UBES #1001028, VOL 00, ISS 0

On a Threshold Double Autoregressive Model

Dong Li, Shiqing Ling, and Rongmao Zhang

QUERY SHEET

This page lists questions we have about your paper. The numbers displayed at left can be found in the text of the paper for reference. In addition, please review your paper as a whole for correctness.

Q1. Author: Please define acronyms "ARCH," "GARCH," "ARMA," and "CHARMA."
Q2. Author: As per the journal style only a maximum of six keywords are allowed, so please delete the extra words.
Q3. Author: Please provide publisher location in reference "Francq and Zakoïan (2010)."
Q4. Author: Please provide publisher location in reference "Kushner (1984)."
Q5. Author: Please provide publisher details (name and location) in reference "Merlevède et al. (2009)."
Q6. Author: Please provide publisher location in reference "Pollard (1984)."
Q7. Author: Please provide publisher location in reference "Tsay (2010)."

TABLE OF CONTENTS LISTING

The table of contents for the journal will list your paper exactly as it appears below:

Supplementary materials for this article are available online. Please go to http://tandfonline.com/r/JBES

On a Threshold Double Autoregressive Model

Dong Li

Mathematical Sciences Center and Tsinghua Center for Statistics Science, Tsinghua University, Beijing 100084, China (*dongli@math.tsinghua.edu.cn*)

Shiqing Ling

Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong (*maling@ust.hk*)

Rongmao Zhang

Department of Mathematics, Zhejiang University, Hangzhou 310027, China (*rmzhang@zju.edu.cn*)

This article first proposes a score-based test for a double autoregressive model against a threshold double autoregressive (AR) model. It is an asymptotically distribution-free test and is easy to implement in practice. The article further studies the quasi-maximum likelihood estimation of a threshold double autoregressive model. It is shown that the estimated threshold is n-consistent and converges weakly to a functional of a two-sided compound Poisson process and the remaining parameters are asymptotically normal. Our results include the asymptotic theory of the estimator for threshold AR models with ARCH errors and threshold ARCH models as special cases, each of which is also new in literature. Two portmanteau-type statistics are also derived for checking the adequacy of fitted model when either the error is nonnormal or the threshold is unknown. Simulation studies are conducted to assess the performance of the score-based test and the estimator in finite samples. The results are illustrated with an application to the weekly closing prices of Hang Seng Index. This article also includes the weak convergence of a score-marked empirical process on the space $\mathbb{D}(\overline{\mathbb{R}})$ under an α-mixing assumption, which is independent of interest.

KEY WORDS: Asymptotic normality; Compound Poisson process; Continuous TAR model; Portmanteau-type test; Quasi-maximum likelihood estimation; Score test; Threshold ARCH model; Threshold double AR model; Volatility.

1. INTRODUCTION

Generally speaking, the conditional mean function and the conditional variance function (i.e., the volatility or diffusion) of a time series are most important in practice. A lot of time series models have been suggested in the literature, see Tong (1990). The threshold autoregressive (TAR) model proposed by Tong (1978) has been widely investigated for the conditional mean function and it has been applied in a wide range of fields such as economics, econometrics, finance, etc. A comprehensive survey on TAR models is available in Tong (1990, 2011) and Hansen (2011). The ARCH-type models proposed by Engle (1982) and Bollerslev (1986) are commonly used in modeling the conditional variance functions in economic and financial time series. An overall review on GARCH models was given in Francq and Zakoïan (2010). The TAR model simply with the plug-in GARCH model, called TAR/GARCH model, has been used for a full specification of time series, see Li and Lam (1995), Li and Li (1996), and Tsay (2010). In this model, the driving random component in the GARCH part is not observable, but rather to the innovations of the TAR model. One cannot directly measure on the market volatilities via its observations. Its structure is generally unclear, except for a special case in Ling (1999), such that there is not any theoretical support to the statistical inference of this model up to date. These disadvantages can be avoid in an alternative class of ARCH-type models proposed in the literature, such as of ARMA-ARCH models in Weiss (1986) and CHARMA models in Tsay (1987).

Following Weiss (1986), in this article we consider a class of self-exciting TAR models with conditional heteroscedasticity, called threshold double autoregressive (TDAR) models. Specifically, a time series $\{y_t\}$ is said to be a TDAR model of order $(p_1, p_2; q_1, q_2)$ (hereafter abbreviated as TDAR$(p_1, p_2; q_1, q_2)$) if it satisfies the following equation:

$$y_t = \begin{cases} \phi_{10} + \sum_{j=1}^{p_1} \phi_{1j}\, y_{t-j} + \varepsilon_t \sqrt{\alpha_{10} + \sum_{j=1}^{q_1} \alpha_{1j}\, y_{t-j}^2}, \\ \quad \text{if } y_{t-d} \le r, \\ \\ \phi_{20} + \sum_{j=1}^{p_2} \phi_{2j}\, y_{t-j} + \varepsilon_t \sqrt{\alpha_{20} + \sum_{j=1}^{q_2} \alpha_{2j}\, y_{t-j}^2}, \\ \quad \text{if } y_{t-d} > r, \end{cases} \tag{1.1}$$

where ϕ_{ij}'s and α_{ij}'s are the coefficients, r is the threshold parameter, d is a positive integer called the delay parameter, and p_i and q_i are known nonnegative integers. Compared with the TAR/GARCH model, a significant difference of model (1.1) is that the conditional variance is specified in function of the observations. Its expression gives a visible dynamic behavior of the conditional variance and provides a direct way to compute the one-step future volatility. Its structure, such as the strict stationarity and V-uniform ergodicity, was studied by Cline and Pu (2004) under a general setting.

Model (1.1) implies the DAR model as a special case. The related work can be found in Ling (2004, 2007), Chan and

Journal of Business & Economic Statistics
xxx 2015, Vol. 00, No. 0
DOI: 10.1080/07350015.2014.1001028

Peng (2005), Ling and Li (2008), Zhu and Ling (2013), and Chen, Li, and Ling (2014). When α_{ij}'s are zeros, $i = 1, 2$, $j = 1, \ldots, q_i$, model (1.1) reduces to a TAR model. Asymptotic theory on least-square estimates (LSE) of TAR models were developed by Chan (1993) and Li and Ling (2012) when the autoregressive function is discontinuous and by Chan and Tsay (1998) when the autoregressive function is continuous. Under the assumption that the threshold effect is vanishingly small, Hansen (1997, 2000) obtained the distribution- and parameter-free limit of the estimated threshold. Seo and Linton (2007) proposed a smoothed least-square estimation for the TAR/regression model and showed that the estimated threshold is asymptotically normal but its convergence rate is less than n and depends on the bandwidth. When ϕ_{ij}'s are zeros, model (1.1) is a threshold ARCH (TARCH) model, see Rabemananjara and Zakoïan (1993) and Zakoïan (1994). If the threshold were known, it is more or less standard to estimate the parameters in model (1.1). The difficulty is when the threshold is unknown. In this case, no asymptotic theory has been established in literature up to now, even for the simple TARCH model.

In this article, we first study Ling and Tong's (2011), abbreviated to LT(2011), score-based statistic for testing the null DAR model against the alternative TDAR model. Under the null hypothesis, it is shown that the test statistic converges weakly to the maxima of a squared standard Brownian motion. We then study the quasi-maximum likelihood estimator (QMLE) of model (1.1). It is shown that the estimated threshold is n-consistent and converges weakly to a functional of a two-sided compound Poisson process and the remaining parameters are $\sqrt{n}$-consistent and asymptotically normal. Our results include the asymptotic theory of the estimator for TAR models with Weiss'(1986) ARCH errors and for TARCH models as special cases, each of which is also new in literature. Two portmanteau test statistics are derived for checking the adequacy of fitted models. Simulation studies are conducted to assess the power of our test and the performance of the QMLE in finite samples. The results are illustrated with an application to the weekly closing prices of Hang Seng Index.

The remainder of this article is organized as follows. Section 2 gives a score-based test and derives its limiting distribution. Section 3 presents the QMLE and states its asymptotic properties. Section 4 gives portmanteau test statistics. Simulation studies are reported in Section 5 and an empirical example is analyzed in Section 6. All proofs of main theorems are given in Appendices. It includes the weak convergence of a score-marked empirical process under an α-mixing assumption, which is independent of interest.

Throughout the article, some symbols are conventional. C is an absolutely positive constant, which may be different in different places. $I(\cdot)$ is the indicator function. $\mathbb{R}^p$ is the Euclidean space of dimension p and $\|\cdot\|$ denotes the Euclidian norm. $\|\cdot\|_\infty$ is the supremum norm, that is, $\|f\|_\infty = \sup_{x\in\mathbb{R}} |f(x)|$. $\mathbf{I}_m$ is an $m \times m$ identity matrix. Denote $\mathbb{D}(A)$ as the space of real-valued functions on the set A, which are right continuous and have left-hand limits. The space $\mathbb{D}(A)$ is equipped with the Skorohod topology (see Billingsley 1999). $\Longrightarrow$ denotes the weak convergence.

2. A SCORE-BASED TEST FOR DAR AGAINST TDAR MODELS

It is an important step to test for a threshold effect in time series modeling. The likelihood ratio (LR) test was studied by Chan (1990) and Chan and Tong (1990) for AR against TAR models, and by Wong and Li (1997, 2000) for AR-ARCH against TAR-ARCH models, see also Zhang et al. (2011). In this section, we will study a score-based test for DAR against TDAR models.

Under the null hypothesis H_0, we assume that time series $\{y_t\}$ follows a DAR model:

$$y_t = \phi'\mathbf{Y}_{t-1} + \varepsilon_t\sqrt{\boldsymbol{\alpha}'\mathbf{X}_{t-1}}, \tag{2.1}$$

where $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (iid) random variables with zero mean and unit variance, $\phi = (\phi_0, \phi_1, \ldots, \phi_p)'$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_q)'$, $\mathbf{Y}_{t-1} = (1, y_{t-1}, \ldots, y_{t-p})'$, and $\mathbf{X}_{t-1} = (1, y_{t-1}^2, \ldots, y_{t-q}^2)'$. The alternative H_1 is the threshold counterpart of (2.4) like model (1.1). Let $\boldsymbol{\theta} = (\phi', \boldsymbol{\alpha}')'$ be the parameter and Θ be the parameter space, which is compact with $\underline{\alpha} \le \alpha_i \le \bar{\alpha}$ $(i = 0, \ldots, q)$, where $\underline{\alpha}$ and $\bar{\alpha}$ are some positive constants. The true value $\boldsymbol{\theta}_0 = (\phi_0', \boldsymbol{\alpha}_0')'$ is an interior point of Θ. Given data $\{y_{1-p}, \ldots, \sim y_n\}$, under H_0, the conditional quasi-log-likelihood function (ignoring a constant) can be written as

$$L_n(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{n} l_t(\boldsymbol{\theta}) \quad \text{with} \quad l_t(\boldsymbol{\theta}) = \log(\boldsymbol{\alpha}'\mathbf{X}_{t-1}) + \frac{(y_t - \phi'\mathbf{Y}_{t-1})^2}{\boldsymbol{\alpha}'\mathbf{X}_{t-1}}.$$

Denote $\widehat{\boldsymbol{\theta}}_n$ as the QMLE of $\boldsymbol{\theta}_0$, that is, the maximizer of $L_n(\boldsymbol{\theta})$ on Θ. For simplicity, in this section, we assume that ε_t is symmetric. If $\{y_t\}$ is stationary and ergodic with $Ey_t^4 < \infty$, the density of ε_t is positive on $\mathbb{R}$, and $\kappa_4 \equiv E\varepsilon_t^4 < \infty$, Ling (2007) showed that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \Sigma_\infty^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} D_t(\boldsymbol{\theta}_0) + o_p(1),$$

where

$$\begin{aligned} D_t(\boldsymbol{\theta}) &= \Big(\frac{\mathbf{Y}_{t-1}'(y_t - \phi'\mathbf{Y}_{t-1})}{\boldsymbol{\alpha}'\mathbf{X}_{t-1}}, \\ &\quad -\frac{\mathbf{X}_{t-1}'}{2\boldsymbol{\alpha}'\mathbf{X}_{t-1}}\Big[1 - \frac{(y_t - \phi'\mathbf{Y}_{t-1})^2}{\boldsymbol{\alpha}'\mathbf{X}_{t-1}}\Big]\Big)', \\ \Sigma_x &= \operatorname{diag}\left\{E\left(\frac{\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}' I(y_{t-d} \le x)}{\boldsymbol{\alpha}_0'\mathbf{X}_{t-1}}\right),\right. \\ &\quad \left. E\left(\frac{\mathbf{X}_{t-1}\mathbf{X}_{t-1}' I(y_{t-d} \le x)}{2(\boldsymbol{\alpha}_0'\mathbf{X}_{t-1})^2}\right)\right\}, \\ &\quad x \in \overline{\mathbb{R}} \equiv \mathbb{R} \cup \{\pm\infty\}, \end{aligned}$$

for some positive integer d.

To introduce our test statistic, we first define the score marked empirical process

$$T_n(x, \widehat{\boldsymbol{\theta}}_n) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \widehat{\mathbf{U}}^{-1} D_t(\widehat{\boldsymbol{\theta}}_n) I(y_{t-d} \le x), \tag{2.2}$$

where $\widehat{\mathbf{U}} = \mathrm{diag}\{\mathbf{I}_{p+1}, \sqrt{0.5(\widehat{\kappa}_4 - 1)} \sim \mathbf{I}_{q+1}\}, \widehat{\kappa}_4 = \frac{1}{n}\sum_{t=1}^{n}\widehat{\varepsilon}_t^4$, and $1 \leq d \leq \max\{p, q, 1\}$. $T_n(x, \widehat{\boldsymbol{\theta}}_n)$ is precisely the score function in the LR test under H_1. When $\varepsilon_t \sim N(0, 1)$, it was discussed by LT(2011). Our current setting in (2.2) can be applied for the cases when $\varepsilon_t \not\sim N(0, 1)$. LT(2011) established the weak convergence of $T_n(x, \widehat{\boldsymbol{\theta}}_n)$ on the space $\mathbb{D}[a, b]$ for any fixed $b < \infty$. Theorem 1 gives its weak convergence on the space $\mathbb{D}(\overline{\mathbb{R}})$. This improvement is not trivial and is because of a new weak convergence under an α-mixing assumption in Appendix A.

Theorem 1. Under the null H_0, if $\{y_t\}$ from model (2.4) is stationary and geometrically ergodic with $Ey_t^4 < \infty$, the density of ε_t is positive on $\mathbb{R}$, and $\kappa_4 \equiv E\varepsilon_t^4 < \infty$, then

$$T_n(x, \widehat{\boldsymbol{\theta}}_n) \Longrightarrow G_{p+q+2}(x) \text{ in } \mathbb{D}(\overline{\mathbb{R}}),$$

where $\{G_{p+q+2}(x) : x \in \overline{\mathbb{R}}\}$ is a $(p+q+2)$-dimensional Gaussian process with mean zero and covariance kernel $\mathbf{K}_{xy} = \Sigma_{x\wedge y} - \Sigma_x\Sigma_\infty^{-1}\Sigma_y$; almost all paths of $G_{p+q+2}(x)$ are continuous in x.

Ideally, we should use the LR test for the threshold effect. However, as mentioned in LT(2011), the LR test is a quadratic form of $T_n(x, \widehat{\boldsymbol{\theta}}_n)$ and its limiting distribution involves some nuisance parameters. Except Chan and Tong (1990) for the AR model with iid normal errors, we need to use the simulation approach to obtain its critical case by case; see, for example, Wong and Li (1997, 2000). A possible way is to use a transformation of $T_n(x, \widehat{\boldsymbol{\theta}}_n)$. A general Gaussian process cannot be transformed into a Brownian motion by a simple scaling and linear transformation as a referee pointed out. However, LT(2011) observed that $\Sigma_x^{-1}T_n(x, \widehat{\boldsymbol{\theta}}_n) \Longrightarrow G^*(x)$ under H_0, where $G^*(x)$ is a vector Gaussian process in $\overline{\mathbb{R}}$ with mean zero and covariance kernel $\mathbf{K}_{xy}^* = \Sigma_{x\vee y}^{-1} - \Sigma_\infty^{-1}$, and it has independent increments. Because of this feature, LT(2011) showed that, for any nonzero constant vector β, by a time-change technique, the process $B(\tau) \equiv \beta' G^*(x)/\sqrt{\sigma_a}$ is a standard Brownian motion on $\tau = \sigma_x/\sigma_a \in [0, 1]$, where $\sigma_x = \beta'(\Sigma_x^{-1} - \Sigma_\infty^{-1})\beta$.

Following LT(2011), we now define our score-based test statistic as follows:

$$S_n^a = \max_{x \geq a} \frac{\left[\boldsymbol{\beta}'\widehat{\Sigma}_{nx}^{-1}T_n(x, \widehat{\boldsymbol{\theta}}_n)\right]^2}{\boldsymbol{\beta}'(\widehat{\Sigma}_{na}^{-1} - \widehat{\Sigma}_{n,\infty}^{-1})\boldsymbol{\beta}}, \tag{2.3}$$

where a is a fixed constant, $\boldsymbol{\beta}$ is a nonzero $p \times 1$ constant vector, and

$$\widehat{\Sigma}_{nx} = \mathrm{diag}\left\{\frac{1}{n}\sum_{t=1}^{n}\frac{\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}' I(y_{t-d} \leq x)}{\widehat{\boldsymbol{\alpha}}_n'\mathbf{X}_{t-1}}, \frac{1}{2n}\sum_{t=1}^{n}\frac{\mathbf{X}_{t-1}\mathbf{X}_{t-1}' I(y_{t-d} \leq x)}{(\widehat{\boldsymbol{\alpha}}_n'\mathbf{X}_{t-1})^2}\right\}.$$

The range of maxima in S_n^a is $[a, \infty]$, while the one in LT(2011) is $[a, b]$ for any fixed $b < \infty$. Our test avoids to select the constant b as in LT(2011). By Theorem 1 and the continuous mapping theorem, we have the following result:

Theorem 2. If the assumptions in Theorem 1 hold, then, for any $p \times 1$ nonzero constant vector $\boldsymbol{\beta}$, any fixed $a \in \mathbb{R}$ and any $x \in \mathbb{R}$, it follows that

$$\lim_{n\to\infty} P\left(S_n^a \leq x\right) = P\left(\max_{\tau\in[0,1]} B^2(\tau) \leq x\right),$$

where $B(\tau)$ is a standard Brownian motion on $\mathbb{C}[0, 1]$.

Choosing the constant C_α such that $P(\max_{\tau\in[0,1]} B^2(\tau) \geq C_\alpha) = \alpha$ can provide an approximate critical value of S_n^a for rejecting the null H_0 at the significance level α. Here, $C_{0.1} = 3.83$, $C_{0.05} = 5.00$, and $C_{0.01} = 7.63$ from the formula in Shorack and Wellner (1986, p. 34)

$$P\left(\max_{\tau\in[0,1]} B^2(\tau) \geq x\right) = 1 - \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{2k+1}\exp\left(-\frac{(2k+1)^2\pi^2}{8x}\right).$$

There is no universal criterion for the choice of $\boldsymbol{\beta}$. A simple choice for β is $(1, \ldots, 1)'$, that is, we put equal weight on each component of $\widehat{\Sigma}_{nx}^{-1}T_n(x, \widehat{\boldsymbol{\theta}}_n)$. The optimal choice of β still remains open. a is usually taken as the lower quantile of data so that $\widehat{\Sigma}_{na}^{-1}$ exits. The simulation studies in Section 5 show that S_n^a has a good power empirically when a is around the $5(p+q+2)\%$ quantile of data.

Our test provides an easy and simple way to implement in practice. But it may result in loss of power under some directions as a referee pointed out. It is a compromise to the difficulty in the LR test. LT(2011) showed that S_n^a has a nontrivial local power under a general local alternative. For the following specific local threshold alternative H_{1n}:

$$y_t = \phi_0'\mathbf{Y}_{t-1} + \frac{\mathbf{h}_1'\mathbf{Y}_{t-1}I(y_{t-d} \leq x)}{\sqrt{n}} + \varepsilon_t\sqrt{\boldsymbol{\alpha}_0'\mathbf{X}_{t-1} + \frac{\mathbf{h}_2'\mathbf{X}_{t-1}I(y_{t-d} \leq x)}{\sqrt{n}}}, \tag{2.4}$$

with $\varepsilon_t \sim N(0, 1)$, similar to Theorem 3.3 of LT(2011), we can show that, under H_{1n},

$$\lim_{n\to\infty} P\left(S_n^a \leq x\right) = P\left(\max_{\tau\in[0,1]}[m_\tau + B(\tau)]^2 \leq x\right),$$

where $m_\tau = \beta'(\Sigma_r^{-1} - \Sigma_\infty^{-1})\Sigma_r\mathbf{u}$, where $\mathbf{u} = (\mathbf{h}_1', \mathbf{h}_2')'$, $r = F_y^{-1}(\tau)$, and $F_y(x)$ is the distribution of y_t under H_0. Thus, our test has a nontrivial local power unless $m_\tau = 0$, which unlikely happens. In particular, for the TAR(1) model, it is equivalent to the LR test in Chan (1990). It is expected to be useful for testing the presence of threshold effect, see our simulation in Section 5.

3. THE QMLE AND ASYMPTOTICS OF TDAR MODEL

Assume that $\{y_1, \ldots, y_n\}$ is a sample from model (1.1). Given the initial values $\{y_{1-p}, \ldots, y_0\}$, where $p = \max\{p_1, p_2, q_1, q_2\}$, the conditional log-likelihood function (omitting a constant) is defined as

$$L_n(\boldsymbol{\theta}) = \sum_{t=1}^{n}\ell_t(\boldsymbol{\theta}) \quad \text{with} \quad \ell_t(\boldsymbol{\theta}) = -\frac{1}{2}\log h_t(\boldsymbol{\theta}) - \frac{1}{2}\frac{u_t^2(\boldsymbol{\theta})}{h_t(\boldsymbol{\theta})},$$

where $\boldsymbol{\theta} = (\boldsymbol{\lambda}', r)' \equiv (\phi_1', \boldsymbol{\alpha}_1', \phi_2', \boldsymbol{\alpha}_2', r)'$ is the parameter with $\phi_i = (\phi_{i0}, \phi_{i1}, \ldots, \phi_{ip_i})'$ and $\boldsymbol{\alpha}_i = (\alpha_{i0}, \alpha_{i1}, \ldots, \alpha_{iq_i})'$, and

$$u_t(\boldsymbol{\theta}) = y_t - \mu_t(\boldsymbol{\theta}), \quad \mu_t(\boldsymbol{\theta}) = (\phi_1'\mathbf{Y}_{1,t-1})I(y_{t-d} \leq r) + (\phi_2'\mathbf{Y}_{2,t-1})I(y_{t-d} > r),$$

$$h_t(\boldsymbol{\theta}) = (\boldsymbol{\alpha}_1'\mathbf{X}_{1,t-1})I(y_{t-d} \le r) + (\boldsymbol{\alpha}_2'\mathbf{X}_{2,t-1})I(y_{t-d} > r), \tag{3.1}$$

with $\mathbf{Y}_{i,t-1} = (1, y_{t-1}, \ldots, y_{t-p_i})'$, and $\mathbf{X}_{i,t-1} = (1, y_{t-1}^2, \ldots, y_{t-q_i}^2)'$ for $i = 1, 2$.

In practice, d is unknown and can be estimated consistently by an analogous procedure in Chan (1993), Li and Ling (2012), and Li, Ling, and Li (2013). For simplicity, we assume that d is known and $1 \le d \le \max(p, 1)$. Let $\boldsymbol{\Theta}$ be the parameter space. The maximizer $\widehat{\boldsymbol{\theta}}_n = (\widehat{\boldsymbol{\lambda}}_n', \widehat{r}_n)'$ of $L_n(\boldsymbol{\theta})$ on $\boldsymbol{\Theta}$ is called a QMLE of the true value $\boldsymbol{\theta}_0 = (\boldsymbol{\lambda}_0', r_0)' \in \Theta$. That is, $\widehat{\boldsymbol{\theta}}_n$ is defined by $\widehat{\boldsymbol{\theta}}_n = \arg\max_{\theta\in\boldsymbol{\Theta}} L_n(\boldsymbol{\theta})$. Due to the discontinuity of $L_n(\boldsymbol{\theta})$ in r, one can take two steps to find $\widehat{\boldsymbol{\theta}}_n$:

- For each fixed r, maximize $L_n(\boldsymbol{\theta})$ and get its maximizer $\widehat{\boldsymbol{\lambda}}_n(r)$.
- Since the profile log-likelihood $L_n^*(r) \equiv L_n(\widehat{\boldsymbol{\lambda}}_n(r), r)$ is a piecewise constant function and only takes finite possible values, one can get the maximizer $\widehat{r}_n$ of $L_n^*(r)$ by the enumeration approach and then obtain the estimator $\widehat{\boldsymbol{\theta}}_n = (\widehat{\boldsymbol{\lambda}}_n(\widehat{r}_n)', \widehat{r}_n)'$.

Generally, there exist infinitely many r such that $L_n(\cdot)$ attains its global maximum. One can choose the smallest r as an estimator of r_0, for example. According to this procedure, $\widehat{\boldsymbol{\theta}}_n$ is the QMLE of $\boldsymbol{\theta}_0$, that is, $L_n(\widehat{\boldsymbol{\theta}}_n) = \max_{\theta\in\boldsymbol{\Theta}} L_n(\boldsymbol{\theta})$.

In applications, the order $(p_1, p_2; q_1, q_2)$ is unknown and needs to be determined. It can be selected by the Akaike information criterion (AIC) or Bayesian information criterion (BIC) as follows:

$$\mathrm{AIC}(\{p_i; q_i\}) \quad = -2L_n(\widehat{\boldsymbol{\theta}}_n) + 2(p_1 + p_2 + q_1 + q_2 + 5);$$
$$\mathrm{BIC}(\{p_i; q_i\}) = -2L_n(\widehat{\boldsymbol{\theta}}_n) + (p_1 + p_2 + q_1 + q_2 + 5)\log n.$$

Without loss of generality, in what follows, we assume that the order $(p_1, p_2; q_1, q_2)$ is known. To state asymptotic properties of $\widehat{\boldsymbol{\theta}}_n$, we first give two assumptions on the error $\{\varepsilon_t\}$ and the parameter space $\boldsymbol{\Theta}$ as follows.

Assumption 1. $\{\varepsilon_t\}$ is iid with zero mean and unit variance, and has a positive and continuous density $f(x)$ on $\mathbb{R}$.

Assumption 2. The parameter space $\boldsymbol{\Theta} = \{\boldsymbol{\theta} \in \mathbb{R}^{p_1+p_2+q_1+q_2+5} : \phi_1 \ne \phi_2 \text{ or } \boldsymbol{\alpha}_1 \ne \boldsymbol{\alpha}_2, \alpha_{ij} > 0, i = 1, 2, j = 0, 1, \ldots, q_i\}$ is compact.

The following theorem states the strong consistency of $\widehat{\boldsymbol{\theta}}_n$.

Theorem 3. Suppose that Assumptions 1–2 hold and $\{y_t\}$ is strictly stationary and ergodic with $Ey_t^2 < \infty$. Then, $\widehat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ almost surely (a.s.), as $n \to \infty$.

We should mention that there is no requirement for the moment of y_t in Theorem 3 if $p_1 = p_2 = q_1 = q_2$. Since the compactness of Θ, there exists a positive constant $\underline{\alpha}$ such that $\alpha_{ij} \ge \underline{\alpha} > 0$. Thus, $\underline{\alpha}(1 + \sum_{i=1}^p y_{t-i}^2)$ can control the log-likelihood and the score functions such that they are bounded, see Remark 3.2 in Ling (2007). Similar phenomenon can be also found in Ling (2004) and Ling and Li (2008).

Let $\mathbf{Z}_t = (y_t, \ldots, y_{t-p+1})'$. Then $\{\mathbf{Z}_t\}$ is a Markov chain. Denote its l-step transition probability by $\mathcal{P}^l(\mathbf{z}, A)$, where $\mathbf{z} \in \mathbb{R}^p$ and A is a Borel set. To obtain the convergence rate of $\widehat{r}_n$ and the asymptotic normality of $\widehat{\boldsymbol{\lambda}}_n \equiv \widehat{\boldsymbol{\lambda}}_n(\widehat{r}_n)$, we need three more assumptions as follows.

Assumption 3. $\{\mathbf{Z}_t\}$ admits a unique invariant measure $\Pi(\cdot)$ such that there exist $K > 0$ and $\rho \in [0, 1)$, for any $\mathbf{z} \in \mathbb{R}^p$ and any $m \ge 1$, $\|\mathcal{P}^m(\mathbf{z}, \cdot) - \Pi(\cdot)\|_{\mathrm{v}} \le K(1 + \|\mathbf{z}\|^2)\rho^m$, where $\|\cdot\|_{\mathrm{v}}$ denotes the total variation norm.

This assumption is on the V-uniform ergodicity of model (1.1) with $V(\mathbf{z}) = K(1 + \|\mathbf{z}\|^2)$, under which $\{y_t\}$ is strictly stationary if the initial value $\mathbf{Z}_0$ follows the invariant measure Π. Without loss of generality, in what follows we assume that $\mathbf{Z}_0 \sim \Pi$. Assumption 3 is stronger than that $\{y_t\}$ is geometrically ergodic. From Corollary 2.2 in Cline and Pu (2004), Assumption 3 holds if Assumption 1 holds with $\sup_{x\in\mathbb{R}}\{(1 + |x|)f(x)\} < \infty$ and

$$\left\{\sum_{j=1}^p \max(|\phi_{1j}|, |\phi_{2j}|)\right\}^2 + \sum_{j=1}^p \max(\alpha_{1j}, \alpha_{2j}) < 1,$$

where $\phi_{ij} = 0$ for $j > p_i$ and $\alpha_{ij} = 0$ for $j > q_i$, $i = 1, 2$.

Assumption 4. $\kappa_4 \equiv E(\varepsilon_t^4) < \infty$ and $Ey_t^4 < \infty$.

Assumption 5. There exist nonrandom vectors $\mathbf{w} = (1, w_1, \ldots, w_p)'$ with $w_d = r_0$ and $\mathbf{W} = (1, W_1, \ldots, W_p)'$ with $W_d = r_0^2$ such that

$$\left\{(\phi_{10} - \phi_{20})'\mathbf{w}\right\}^2 + \left\{(\boldsymbol{\alpha}_{10} - \boldsymbol{\alpha}_{20})'\mathbf{W}\right\}^2 > 0,$$

where the vectors ϕ_{i0}'s and $\boldsymbol{\alpha}_{i0}$'s have been extended by adding zero entries such that they are $(p+1)$-dimensional vectors for simplifying notations, that is, $\phi_{ij,0} = 0$ for $j > p_i$ and $\alpha_{ij,0} = 0$ for $j > q_i$, $i = 1, 2$. (In what follows, we use this convention.)

Assumption 5 is similar to the Condition 4 in Chan (1993) and implies that either the conditional mean function $\mu_t(\boldsymbol{\theta})$ or volatility function $h_t(\boldsymbol{\theta})$ in model (1.1) is discontinuous over the hyperplane $y_{t-d} = r_0$. It is a necessary condition for the n-convergence rate of $\widehat{r}_n$. If $\boldsymbol{\alpha}_{10} = \boldsymbol{\alpha}_{20}$, then Assumption 5 is equivalent to $(\phi_{10} - \phi_{20})'\mathbf{w} \ne 0$, which is exactly the Condition 4 in Chan (1993) that $\mu_t(\boldsymbol{\theta})$ is discontinuous. The discontinuity of $\mu_t(\boldsymbol{\theta})$ plays a key role in obtaining the convergence rate of the estimated threshold in TAR models; see Chan (1993) and Chan and Tsay (1998). In Assumption 5, w_d and W_d may not be components of $\mathbf{w}$ and $\mathbf{W}$ if $d > p$. In this case, Assumption 5 is identical to $\|\phi_{10} - \phi_{20}\| + \|\boldsymbol{\alpha}_{10} - \boldsymbol{\alpha}_{20}\| > 0$, which is necessary and sufficient for the identification of the threshold. Both $\mu_t(\boldsymbol{\theta})$ and $h_t(\boldsymbol{\theta})$ are continuous over the hyperplane $y_{t-d} = r_0$ if and only if

$$\phi_{10} + \phi_{1d}\, r_0 = \phi_{20} + \phi_{2d}\, r_0, \quad \phi_{1j} = \phi_{2j},$$
$$\alpha_{10} + \alpha_{1d}\, r_0^2 = \alpha_{20} + \alpha_{2d}\, r_0^2, \quad \alpha_{1j} = \alpha_{2j}, \sim j \ne d.$$

In this case, we call model (1.1) *continuous TDAR model*. For continuous TDAR models, the theory of estimation is challenging and we will study this case in a separate article.

Theorem 4. If Assumptions 1–5 hold and $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$, then

(i) $n(\widehat{r}_n - r_0) = O_p(1)$;

(ii) $\sqrt{n} \sup_{|r-r_0|\le B/n} \left\|\widehat{\boldsymbol{\lambda}}_n(r) - \widehat{\boldsymbol{\lambda}}_n(r_0)\right\| = o_p(1)$

for any fixed constant $0 < B < \infty$,

where $\widehat{\boldsymbol{\lambda}}_n(r)$ is the QMLE of the coefficients when r is known. Further, it follows that

$$\sqrt{n}(\widehat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0) = \sqrt{n}(\widehat{\boldsymbol{\lambda}}_n(r_0) - \boldsymbol{\lambda}_0) + o_p(1) \Longrightarrow N(\mathbf{0}, \sim \Omega^{-1} \Sigma\Omega^{-1}) \quad \text{as } n \to \infty,$$

where $\Omega = \operatorname{diag}(\mathbf{A}_1, 0.5\mathbf{B}_1, \mathbf{A}_2, 0.5\mathbf{B}_2)$, $\Sigma = \operatorname{diag}(\Sigma_1, \Sigma_2)$ with

$$\Sigma_i = \begin{pmatrix} \mathbf{A}_i & \frac{\kappa_3}{2}\mathbf{D}_i \\ \frac{\kappa_3}{2}\mathbf{D}_i^{\tau} & \frac{\kappa_4-1}{4}\mathbf{B}_i \end{pmatrix}, \quad i = 1, 2,$$

where $\kappa_3 = E(\varepsilon_1^3)$,

$$\mathbf{A}_i = E\left\{\frac{\mathbf{Y}_{i,t-1}\mathbf{Y}_{i,t-1}'}{\boldsymbol{\alpha}_{i0}'\mathbf{X}_{i,t-1}} g_i(r_0)\right\},$$

$$\mathbf{B}_i = E\left\{\frac{\mathbf{X}_{i,t-1}\mathbf{X}_{i,t-1}'}{(\boldsymbol{\alpha}_{i0}'\mathbf{X}_{i,t-1})^2} g_i(r_0)\right\}, \quad \text{and}$$

$$\mathbf{D}_i = E\left\{\frac{\mathbf{Y}_{i,t-1}\mathbf{X}_{i,t-1}'}{(\boldsymbol{\alpha}_{i0}'\mathbf{X}_{i,t-1})^{3/2}} g_i(r_0)\right\}$$

with $g_1(r_0) = I(y_{t-d} \le r_0)$ and $g_2(r_0) = I(y_{t-d} > r_0)$.

If $\varepsilon_1 \sim N(0, 1)$, then $\widehat{\boldsymbol{\theta}}_n$ is the maximum likelihood estimator of $\boldsymbol{\theta}_0$ and $\Omega^{-1}\Sigma\Omega^{-1} = \Omega^{-1}$. If ε_1 is symmetric, then $\kappa_3 = 0$ and $\Omega^{-1}\Sigma\Omega^{-1} = \operatorname{diag}\{\mathbf{A}_1^{-1}, (\kappa_4 - 1)\mathbf{B}_1^{-1}, \mathbf{A}_2^{-1}, (\kappa_4 - 1)\mathbf{B}_2^{-1}\}$.

To describe the limiting distribution of $n(\widehat{r}_n - r_0)$, we consider the limiting behavior of a sequence of normalized profile log-likelihood processes defined by

$$\widetilde{L}_n(z) = -2\big\{L_n\big(\widehat{\boldsymbol{\lambda}}_n(r_0 + z/n), r_0 + z/n\big) - L_n\big(\widehat{\boldsymbol{\lambda}}_n(r_0), r_0\big)\big\}, \quad z \in \mathbb{R}. \tag{3.2}$$

Using Theorem 4 and Taylor's expansion, it is straightforward to show that $\widetilde{L}_n(z)$ can be approximated in $\mathbb{D}(\mathbb{R})$ by

$$\wp_n(z) = I(z < 0)\sum_{t=1}^{n} \zeta_{1t} I\big(r_0 + z/n < y_{t-d} \le r_0\big) + I(z \ge 0)\sum_{t=1}^{n} \zeta_{2t} I\big(r_0 < y_{t-d} \le r_0 + z/n\big),$$

where

$$\zeta_{1t} = \log\frac{\boldsymbol{\alpha}_{20}'\mathbf{X}_{t-1}}{\boldsymbol{\alpha}_{10}'\mathbf{X}_{t-1}} + \frac{\left\{(\phi_{10} - \phi_{20})'\mathbf{Y}_{t-1} + \varepsilon_t\sqrt{\boldsymbol{\alpha}_{10}'\mathbf{X}_{t-1}}\right\}^2}{\boldsymbol{\alpha}_{20}'\mathbf{X}_{t-1}} - \varepsilon_t^2,$$

$$\zeta_{2t} = \log\frac{\boldsymbol{\alpha}_{10}'\mathbf{X}_{t-1}}{\boldsymbol{\alpha}_{20}'\mathbf{X}_{t-1}} + \frac{\left\{(\phi_{10} - \phi_{20})'\mathbf{Y}_{t-1} - \varepsilon_t\sqrt{\boldsymbol{\alpha}_{20}'\mathbf{X}_{t-1}}\right\}^2}{\boldsymbol{\alpha}_{10}'\mathbf{X}_{t-1}} - \varepsilon_t^2. \tag{3.3}$$

We further define a two-sided compound Poisson process $\wp(z)$ as

$$\wp(z) = I(z < 0)\wp_1(|z|) + I(z \ge 0)\wp_2(z), \ \sim z \in \mathbb{R}, \tag{3.4}$$

where $\{\wp_1(z), z \ge 0\}$ and $\{\wp_2(z), z \ge 0\}$ are two independent compound Poisson processes, both with jump rate $\pi(r_0)$, which is the value of the density $\pi(x)$ of y_1 at $x = r_0$, $\wp_1(0) = \wp_2(0) = 0$ a.s. and the distributions of jump being given by the conditional distribution of $\zeta_1 \doteq \zeta_{1t}$ given $y_{t-d} = r_0^-$ and the conditional distribution of $\zeta_2 \doteq \zeta_{2t}$ given $y_{t-d} = r_0^+$, respectively. We work with the left-continuous version for $\wp_1(z)$ and the right-continuous version for $\wp_2(z)$. The former conditional distribution is the limiting conditional distribution of ζ_{1t} given $r_0 - \delta < y_{t-d} \le r_0$ as $\delta \downarrow 0$ and the latter that of ζ_{2t} given $r_0 < y_{t-d} \le r_0 + \delta$ as $\delta \downarrow 0$. Clearly, $\wp(z)$ goes to infinity a.s. as $z \to \pm\infty$ since $E\zeta_1 > 0$ and $E\zeta_2 > 0$ by Assumption 5 and an elementary inequality $\log(1/x) + x - 1 > 0$ for $x > 0$ unless $x = 1$. Thus, there exists a unique random interval $[M_-, M_+)$ at which the process $\wp(z)$ attains its global minimum. The following theorem states that $n(\widehat{r}_n - r_0)$ converges weakly to a functional of the compound Poisson process defined in (3.4).

Theorem 5. If Assumptions 1–5 hold, then $n(\widehat{r}_n - r_0) \Longrightarrow M_-$. Furthermore, $n(\widehat{r}_n - r_0)$ is asymptotically independent of $\sqrt{n}\big(\widehat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0\big)$, which is always $N(\mathbf{0}, \sim \Omega^{-1}\Sigma\Omega^{-1})$ asymptotically.

When $\alpha_{ij} = 0$, $i = 1, 2$, $j = 1, \ldots, q_i$, model (1.1) reduces to a TAR model. Further, when $\alpha_{10} = \alpha_{20}$, Theorem 5 reduces to the asymptotic theory of the LSE of $\boldsymbol{\theta}_0$ in Chan (1993) for the TAR model. When $\alpha_{10} \neq \alpha_{20}$ and $\mu_t(\boldsymbol{\theta})$ is discontinuous, since our estimator is the QMLE, $\widehat{\boldsymbol{\lambda}}_n$ is more efficient than the LSE of $\boldsymbol{\lambda}_0$ in Chan (1993). Furthermore, $\widehat{r}_n$ has the same convergence rate as the LSE of r_0 in Chan (1993), but the jump sizes in the related compound Poisson processes are different. When $\alpha_{10} \neq \alpha_{20}$ and $\mu_t(\boldsymbol{\theta})$ is continuous, Chan and Tsay (1998) studied the LSE and showed that $\widehat{r}_n$ is $\sqrt{n}$-consistent and $\widehat{r}_n$ and $\widehat{\boldsymbol{\lambda}}_n$ are asymptotically correlated. However, Theorem 5 in this case showed that, based on our QMLE, $\widehat{r}_n$ is n-consistent and asymptotically independent of $\widehat{\boldsymbol{\lambda}}_n$. This fact is quite surprising because the LSE and the QMLE result in sharply different convergence rate of the estimated threshold.

When $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2$, Theorem 5 gives the asymptotic theory for the TAR model with ARCH errors. The corresponding parameter is $\boldsymbol{\theta} = (\boldsymbol{\lambda}', r)'$ with $\boldsymbol{\lambda} = (\phi_1', \phi_2', \boldsymbol{\alpha}')'$, and

$$\Omega^{-1}\Sigma\Omega^{-1} = \begin{pmatrix} \mathbf{A}_1^{-1} & \mathbf{0} & \kappa_3\mathbf{A}_1^{-1}\mathbf{D}_1(\mathbf{B}_1 + \mathbf{B}_2)^{-1} \\ \mathbf{0} & \mathbf{A}_2^{-1} & \kappa_3\mathbf{A}_2^{-1}\mathbf{D}_2(\mathbf{B}_1 + \mathbf{B}_2)^{-1} \\ \kappa_3(\mathbf{B}_1 + \mathbf{B}_2)^{-1}\mathbf{D}_1'\mathbf{A}_1^{-1} & \kappa_3(\mathbf{B}_1 + \mathbf{B}_2)^{-1}\mathbf{D}_2'\mathbf{A}_2^{-1} & (\kappa_4 - 1)(\mathbf{B}_1 + \mathbf{B}_2)^{-1} \end{pmatrix},$$

where $\mathbf{A}_i$, $\mathbf{B}_i$, and $\mathbf{D}_i$ are defined in Theorem 4 with replacing $\boldsymbol{\alpha}_{i0}$'s by $\boldsymbol{\alpha}_0$. When all $\phi_{ij} = 0$, Theorem 5 gives the asymptotic theory for the TARCH model. The corresponding parameter is $\boldsymbol{\theta} = (\boldsymbol{\lambda}', r)'$ with $\boldsymbol{\lambda} = (\boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2')'$, and $\Omega^{-1}\Sigma\Omega^{-1} = (\kappa_4 - 1)\operatorname{diag}(\mathbf{B}_1^{-1}, \mathbf{B}_2^{-1})$. Even for the last special cases, our results are new in literature since the threshold parameter is assumed to be known in Rabemananjara and Zakoïan (1993), Zakoïan (1994), Li and Li (1996).

4. MODEL DIAGNOSTIC CHECKING

This section studies the asymptotic distributions of residual and squared residual autocorrelation functions (ACF) of model

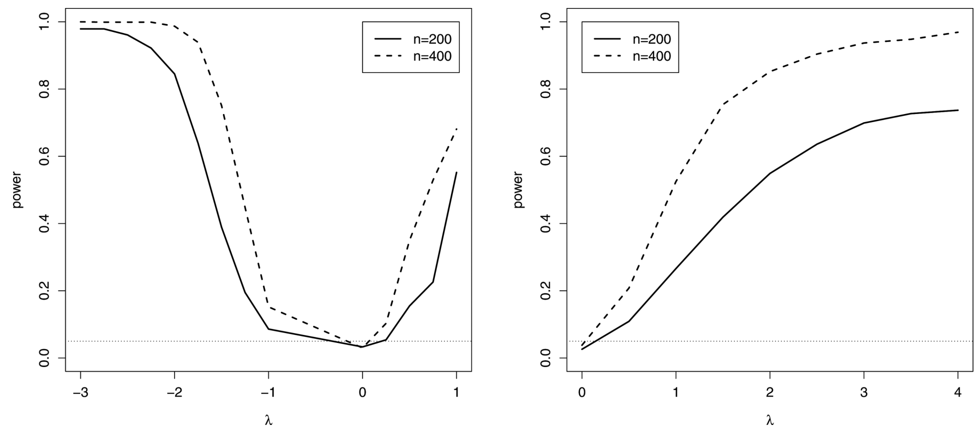

Figure 1. Powers of the test statistic S_n^a at significance level 0.05 based on 1000 simulations. The left panel is the power of the test of H_0 against the alternative (I). The right one is for the test of H_0 against the alternative (II).

Table 1. Simulation results for model (5.1) with $\boldsymbol{\theta}_0 = (1, -0.6, 1, 0.5, -1, -0.2, 0.5, 0.3, 0)'$

n		ϕ_{10}	ϕ_{11}	α_{10}	α_{11}	ϕ_{20}	ϕ_{21}	α_{20}	α_{21}	r
					$N(0, 1)$					
	EM	1.0477	−0.5741	0.8650	0.4786	−1.0173	−0.1935	0.4180	0.2923	−0.0528
100	ESD	0.3542	0.2547	0.4112	0.2148	0.2555	0.1632	0.2288	0.1082	0.1242
	ASD	0.3203	0.2363	0.3965	0.2116	0.2361	0.1550	0.2182	0.1029	0.1012
	EM	1.0253	−0.5851	0.9398	0.4865	−1.0050	−0.1983	0.4596	0.2939	−0.0250
200	ESD	0.2337	0.1664	0.2931	0.1547	0.1692	0.1086	0.1579	0.0749	0.0548
	ASD	0.2239	0.1670	0.2768	0.1501	0.1639	0.1088	0.1511	0.0725	0.0506
	EM	1.0227	−0.5909	0.9734	0.4988	−1.0135	−0.1970	0.4861	0.2971	−0.0127
400	ESD	0.1605	0.1182	0.1977	0.1069	0.1132	0.0771	0.1088	0.0506	0.0256
	ASD	0.1575	0.1171	0.1951	0.1051	0.1152	0.0764	0.1064	0.0510	0.0253
	EM	1.0042	−0.6006	0.9973	0.4926	−1.0026	−0.1996	0.4946	0.2971	−0.0061
800	ESD	0.1080	0.0811	0.1391	0.0750	0.0830	0.0540	0.0778	0.0377	0.0140
	ASD	0.1110	0.0825	0.1376	0.0741	0.0813	0.0539	0.0751	0.0360	0.0127
					st_5					
	EM	1.0114	−0.5931	0.8219	0.4231	−1.0323	−0.2003	0.3254	0.2828	−0.0591
100	ESD	0.3659	0.2602	0.5377	0.2766	0.2791	0.1666	0.3446	0.1553	0.1753
	ASD	0.3382	0.2506	0.6257	0.3338	0.2503	0.1620	0.3516	0.1589	0.1323
	EM	0.9959	−0.5960	0.9006	0.4594	−1.0064	−0.1999	0.4243	0.2804	−0.0295
200	ESD	0.2421	0.1773	0.4446	0.2270	0.1790	0.1154	0.2686	0.1245	0.0866
	ASD	0.2370	0.1751	0.4366	0.2317	0.1754	0.1132	0.2434	0.1106	0.0662
	EM	0.9960	−0.6058	0.9034	0.4839	−1.0078	−0.1958	0.4477	0.2903	−0.0136
400	ESD	0.1673	0.1233	0.3450	0.1821	0.1236	0.0816	0.2042	0.0954	0.0456
	ASD	0.1664	0.1230	0.3261	0.1738	0.1238	0.0800	0.1826	0.0833	0.0331
	EM	0.9993	−0.6041	0.9535	0.4798	−1.0002	−0.2027	0.4736	0.2905	−0.0077
800	ESD	0.1137	0.0843	0.2469	0.1259	0.0871	0.0563	0.1424	0.0690	0.0172
	ASD	0.1171	0.0865	0.2490	0.1329	0.0870	0.0562	0.1390	0.0636	0.0165
					Dexp					
	EM	1.0486	−0.5790	0.8500	0.4320	−1.0358	−0.1845	0.3582	0.2618	−0.0719
100	ESD	0.3933	0.2770	0.5795	0.2883	0.2795	0.1700	0.3675	0.1440	0.2169
	ASD	0.3568	0.2598	0.6281	0.3254	0.2643	0.1658	0.3531	0.1524	0.1527
	EM	1.0134	−0.5929	0.9154	0.4657	−1.0238	−0.1893	0.4193	0.2888	−0.0331
200	ESD	0.2586	0.1855	0.4561	0.2337	0.1870	0.1120	0.3035	0.1177	0.1033
	ASD	0.2495	0.1806	0.4531	0.2340	0.1854	0.1158	0.2554	0.1101	0.0763
	EM	1.0055	−0.5981	0.9454	0.4881	−1.0089	−0.1962	0.4522	0.2991	−0.0182
400	ESD	0.1762	0.1242	0.3466	0.1752	0.1275	0.0829	0.2072	0.0858	0.0424
	ASD	0.1750	0.1267	0.3324	0.1719	0.1303	0.0815	0.1876	0.0812	0.0382
	EM	1.0008	−0.5995	0.9819	0.4917	−1.0089	−0.1986	0.4724	0.2968	−0.0087
800	ESD	0.1249	0.0883	0.2442	0.1278	0.0924	0.0578	0.1389	0.0590	0.0198
	ASD	0.1231	0.0891	0.2352	0.1218	0.0918	0.0575	0.1328	0.0577	0.0191

Table 2. Empirical quantiles of M_{-}

α	0.5%	1%	2.5%	5%	95%	97.5%	99%	99.5%
$N(0, 1)$	−45.02	−38.20	−30.38	−24.25	5.77	12.50	21.54	28.81
st_5	−52.47	−46.91	−37.16	−29.66	8.75	19.23	33.56	46.25
Dexp	−65.14	−56.61	−44.80	−34.44	11.86	22.93	37.78	51.14

(1.1) and then uses them to construct test statistics for model checking. When the threshold is known, the related work can be found in Li and Mak (1994) and Li and Li (1996).

Let $\varepsilon_t(\boldsymbol{\lambda}, r) \equiv \varepsilon_t(\boldsymbol{\theta}) = u_t(\boldsymbol{\theta})/\sqrt{h_t(\boldsymbol{\theta})}$, where $u_t(\boldsymbol{\theta})$ and $h_t(\boldsymbol{\theta})$ are defined in (3.1). Clearly, the residual $\widehat{\varepsilon}_t = \varepsilon_t(\widehat{\boldsymbol{\lambda}}(\widehat{r}_n), \widehat{r}_n)$. Similarly, define the residual $\widetilde{\varepsilon}_t$ by $\widetilde{\varepsilon}_t \equiv \varepsilon_t(\widehat{\boldsymbol{\lambda}}(r_0), r_0)$ when r_0 is known. We first define the lag k residual ACF as follows:

$$\widehat{\rho}_k = \frac{1}{n}\sum_{t=k+1}^{n}(\widehat{\varepsilon}_t - \overline{\widehat{\varepsilon}})(\widehat{\varepsilon}_{t-k} - \overline{\widehat{\varepsilon}}), \quad k = 1, 2, \ldots,$$

where $\overline{\widehat{\varepsilon}} = n^{-1}\sum_{t=1}^{n}\widehat{\varepsilon}_t$. Similarly, we can define $\widetilde{\rho}_k$ for $\{\widetilde{\varepsilon}_t\}$. Denote $\widehat{\boldsymbol{\rho}} = (\widehat{\rho}_1, \ldots, \widehat{\rho}_m)'$ and $\widetilde{\boldsymbol{\rho}} = (\widetilde{\rho}_1, \ldots, \widetilde{\rho}_m)'$, where m is a fixed positive integer. We have the following theorem:

Theorem 6. Suppose that Assumptions 1–5 hold. Then, $\sqrt{n}\|\widehat{\boldsymbol{\rho}} - \widetilde{\boldsymbol{\rho}}\| = o_p(1)$. Furthermore,

$$\sqrt{n}\widehat{\boldsymbol{\rho}} \Longrightarrow N(\mathbf{0}, \Upsilon),$$

where $\Upsilon = \mathbf{I}_m - \mathbf{T}\Omega^{-1}(2\Omega - \Sigma)\Omega^{-1}\mathbf{T}' + \frac{\kappa_3}{2}(\mathbf{T}\Omega^{-1}\mathbf{S}' + \mathbf{S}\Omega^{-1}\mathbf{T}')$, $\mathbf{T} = (T_1, \ldots, T_m)'$, and $\mathbf{S} = (S_1, \ldots, S_m)'$ with

$$T_k = E\left\{\frac{u_{t-k}}{\sqrt{h_t h_{t-k}}}\frac{\partial u_t}{\partial \boldsymbol{\lambda}}\right\}_{|\boldsymbol{\theta}=\boldsymbol{\theta}_0} \quad \text{and}$$

$$S_k = E\left\{\frac{1}{h_t}\frac{u_{t-k}}{\sqrt{h_{t-k}}}\frac{\partial h_t}{\partial \boldsymbol{\lambda}}\right\}_{|\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

Here and in what follows, $u_t = u_t(\boldsymbol{\theta})$ and $h_t = h_t(\boldsymbol{\theta})$.

Following Li and Mak (1994), we define the lag k squared residual ACF as follows:

$$\widehat{\rho}_k^* = \frac{1}{n}\sum_{t=k+1}^{n}(\widehat{\varepsilon}_t^2 - \overline{\widehat{\varepsilon}^2})(\widehat{\varepsilon}_{t-k}^2 - \overline{\widehat{\varepsilon}^2}), \quad k = 1, 2, \ldots,$$

where $\overline{\widehat{\varepsilon}^2} = n^{-1}\sum_{t=1}^{n}\widehat{\varepsilon}_t^2$. Similarly, we define $\widetilde{\rho}_k^*$ for $\{\widetilde{\varepsilon}_t^2\}$. Denote $\widehat{\boldsymbol{\rho}}^* = (\widehat{\rho}_1^*, \ldots, \widehat{\rho}_m^*)'$ and $\widetilde{\boldsymbol{\rho}}^* = (\widetilde{\rho}_1^*, \ldots, \widetilde{\rho}_m^*)'$. We have the following theorem:

Theorem 7. Suppose that Assumptions 1–5 hold. Then, $\sqrt{n}\|\widehat{\boldsymbol{\rho}}^* - \widetilde{\boldsymbol{\rho}}^*\| = o_p(1)$ and

$$\sqrt{n}\widehat{\boldsymbol{\rho}}^* \Longrightarrow N(\mathbf{0}, \mathbf{V}),$$

where $\mathbf{V} = \mathbf{I}_m - (\kappa_4 - 1)^{-2}\mathbf{D}\Omega^{-1}\{(\kappa_4 - 1)\Omega - \Sigma\}\Omega^{-1}\mathbf{D}' - \kappa_3(\kappa_4 - 1)^{-2}(\mathbf{D}\Omega^{-1}\mathbf{J}' + \mathbf{J}\Omega^{-1}\mathbf{D}')$, $\mathbf{D} = (D_1, \ldots, D_m)'$, and $\mathbf{J} = (J_1, \ldots, J_m)'$ with

$$D_k = E\left\{\frac{1}{h_t}\frac{\partial h_t}{\partial \boldsymbol{\lambda}}\left(\frac{u_{t-k}^2}{h_{t-k}} - 1\right)\right\}_{|\boldsymbol{\theta}=\boldsymbol{\theta}_0} \quad \text{and}$$

$$J_k = E\left\{\frac{1}{\sqrt{h_t}}\frac{\partial u_t}{\partial \boldsymbol{\lambda}}\left(\frac{u_{t-k}^2}{h_{t-k}} - 1\right)\right\}_{|\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

Using Theorem 4, the proofs of Theorems 6 and 7 are straightforward and hence the details are omitted. In practice, Υ and $\mathbf{V}$ are replaced by their sample averages, denoted by $\widehat{\Upsilon}$ and $\widehat{\mathbf{V}}$, respectively. By the previous two theorems, we can construct the Ljung–Box test and the Li–Mak test as follows:

$$Q_m = n\widehat{\boldsymbol{\rho}}'\,\widehat{\Upsilon}^{-1}\widehat{\boldsymbol{\rho}} \sim \chi_m^2 \quad \text{and} \quad Q_m^* = n\widehat{\boldsymbol{\rho}}^{*\prime}\widehat{\mathbf{V}}^{-1}\widehat{\boldsymbol{\rho}}^* \sim \chi_m^2,$$

as n is large. Generally, m is taken 6 and 12, see Tse (2002) for a discussion on the choice of m.

5. SIMULATION STUDIES

We first examine the performance of S_n^a in finite samples. Under the null H_0, $\{y_t\}$ follows a DAR(1) model: $y_t = 0.2y_{t-1} + \varepsilon_t\sqrt{0.2 + 0.2y_{t-1}^2}$, where ε_t is iid $N(0, 1)$. The alternative models are

(I) $y_t = 0.2y_{t-1} + \lambda y_{t-1}I(y_{t-1} \le -1) + \varepsilon_t\sqrt{0.2 + 0.2y_{t-1}^2}$ with $-3 \le \lambda \le 1$;

Table 3. Coverage probabilities

ε_t	α	100	200	400	800
	0.01	0.979	0.986	0.989	0.984
$N(0, 1)$	0.05	0.932	0.940	0.944	0.946
	0.10	0.880	0.893	0.900	0.887
	0.01	0.970	0.980	0.984	0.987
st_5	0.05	0.906	0.925	0.934	0.949
	0.10	0.859	0.871	0.884	0.886
	0.01	0.970	0.969	0.987	0.991
Dexp	0.05	0.919	0.922	0.942	0.945
	0.10	0.845	0.878	0.886	0.892

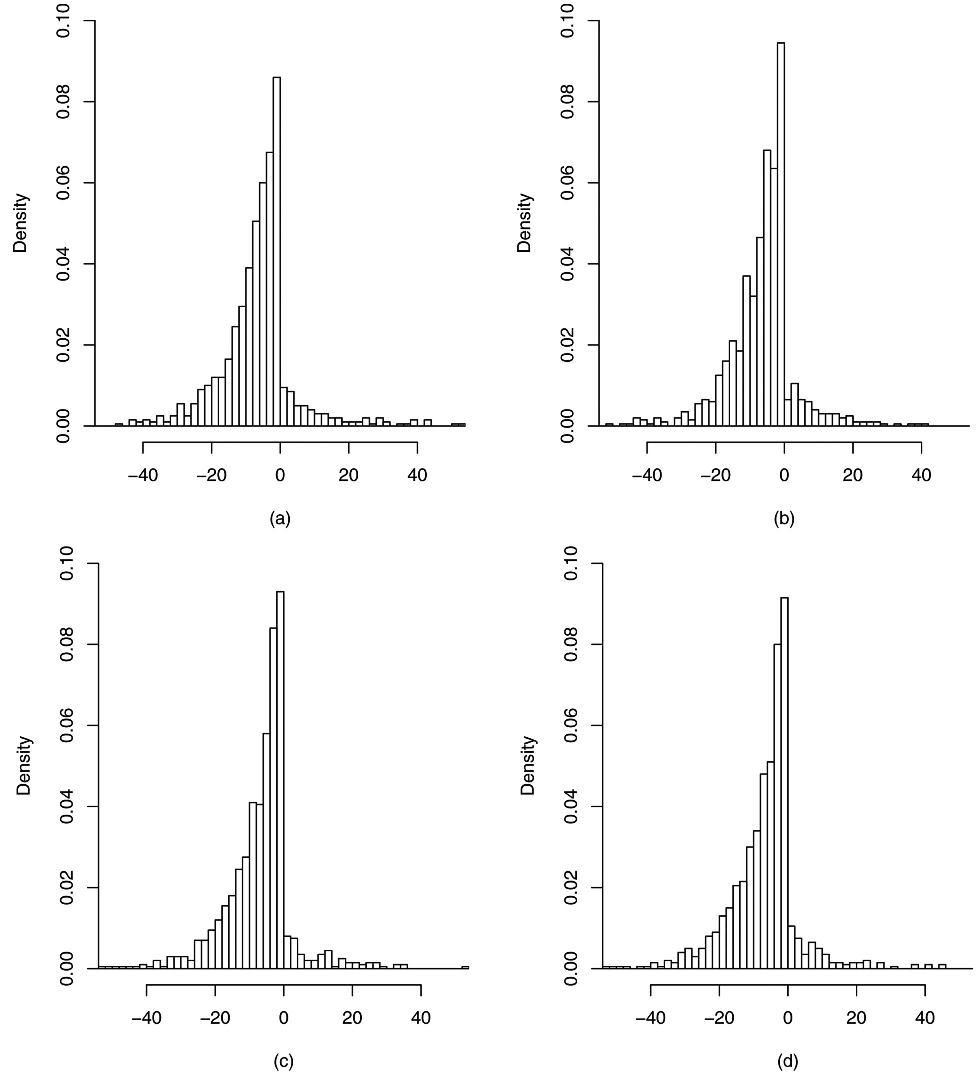

Figure 2. The densities of $n(\widehat{r}_n - r_0)$ when $n = 100$ (a), 200 (b), 400 (c), and 800 (d), respectively, for $\varepsilon_t \sim N(0, 1)$.

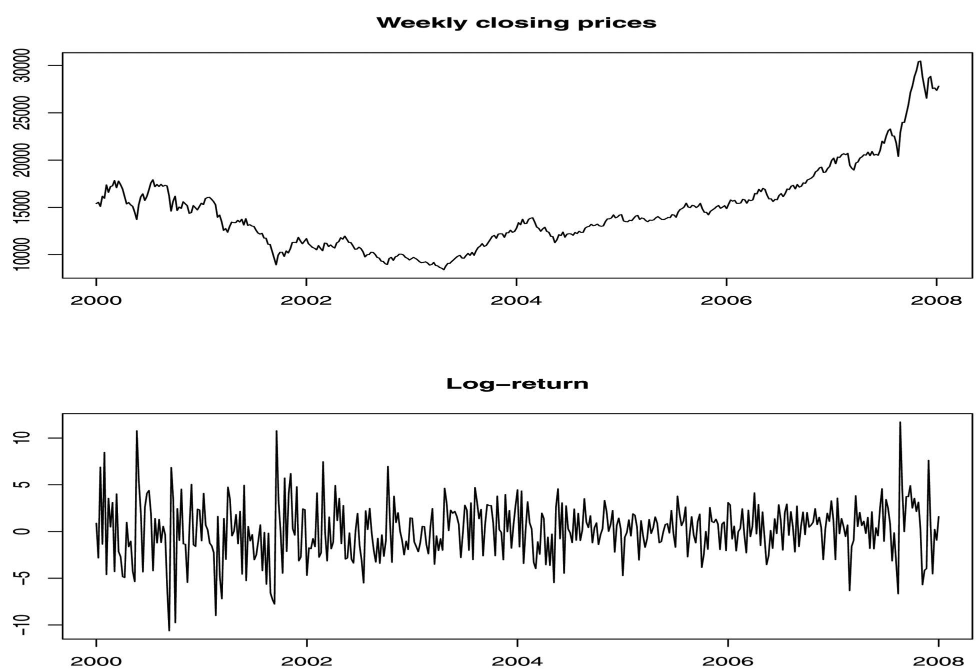

Figure 3. Time plots of the weekly closing prices and the log-returns for Hang Seng Index.

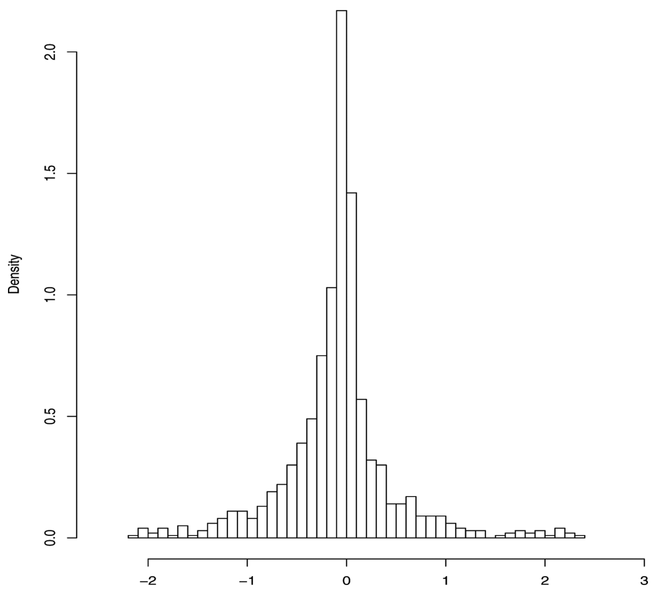

Figure 4. The density function of $\widehat{r}_n$.

(II) $y_t = 0.2y_{t-1} + \varepsilon_t\sqrt{0.2 + 0.2y_{t-1}^2 + \lambda y_{t-1}^2 I(y_{t-1} \le -1)}$ with $0 \le \lambda \le 4$.

We use the sample size $n = 200$ and 400, and 1000 replications. We take a as $5(p+q+2)\%$ quantile of data $\{y_1, \ldots, y_n\}$ and $\beta = (1, \ldots, 1)'$ in S_n^a. The significance level α is 0.05. The sizes of our test are 0.038 and 0.041 when $n = 200$ and 400, respectively. They are close to its nominal values, but there is a little conservation. Figure 1 illustrates the power of the test S_n^a in (2.3) with varying λ. From Figure 1, we can see that our test is powerful, especially when $|\lambda|$ increases.

To assess the performance of the QMLE in finite samples, we use sample sizes $n = 100$, 200, 400, and 800, each with replications 1000 for the following model:

$$y_t = \begin{cases} 1 - 0.6y_{t-1} + \varepsilon_t\sqrt{1 + 0.5y_{t-1}^2}, & \text{if } y_{t-1} \le 0, \\ -1 - 0.2y_{t-1} + \varepsilon_t\sqrt{0.5 + 0.3y_{t-1}^2}, & \text{if } y_{t-1} > 0. \end{cases} \tag{5.1}$$

ε_t takes $N(0, 1)$, standardized Student's t_5-distribution (st_5) and standardized double exponential distribution (Dexp, also called standardized Laplace distribution), respectively. Table 1 summarizes the empirical means (EM), the empirical standard deviations (ESD), and the asymptotic standard deviations (ASD). Here, the asymptotic standard deviations of $\widehat{\boldsymbol{\lambda}}_n$ and $\widehat{r}_n$ are computed by using Σ and Ω in Theorem 4 and by resampling method in Li and Ling (2012), respectively. From Table 1, we see that the consistency of the estimators is shown by the empirical means and the closeness of the empirical standard deviations to the asymptotic standard deviations. We also see that the values of the empirical standard deviations for $\widehat{r}_n$ are about halved each time when the value of n is doubled. This partially illustrates the n-consistency of $\widehat{r}_n$, under which the estimator of the threshold would approach the true value much faster than the coefficient parameter estimators do.

We now study the coverage probabilities of r_0. Using the resampling method in Li and Ling (2012), we first obtain the empirical quantiles of M_- by 10,000 replications. Table 2 gives the values for different significance level α when ε_t takes $N(0, 1)$, st_5, and Dexp. Based on the values in Table 2, the coverage probabilities of r_0 are reported in Table 3. We can see that the coverage probability is rather accurate when the sample size n is 400. To see the overall feature of the estimated threshold, Figure 2 displays the densities of $n(\widehat{r}_n - r_0)$ for different sample sizes.

6. AN EMPIRICAL EXAMPLE

The purpose of this section is to analyze the log-return of the weekly closing prices of Hang Seng Index over the period January 2000–December 2007 with 418 observations in total. Let P_t be the weekly closing price at time t. The log-return y_t is defined as $y_t = 100(\log P_t - \log P_{t-1})$. Figure 3 shows time plots of $\{P_t\}$ and $\{y_t\}$, respectively.

The p-value of Tsay's test (Tsay 1986) is 0.038, which suggests that $\{y_t\}$ contains the nonlinearity at the significant level 0.05. The p-values of the McLeod–Li test (first 36 lags) are all less than 10^{-6}, which indicates that $\{y_t\}$ has the ARCH effect. Tsay's test and McLeod–Li's test can be implemented in the R package TSA. Further, our score-based test shows that it may exist the threshold effect since the value of S_n^a is 7.139 for $p = 2$, $q = 3$, and $d = 3$. Thus, linear ARMA model is inappropriate to fit $\{y_t\}$. To capture the nonlinearity and asymmetry contained in $\{y_t\}$, we employ TDAR models. Based on the AIC, we obtain the following model:

$$y_t = \begin{cases} \underset{(0.317)}{-0.238} - \underset{(0.149)}{0.154}y_{t-1} + \underset{(0.088)}{0.264}y_{t-2} + \varepsilon_t\sigma_t, & \text{if } y_{t-1} \le \underset{(0.423)}{0}, \\ \underset{(0.250)}{-0.104} + \underset{(0.092)}{0.096}y_{t-1} - \underset{(0.061)}{0.068}y_{t-2} + \varepsilon_t\sigma_t, & \text{if } y_{t-1} > 0, \end{cases} \tag{6.1}$$

with

$$\sigma_t^2 = \begin{cases} \underset{(1.102)}{4.402} + \underset{(0.165)}{0.513}y_{t-1}^2 + \underset{(0.124)}{0.178}y_{t-2}^2 + \underset{(0.085)}{0.105}y_{t-3}^2, & \text{if } y_{t-1} \le 0, \\ \underset{(0.658)}{4.000} + \underset{(0.059)}{0.075}y_{t-2}^2 + \underset{(0.078)}{0.134}y_{t-3}^2, & \text{if } y_{t-1} > 0, \end{cases}$$

where the values in parentheses are the corresponding standard deviations calculated from Theorem 4, and the estimated delay lag d is 1. The estimator of the threshold is 0 in the sense that we use 4 decimal places. By using the resampling method in Li and Ling (2012) with 1000 replications, we get the asymptotic standard deviation 0.423 and a 95% confidence interval $[-1.338, 1.190]$ of the threshold. Figure 4 gives the density of $\widehat{r}_n$. The value of the log-likelihood is 613.41. To check the adequacy of the fit, the Ljung–Box test statistic Q_m and the McLeod–Li test statistic Q_m^* in Section 4 are used with $m = 6, 12$. The p-values of Q_6, Q_{12}, Q_6^*, and Q_{12}^* are 0.72, 0.45, 0.84, and 0.53, respectively. These p-values suggest that the fit is adequate at the significance level 0.05.

Model (6.1) clearly describes the asymmetric dynamic behavior of the log-returns in response to the past log-returns. The last log-return y_{t-1} always has a positive contribution to the current log-return y_t. Specifically, when y_{t-1} is negative (i.e., the market is dropping down), see Figure 5(a), there is a larger rebound force that pulls the current log-return y_t up since its

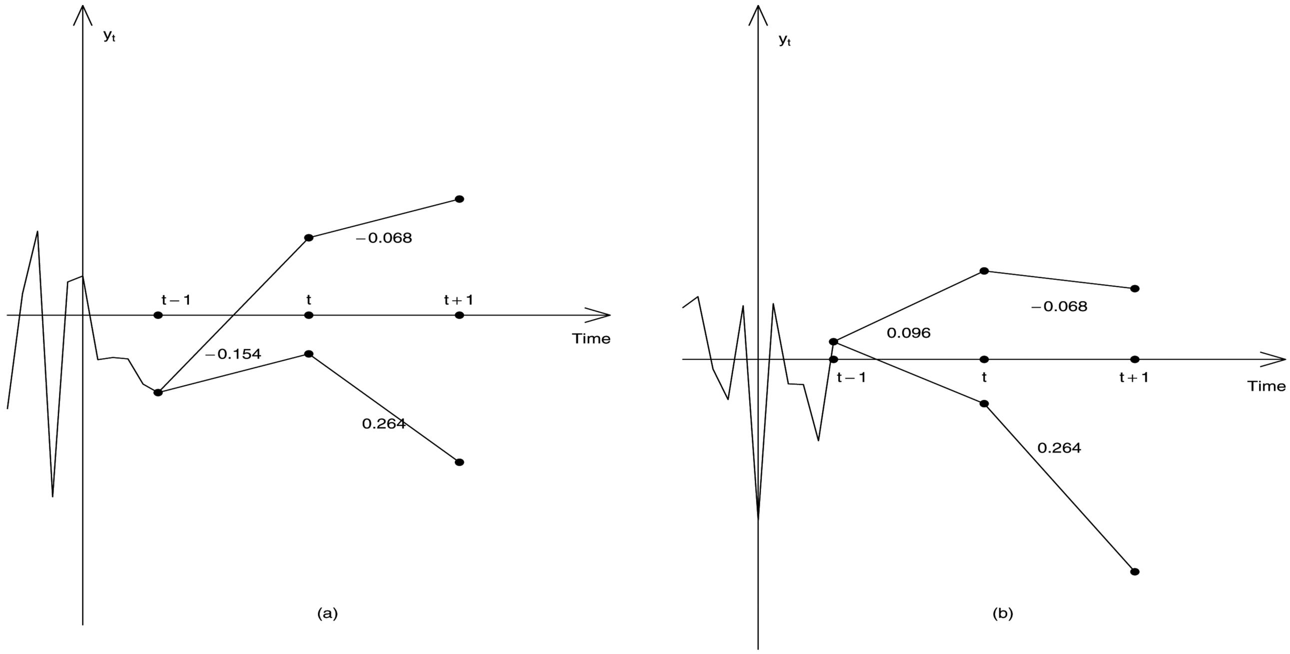

Figure 5. An illustration of model (6.1).

coefficient is -0.154. If the rebound successes (i.e., $y_t > 0$), then the persistent effect of y_{t-1} against to y_{t+1} will fade since its coefficient is -0.068. However, if the rebound fails (i.e., $y_t < 0$), then the persistent effect of y_{t-1} against to y_{t+1} will cause a sharp drop since its coefficient is 0.264. This may be because the market is weak and its investors loss their confidence. When y_{t-1} is positive, there is an analogous illustration, see Figure 5(b). The equation σ_t^2 in (6.1) reflects two different volatilities when the stock market is up and down, respectively. Given $|y_{t-1}|$, the uncertainty of the market will become larger if the market is down. This may be the leverage effect in the stock market.

APPENDIX A: PROOF OF THEOREM 1

A.1 Weak Convergence of a General Marked Empirical Process

Let $\mathcal{F}_t$ be the σ-field. Assume $\mathbf{Z}_t$ and ξ_t, $t = 0, \pm 1, \ldots$, are $\mathcal{F}_t$-measurable $p \times 1$ random vectors and univariate random variables, respectively. We consider the general marked empirical process

$$W_n(x, \tau) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[n\tau]} \mathbf{Z}_t I(\xi_{t-d} \le x), \quad (x, \tau) \in [-\infty, \infty] \times [0, 1], \tag{A.1}$$

where d is a positive integer.

Theorem 8. Let $\mathbf{K}_x \equiv E\{\mathbf{Z}_t \mathbf{Z}_t' I(\xi_{t-d} \le x)\}$. Assume (i) $\{(\mathbf{Z}_t, \xi_{t-d})\}$ is an α-mixing process with geometric rate; (ii) $E(\mathbf{Z}_t | \mathcal{F}_{t-1}) = 0$ and $0 < E\Big[\|\mathbf{Z}_t\|^2 (\log \|\mathbf{Z}_t\|)^5\Big] < \infty$; (iii) $\mathbf{K}_x$ and $\mathbf{K}_x - \mathbf{K}_y$ are positive definite for any $x, y \in \mathbb{R}$ with $x > y$. Then, $W_n(x, \tau) \Longrightarrow G(x, \tau)$ in $\mathbb{D}([-\infty, \infty] \times [0, 1])$, where $\{G(x, \tau) : (x, \tau) \in [-\infty, \infty] \times [0, 1]\}$ is a Gaussian process with mean zero and covariance kernel $\mathrm{cov}(G(x, \tau_1), G(y, \tau_2)) = (\tau_1 \wedge \tau_2) \mathbf{K}_{x \wedge y}$; almost all paths of $G(x, \tau)$ are continuous in x and τ.

Proof. First, since $\{\mathbf{Z}_t I(\xi_{t-d} \le x)\}$ is a sequence of martingale difference, the convergence of the finite-dimensional distribution can be shown by Crámer–Wold device and the martingale central limit theorem; see, for example, Billingsley (1999).

Next, we use a bracketing technique to show the tightness of $W_n(x, \tau)$. Denote $\Gamma_{(x,\tau)}(a) = a_1 I(a_2 \le x) I(a_3 \le \tau)$ for $a = (a_1, a_2, a_3) \in \mathbb{R}^3$ and

$$\mathcal{F} = \{\Gamma_{(x,\tau)} : x \in \mathbb{R}, \tau \in [0, 1]\}.$$

Let $X_{nt} = (\mathbf{Z}_t / \sqrt{n}, \sim t/n, \sim \xi_{t-d})$, then

$$W_n(x, \tau) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbf{Z}_t I(t/n \le \tau) I(\xi_{t-d} \le x) = \sum_{t=1}^{n} \Gamma_{(x,\tau)}(X_{nt}).$$

Adopt the convention $I(a \le x \le b) = -I(b \le x \le a)$ if $a \ge b$. Then, for any $(x_1, \tau_1), (x_2, \tau_2) \in [-\infty, \sim \infty] \times [0, 1]$, we have

$$\begin{aligned}
&E\|W_n(x_1, \tau_1) - W_n(x_2, \tau_2)\|^2 \\
&= \frac{1}{n} E\left\| \sum_{t=1}^{[n\tau_1]} \mathbf{Z}_t I(\xi_{t-d} \le x_1) - \sum_{t=1}^{[n\tau_2]} \mathbf{Z}_t I(\xi_{t-d} \le x_1) + \sum_{t=1}^{[n\tau_2]} \mathbf{Z}_t I(\xi_{t-d} \right. \\
&\quad \left. \le x_1) - \sum_{t=1}^{[n\tau_2]} \mathbf{Z}_t I(\xi_{t-d} \le x_2) \right\|^2 \\
&\le \frac{2}{n} E\left\| \sum_{t=[n\tau_2]}^{[n\tau_1]} \mathbf{Z}_t I(\xi_{t-d} \le x_1) \right\|^2 + \frac{2}{n} E\left\| \sum_{t=1}^{[n\tau_2]} \mathbf{Z}_t \{ I(\xi_{t-d} \le x_1) \right. \\
&\quad \left. - I(\xi_{t-d} \le x_2)\} \right\|^2 \\
&= 2|\tau_1 - \tau_2| E\{\|\mathbf{Z}_t\|^2 I(\xi_{t-d} \le x_1)\} \\
&\quad + 2\tau_2 E\{\|\mathbf{Z}_t\|^2 |I(x_2 \le \xi_{t-d} \le x_1)|\} \\
&\le 2(E\|\mathbf{Z}_1\|^2)\{|\tau_1 - \tau_2| + |G(x_1) - G(x_2)|\},
\end{aligned}$$

where $G(x) = E\{\|\mathbf{Z}_t\|^2 I(\xi_{t-d} \le x)\} / (E\|\mathbf{Z}_t\|^2)$. This implies that under the pseudo-metric

$$d((x_1, \tau_1), (x_2, \tau_2)) = \sqrt{2E\|\mathbf{Z}_1\|^2} \{|\tau_1 - \tau_2| + |G(x_1) - G(x_2)|\}^{1/2},$$

the brackets number $N(\varepsilon, \mathcal{F}, L_2)$, that is, the minimum number of ε-brackets to cover $\mathcal{F}$ (see van der Vaart 1998, p. 270), is of order ε^{-4}. Thus, for any finite $\delta > 0$, we have that the integral of the bracketing entropy

$$\int_0^\delta \sqrt{\log N(\varepsilon, \mathcal{F}, L_2)}\, d\varepsilon \le C \int_0^\delta \sqrt{\log(1/\varepsilon)}\, d\varepsilon < \infty.$$

Fixed q_0 such that $4\delta \le 2^{-q_0} \le 8\delta$. Let $P_q = \{\Gamma_{(x,\tau)} : (x, \tau) \in B_{qi}, 1 \le i \le N_q\}$, $q \ge q_0$, be a nested sequence of finite partitions of $\mathcal{F}$ such that

$$\sum_{q=q_0}^{\infty} 2^{-q} \sqrt{\log N_q} < \int_0^\delta \sqrt{\log N(\varepsilon, \mathcal{F}, L_2)}\, d\varepsilon,$$

$$\begin{aligned} E\Lambda^2(B_{qi}) := {} & \frac{1}{n} E \sum_{t=1}^{n} \sup_{(x,\tau_1),(x,\tau_2)\in B_{qi}} \{\|\mathbf{Z}_t\|^2 I(\xi_{t-d} \le x) | I(\tau_2 \\ & \le t/n \le \tau_1)|\} \\ & + \frac{1}{n} E \sum_{t=1}^{n} \sup_{(x_1,\tau),(x_2,\tau)\in B_{qi}} \{\|\mathbf{Z}_t\|^2 I(\xi_{t-d} \le x) I(t/ \\ & n \le \tau) | I(x_1 \le \xi_{t-d} \le x_2)|\} \\ \le {} & 2^{-2q}. \end{aligned} \tag{A.2}$$

This can be obtained as in Lemma 19.34 of van der Vaart (1998, p. 286). For each q, we choose a fixed element $(x_{qi}, \sim \tau_{qi}) \in B_{qi}$ and set

$$(\pi_q x, \pi_q \tau) = (x_{qi}, \tau_{qi}) \quad \text{and} \quad (B_q x, B_q \tau) = B_{qi}, \text{if}(x, \tau) \in B_{qi}.$$

Then, using the Bernstein-type inequality (2.3) in Merlevède, Peligrad, and Rio (2009) and truncating $\mathbf{Z}_t$ by $\sqrt{n}/(\log n)^2$ instead of $\sqrt{n}$ in the proof of Theorem 2.5.6 in van der Vaart and Wellner (1996), the proof is concluded.

A.2 Proof of Theorem 1

Under the conditions of Theorem 1, it is not hard to get

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \|D_t(\widehat{\boldsymbol{\theta}}_n) D_t'(\widehat{\boldsymbol{\theta}}_n) - D_t(\boldsymbol{\theta}_0) D_t'(\boldsymbol{\theta}_0)\| = o_p(1).$$

Using this equality, we then have

$$\sup_{x\in\overline{\mathbb{R}}} \|\widehat{\Sigma}_{nx} - \Sigma_x\| \le \sup_{x\in\overline{\mathbb{R}}} \|\Xi(x)\| + o_p(1),$$

where

$$\Xi(x) = \frac{1}{n} \sum_{t=1}^{n} D_t(\boldsymbol{\theta}_0) D_t'(\boldsymbol{\theta}_0) I(y_{t-d} \le x) - \Sigma_x.$$

By Theorem 2 in Pollard (1984, p. 8), we can get $\sup_{x\in\overline{\mathbb{R}}} \|\Xi(x)\| = o_p(1)$. Thus,

$$\sup_{x\in\overline{\mathbb{R}}} \|\widehat{\Sigma}_{nx} - \Sigma_x\| = o_p(1). \tag{A.3}$$

By the Taylor expansion and (A.3), it follows that

$$\sup_{x\in\overline{\mathbb{R}}} \left\| T_n(x, \widehat{\boldsymbol{\theta}}_n) - \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbf{U}^{-1} D_t(\boldsymbol{\theta}_0) I(y_{t-d} \le x) + \mathbf{U}^{-1} \right.$$
$$\left. \Sigma_x \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \right\| = o_p(1),$$

where $\mathbf{U} = \operatorname{diag}\{\mathbf{I}_{p+1}, \sqrt{0.5(\kappa_4 - 1)}\mathbf{I}_{q+1}\}$. Thus, $T_n(x, \widehat{\boldsymbol{\theta}}_n)$ has the same asymptotical behavior as

$$\begin{aligned} & \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^{n} \mathbf{U}^{-1} D_t(\boldsymbol{\theta}_0) I(y_{t-d} \le x) - \mathbf{U}^{-1} \Sigma_x \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\ = {} & \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^{n} \left[\mathbf{U}^{-1} D_t(\boldsymbol{\theta}_0)\right] I(y_{t-d} \le x) - \Sigma_x \Sigma_\infty^{-1} \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^{n} [\mathbf{U}^{-1} D_t(\boldsymbol{\theta}_0)] \end{aligned}$$

since $\Sigma_x \Sigma_\infty^{-1}$ and $\mathbf{U}^{-1}$ are block diagonal and commutative. Let $\mathbf{Z}_t = \mathbf{U}^{-1} D_t(\boldsymbol{\theta}_0)$ and $\xi_{t-d} = y_{t-d}$. Applying Theorem 8 with $\tau = 1$, then Theorem 1 holds.

APPENDIX B: PROOFS OF THEOREMS 3–5

B.1 Proof of Theorem 3

Let $\beta(\boldsymbol{\theta}) = E\{\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}_0)\}$. For any given open neighborhood V of $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ and any $\boldsymbol{\theta} \in V^c \cap \boldsymbol{\Theta}$, a conditional argument yields that

$$\begin{aligned} -2\beta(\boldsymbol{\theta}) = {} & E\{K_{1t} I(y_{t-d} \le r_0) + K_{2t} I(r_0 < y_{t-d} \le r) \\ & + K_{3t} I(y_{t-d} > r)\}, \end{aligned}$$

where

$$K_{1t} = \log \frac{\boldsymbol{\alpha}_1' \mathbf{X}_{t-1}}{\boldsymbol{\alpha}_{10}' \mathbf{X}_{t-1}} + \frac{\boldsymbol{\alpha}_{10}' \mathbf{X}_{t-1}}{\boldsymbol{\alpha}_1' \mathbf{X}_{t-1}} - 1 + \frac{\{(\phi_{10} - \phi_1)' \mathbf{Y}_{t-1}\}^2}{\boldsymbol{\alpha}_1' \mathbf{X}_{t-1}},$$

$$K_{2t} = \log \frac{\boldsymbol{\alpha}_1' \mathbf{X}_{t-1}}{\boldsymbol{\alpha}_{20}' \mathbf{X}_{t-1}} + \frac{\boldsymbol{\alpha}_{20}' \mathbf{X}_{t-1}}{\boldsymbol{\alpha}_1' \mathbf{X}_{t-1}} - 1 + \frac{\{(\phi_{20} - \phi_1)' \mathbf{Y}_{t-1}\}^2}{\boldsymbol{\alpha}_1' \mathbf{X}_{t-1}},$$

$$K_{3t} = \log \frac{\boldsymbol{\alpha}_2' \mathbf{X}_{t-1}}{\boldsymbol{\alpha}_{20}' \mathbf{X}_{t-1}} + \frac{\boldsymbol{\alpha}_{20}' \mathbf{X}_{t-1}}{\boldsymbol{\alpha}_2' \mathbf{X}_{t-1}} - 1 + \frac{\{(\phi_{20} - \phi_2)' \mathbf{Y}_{t-1}\}^2}{\boldsymbol{\alpha}_2' \mathbf{X}_{t-1}}.$$

Observe that all $K_{it} \ge 0$ a.s. by an elementary inequality $\log(1/x) + x - 1 > 0$ for $x > 0$ unless $x = 1$. Hence, $\beta(\boldsymbol{\theta}) < 0$. The remainder is similar to that of Theorem 2.1 in Li, Ling, and Li (2013) and hence it is omitted.

B.2 Proof of Theorem 4

(i) We only prove the case $p = 1$. When $p > 1$, using the technique in Chan (1993, p. 529), the proof would go through with a minor modification. Since $\widehat{\boldsymbol{\theta}}_n$ is strongly consistent, we restrict the parameter space to a neighborhood $V_\delta = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : \|\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\| < \delta,\ |r - r_0| < \delta\}$ of $\boldsymbol{\theta}_0$ for some $0 < \delta < 1$ to be determined later. Then, it suffices to prove that there exist constants $B > 0$ and $\gamma > 0$ such that, for any $\varepsilon > 0$,

$$P\left(\sup_{\substack{B/n < |r - r_0| \le \delta \\ \theta \in V_\delta}} \frac{L_n(\boldsymbol{\lambda}, r) - L_n(\boldsymbol{\lambda}, r_0)}{nG(|r - r_0|)} < -\gamma \right) > 1 - \varepsilon, \tag{A.1}$$

as n is large enough, where $G(u) = P(r_0 < y_0 \le r_0 + u)$. Writing $r = r_0 + u$ for some $u \ge 0$. By a calculation, it follows that

$$\begin{aligned} \frac{2\{L_n(\boldsymbol{\lambda}, r) - L_n(\boldsymbol{\lambda}, r_0)\}}{nG(u)} = {} & \frac{-1}{nG(u)} \sum_{t=1}^{n} \zeta_{2t} I(r_0 < y_{t-1} \le r_0 + u) \\ & + O_p(\sqrt{\delta}) \\ = {} & -K_4 \frac{G_n(u)}{G(u)} + K_5 \frac{\sum_{t=1}^{n} \varepsilon_t I(r_0 < y_{t-1} \le r_0 + u)}{nG(u)} \\ & + K_6 \frac{\sum_{t=1}^{n} (\varepsilon_t^2 - 1) I(r_0 < y_{t-1} \le r_0 + u)}{nG(u)} + O_p(\sqrt{\delta}), \end{aligned}$$

where $G_n(u) = \frac{1}{n} \sum_{t=1}^{n} I(r_0 < y_{t-1} \le r_0 + u)$,

$$K_4 = \log \frac{\boldsymbol{\alpha}_{10}' \mathbf{X}}{\boldsymbol{\alpha}_{20}' \mathbf{X}} + \frac{\boldsymbol{\alpha}_{20}' \mathbf{X}}{\boldsymbol{\alpha}_{10}' \mathbf{X}} - 1 + \frac{\{(\phi_{20} - \phi_{10})' \mathbf{Y}\}^2}{\boldsymbol{\alpha}_{10}' \mathbf{X}},$$

$$K_5 = \frac{2\{(\phi_{10} - \phi_{20})' \mathbf{Y}\} \sqrt{\boldsymbol{\alpha}_{20}' \mathbf{X}}}{\boldsymbol{\alpha}_{10}' \mathbf{X}}, \quad \text{and} \quad K_6 = \frac{(\boldsymbol{\alpha}_{10} - \boldsymbol{\alpha}_{20})' \mathbf{X}}{\boldsymbol{\alpha}_{10}' \mathbf{X}}$$

with $\mathbf{Y} = (1, r_0)'$ and $\mathbf{X} = (1, r_0^2)'$. Similar to Claim 2 in Chan (1993), for any $\varepsilon > 0$ and $\eta > 0$, there exists a positive constant B such that as n is large enough

$$P\left(\sup_{B/n < u \le \delta} \left| \frac{G_n(u)}{G(u)} - 1 \right| < \eta \right) > 1 - \varepsilon,$$

$$P\left(\sup_{B/n<u\le\delta}\left|\frac{\sum_{t=1}^{n}\varepsilon_t I(r_0<y_{t-1}\le r_0+u)}{nG(u)}\right|<\eta\right)>1-\varepsilon,$$

$$P\left(\sup_{B/n<u\le\delta}\left|\frac{\sum_{t=1}^{n}(\varepsilon_t^2-1) I(r_0<y_{t-1}\le r_0+u)}{nG(u)}\right|<\eta\right)>1-\varepsilon.$$

Note that $K_4>0$ by Assumption 5. Choosing δ small enough and $\gamma=K_4/4$, (B.1) holds and so does (i).

The proof of (ii) is similar to that of Theorem 2.2 in Li, Ling, and Li (2013). It is trivial and hence it is omitted.

B.3 Proof of Theorem 5

Without loss of generality, we assume that ζ_{it}, defined in (3.3), is bounded. Otherwise, we can truncate it using the technique in Li, Ling, and Li (2013) and consider a new process made up of the truncated random variables. Consider the weak convergence of the process $\wp_n(z)$ on the interval $[0, T]$. The tightness of $\wp_n(z)$ can be easily shown by Theorem 5 in Kushner (1984, p. 32). The key step is to describe convergence of finite-dimensional distributions. To this end, for any $0\le z_1\le z_2<z_3\le z_4\le T$ and for any constants c_1 and c_2, the linear combination of the increments of $\wp_n(z)$ is

$$S_n\equiv c_1\{\wp_n(z_2)-\wp_n(z_1)\}+c_2\{\wp_n(z_4)-\wp_n(z_3)\}=\textstyle\sum_{t=1}^{n}J_t^{\epsilon},$$

where $J_t^{\epsilon}=\zeta_{2t}\{c_1I_t(z_1,z_2)+c_2I_t(z_3,z_4)\}$, $\epsilon=1/n$, and $I_t(u,v)=I(r_0+u\epsilon<y_{t-1}\le r_0+v\epsilon)$. We first verify Assumptions A.1–A.3 in Li, Ling, and Li (2013) for J_t^{ϵ}. By Assumption 3, it follows that

$$\lim_{n\to\infty}\epsilon^{-1}P_k^{\epsilon}(J_n^{\epsilon}\ne 0)=\pi(r_0)\{(z_2-z_1)+(z_4-z_3)\}. \tag{A.2}$$

By Assumption 3 again, for any Borel set B, it follows that

$$Q^*(B)=\lim_{n\to\infty}P(J_n^{\epsilon}\in B|J_n^{\epsilon}\ne 0)=wQ_1^*(B)+(1-w)Q_2^*(B), \tag{A.3}$$

where $w=(z_2-z_1)/\{(z_2-z_1)+(z_4-z_3)\}$ and $Q_i^*(B)=P(c_i\zeta_{2t}\in B)$, $i=1,2$. By a conditional argument, for any $f\in\widehat{C}_0^2$, a space of functions with compact support and continuous second derivative, and a scalar x,

$$\begin{aligned}E_k^{\epsilon}\{f(x+J_n^{\epsilon})-f(x)|J_n^{\epsilon}\ne 0\}&=E\{f(x+J_n^{\epsilon})-f(x)|J_n^{\epsilon}\ne 0\}\\&\to\int\{f(x+u)-f(x)\}Q^*(du),\end{aligned} \tag{A.4}$$

as $n\to\infty$. By (A.2)–(A.4), Assumptions A.1–A.3 in Li, Ling, and Li (2013) hold. Furthermore, by their Theorem A.1, we have that S_n converges weakly to a compound Poisson random variable J with jump rate $\pi(r_0)\{(z_2-z_1)+(z_4-z_3)\}$ and the jump distribution Q^*. The characteristic function $f_J(t)$ of J is equal to that of $c_1\{\wp(z_2)-\wp(z_1)\}+c_2\{\wp(z_4)-\wp(z_3)\}$, where $\wp(z)$ is defined in (3.4). Thus, $\widetilde{L}_n(z)$, defined in (3.2), converges weakly to $\wp(z)$ as $n\to\infty$. The remainder of the proof is similar to that of Theorem 2 in Chan (1993).

ACKNOWLEDGMENTS

The authors thank the editor, the associate editor, and two anonymous referees for their helpful comments that improved the presentation. Li's research is supported in part by the Start-up Fund of Tsinghua University (No. 553310013) and the National Natural Science Foundation of China (NSFC) under grant (No. 11401337). Ling's research is supported in part by Hong Kong Research Grants Commission Grants HKUST641912, HKUST602609, and HKUST603413. Zhang's research is supported in part by NSFC grants No. 11371318 and No. 11171074, the Fundamental Research Funds for the Central Universities and Scientific Research Fund of Zhejiang Provincial Education Department (Y201009944).

[Received June 2012. Revised September 2014.]

REFERENCES

Billingsley, P. (1999), Convergence of Probability Measures (2nd ed.), New York: Wiley. [2,10]

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327. [1]

Chan, K. S. (1990), "Testing for Threshold Autoregression," *The Annals of Statistics*, 18, 1886–1894. [2,3]

——— (1993), "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model," *The Annals of Statistics*, 21, 520–533. [2,4,5,11,12]

Chan, K. S., and Tong, H. (1990), "On Likelihood Ratio Tests for Threshold Autoregression," *Journal of the Royal Statistical Society*, Series B, 52, 469–476. [2,3]

Chan, K. S., and Tsay, R. S. (1998), "Limiting Properties of the Least Squares Estimator of a Continuous Threshold Autoregressive Model," *Biometrika*, 85, 413–426. [2,4,5]

Chan, N. H., and Peng, L. (2005), "Weighted Least Absolute Deviation Estimation for an AR(1) Process With ARCH(1) Errors," *Biometrika*, 92, 477–484. [2]

Chen, M., Li, D., and Ling, S. (2014), "Non-Stationarity and Quasi-Maximum Likelihood Estimation on a Double Autoregressive Model," *Journal of Time Series Analysis*, 35, 189–202. [2]

Cline, D. B. H., and Pu, H. H. (2004), "Stability and the Lyapounov Exponent of Threshold AR-ARCH Models," *The Annals of Applied Probability*, 14, 1920–1949. [1,4]

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007. [1]

Francq, C., and Zakoïan, J.-M. (2010), GARCH Models: Structure, Statistical Inference and Financial Applications, New York: Wiley. [1] Q3

Hansen, B. E. (1997), "Inference in TAR Models," *Studies in Nonlinear Dynamics and Econometrics*, 2, 1–14. [2]

——— (2000), "Sample Splitting and Threshold Estimation," *Econometrica*, 68, 575–603. [2]

——— (2011), "Threshold Autoregression in Economics," *Statistics and Its Interface*, 4, 123–128. [1]

Kushner, H. J. (1984), Approximation and Weak Convergence Methods for Random Processes, With Applications to Stochastic Systems Theory, Cambridge, MA: MIT Press. [12] Q4

Li, C. W., and Li, W. K. (1996), "On a Double Threshold Autoregressive Heteroscedastic Time Series Model," *Journal of Applied Econometrics*, 11, 253–274. [1,5]

Li, D., and Ling, S. (2012), "On the Least Squares Estimation of Multiple-regime Threshold Autoregressive Models," *Journal of Econometrics*, 167, 240–253. [2,4,9]

Li, D., Ling, S., and Li, W. K. (2013), "Asymptotic Theory on the Least Squares Estimation of Threshold Moving-Average Models," *Econometric Theory*, 29, 482–516. [4,11,12]

Li, W. K., and Lam, K. (1995), "Modeling Asymmetry in Stock Returns by a Threshold ARCH Model," *Journal of the Royal Statistical Society*, Series D, 44, 333–341. [1]

Li, W. K., and Mak, T. K. (1994), "On the Squared Residual Autocorrelations in Non-Linear Time Series With Conditional Heteroskedasticity," *Journal of Time Series Analysis*, 15, 627–636. [7]

Ling, S. (1999), "On the Probabilistic Properties of a Double Threshold ARMA Conditional Heteroskedastic Model," *Journal of Applied Probability*, 36, 688–705. [1]

——— (2004), "Estimation and Testing Stationarity for Double Autoregressive Models," *Journal of the Royal Statistical Society*, Series B, 66, 63–78. [1]

——— (2007), "A Double AR(p) Model: Structure and Estimation," *Statistica Sinica*, 17, 161–175. [1]

Ling, S., and Li, D. (2008), "Asymptotic Inference for a Nonstationary Double AR(1) Model," *Biometrika*, 95, 257–263. [2,4]

Ling, S., and Tong, H. (2011), "Score Based Goodness-of-Fit Tests for Time Series," *Statistica Sinica*, 21, 1807–1829. [2]

Merlevède, F., Peligrad, M., and Rio, E. (2009), Bernstein Inequality and Moderate Deviations Under Strong Mixing Conditions (*IMS Collections, High Dimensional Probability V*), pp. 273–292. [11] Q5

Pollard, D. (1984), Convergence of Stochastic Processes, Springer-Verlag. [11] Q6

Rabemananjara, R., and Zakoïan, J.-M. (1993), "Threshold ARCH Models and Asymmetries in Volatility," *Journal of Applied Econometrics*, 8, 31–49. [2,5]

Seo, M. H., and Linton, O. (2007), “A Smoothed Least Squares Estimator for Threshold Regression Models,” *Journal of Econometrics*, 141, 704–735. [2]

Shorack, G. R., and Wellner, J. A. (1986), Empirical Processes With Applications to Statistics, New York: Wiley. [3]

Tong, H. (1978), “On a Threshold Model,” in Pattern Recognition and Signal Processing, ed. C. H. Chen, Amsterdam: Sijthoff and Noordhoff, pp. 575–586. [1]

——— (1990), Non-Linear Time Series: A Dynamical System Approach, New York: Oxford University Press. [1]

——— (2011), “Threshold Models in Time Series Analysis — 30 Years On,” *Statistics and Its Interface*, 4, 107–118. [1]

Tsay, R. S. (1986), “Nonlinearity Test for Time Series,” *Biometrika*, 73, 461–466. [xxxx]

——— (1987), “Conditional Heteroscedastic Time Series Models,” *Journal of the American Statistical Association*, 82, 590–604. [9]

Tsay, R. S. (2010), Analysis of Financial Time Series (3rd ed.), New York: Wiley. [1]

Q7

Tse, Y. K. (2002), “Residual-Based Diagnostics for Conditional Heteroscedasticity Models,” *Econometrics Journal*, 5, 358–373. [7]

van der Vaart, A. W. (1998), Asymptotic Statistics. Cambridge: Cambridge University Press. [11]

van der Vaart, A. W., and Wellner, J. A. (1996), Weak Convergence and Empirical Processes: With Applications to Statistics, New York: Springer-Verlag. [11]

Weiss, A. A. (1986), “Asymptotic Theory for ARCH Models: Estimation and Testing,” *Econometrics Theory*, 2, 107–131. [1]

Wong, C. S., and Li, W. K. (1997), “Testing for Threshold Autoregression With Conditional Heteroscedasticity,” *Biometrika*, 84, 407–418. [2,3]

——— (2000), “Testing for Double Threshold Autoregressive Conditional Heteroscedastic Model,” *Statistica Sinica*, 10, 173–189. [2,3]

Zakoïan, J.-M. (1994), “Threshold Heteroskedastic Models,” *Journal of Economic Dynamics and Control*, 18, 931–955. [2,5]

Zhang, X., Wong, H., Li, Y., and Ip, W.-C. (2011), “A Class of Threshold Autoregressive Conditional Heteroscedastic Models,” *Statistics and Its Interface*, 4, 149–157. [2]

Zhu, K., and Ling, S. (2013), “Quasi-maximum Exponential Likelihood Estimators for a Double AR(p) Model,” *Statistica Sinica*, 23, 251–270. [2]