

A class of threshold autoregressive conditional heteroscedastic models

XINGFA ZHANG, HEUNG WONG*, YUAN LI AND WAI-CHEUNG IP

This paper generalizes Ling’s (2007) double AR(p) model by considering a threshold effect in the mean equation. Provided the threshold is known, consistency and asymptotic normality of the quasi maximum likelihood estimators for the model are proved under weak conditions. Based on the Lagrange Multiplier principle, a threshold effect test is studied and its asymptotic null distribution is shown to be a functional of a zero-mean Gaussian process. Approximate methods are given to compute the upper percentage points and simulation results show that they perform well. From the empirical studies, we know that the original model can be improved when the threshold effect is considered.

KEYWORDS AND PHRASES: Threshold AR(p) model, Quasi maximum likelihood estimator, Asymptotic normality, Lagrange multiplier test.

1. INTRODUCTION

In a recent paper, Ling (2007) considered the double AR(p) or DAR(p) model, which has the form

$$\begin{aligned}
 (1) \quad & y_t = \theta_1 y_{t-1} + \dots + \theta_p y_{t-p} + \varepsilon_t, \\
 & \varepsilon_t = e_t \sqrt{h_t}, e_t \sim i.i.d(0, 1), \\
 & h_t = \omega + a_1 y_{t-1}^2 + \dots + a_p y_{t-p}^2,
 \end{aligned}$$

where $\omega, a_i > 0, t \in \mathcal{N} \equiv \{-p, \dots, 0, 1, 2, \dots\}$, y_s is independent of $\{e_t\}$ for $t > s$. Let \mathcal{F}_t be the σ -field generated by $\{e_t, \dots, e_1, y_0, \dots, y_{-p}\}, t \in \mathcal{N}$, then we have $\text{var}(y_t | \mathcal{F}_{t-1}) = \omega + a_1 y_{t-1}^2 + \dots + a_p y_{t-p}^2$. As mentioned in the paper, model (1) is a special case of ARMA-ARCH models in Weiss (1986), but it differs from Engle’s (1982) ARCH model if $\theta_i \neq 0$. The difference lies in the specifications of the conditional variance: Engle’s (1982) conditional variance is driven by the unobserved errors while model (1)’s depends on the past observations. Such a specification of the conditional variance brings both novelty and difficulty. The novel results acquired by Ling (2004, 2007) were that the quasi maximum likelihood estimators can be still consistent and asymptotically normal when $E y_t^2 = \infty$, which usually does not hold any more for the classical AR(p) model with

i.i.d errors. Difficulty lies in finding the conditions under which the series generated from the model is stationary and geometrically ergodic. Though Ling (2007) has obtained a sufficient and necessary condition on stationarity and ergodicity for model (1) when $e_t \sim i.i.d. N(0, 1)$, it is still an open problem for general cases.

When $p = 1$, model (1) becomes the DAR(1) model whose theoretical results and practical application have been well discussed by Ling (2004), Ling and Li (2008) for both stationary and nonstationary cases. When the DAR(1) model was applied to the US 3-month treasure bill rate series in Ling (2004), it was found that the model is superior to the usual AR(1) and seems to be able to get a more reliable statistical inference compared to the usual AR(1)-GARCH(1, 1) model. Nevertheless, since financial data usually present some asymmetric effect or nonlinear relationship, it is helpful to take these factors into account. A well-known tool to deal with this is the threshold autoregressive model because of its ability to capture some important characteristics such as jumps and limit cycles (Tong and Lim, 1980; Tong, 1990; Li and Lam, 1995). Consequently, it is worthwhile to consider a generalized DAR(p) model, which is piecewise linear in the mean function.

In this paper we consider the following threshold autoregressive conditional heteroscedastic (TARCH) model:

$$\begin{aligned}
 (2) \quad & y_t = \theta_0 + \sum_{i=1}^p \theta_i y_{t-i} + I(y_{t-d} \leq r) (\phi_0 + \sum_{i=1}^p \phi_i y_{t-i}) \\
 & + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t}, e_t \sim i.i.d(0, 1), \\
 & h_t = \omega + a_1 y_{t-1}^2 + \dots + a_m y_{t-m}^2,
 \end{aligned}$$

where $\omega, a_i > 0, t \in \mathcal{N} \equiv \{-m, \dots, 0, 1, 2, \dots\}$, y_s is independent of $\{e_t\}$ for $t > s$; $I(\cdot)$ is the indicator function and r is the threshold parameter. For simplicity, the nonnegative integers p, d, m are assumed known and satisfy $0 \leq p \leq m, 1 \leq d \leq m$. The threshold parameter r is assumed to have a known bounded numerical range \tilde{R} , usually a finite interval. When $\theta_0 = \phi_0 = \phi_i = 0, p = m$, model (2) is reduced to Ling’s (2007) DAR(p) model. Model (2) also bears resemblance to Li and Lam’s (1995) TARCH model. The difference is that: the former belongs to Weiss’ ARCH-type models while the latter is an Engle’s ARCH-type model. Moreover, we relax the distribution of the process $\{\varepsilon_t\}$ to the more general cases instead of the original normal distribution.

*Corresponding author.

The paper is arranged as follows. In Section 2, we discuss the quasi maximum likelihood estimator (QMLE), threshold effect test and some associated asymptotic properties. Simulations and empirical studies are shown in Section 3. We conclude the paper in Section 4 and all proofs are put in the Appendix.

2. ESTIMATION AND THRESHOLD EFFECT TEST

2.1 Parametric estimation

We assume the threshold parameter r in model (2) is known. In practice, as that has been done in the subsequent Section 3.2, we can adopt the idea of Li and Lam (1995) to estimate r . Let $\psi = (\theta^\tau, \phi^\tau, a^\tau)^\tau$, $\theta = (\theta_0, \dots, \theta_p)^\tau$, $\phi = (\phi_0, \dots, \phi_p)^\tau$, $a = (\omega, a_1, \dots, a_m)^\tau$ and $\psi \in \Psi$, which is a bounded parameter space for model (2). All through this article, the superscript τ denotes the transpose of a vector or a matrix. Suppose that the true parameter $\psi_0 = (\theta_0^\tau, \phi_0^\tau, a_0^\tau)^\tau$ is an interior point of Ψ . Without loss of generality, we consider Ψ as a neighborhood of ψ_0 . We need to estimate ψ based on the observations $\{y_t\}_{t=1}^T$ and initial values y_0, \dots, y_{1-m} .

Consider the following quasi conditional log-likelihood (apart from a constant term)

$$(3) \quad L_T(\psi) = \sum_{t=1}^T l_t(\psi) = \sum_{t=1}^T \left[-\frac{1}{2} \log h_t - \frac{1}{2} \frac{\varepsilon_t^2}{h_t} \right],$$

and we have

$$(4) \quad \frac{\partial l_t}{\partial \psi} = -\frac{1}{2} \left(1 - \frac{\varepsilon_t^2}{h_t} \right) \frac{1}{h_t} \frac{\partial h_t}{\partial \psi} - \frac{\varepsilon_t}{h_t} \frac{\partial \varepsilon_t}{\partial \psi},$$

$$(5) \quad \begin{aligned} \frac{\partial^2 l_t}{\partial \psi \partial \psi^\tau} &= \frac{1}{2h_t^2} \left(1 - \frac{\varepsilon_t^2}{h_t} \right) \frac{\partial h_t}{\partial \psi} \frac{\partial h_t}{\partial \psi^\tau} + \frac{\varepsilon_t}{h_t^2} \frac{\partial h_t}{\partial \psi} \frac{\partial \varepsilon_t}{\partial \psi^\tau} \\ &\quad - \frac{\varepsilon_t^2}{2h_t^3} \frac{\partial h_t}{\partial \psi} \frac{\partial h_t}{\partial \psi^\tau} - \frac{1}{h_t} \frac{\partial \varepsilon_t}{\partial \psi} \frac{\partial \varepsilon_t}{\partial \psi^\tau} + \frac{\varepsilon_t}{h_t^2} \frac{\partial \varepsilon_t}{\partial \psi} \frac{\partial h_t}{\partial \psi^\tau}, \end{aligned}$$

where

$$(6) \quad \begin{aligned} \frac{\partial \varepsilon_t}{\partial \theta} &= -(1, y_{t-1}, \dots, y_{t-p})^\tau, \\ \frac{\partial \varepsilon_t}{\partial \phi} &= -I(y_{t-d} \leq r)(1, y_{t-1}, \dots, y_{t-p})^\tau, \\ \frac{\partial h_t}{\partial \phi} &= \frac{\partial h_t}{\partial \theta} = \mathbf{0}, \quad \frac{\partial \varepsilon_t}{\partial a} = \mathbf{0}, \quad \frac{\partial h_t}{\partial a} = (1, y_{t-1}^2, \dots, y_{t-m}^2)^\tau. \end{aligned}$$

For convenience of notations, we put $h_t = h_t(\psi)$, $\varepsilon_t = \varepsilon_t(\psi)$, $h_{0t} = h_t(\psi_0)$, $\varepsilon_{0t} = \varepsilon_t(\psi_0)$, $\varsigma = E\varepsilon_t^4 - 1$, $Y_{1t} = (1, y_{t-1}, \dots, y_{t-p})^\tau$, $Y_{2t} = (1, y_{t-1}^2, \dots, y_{t-m}^2)^\tau$, then the following theorem holds under Assumptions 1–2 in the Appendix.

Theorem 2.1. *For model (2) with known threshold and the considered quasi log-likelihood function $L_T(\psi)$ given by (3),*

under the Assumptions 1–2 in the Appendix, there exists a fixed open neighborhood $U(\psi_0) \subset \Psi$ such that with probability one, as $T \rightarrow \infty$, $L_T(\psi)$ has a unique maximum point $\hat{\psi}_T$ in U . Furthermore, $\sqrt{T}(\hat{\psi}_T - \psi_0) \xrightarrow{L} N(0, \Omega_I^{-1} \Omega_S \Omega_I^{-1})$, where Ω_S, Ω_I are respectively given as

$$E \begin{pmatrix} \frac{4}{h_{0t}} Y_{1t} Y_{1t}^\tau & \frac{4I(y_{t-d} \leq r)}{h_{0t}} Y_{1t} Y_{1t}^\tau & \mathbf{0} \\ \frac{4I(y_{t-d} \leq r)}{h_{0t}} Y_{1t} Y_{1t}^\tau & \frac{4I(y_{t-d} \leq r)}{h_{0t}} Y_{1t} Y_{1t}^\tau & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{\varsigma}{h_{0t}^2} Y_{2t} Y_{2t}^\tau \end{pmatrix}$$

and

$$E \begin{pmatrix} \frac{2}{h_{0t}} Y_{1t} Y_{1t}^\tau & \frac{2I(y_{t-d} \leq r)}{h_{0t}} Y_{1t} Y_{1t}^\tau & \mathbf{0} \\ \frac{2I(y_{t-d} \leq r)}{h_{0t}} Y_{1t} Y_{1t}^\tau & \frac{2I(y_{t-d} \leq r)}{h_{0t}} Y_{1t} Y_{1t}^\tau & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{h_{0t}^2} Y_{2t} Y_{2t}^\tau \end{pmatrix}.$$

Remark 1. Through the proof in the Appendix, it is known that $E y_t^2 < \infty$ is not required to guarantee the validity of the theorem, which is consistent with Ling (2007). The matrices Ω_I, Ω_S can be calculated by the relevant sample means after the parameters are estimated.

2.2 Threshold effect test

In this section, we consider the test for the threshold effect, i.e., to test

$$H_0 : \phi_0 = \phi_1 = \dots = \phi_p = 0.$$

Such a test is nonstandard because the threshold parameter r is absent under H_0 . From (5) and (6), we have

$$E\{\partial^2 L_T / (\partial \phi \partial a^\tau)\} = E\{\partial^2 L_T / (\partial \theta \partial a^\tau)\} = \mathbf{0}.$$

Following Davies (1977, 1987), the LM test statistic for our null hypothesis is

$$(7) \quad S = \sup_{r \in \bar{R}} \eta_r^\tau (C_r - L_r^\tau C^{-1} L_r)^{-1} \eta_r,$$

where

$$\begin{aligned} \eta_r &= T^{-\frac{1}{2}} \frac{\partial L_T(\psi)}{\partial \phi} \Big|_{\hat{\theta}_T, \hat{a}_T, \phi=0}, \\ C &= -\frac{1}{T} E \left(\frac{\partial^2 L_T(\psi)}{\partial \theta \partial \theta^\tau} \right) \Big|_{\hat{\theta}_T, \hat{a}_T, \phi=0}, \\ C_r &= -\frac{1}{T} E \left(\frac{\partial^2 L_T(\psi)}{\partial \phi \partial \phi^\tau} \right) \Big|_{\hat{\theta}_T, \hat{a}_T, \phi=0}, \\ L_r &= -\frac{1}{T} E \left(\frac{\partial^2 L_T(\psi)}{\partial \theta \partial \phi^\tau} \right) \Big|_{\hat{\theta}_T, \hat{a}_T, \phi=0}. \end{aligned}$$

Here $\hat{\theta}_T, \hat{a}_T$ are the QMLE under the null hypothesis, and the above estimators are consistent in terms of Theorem 3.1 in Ling (2007). Under the framework of the Lagrange Multiplier test (Silvey, 1959), the above quantities η_r, C, C_r, L_r are asymptotically convergent to the ones that are evaluated

at the true values for θ and a . With an abuse of notation, in the rest of this article, η_r, C, C_r, L_r stand for the quantities evaluated at the true value of θ and a under H_0 . Then we can state the following theorem:

Theorem 2.2. *Suppose Assumptions 1–3 in the Appendix hold, then the asymptotic distribution of $\{\eta_r\}$ is identical to that of a $(p+1)$ -dimensional Gaussian process $\{\xi_r\}$ indexed by the threshold parameter $r \in R$. For $r, s \in R$, we have*

$$\begin{aligned}\xi_r &\sim N_{p+1}(0, C_r - L_r^T C^{-1} L_r), \\ \text{cov}(\xi_r, \xi_s) &= C_{\min(r,s)} - L_r^T C^{-1} L_s.\end{aligned}$$

Also, the asymptotic null distribution of the LM test statistic S in (7) is given by the distribution of $\sup_{r \in \tilde{R}} \xi_r^T (C_r - L_r^T C^{-1} L_r)^{-1} \xi_r$.

Remark 2. Theorem 2.2 is similar to Wong and Li's (1997) Theorem, but concerns different situations. Moreover, our Assumptions 1–3 are weaker in contrast with theirs (e.g., $E\varepsilon_t^4 < \infty$ is a little stronger than $Ee_t^4 < \infty$). The proof is a generalization of Chan (1990), Wong and Li (1997), which is given in the Appendix.

In practice, it is necessary to estimate the upper percentage points of the asymptotic null distribution for S . For model (2), note that $C_r = L_r, C_r$ and $C - C_r$ are positive definite. Then there exist an invertible matrix Q and a diagonal matrix $D = \text{diag}\{\lambda_1(r), \dots, \lambda_{p+1}(r)\}$ such that QCQ^T is an identity matrix and $QC_rQ^T = D$, with all $\{\lambda_i(r)\}$ being strictly between 0 and 1. Let $Q\xi_r = (B_{1r}, \dots, B_{p+1,r})^T$. Then B_{ir} 's are independent Gaussian processes with mean zero and

$$\text{cov}(B_{ir}, B_{is}) = \lambda_i\{\min(r, s)\} - \lambda_i(r)\lambda_i(s).$$

As a result,

$$\begin{aligned}\xi_r^T (C_r - L_r^T C^{-1} L_r)^{-1} \xi_r \\ = \left\{ \frac{B_{1r}^2}{\lambda_1(r) - \lambda_1^2(r)} + \dots + \frac{B_{p+1,r}^2}{\lambda_{p+1}(r) - \lambda_{p+1}^2(r)} \right\}.\end{aligned}$$

When $p = 0$, we need to compute

$$(8) \quad \text{pr} \left\{ \sup_{\beta_1 \leq \lambda_1(r) \leq \beta_2} \frac{B_{1r}^2}{\lambda_1(r) - \lambda_1^2(r)} > z^2 \right\}, (0 < \beta_1 < \beta_2 < 1)$$

for a given z , where $\beta_1 = \min\{\lambda_1(r)\}$ and $\beta_2 = \max\{\lambda_1(r)\}$ for $r \in \tilde{R}$. For general $p > 0$ cases, we want to evaluate

$$(9) \quad \text{pr} \left\{ \sup_{r \in \tilde{R}} \left[\frac{B_{1r}^2}{\lambda_1(r) - \lambda_1^2(r)} + \dots + \frac{B_{p+1,r}^2}{\lambda_{p+1}(r) - \lambda_{p+1}^2(r)} \right] > y \right\}.$$

Based on Chan and Tong (1990), Chan (1991), by using techniques similar to Wong and Li (1997), we have that the probability in (8) can be approximated by

$$(10) \quad \begin{aligned}\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{z^2}{2}\right) \left(\gamma z - \frac{\gamma}{z} + \frac{1}{z}\right), \\ \gamma = \frac{1}{2} \log \left\{ \frac{\beta_2(1 - \beta_1)}{\beta_1(1 - \beta_2)} \right\},\end{aligned}$$

and the probability in (9) can be approximated by

$$(11) \quad 1 - \exp \left\{ -2\chi_{p+1}^2(y) \left(\frac{y}{p+1} - 1\right) \sum_{i=1}^{p+1} \int_{\tilde{R}} \frac{dt_i}{dr} dr \right\},$$

where $\chi_{p+1}^2(\cdot)$ means the probability density function of the χ^2 -distribution with $(p+1)$ degree of freedom and $t_i = \frac{1}{2} \log\{\lambda_i(r)/(1 - \lambda_i(r))\}$.

It remains to find the matrix Q or the $\lambda_i(r)$'s. Note that $C, C - C_r$ are positive definite matrices. Let $C^{-\frac{1}{2}}$ be the inverse of the matrix $C^{\frac{1}{2}}$ that satisfies $C^{\frac{1}{2}}C^{\frac{1}{2}} = C$, then $C^{-\frac{1}{2}}(C - C_r)C^{-\frac{1}{2}}$ is also positive definite. Denote the eigenvalues of $C^{-\frac{1}{2}}(C - C_r)C^{-\frac{1}{2}}$ by $(\delta_1(r), \dots, \delta_{p+1}(r))^\tau$ and accordingly there exists an orthogonal matrix Q_1 satisfying $Q_1^T C^{-\frac{1}{2}}(C - C_r)C^{-\frac{1}{2}} Q_1 = \text{diag}\{\delta_1(r), \dots, \delta_{p+1}(r)\}$, namely,

$$I - Q_1^T C^{-\frac{1}{2}} C_r C^{-\frac{1}{2}} Q_1 = \text{diag}\{\delta_1(r), \dots, \delta_{p+1}(r)\}.$$

Here, I means an identity matrix. Define $Q = Q_1^T C^{-\frac{1}{2}}$, then we have

$$QCQ^T = I, QC_rQ^T = \text{diag}\{1 - \delta_1(r), \dots, 1 - \delta_{p+1}(r)\}.$$

By definition, it is known that $\{1 - \delta_i(r)\}$'s are exactly the eigenvalues of $C^{-\frac{1}{2}} C_r C^{-\frac{1}{2}}$. Hence, for evaluating the probabilities in (8–9), we firstly need to estimate the eigenvalues of the matrix $C^{-\frac{1}{2}} C_r C^{-\frac{1}{2}}$.

3. SIMULATIONS AND EMPIRICAL STUDIES

3.1 Simulations

This section examines the performance of the proposed LM test in finite samples through Monte Carlo simulations. We give examples for $d = 1, p \leq m \leq 2$ as the following:

$$M1 : y_t = 0.1 + e_t \sqrt{h_t}, h_t = 0.2 + 0.16y_{t-1}^2 + 0.09y_{t-2}^2;$$

$$M2 : y_t = 0.36y_{t-1} + e_t \sqrt{h_t}, h_t = 0.3 + 0.49y_{t-1}^2;$$

$$M3 : y_t = 0.3y_{t-1} + 0.1y_{t-2} + e_t \sqrt{h_t}, \\ h_t = 0.5 + 0.1y_{t-1}^2 + 0.1y_{t-2}^2;$$

$$M4 : y_t = 0.1 + 0.3y_{t-1} + e_t \sqrt{h_t}, h_t = 0.5 + 0.32y_{t-1}^2;$$

Table 1. Results of the simulation experiments for assessing the empirical sizes and power with and without intercept

| Model | p m d | Sample size | Empirical sizes and power | | | |
|-------|-------|-------------|---------------------------|------|------|------|
| | | | 10.0% | 5.0% | 2.5% | 1.0% |
| M1 | 0 2 1 | $T = 100$ | 5.9 | 3.1 | 1.4 | 0.3 |
| | | $T = 300$ | 5.4 | 2.7 | 1.3 | 0.5 |
| | | $T = 500$ | 6.0 | 3.1 | 1.8 | 0.7 |
| M2† | 1 1 1 | $T = 100$ | 6.7 | 3.2 | 1.3 | 0.8 |
| | | $T = 300$ | 9.8 | 5.1 | 2.2 | 0.7 |
| | | $T = 500$ | 9.9 | 5.0 | 2.3 | 1.2 |
| M3† | 2 2 1 | $T = 100$ | 13.8 | 8.1 | 6.5 | 4.6 |
| | | $T = 300$ | 10.7 | 5.1 | 3.1 | 2.1 |
| | | $T = 500$ | 12.9 | 6.3 | 2.8 | 0.9 |
| M4 | 1 1 1 | $T = 100$ | 17.9 | 14.6 | 11.7 | 10.3 |
| | | $T = 300$ | 7.1 | 4.3 | 3.0 | 2.0 |
| | | $T = 500$ | 7.2 | 3.7 | 1.7 | 1.0 |
| M5 | 0 2 1 | $T = 100$ | 51.2 | 37.5 | 27.2 | 16.9 |
| | | $T = 300$ | 97.5 | 94.8 | 90.8 | 85.3 |
| | | $T = 500$ | 100.0 | 99.8 | 99.8 | 99.1 |
| M6 | 1 1 1 | $T = 100$ | 89.8 | 83.3 | 77.2 | 67.5 |
| | | $T = 300$ | 89.3 | 82.2 | 75.4 | 64.5 |
| | | $T = 500$ | 90.3 | 83.9 | 76.9 | 66.8 |
| M7† | 1 1 1 | $T = 100$ | 17.1 | 9.6 | 5.2 | 2.4 |
| | | $T = 300$ | 49.7 | 37.8 | 28.0 | 15.3 |
| | | $T = 500$ | 71.1 | 61.4 | 50.2 | 37.8 |
| M8† | 2 2 1 | $T = 100$ | 30.1 | 21.9 | 17.5 | 14.9 |
| | | $T = 300$ | 59.6 | 38.3 | 27.1 | 16.7 |
| | | $T = 500$ | 58.36 | 46.2 | 35.9 | 24.5 |

Note: †testing with no intercept; number of replications=1000.

$$\begin{aligned}
 M5 : y_t &= 0.6 - 0.4I(y_{t-1} \leq 0) + e_t\sqrt{h_t}, \\
 h_t &= 0.3 + 0.15y_{t-1}^2 + 0.1y_{t-2}^2; \\
 M6 : y_t &= 0.1 + 0.1y_{t-1} - I(y_{t-1} \leq 0)(0.2 + 0.2y_{t-1}) \\
 &\quad + e_t\sqrt{h_t}, h_t = 0.05 + 0.36y_{t-1}^2; \\
 M7 : y_t &= 0.4y_{t-1} - 0.3y_{t-1}I(y_{t-1} \leq 0) + e_t\sqrt{h_t}, \\
 h_t &= 0.25 + 0.4y_{t-1}^2; \\
 M8 : y_t &= 0.1y_{t-1} + 0.1y_{t-2} + (0.2y_{t-1} - 0.15y_{t-2}) \\
 &\quad \times I(y_{t-1} \leq 0) + e_t\sqrt{h_t}, h_t \\
 &= 0.5 + 0.16y_{t-1}^2 + 0.1y_{t-2}^2;
 \end{aligned}$$

where $e_t \sim i.i.d. N(0, 1)$. M1–M4 are used to check the empirical size and M5–M8 are adopted to demonstrate the power of the test. We conducted 1000 replications with sample sizes $T=100, 300$ and 500 for each of the above examples. Following Wong and Li (1997), we chose \tilde{R} , the numerical range for the threshold, to be the intervals between the 10th percentile and 90th percentile of y_t . The empirical sizes or power at the nominal upper 10%, 5%, 2.5% and 1% points are listed in Table 1.

Table 1 shows that both sizes and powers behave well. Empirical sizes in each case get closer to the nominal level (especially at the nominal levels of 2.5% and 1%) and the test gets more powerful with increasing sample size n .

3.2 Empirical studies

Ling (2004) applied the DAR(1) model to the US 3-month treasury bill rate series from July 1972 to August 2001 and found that the model fitted the data well, compared to the common AR(1) model. For comparison, we also consider the same set of data except for a longer period from January 1951 to October 2008 (totally 694 observations).

We take x_t to be the logarithms of the observed series and $y_t = x_t - x_{t-1}$. Based on Ling (2004), it is reasonable to apply the model (2) with $p = m = d = 1$ and $\theta_0 = \phi_0 = 0$ to the considered data, which has the form

$$\begin{aligned}
 y_t &= \theta y_{t-1} + \phi y_{t-1}I(y_{t-1} \leq r) + e_t\sqrt{\omega + a_1 y_{t-1}^2}, \\
 (12) \quad e_t &\sim i.i.d(0, 1).
 \end{aligned}$$

Before fitting the data by model (12), we first test whether $\phi = 0$ is significant. With the numerical range \tilde{R} being the interval between the 10th percentile and 90th percentile of y_t , the p value for the considered test is 0.0194, which shows that it is reasonable to introduce the threshold part. To estimate the threshold parameter r , we adopt the idea of Li and Lam (1995). Denote the potential candidates for r by $\mathcal{R} = \{r_1, r_2, \dots, r_L\}$, the estimation of r is performed by considering

$$\max_{r \in \mathcal{R}} L_T(\hat{\psi}_T(r)),$$

where $\hat{\psi}_T(r)$ is the maximizer of the quasi log-likelihood given by (3) with the threshold parameter r being fixed.

For comparison, we use $\{y_t\}_1^{640}$ to estimate model (12) and leave $\{y_t\}_{641}^{693}$ for forecasting. Take \mathcal{R} as a series of evenly spaced points in \tilde{R} with the step length being 0.001. Then we get $\hat{r} = -0.0422$, based on which, the model is as follows:

$$(13) \quad \begin{aligned} y_t &= \underset{(0.0058)}{0.3149}y_{t-1} + \underset{(0.0147)}{0.2188}y_{t-1}I(y_{t-1} \leq -0.0422) \\ &\quad + e_t\sqrt{h_t}, \\ h_t &= \underset{(0.0000)}{0.0022} + \underset{(0.0404)}{0.7656}y_{t-1}^2, e_t \sim i.i.d(0, 1). \end{aligned}$$

The values in parentheses are the corresponding standard deviations which were calculated based on Theorem 2.1. We also estimate the DAR(1) model based on Ling (2004):

$$(14) \quad \begin{aligned} y_t &= \underset{(0.0035)}{0.4009}y_{t-1} + e_t\sqrt{h_t}, \\ h_t &= \underset{(0.0000)}{0.0021} + \underset{(0.0403)}{0.7740}y_{t-1}^2, e_t \sim i.i.d(0, 1). \end{aligned}$$

Moreover, we have the estimations:

For (13),

$$\begin{aligned} E(\log |0.3149 + e_t\sqrt{0.7656}|) &= -0.8957 < 0, \\ E(\log |0.3149 - e_t\sqrt{0.7656}|) &= -0.9483 < 0. \end{aligned}$$

For (14),

$$\begin{aligned} E(\log |0.4009 + e_t\sqrt{0.7740}|) &= -0.8236 < 0, \\ E(\log |0.4009 - e_t\sqrt{0.7740}|) &= -0.8381 < 0. \end{aligned}$$

Note that (12) and DAR(1) model can be rewritten respectively as

$$\begin{aligned} y_t &= \theta y_{t-1} + e_t\sqrt{a_1}|y_{t-1}| + \phi y_{t-1}I(y_{t-1} \leq r) \\ &\quad + e_t \left(\frac{\omega}{\sqrt{\omega + a_1 y_{t-1}^2} + \sqrt{a_1}|y_{t-1}|} \right), \\ y_t &= \theta y_{t-1} + e_t\sqrt{a_1}|y_{t-1}| + e_t \left(\frac{\omega}{\sqrt{\omega + a_1 y_{t-1}^2} + \sqrt{a_1}|y_{t-1}|} \right). \end{aligned}$$

Both $\phi y_{t-1}I(y_{t-1} \leq r)$ and $\omega/(\sqrt{\omega + a_1 y_{t-1}^2} + \sqrt{a_1}|y_{t-1}|)$ are $o(|y_{t-1}|)$ as y_{t-1} goes to infinity. We know from Example 4.1 in Cline and Pu (2004) that the estimated parameters for the above models satisfy the geometric ergodicity conditions. The statistic $Q(M)$ in Li and Mak (1994) with $M = 3, 6, 12$ are used for checking the adequacy of the model (13) and their values are $Q(3) = 0.8764 < \chi_{3,0.95}^2 = 7.815$, $Q(6) = 5.5123 < \chi_{6,0.95}^2 = 12.592$, $Q(12) = 16.8462 < \chi_{12,0.95}^2 = 21.026$, which suggests that model (13) is adequate for the considered data at a 5% significance level. The value of the log-likelihood for model (13) is 1436 and that for model (14) is 1434.4.

Table 2. Percentiles of difference series between upper and lower bounds

| Difference series | Percentiles | | | | |
|-----------------------|-------------|--------|--------|--------|--------|
| | 10% | 25% | 50% | 75% | 90% |
| $\{u_{at} - l_{at}\}$ | 0.1873 | 0.1978 | 0.2418 | 0.3487 | 0.6746 |
| $\{u_{bt} - l_{bt}\}$ | 0.1875 | 0.1981 | 0.2425 | 0.3502 | 0.6781 |

Now we apply model (13–14) for one step ahead forecast for $\{y_t\}_{642}^{693}$. We have

$$\begin{aligned} \text{for (13), } RMSE &= \sqrt{\frac{1}{52} \sum_{t=642}^{693} (y_t - \hat{y}_t)^2} = 0.1253. \\ \text{for (14), } RMSE &= \sqrt{\frac{1}{52} \sum_{t=642}^{693} (y_t - \hat{y}_t)^2} = 0.1281. \end{aligned}$$

We also computed the one-step ahead forecast intervals with 95% confidence level for each case. Denote u_{at}, u_{bt} as the upper bounds series, which are calculated respectively according to (13) and (14). Similarly l_{at} and l_{bt} denote the corresponded lower bounds. We list the percentiles of the difference series between upper and lower bounds in Table 2. It can be seen from the table that model (13) generates slightly narrower confidence intervals. In terms of the log-likelihood values, the RMSEs and the distance between the estimated bounds, we know that model (13) is superior to model (14).

It makes sense to consider Li and Lam's (1995) TAR(1) model with order $p_1 = p_2 = 1, d = 1$ for the data. The model is

$$(15) \quad \begin{aligned} y_t &= \theta y_{t-1} + \phi y_{t-1}I(y_{t-1} \leq r) + \varepsilon_t, \varepsilon_t = e_t\sqrt{h_t}, \\ h_t &= \omega + a_1\varepsilon_{t-1}^2, e_t \sim i.i.d. N(0, 1). \end{aligned}$$

Before fitting the data by model (15), we apply the method in the special case (A) of Wong and Li (1997) to test whether ϕ is significantly different from zero. The p value for the considered test is computed as 0.5591 by choosing \tilde{R} as the interval between the 10th percentile and 90th percentile of y_t , which suggests $\phi = 0$ in (15). Then we have

$$y_t = \theta y_{t-1} + \varepsilon_t, \varepsilon_t = e_t\sqrt{\omega + a_1\varepsilon_{t-1}^2}, e_t \sim i.i.d. N(0, 1).$$

Based on observations $\{y_t\}_1^{640}$, we obtain

$$(16) \quad \begin{aligned} y_t &= \underset{(0.0259)}{0.2603}y_{t-1} + \varepsilon_t, \varepsilon_t = e_t\sqrt{h_t}, \\ h_t &= \underset{(0.0001)}{0.0022} + \underset{(0.0704)}{0.8567}\varepsilon_{t-1}^2, e_t \sim i.i.d. N(0, 1). \end{aligned}$$

In terms of (16), we construct one step ahead of the forecast for $\{y_t\}_{642}^{693}$ and it can be found that the root mean squared error is 0.1330 which is larger than that of model (13) and (14) respectively. Thus model (13) seems to be more reasonable than (16) for the considered data.

4. CONCLUSIONS

This paper considers a class of the threshold ARCH model by adding a threshold effect in the mean equation of the DAR(p) model proposed by Ling (2007). The asymptotic theory for the QMLE of the considered model is proved. A LM test is proposed for testing the threshold effect and approximate methods are given to tabulate the upper percentage points of the asymptotical null distribution. From the simulation results, we found the approximation methods perform well and via the empirical studies, we know our model has improvement over existing models when we introduce a threshold effect for the considered data.

ACKNOWLEDGEMENTS

The authors are grateful to the two referees and the Co-Editor for their useful comments, which led to improvements in the presentation of the paper. The first two and the fourth authors were supported by research grants from the Research Committee of The Hong Kong Polytechnic University. The third author's work was partially supported by National Natural Science Foundation of China (NSFC 10971042).

APPENDIX

A.1 Assumptions and lemma

We make the following assumptions for model (2).

Assumption 1. The series $\{y_t\}$ generated from model (2) is strictly stationary and geometrically ergodic for the considered parameter space Ψ .

Assumption 2. The i.i.d $(0, 1)$ process $\{e_t\}$ satisfies $Ee_t^4 < \infty$, and is absolutely continuous with a continuous symmetric probability density function which is positive everywhere.

Assumption 3. The process $\{y_t\}$ is ρ -mixing with an exponential decreasing rate, i.e., there exists a μ between 0 and 1 such that $\rho(m) = O(\mu^m)$, $m \in \mathcal{N}$, where $\rho(m) = \sup |\text{corr}(f, g)|$, the supremum being over all square integrable f and g which are measurable with respect to $\{y_t, t \leq 0\}$ and $\{y_t, t \geq m\}$, respectively.

Remark 3. For judging the geometrical ergodicity required in Assumption 1, we can make use of Cline and Pu (2004)(e.g., Corollary 2.2, Theorem 3.5 and Example 4.1); Part of the conditions in the Assumptions 2–3 have also been adopted by Chan (1990) to weaken the condition of normality.

Lemma. Denote $L_T(\psi)$ as a function of the observations y_1, \dots, y_T and the parameter $\psi \in \Psi \subseteq R^k$. Suppose ψ_0 is an interior point of Ψ . Assume $L_T(\cdot) : R^k \rightarrow R$ is three times continuously differentiable in ψ and that

$$A1 \text{ As } T \rightarrow \infty, \sqrt{T} \partial L_T(\psi_0) / \partial \psi \xrightarrow{L} N(0, \Omega_S), \Omega_S > 0.$$

$$A2 \text{ As } T \rightarrow \infty, \partial^2 L_T(\psi_0) / \partial \psi \partial \psi^\tau \xrightarrow{P} \Omega_I > 0.$$

$$A3 \max_{i,j,k=1,\dots,p+2} \sup_{\psi \in N(\psi_0)} |\partial^3 L_T(\psi) / \partial \psi_i \partial \psi_j \partial \psi_k| \leq c_T.$$

Here $N(\psi_0)$ is a neighborhood of ψ_0 and $0 \leq c_T \xrightarrow{P} c, 0 < c < \infty$. Then there exists a fixed open neighborhood $U(\psi_0) \subseteq N(\psi_0)$ such that

B1 As $T \rightarrow \infty$, with probability one that there exists a minimum point $\hat{\psi}_T$ of $L_T(\psi)$ in $U(\psi_0)$ and $L_T(\psi)$ is convex in $U(\psi_0)$. Moreover, $\hat{\psi}_T$ is unique and solves $\partial L_T(\hat{\psi}_T) / \partial \psi = 0$.

$$B2 \text{ As } T \rightarrow \infty, \hat{\psi}_T - \psi_0 \xrightarrow{P} 0, \sqrt{T}(\hat{\psi}_T - \psi_0) \xrightarrow{L} N(0, \Omega_I^{-1} \Omega_S \Omega_I^{-1}).$$

Proof. See Lemma 1 in Jensen and Rahbek (2004). \square

A.2 Proof of Theorem 2.1

Let

$$(17) \quad L_T^*(\psi) = \frac{1}{T} \sum_{t=1}^T \left[\log h_t + \frac{\varepsilon_t^2}{h_t} \right] = \frac{1}{T} \sum_{t=1}^T l_t^*(\psi),$$

and it can be shown:

$$(18) \quad \frac{\partial l_t^*}{\partial \psi} = \left(1 - \frac{\varepsilon_t^2}{h_t} \right) \frac{1}{h_t} \frac{\partial h_t}{\partial \psi} + \frac{2\varepsilon_t}{h_t} \frac{\partial \varepsilon_t}{\partial \psi},$$

$$(19) \quad \frac{\partial^2 l_t^*}{\partial \psi \partial \psi^\tau} = \frac{1}{h_t^2} \left(\frac{\varepsilon_t^2}{h_t} - 1 \right) \frac{\partial h_t}{\partial \psi} \frac{\partial h_t}{\partial \psi^\tau} - \frac{2\varepsilon_t}{h_t^2} \frac{\partial h_t}{\partial \psi} \frac{\partial \varepsilon_t}{\partial \psi^\tau} + \frac{\varepsilon_t^2}{h_t^3} \frac{\partial h_t}{\partial \psi} \frac{\partial h_t}{\partial \psi^\tau} + \frac{2}{h_t} \frac{\partial \varepsilon_t}{\partial \psi} \frac{\partial \varepsilon_t}{\partial \psi^\tau} - \frac{2\varepsilon_t}{h_t^2} \frac{\partial \varepsilon_t}{\partial \psi} \frac{\partial h_t}{\partial \psi^\tau},$$

and

$$(20) \quad \frac{\partial^3 l_t^*}{\partial \psi_i \partial \psi_j \partial \psi_k} = \left[2 \left(1 - \frac{3\varepsilon_t^2}{h_t} \right) \frac{1}{h_t^3} \frac{\partial h_t}{\partial \psi_i} \frac{\partial h_t}{\partial \psi_j} \frac{\partial h_t}{\partial \psi_k} \right] + \left[\left(\frac{4\varepsilon_t}{h_t^3} \frac{\partial h_t}{\partial \psi_i} \frac{\partial h_t}{\partial \psi_j} \frac{\partial \varepsilon_t}{\partial \psi_k} + \frac{4\varepsilon_t}{h_t^3} \frac{\partial h_t}{\partial \psi_i} \frac{\partial \varepsilon_t}{\partial \psi_j} \frac{\partial h_t}{\partial \psi_k} + \frac{4\varepsilon_t}{h_t^3} \frac{\partial \varepsilon_t}{\partial \psi_i} \frac{\partial h_t}{\partial \psi_j} \frac{\partial h_t}{\partial \psi_k} \right) \right] - \left[\left(\frac{2}{h_t^2} \frac{\partial h_t}{\partial \psi_i} \frac{\partial \varepsilon_t}{\partial \psi_j} \frac{\partial \varepsilon_t}{\partial \psi_k} + \frac{2}{h_t^2} \frac{\partial \varepsilon_t}{\partial \psi_i} \frac{\partial \varepsilon_t}{\partial \psi_j} \frac{\partial h_t}{\partial \psi_k} + \frac{2}{h_t^2} \frac{\partial \varepsilon_t}{\partial \psi_i} \frac{\partial h_t}{\partial \psi_j} \frac{\partial \varepsilon_t}{\partial \psi_k} \right) \right] \triangleq l_{1t} + l_{2t} - l_{3t}.$$

Here, $l_{it}, i = 1, 2, 3$ mean the corresponding quantities expressed in the above three square brackets. To prove Theorem 2.1, we just need to verify A1–A3 described in the above lemma.

Recall $h_t = h_t(\psi)$, $\varepsilon_t = \varepsilon_t(\psi)$, $h_{0t} = h_t(\psi_0)$, $\varepsilon_{0t} = \varepsilon_t(\psi_0)$, $\varsigma = Ee_t^4 - 1$. From the above (17–18), we know

$$\begin{aligned} & \sqrt{T} \frac{\partial L_T^*(\psi_0)}{\partial \psi} \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \left[\left(1 - \frac{\varepsilon_{0t}^2}{h_{0t}^2}\right) \frac{1}{h_{0t}} \frac{\partial h_t}{\partial \psi} + \frac{2\varepsilon_{0t}}{h_{0t}} \frac{\partial \varepsilon_t}{\partial \psi} \right] \\ &\triangleq \frac{1}{\sqrt{T}} \sum_{t=1}^T S_t. \end{aligned}$$

Consider any non-zero vector $c = (c_1, \dots, c_q)^\tau$, $q = 2p + m + 3$, we have

$$\sqrt{T} c^\tau \frac{\partial L_T^*(\psi_0)}{\partial \psi} = \sum_{t=1}^T \left(\frac{1}{\sqrt{T}} c^\tau S_t \right) \triangleq \sum_{t=1}^T W_t.$$

Given the information set up to time $t - 1$, $\mathcal{F}_{t-1} = \sigma(e_{t-1}, \dots, e_1, y_0, \dots, y_{-m+1})$, then we know $\{W_t\}$ is a martingale difference with respect to the information set and $E(W_t^2 | \mathcal{F}_{t-1}) = c^\tau \frac{1}{T} E(S_t S_t^\tau | \mathcal{F}_{t-1}) c$. It is not difficult to get

$$(21) \quad E(S_t S_t^\tau | \mathcal{F}_{t-1}) = \frac{\varsigma}{h_{0t}^2} \frac{\partial h_t}{\partial \psi} \frac{\partial h_t}{\partial \psi^\tau} + \frac{4}{h_{0t}} \frac{\partial \varepsilon_t}{\partial \psi} \frac{\partial \varepsilon_t}{\partial \psi^\tau} \triangleq \Omega_{S,t}.$$

Consequently, we have

$$\sum_{t=1}^T E(W_t^2 | \mathcal{F}_{t-1}) = c^\tau \left(\frac{1}{T} \sum_{t=1}^T \Omega_{S,t} \right) c \xrightarrow{P} c^\tau \Omega_S c,$$

where

$$(22) \quad \Omega_S = E(\Omega_{S,t}) = E\left(\frac{\varsigma}{h_{0t}^2} \frac{\partial h_t}{\partial \psi} \frac{\partial h_t}{\partial \psi^\tau} + \frac{4}{h_{0t}} \frac{\partial \varepsilon_t}{\partial \psi} \frac{\partial \varepsilon_t}{\partial \psi^\tau} \right).$$

Furthermore, given any $\delta > 0$, we have

$$\begin{aligned} & \sum_{t=1}^T E[W_t^2 I(|W_t| \geq \delta)] \\ &= \frac{1}{T} \sum_{t=1}^T E[c^\tau S_t S_t^\tau c I(|c^\tau S_t S_t^\tau c| \geq \delta^2 T)] \\ &= E[c^\tau S_1 S_1^\tau c I(|c^\tau S_1 S_1^\tau c| \geq \delta^2 T)] \longrightarrow 0. \end{aligned}$$

The above limit can be explained by the fact that $E\Omega_{S,t} < \infty$. By the martingale central limit theorem, see, for example, Theorem 35.12 in Billingsley (1995), we proved that $\sum_{t=1}^T W_t \xrightarrow{L} N(0, c^\tau \Omega_S c)$, which means

$$(23) \quad \sqrt{T} \frac{\partial L_T^*(\psi_0)}{\partial \psi} \xrightarrow{L} N(0, \Omega_S),$$

namely condition A1 is satisfied.

Applying the double expectation formula we can get

$$(24) \quad \begin{aligned} E\left(\frac{\partial^2 L_t^*(\psi_0)}{\partial \psi \partial \psi^\tau} \right) &= E\left(\frac{1}{h_{0t}^2} \frac{\partial h_t}{\partial \psi} \frac{\partial h_t}{\partial \psi^\tau} + \frac{2}{h_{0t}} \frac{\partial \varepsilon_t}{\partial \psi} \frac{\partial \varepsilon_t}{\partial \psi^\tau} \right) \triangleq \Omega_I, \\ \frac{\partial^2 L_t^*(\psi_0)}{\partial \psi \partial \psi^\tau} &= \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 L_t^*(\psi_0)}{\partial \psi \partial \psi^\tau} \xrightarrow{P} \Omega_I, \end{aligned}$$

which means A2 holds.

For each $\psi \in \Psi$, from (6) and (20) it is not difficult to show there exist finite positive constants C_1, C_2, C_3 satisfying $|l_{1t}| \leq C_1(1 + \frac{\varepsilon_t^2}{h_t^2})$, $|l_{2t}| \leq C_2(1 + \frac{\varepsilon_t^2}{h_t^2})$, $|l_{3t}| \leq C_3$. Similar to the proof of (ii) in Ling's (2007) lemma B.2, we can further show that there are two constants C_4, C_5 depending only on the parametric space Ψ such that $1 + \sup_{\psi \in \Psi} \frac{\varepsilon_t^2(\psi)}{h_t(\psi)} \leq C_4 e_t^2 + C_5$. Put $w_t = (C_1 + C_2)(C_4 e_t^2 + C_5) + C_3$. Then we know w_t is a stationary, nonnegative series and satisfies

$$\max_{i,j,k=1,\dots,p+2} \sup_{\psi \in N(\psi_0)} \left| \frac{\partial^3 L_T^*(\psi)}{\partial \psi_i \partial \psi_j \partial \psi_k} \right| \leq \frac{1}{T} \sum_{t=1}^T w_t \xrightarrow{P} E w_t < \infty.$$

Hence A3 is satisfied. In conjunction with (22) and (24), we establish Theorem 2.1.

A.3 Proof of Theorem 2.2

Define

$$\eta_\infty = T^{-\frac{1}{2}} \frac{\partial L_T(\psi)}{\partial \theta} = \frac{1}{\sqrt{T}} \sum_{t=1}^T \left[-\frac{\varepsilon_t}{h_t} \frac{\partial \varepsilon_t}{\partial \theta} \right],$$

and let $u_r = (\eta_\infty^\tau, \eta_r^\tau)^\tau$, $c^* = (c_1^\tau, c_2^\tau)$. Here, c^* is any nonzero constant vector and $c_1 = (c_{10}, \dots, c_{1p})^\tau$, $c_2 = (c_{20}, \dots, c_{2p})^\tau$. Consider

$$\begin{aligned} c^{*\tau} u_r &= c_1^\tau \eta_\infty + c_2^\tau \eta_r \\ &= \sum_{t=1}^T \frac{1}{\sqrt{T}} \left[-\frac{\varepsilon_t}{h_t} (c_1^\tau \frac{\partial \varepsilon_t}{\partial \theta} + c_2^\tau \frac{\partial \varepsilon_t}{\partial \phi}) \right] = \sum_{t=1}^T U_t, \end{aligned}$$

then we know $\{U_t\}$ is a martingale difference with respect to \mathcal{F}_{t-1} . Using analogous discussion to (23), we can show

$$u_r \xrightarrow{L} N \left\{ 0, \begin{pmatrix} C & L_r \\ L_r^\tau & C_r \end{pmatrix} \right\}.$$

Hence, on condition that $\eta_\infty = 0$, we have

$$\eta_r \xrightarrow{L} N \{0, (C_r - L_r^\tau C^{-1} L_r)\}.$$

For $r \neq s$, let $u_{r,s} = (\eta_\infty^\tau, \eta_r^\tau, \eta_s^\tau)^\tau$, then it can be similarly obtained that

$$u_{r,s} \xrightarrow{L} N \left\{ 0, \begin{pmatrix} C & L_r & L_s \\ L_r^\tau & C_r & C_{\min(r,s)} \\ L_s^\tau & C_{\min(r,s)}^\tau & C_s \end{pmatrix} \right\},$$

and conditionally, $(\eta_r, \eta_s)'$ converges in distribution to

$$N \left\{ 0, \begin{pmatrix} C_r & C_{\min(r,s)} \\ C_{\min(r,s)}^\tau & C_s \end{pmatrix} - \begin{pmatrix} L_r^\tau \\ L_s^\tau \end{pmatrix} C^{-1} \begin{pmatrix} L_r & L_s \end{pmatrix} \right\}. \quad (27)$$

Hence, (η_r, η_s) asymptotically follows a joint normal distribution with the covariance being $C_{\min(r,s)} - L_r^\tau C^{-1} L_s$.

Let $b > 0$, $D_k(-\infty, \infty)(D_k[-b, b])$ denote the function spaces with each element $f : R([-b, b]) \rightarrow R^k$ being right continuous and having left-hand limit. Equip $D_k(-\infty, \infty)(D_k[-b, b])$ with the topology of uniform convergence over compact sets. Let $C_k(-\infty, \infty)$ be the subspace of $D_k(-\infty, \infty)$ consisting of functions continuous everywhere. More details on these spaces can be found in Pollard (1984). Now, $\{\eta_r, -\infty < r < \infty\}$ lives on $D_{q+1}(-\infty, \infty)$.

Subsequently, we show that η_r converges weakly to $\{\xi_r\}$ in $D_{p+1}(-\infty, \infty)$ and each realization of $\{\xi_r\}$ belongs to $C_{p+1}(-\infty, \infty)$ almost surely. It suffices to verify the tightness of $\{\eta_r, -b \leq r \leq b\}$ componentwise. Without loss of generality, consider the last component of $\{\eta_r, -b \leq r \leq b\}$. It is tight if and only if

$$g_T(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^T \left[\frac{\varepsilon_t}{h_t} y_{t-p} I(y_{t-d} \leq r) \right]$$

is tight.

Let $-b \leq s \leq r \leq b$ be two arbitrary numbers, M_i, K_i ($i=1, 2$) be constants independent of T . Then

$$g_T(r) - g_T(s) = \frac{1}{\sqrt{T}} \sum_{t=1}^T \left[\frac{\varepsilon_t}{h_t} y_{t-p} I(s < y_{t-d} \leq r) \right].$$

For $i = 1, \dots, p, \delta = 1, 2, 3, 4$, note $d, p \leq m, y_{t-i}^2/h_t = O(1)$, then we have

$$\begin{aligned} & E \left\{ \left| \frac{\varepsilon_t}{h_t} y_{t-i} I(s < y_{t-d} \leq r) \right|^\delta \right\} \\ & \leq E \left\{ \left(1 + \frac{y_{t-i}^4}{h_t^2} \right) I(s < y_{t-d} \leq r) E \left(\left(1 + \frac{\varepsilon_t^4}{h_t^2} \right) | \mathcal{F}_{t-1} \right) \right\} \\ (25) \quad & \leq M_1 E \{ I(s < y_{t-d} \leq r) \} \leq M_2 (r - s). \end{aligned}$$

Let $\zeta_t = \frac{1}{\sqrt{T}} \frac{\varepsilon_t}{h_t} y_{t-p} I(s < y_{t-d} \leq r)$. Applying the Assumption 3 and Lemma 3.6 in Peligrad (1982), we have

$$\begin{aligned} E |g_T(r) - g_T(s)|^4 & \leq K_1 (T^{\frac{1}{4}} \|\zeta_t\|_4 + T^{\frac{1}{2}} \|\zeta_t\|_2)^4 \\ (26) \quad & \leq K_2 [(r - s)/T + (r - s)^2]. \end{aligned}$$

Here $\|\cdot\|_\delta$ means the usual L^δ norm. The second line in the above inequalities follows from (25). For $[-b, b]$, consider a partition $\{-b = r_0 < r_1 < \dots < r_L = b\}$ with $u > 0, r_j = r_{j-1} + u, 0 \leq j \leq L - 1$ and $r_L - r_{L-1} \leq u$. Define

$$\kappa_{t,i} = \frac{1}{\sqrt{T}} \frac{|\varepsilon_t y_{t-p}|}{h_t} I(r_{i-1} < y_{t-d} \leq r_i),$$

then, $\forall i$, for $r_{i-1} \leq r \leq r_i$, we have

$$|g_T(r) - g_T(s)| \leq \sum_{t=1}^T \kappa_{t,i}. \quad (27)$$

Based on (25), it is not difficult to show

$$(28) \quad \sup_i \sum_{t=1}^T \kappa_{t,i} = u O_p(\sqrt{T}).$$

In terms of (26–28), by applying the similar discussion used in the proof of Theorem 22.1 in Billingsley (1968), we can show the tightness of $\{g_T(r), -b \leq r \leq b\}$.

Received 27 March 2010

REFERENCES

- [1] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*, Wiley, New York. [MR0233396](#)
- [2] BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. Wiley, New York. [MR1324786](#)
- [3] CHAN, K. S. (1990). Testing for threshold autoregression. *The Annals of Statistics* **18** 1886–1894. [MR1074443](#)
- [4] CHAN, K. S. (1991). Percentage points of likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society B* **53** 691–696. [MR1125726](#)
- [5] CHAN, K. S. and TONG, H. (1990). On likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society B* **52** 469–476. [MR1086798](#)
- [6] CLINE, D. B. H. and PU, H. H. (2004). Stability and the Lyapunov exponent of threshold AR-ARCH models. *The Annals of Applied Probability* **14** 1920–1949. [MR2099657](#)
- [7] DAVIES, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64** 247–254. [MR0501523](#)
- [8] DAVIES, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74** 33–43. [MR0885917](#)
- [9] ENGLE, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* **50**(4) 987–1007. [MR0666121](#)
- [10] JENSEN, S. T AND RAHBEK, A. (2004). Asymptotic inference for non-stationary GARCH. *Econometric theory* **20** 1203–1226. [MR2101955](#)
- [11] LI, W. K. and LAM, K. (1995). Modelling the asymmetry in stock returns using threshold ARCH model. *Statistician* **44** 333–341.
- [12] 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. [MR1312326](#)
- [13] LING, S. (2004). Estimation and testing of stationarity for double autoregressive models. *Journal of the Royal Statistical Society B* **66** 63–78. [MR2035759](#)
- [14] LING, S. (2007). A double AR(p) model: structure and estimation. *Statistica Sinica* **17** 161–175. [MR2352507](#)
- [15] LING, S. and LI, D. (2008). Asymptotic inference for a non-stationary double AR(1) model. *Biometrika* **95** 257–263. [MR2409729](#)
- [16] PELIGRAD, M. (1982). Invariance principles for mixing sequences of random variables. *The Annals of Probability* **10** 968–981. [MR0672297](#)

- [17] POLLARD, D. (1984). *Convergence of Stochastic Process*. Springer, New York. [MR0762984](#)
- [18] SILVEY, S. D. (1959). The Lagrangian Multiplier test. *The Annals of Mathematical Statistics* **30** 389–407. [MR0104307](#)
- [19] TONG, H. (1990). *Nonlinear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford. [MR1079320](#)
- [20] TONG, H. and LIM, K. S. (1980). Threshold autoregression, limit cycles and cyclical data (with Discussion). *Journal of the Royal Statistical Society B* **42** 245–292.
- [21] WEISS, A. A. (1986). Asymptotic theory for ARCH models: estimation and testing. *Econometric theory* **2** 101–131.
- [22] WONG, C. S. and LI, W. K. (1997). Testing for threshold autoregression with conditional heteroscedasticity. *Biometrika* **84** 407–418. [MR1467056](#)
- Xingfa Zhang
 Department of Applied Mathematics
 The Hong Kong Polytechnic University
 Hung Hom, Kowloon, Hong Kong
 E-mail address: xingfazhang@hotmail.com
- Heung Wong
 Department of Applied Mathematics
 The Hong Kong Polytechnic University
 Hung Hom, Kowloon, Hong Kong
 E-mail address: mathwong@polyu.edu.hk
- Yuan Li
 School of Mathematics and Information Science
 Guangzhou University, Guangzhou, China
 E-mail address: yli57@hotmail.com
- Wai-Cheung Ip
 Department of Applied Mathematics
 The Hong Kong Polytechnic University
 Hung Hom, Kowloon, Hong Kong
 E-mail address: mathipwc@polyu.edu.hk