgments

The authors are very grateful to the two referees, the associate editor, and the co-editor for their constructive suggestions and comments, which have led to a substantial improvement of our paper. Zhang acknowledges that her work was mainly carried out during studying in Tsinghua University and is from Chapter 2 of her PhD thesis. Li's research is supported in part by the NSFC (No.71973077) and the Tsinghua University Initiative Scientific Research Program (No.2019Z07L01009). The authors report there are no competing interests to declare.

References

Brachner, C., Fasen, V. and Lindner, A. (2009). Extremes of autoregressive threshold processes, *Advances in Applied Probability* **41**(2): 428–451.

Cai, Z. (2002). Regression quantiles for time series, *Econometric Theory* **18**(1): 169–192.

Cai, Z. and Tiwari, R. C. (2000). Application of a local linear autoregressive model to BOD time series, *Environmetrics* **11**(3): 341–350.

Campbell, J. Y., Grossman, S. J. and Wang, J. (1993). Trading volume and serial correlation in stock returns, *The Quarterly Journal of Economics* **108**(4): 905–939.

Chan, K.-S. (1989). A note on the geometric ergodicity of a Markov chain, *Advances in Applied Probability* **21**(03): 702–704.

Chan, K.-S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model, *The Annals of Statistics* **21**(1): 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**(03): 666–678.

Chan, K.-S., Tong, H. and Stenseth, N. C. (2009). Analyzing short time series data from periodically fluctuating rodent populations by threshold models: A nearest block bootstrap approach, *Science in China Series A: Mathematics* **52**(6): 1085–1106.

Chan, W. S., Wong, A. C. S. and Tong, H. (2004). Some nonlinear threshold autoregressive time series models for actuarial use, *North American Actuarial Journal* **8**(4): 37–61.

Chen, C. W. S., So, M. K. P. and Liu, F. C. (2011). A review of threshold time series models in finance, *Statistics and Its Interface* **4**(2): 167–181.

Chen, H., Chong, T. T. L. and Bai, J. (2012). Theory and applications of TAR model with two threshold variables, *Econometric Reviews* **31**(2): 142–170.

Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Processes*, Chapman & Hall/CRC, Boca Raton, FL.

De Gooijer, J. G. (2001). Cross-validation criteria for SETAR model selection, *Journal of Time Series Analysis* **22**(3): 267–281.

Francq, C. and Zakoian, J. M. (2019). *GARCH Models: Structure, Statistical Inference and Financial Applications*, second edn, John Wiley & Sons, Hoboken, NJ.

Genest, C. and Rémillard, B. (2004). Test of independence and randomness based on the empirical copula process, *Test* **13**(2): 335–369.

Hall, P., Wolff, R. C. L. and Yao, Q. (1999). Methods for estimating a conditional distribution function, *Journal of the American Statistical Association* **94**(445): 154–163.

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

Hansen, B. E. (2011). Threshold autoregression in economics, *Statistics and Its Interface* **4**(2): 123–127.

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., Tong, H. and Yang, G. (2019). On brownian motion approximation of compound poisson processes with applications to threshold models, *Advances in Decision Sciences* **23**(2): 164–191.

Li, D. and Tong, H. (2016). Nested sub-sample search algorithm for estimation of threshold models, *Statistica Sinica* **26**(4): 1543–1554.

Llorente, G., Michaely, R., Saar, G. and Wang, J. (2002). Dynamic volume-return relation of individual stocks, *The Review of Financial Studies* **15**(4): 1005–1047.

McQuarrie, A., Shumway, R. and Tsai, C.-L. (1997). The model selection criterion AICu, *Statistics & Probability Letters* **34**(3): 285–292.

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

Peña, D. and Rodriguez, J. (2005). Detecting nonlinearity in time series by model selection criteria, *International Journal of Forecasting* **21**(4): 731–748.

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer, New York.

Samia, N. I., Chan, K.-S. and Stenseth, N. C. (2007). A generalized threshold mixed model for analyzing nonnormal nonlinear time series, with application to plague in Kazakhstan, *Biometrika* **94**(1): 101–118.

Schumpeter, J. A. (1939). *Business Cycles*, McGraw-Hill, New York.

Seijo, E. and Sen, B. (2011). A continuous mapping theorem for the smallest argmax functional, *Electronic Journal of Statistics* **5**: 421–439.

Stenseth, N. C. (2009). The importance of TAR-modelling for understanding the structure of ecological dynamics: The hare-lynx population cycles as an example, *Exploration of a Nonlinear World: An Appreciation of Howell Tong's Contributions to Statistics*, World Scientific, pp. 365–374.

Stenseth, N. C., Chan, K.-S., Tong, H., Boonstra, R., Boutin, S., Krebs, C. J., Post, E., O'Donoghue, M., Yoccoz, N. G., Forchhammer, M. C. and Hurrell, J. W. (1999). Common dynamic structure of Canada lynx populations within three climatic regions, *Science* **285**(5430): 1071–1073.

Tiao, G. C. and Tsay, R. S. (1994). Some advances in non-linear and adaptive modelling in time-series, *Journal of Forecasting* **13**(2): 109–131.

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

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

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

Wong, C. S. and Li, W. K. (1998). A note on the corrected Akaike information criterion for threshold autoregressive models, *Journal of Time Series Analysis* **19**(1): 113–124.

Yu, K. and Jones, M. C. (1998). Local linear quantile regression, *Journal of the American Statistical Association* **93**(441): 228–237.

Yu, P. and Phillips, P. C. (2018). Threshold regression asymptotics: From the compound Poisson process to two-sided Brownian motion, *Economics Letters* **172**: 123–126.