Statistica Sinica **18**(2008), 647-665

TESTING FOR THRESHOLD MOVING AVERAGE WITH CONDITIONAL HETEROSCEDASTICITY

Guodong Li and Wai Keung Li

Nanyang Technological University and University of Hong Kong

Abstract: The recent paper by Ling and Tong (2005) considered a quasi-likelihood ratio test for the threshold in moving average models with *i.i.d.* errors. This article generalizes their results to the case with GARCH errors, and a new quasi-likelihood ratio test is derived. The generalization is not direct since the techniques developed for TMA models heavily depend on the property of p-dependence that is no longer satisfied by the time series models with conditional heteroscedasticity. The new test statistic is shown to converge weakly to a functional of a centered Gaussian process under the null hypothesis of no threshold, and it is also proved that the test has nontrivial asymptotic power under local alternatives. Monte Carlo experiments demonstrate the necessity of our test when a moving average time series has a time varying conditional variance. As further support, two data examples are reported.

Key words and phrases: Conditional heteroscedasticity, Gaussian process, likelihood ratio test, MA-GARCH model, threshold MA-GARCH model.

1. Introduction

The threshold time series models were first proposed by Tong (1978) and Tong and Lim (1980) in the form of threshold autoregressive (TAR) models and have become a standard class of nonlinear time series models, see Tong (1990), Ling and Tong (2005), and references therein. On the other hand, since Engle (1982), it has been widely accepted by most economists and statisticians that many financial and economic time series have a time varying conditional variance and the generalized autoregressive conditional heteroscedasticity (GARCH) models proposed by Bollerslev (1986) are usually considered to model this phenomena. Bollerslev, Chou and Kroner (1992) also showed that ignoring the conditional heteroscedastic effect in time series models would lead to inefficient estimates and suboptimal statistical inferences. Combining the above two ideas, a second-generation class of models have been widely discussed recently, e.g. threshold AR-ARCH models in Li and Lam (1995), double threshold AR-ARCH models in Liu, Li and Li (1997), double threshold ARMA-GARCH models in Ling (1999), and others.

In the literature, it is an interesting problem to test whether or not a threshold time series model provides a better fit to the data than a model without threshold. For this type of tests, the threshold parameter is usually assumed to be unknown in the alternative hypothesis and is absent in the null hypothesis. Under this circumstance, the threshold parameter is a nuisance parameter and it makes the testing problem nonstandard, see Davies (1977, 1987). Chan (1990, 1991) and Chan and Tong (1990) first considered this problem and suggested a likelihood ratio test for the threshold in AR models. A Wald test was studied by Hansen (1996) for TAR models, and was extended to the case with a unit root by Caner and Hansen (2001). For the extension from common TAR models to conditional heteroscedastic versions, Wong and Li (1997, 2000) considered Lagrange multiplier tests for (double) TAR-ARCH models. It is well known that MA models are as important as AR models in the linear case, and are usually considered in time series modeling from the point of view of parsimony, see Tsay (1987). However, until Ling and Tong (2005), the development of the threshold moving average (TMA) models had been hindered by the unavailability of an invertibility condition, which is vital for making statistical inferences. For TMA models, Ling and Tong (2005) derived the condition of invertibility, and investigated the quasi-likelihood ratio test for threshold in MA models. However, it is still an open problem on how to test for the threshold structure in MA models when the time series has a time varying conditional variance.

As we know, the quasi-likelihood ratio test will perform best when the true distribution is the assumed one. However, for most time series in finance and economics, it is a more reasonable assumption that their conditional distributions have a time varying variance than that they have a constant variance, see Engle (1982) and Bollerslev (1986). Hence, we may expect that the test of Ling and Tong (2005) is too sensitive for TMA models with conditional heteroscedasticity, and the simulation results in Section 4 demonstrate that it even has no reliable sizes when the effect of conditional heteroscedasticity is ignored. This article generalizes the results of Ling and Tong (2005), and derives a new quasi-likelihood ratio test statistic for threshold moving average with GARCH errors. Under the null hypothesis of no threshold, the test statistic is shown to converge weakly to a functional of a zero-mean Gaussian process, and the test also has nontrivial asymptotic power under a sequence of local alternatives. The generalization is not direct since the techniques developed for TMA models heavily depend on an exclusive property of p-dependence that is no longer satisfied for the MA-GARCH or TMA-GARCH models. Furthermore, as in Ling and Tong (2005), the techniques in this article do not involve any mixing conditions. For a general time series model, mixing conditions are difficult to verify though generally assumed, see Chan (1990) and Wong and Li (1997). Our techniques may also be useful in constructing tests for the presence of threshold structure in ARMA models, regarded as a challenging problem by Ling and Tong (2005).

The organization of this article is as follows. Section 2 derives the quasi-likelihood ratio test statistic and its asymptotic distribution under the null hypothesis of no threshold. Under local alternatives, Section 3 shows that the test has nontrivial asymptotic power. Some simulation results are presented in Section 4; this section also gives the modeling and testing results for the centered log return sequences of the S&P 500 weekly closing price and the weekly exchange rate of Japanese Yen against USA dollars. The proofs of the main theorem, stated in Section 2, and two important lemmas are delayed to the appendix.

2. Quasi-Likelihood Ratio Test

Let $\{y_t\}$ be a strictly stationary and ergodic time series generated by the TMA(p,d,q)-GARCH(m,s) model

$$\begin{cases} y_t = \sum_{i=1}^{p} \phi_i e_{t-i} + \sum_{i=1}^{q} \psi_i e_{t-i} I(y_{t-d} \le r) + e_t, \\ e_t = \varepsilon_t h_t^{\frac{1}{2}}, \\ h_t = a_0 + \sum_{i=1}^{m} a_i e_{t-i}^2 + \sum_{i=1}^{s} b_i h_{t-i}, \end{cases} \tag{2.1}$$

where $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (*i.i.d.*) random variables with mean zero, variance one and a finite fourth moment, p, d, q, m, s are known positive integers with $p \ge q$, I is the indicator function, and $r \in R$ is the threshold parameter.

Denote the parameter space by $\Theta = \Theta_\alpha \times \Theta_\beta \times \Theta_\phi \times \Theta_\psi$, where Θ_α, Θ_β, Θ_ϕ and Θ_ψ are compact subsets of R^{m+1}, R^s, R^p and R^q, respectively. Let $\alpha = (a_0, a_1, \ldots, a_m)'$, $\beta = (b_1, \ldots, b_s)'$, $\phi = (\phi_1, \ldots, \phi_p)'$, $\psi = (\psi_1, \ldots, \psi_q)'$, $\gamma = (\alpha', \beta', \phi')'$ and $\lambda = (\gamma', \psi')'$, where λ is termed the parameter vector of model (2.1). Denote the true parameter vector by $\lambda_0 = (\gamma_0', \psi_0')' = (\alpha_0', \beta_0', \phi_0', \psi_0')'$ and assume it to be an interior point of the parameter space Θ. Some assumptions on the parameter space are considered to make sure that model (2.1) has some necessary properties in constructing the quai-likelihood ratio test. We first state some restrictions on the parameters in the GARCH part of model (2.1).

Assumption 2.1. *$a_i > 0, i = 0, 1, \ldots, m$, $b_j > 0, j = 1, \ldots, s$, the polynomials $a_1 x + a_2 x^2 + \cdots + a_m x^m$ and $1 - b_1 x - b_2 x^2 - \cdots - b_s x^s$ are coprime, and the sequence $\{e_t\}$ is strictly stationary and ergodic with a finite fourth moment.*

The positivity of the a_i's and b_j's is a general restriction in estimating the parameters of GARCH models, see Peng and Yao (2003), Berkes and Horvath (2004), and can be replaced by the existence of higher order moments of e_t in the process of our proof. The coprime nature of the two polynomials is necessary to uniquely identify the parameters of GARCH models, i.e., it is indispensable in making the Hessian matrix Ω_r in Lemma 2.1 and Theorem 2.1 positive definite, see Berkes, Horvath and Kokoszka (2003). The conditions of ergodicity, stationarity and finite fourth moment are common assumptions in deriving the asymptotic behavior of tests for the threshold, see Chan (1990), Wong and Li (2000) and Ling and Tong (2005). For details of these probabilistic properties, please refer to Ling (1999).

Given the time series $\{y_t\}$ from model (2.1) with Assumption 2.1, we consider

$$H_0 : \psi_0 = 0 \quad \textit{versus} \quad H_1 : \psi_0 \neq 0 \text{ for some } r \in R.$$

Under H_0, (2.1) reduces to a usual MA-GARCH model and the time series $\{y_t\}$ is always stationary and ergodic. Note that the threshold parameter r is absent in this case. Under H_1, without any further assumptions on the parameter space, Ling (1999) showed that there always exists a strictly stationary solution to (2.1).

For (2.1) (i.e., under H_1), we define the functions $e_t(\lambda, r)$ and $h_t(\lambda, r)$ in terms of the following iterative equations,

$$e_t(\lambda, r) = y_t - \sum_{i=1}^{p} \phi_i e_{t-i}(\lambda, r) - \sum_{i=1}^{q} \psi_i e_{t-i}(\lambda, r) I(y_{t-d} \leq r),$$
$$h_t(\lambda, r) = a_0 + \sum_{i=1}^{m} a_i e_{t-i}^2(\lambda, r) + \sum_{i=1}^{s} b_i h_{t-i}(\lambda, r).$$

However, under H_0, the varying parameters ψ and r disappear from the above two functions. For simplicity, we denote them respectively by $e_t(\gamma)$ and $h_t(\gamma)$, i.e., $e_t(\gamma) = e_t(\lambda, -\infty)$ and $h_t(\gamma) = h_t(\lambda, -\infty)$. Furthermore, for the functions $e_t(\lambda, r)$ and $e_t(\gamma)$ to be meaningful, it is important to consider the invertibility condition of y_t. Ling and Tong (2005) investigated this property for a general TMA model and their results can be extended to our case. We now state the condition of invertibility as Assumption 2.2 below.

Assumption 2.2. $\sum_{i=1}^{p} |\phi_i| < 1$ *and* $\sum_{i=1}^{p} |\phi_i + \psi_i| < 1$, *where* $\psi_i = 0$ *for* $i > q$.

The above assumption is the same as Assumption 2.1 in Ling and Tong (2005), and is also similar to the conditions for the ergodicity of TAR models in Chan and Tong (1985).

Under H_0 and H_1, omitting a negative constant, the quasi-log likelihood functions conditional on $\{y_0, y_{-1}, \ldots\}$ are, respectively,

$$L_{0n}(\gamma) = \sum_{t=1}^{n} l_t(\gamma) \text{ and } L_{1n}(\lambda, r) = \sum_{t=1}^{n} l_t(\lambda, r),$$

where $l_t(\gamma) = l_t(\lambda, -\infty)$, and

$$l_t(\lambda, r) = \frac{e_t^2(\lambda, r)}{h_t(\lambda, r)} + \log h_t(\lambda, r).$$

For a time series, there are only n values available in practice but the quasi-log likelihood functions are all dependent on past observations infinitely far away. Hence, initial values for $\{y_0, y_{-1}, \ldots\}$ are needed. For simplicity, we assume that $y_i = 0, i \leq 0$ and these functions evaluated at these initial values can be denoted respectively by $\tilde{e}_t(\gamma)$, $\tilde{e}_t(\lambda, r)$, $\tilde{h}_t(\gamma)$, $\tilde{h}_t(\lambda, r)$, $\tilde{l}_t(\gamma)$, $\tilde{l}_t(\lambda, r)$, $\tilde{L}_{0n}(\gamma)$ and $\tilde{L}_{1n}(\lambda, r)$. By a method similar to that of Lemma 6.6 in Ling and Tong (2005), we can show that the effect of these initial values is asymptotically ignorable.

Let

$$\tilde{\gamma}_n = \underset{\gamma \in \Theta_\alpha \times \Theta_\beta \times \Theta_\phi}{\operatorname{argmin}} \tilde{L}_{0n}(\gamma) \quad \text{and} \quad \tilde{\lambda}_n(r) = \underset{\lambda \in \Theta}{\operatorname{argmin}} \tilde{L}_{1n}(\lambda, r).$$

We call $\tilde{\gamma}_n$ and $\tilde{\lambda}_n(r)$ the quasi-maximum likelihood estimators for the MA-GARCH model and the TMA-GARCH model, respectively. For a given r, it is well known that the likelihood ratio test statistics for H_0 against H_1 can be defined as

$$\widetilde{LR}_n(r) = -[\tilde{L}_{1n}(\tilde{\lambda}_n(r), r) - \tilde{L}_{0n}(\tilde{\gamma}_n)].$$

In the literature, the threshold parameter r is generally assumed to be unknown and it is natural to consider the supremum on r. However, the quantity $\sup_{r \in R} \widetilde{LR}_n(r)$ will diverge to infinity in probability, see Andrews (1993). In this article, as in Andrews (1993) and Ling and Tong (2005), the supremum of $\widetilde{LR}_n(r)$ on a finite interval $[a, b]$ is considered and the quasi-likelihood ratio test statistic is defined to be

$$LR_n = \sup_{r \in [a,b]} \widetilde{LR}_n(r).$$

To investigate the asymptotic distribution of LR_n, we need another assumption, it is a mild technical condition and includes most continuous distributions.

Assumption 2.3. *ε_t has a continuous and positive density function $f(\cdot)$ on R, and $\sup_{x \in R} x^4 f(x) < \infty$.*

To present the asymptotic results of the test statistic LR_n, we need some more notations in terms of matrices, as follows. For $r, l \in R$:

$$\Gamma_{rl}^{(hh)} = E\{\frac{1}{h_t^2} \frac{\partial h_t(\lambda_0, r)}{\partial \lambda} \frac{\partial h_t(\lambda_0, l)}{\partial \lambda'}\}; \ \Gamma_{rl}^{(eh)} = E\{\frac{1}{h_t^{\frac{3}{2}}} \frac{\partial e_t(\lambda_0, r)}{\partial \lambda} \frac{\partial h_t(\lambda_0, l)}{\partial \lambda'}\};$$

$$\Sigma = E\{\frac{2}{h_t} \frac{\partial e_t(\gamma_0)}{\partial \gamma} \frac{\partial e_t(\gamma_0)}{\partial \gamma'} + \frac{1}{h_t^2} \frac{\partial h_t(\gamma_0)}{\partial \gamma} \frac{\partial h_t(\gamma_0)}{\partial \gamma'}\};$$

$$\Sigma_{1r} = E\{\frac{2}{h_t} \frac{\partial e_t(\gamma_0)}{\partial \gamma} \frac{\partial e_t(\lambda_0, r)}{\partial \psi'} + \frac{1}{h_t^2} \frac{\partial h_t(\gamma_0)}{\partial \gamma} \frac{\partial h_t(\lambda_0, r)}{\partial \psi'}\};$$

$$\Sigma_{rl} = E\{\frac{2}{h_t}\frac{\partial e_t(\lambda_0,r)}{\partial\psi}\frac{\partial e_t(\lambda_0,l)}{\partial\psi'} + \frac{1}{h_t^2}\frac{\partial h_t(\lambda_0,r)}{\partial\psi}\frac{\partial h_t(\lambda_0,l)}{\partial\psi'}\};$$
$$\Omega_r = E\{\frac{2}{h_t}\frac{\partial e_t(\lambda_0,r)}{\partial\lambda}\frac{\partial e_t(\lambda_0,r)}{\partial\lambda'} + \frac{1}{h_t^2}\frac{\partial h_t(\lambda_0,r)}{\partial\lambda}\frac{\partial h_t(\lambda_0,r)}{\partial\lambda'}\}.$$

Note that the matrix Ω_r has the form,

$$\Omega_r = E\{\frac{\partial^2 l_t(\lambda_0,r)}{\partial\lambda\partial\lambda'}\} = \begin{pmatrix}\Sigma & \Sigma_{1r}\\ \Sigma'_{1r} & \Sigma_{rr}\end{pmatrix},$$

and it is not difficult to show that Ω_r is positive definite for each $r \in R$.

With this notation, we first state a basic lemma on the uniform expansion of $\widetilde{LR}_n(r)$ on $[a,b]$. Here, and in the sequel, $o_p(1)$ denotes convergence to zero in probability as $n \to \infty$, $\|\cdot\|$ denotes Euclidean norm.

Lemma 2.1. *If Assumptions* 2.1−2.3 *hold, then under* H_0*, it follows that*

$$(a) \qquad \sup_{r\in[a,b]} \|\tilde{\lambda}_n(r) - \lambda_0\| = o_p(1),$$

$$(b) \qquad \sup_{r\in[a,b]} \|\sqrt{n}[\tilde{\lambda}_n(r) - \lambda_0] + \frac{\Omega_r^{-1}}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\lambda_0,r)}{\partial\lambda}\| = o_p(1),$$

$$(c) \qquad \sup_{r\in[a,b]} \|\widetilde{LR}_n(r) - \frac{1}{2}T'_n(r)(\Sigma_{rr} - \Sigma'_{1r}\Sigma^{-1}\Sigma_{1r})^{-1}T_n(r)\| = o_p(1),$$

where

$$T_n(r) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\lambda_0,r)}{\partial\psi} - \frac{\Sigma'_{1r}\Sigma^{-1}}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\gamma_0)}{\partial\gamma}.$$

It is not difficult to show this lemma by a method similar to that used in Section 6 of Ling and Tong (2005). The proof is omitted. Note that $\{T_n(r), r \in R\}$ is the unique stochastic term in the expansion of $\widetilde{LR}_n(r)$, and hence will play an important role.

Let $R_\nu = [-\nu,\nu]$ and $D^q[R_\nu] = D[R_\nu] \times \cdots \times D[R_\nu]$ (q factors) equipped with the corresponding product Skorohod topology. Weak convergence on $D^q[R]$ is defined as that on $D^q[R_\nu]$ for each $\nu \in (0,\infty)$ as $n \to \infty$, and is denoted by $\Rightarrow$. Let $K_{rl} = \Sigma_{rl} - \Sigma'_{1r}\Sigma^{-1}\Sigma_{1l}$, $\Delta_r = (-\Sigma'_{1r}\Sigma^{-1}, I_{q\times q})$ and $\Gamma_{rl} = 0.5\kappa_4\Gamma_{rl}^{(hh)} - \kappa_3(\Gamma_{rl}^{(eh)} + [\Gamma_{lr}^{(eh)}]')$, where $\kappa_3 = E\varepsilon_t^3$ and $\kappa_4 = E\varepsilon_t^4 - 3$.

Theorem 2.1. *Under Assumptions* 2.1−2.3 *and* H_0*, with* $E\varepsilon_t^8 < \infty$*,* $T_n(r) \Rightarrow \sqrt{2}G_q(r)$ *in* $D^q[R]$*, where* $\{G_q(r), r \in R\}$ *is a* $q \times 1$ *vector Gaussian process with mean zero and covariance kernel* $K^*_{rl} = K_{rl} + \Delta_r\Gamma_{rl}\Delta'_l$*; almost all its paths are continuous.*

When ε_t in model (2.1) is symmetrically distributed, $E[\partial^2 l_t(\lambda_0, r)/\partial\tau\partial\omega] = 0$ and $\kappa_3 = 0$, where $\tau = (\alpha', \beta')'$ and $\omega = (\phi', \psi')'$. Hence we have that $K_{rl} = \Sigma_{rl} - (\Sigma^*_{1r})'(\Sigma^*)^{-1}(\Sigma^*_{1l})$, where

$$\Sigma^* = E\Big\{\frac{2}{h_t}\frac{\partial e_t(\gamma_0)}{\partial\phi}\frac{\partial e_t(\gamma_0)}{\partial\phi'} + \frac{1}{h_t^2}\frac{\partial h_t(\gamma_0)}{\partial\phi}\frac{\partial h_t(\gamma_0)}{\partial\phi'}\Big\},$$

$$\Sigma^*_{1r} = E\Big\{\frac{2}{h_t}\frac{\partial e_t(\gamma_0)}{\partial\phi}\frac{\partial e_t(\lambda_0, r)}{\partial\psi'} + \frac{1}{h_t^2}\frac{\partial h_t(\gamma_0)}{\partial\phi}\frac{\partial h_t(\lambda_0, r)}{\partial\psi'}\Big\},$$

and the derivative functions to τ have disappeared from K_{rl}. If we further assume that ε_t has the same fourth moment as the standard normal distribution, then $\kappa_4 = 0$ and K_{rl} is just the covariance kernel.

As mentioned by Ling and Tong (2005), the stochastic process $\{T_n(r), r \in R\}$ in Theorem 2.1 is a new marked empirical process and our weak convergence result excludes the two points $\pm\infty$. By Lemma 2.1, Theorem 2.1 and the Continuous Mapping Theorem, we can now state the asymptotic distribution of the test statistic LR_n as follows.

Theorem 2.2. *Under the assumptions of Theorem* 2.1*, it follows that*

$$LR_n \to_{\mathcal{L}} \sup_{r\in[a,b]} \{G_q'(r)K_{rr}^{-1}G_q(r)\},$$

as $n \to \infty$, where $\to_{\mathcal{L}}$ denotes convergence in distribution, the matrix K_{rr} and the Gaussian process $\{G_q(r), r \in R\}$ are defined as in Theorem 2.1.

For the case $p = q = 1 < d$, if $\kappa_4 = 0$ and $f(\cdot)$ is symmetric, it holds that $\Sigma_{rl} = \Sigma^*_{1\min\{r,l\}}$. Denote the function $(\Sigma^*)^{-1}\Sigma^*_{1l}$ by $g(l)$. Note that the Gaussian process $\{(\Sigma^*)^{-1/2}G_q(r), r \in R\}$ has mean zero and covariance kernel $g(\min(r, l)) - g(r)g(l)$ with g a monotonic increasing function, $g(-\infty) = 0$ and $g(\infty) = 1$. As in Chan (1990), the supremum in Theorem 2.2 has the same distribution as

$$\sup_{\pi_1\le r\le\pi_2} \frac{B_r^2}{r - r^2},$$

where $\pi_1 = g(a)$, $\pi_2 = g(b)$, and B_r is just a Brownian bridge. This distribution is the same as that of test statistics for change points in Andrews (1993), and critical values can be found in Andrews (1993) or Chan (1991). In practice, we can select the values for (π_1, π_2), e.g., $\pi_1 = 0.05$ and $\pi_2 = 0.95$, and compute LR_n with $a = g^{-1}(\pi_1)$ and $b = g^{-1}(\pi_2)$, where $g^{-1}(\cdot)$ is the inverse function of g. Some guidelines on this can be found in Chan (1990). For other cases, the critical values of LR_n can be obtained via a simulation method and the first experiment in Section 4 provides an overview of this.

3. Asymptotic Power under Local Alternatives

To investigate the asymptotic power of the test proposed in the previous section, we consider the asymptotic behavior of LR_n under

$$H_{1n} : \psi_0 = \frac{h}{\sqrt{n}} \text{ for a constant vector } h \in R^q \text{ and } r = r_0 \in R,$$

where r_0 is a fixed value.

We first introduce some notation. Let $\mathcal{F}^Z$ be the Borel σ-field on $\mathcal{R}^Z$ with $Z = \{0, \pm 1, \pm 2, \ldots\}$, P be a probability measure on $(\mathcal{F}^Z, \mathcal{R}^Z)$ and P_λ^n be the restriction of P on $\mathcal{F}_n$, where $\mathcal{F}_n$ is the σ-field generated by $\{y_n, \ldots, y_1, Y_0\}$ and $Y_0 = \{y_0, y_{-1}, \ldots\}$ is the initial vector. Denote the error functions of model (2.1) by $\varepsilon_t(\lambda, r) = e_t(\lambda, r)/\sqrt{h_t(\lambda, r)}$, where the functions $e_t(\lambda, r)$ and $h_t(\lambda, r)$ are defined in the previous section. Suppose that the error functions $\{\varepsilon_1(\lambda, r_0), \ldots, \varepsilon_t(\lambda, r_0)\}$ are identically independently distributed with density function $f(\cdot)$ under P_λ^n, and are independent of Y_0. From model (2.1), the initial vector Y_0 has the same distribution under P_λ^n and $P_{\lambda_0}^n$. Hence the log-likelihood ratio $\Lambda_n(\lambda_1, \lambda_2)$ of $P_{\lambda_2}^n$ to $P_{\lambda_1}^n$ is

$$\Lambda_n(\lambda_1, \lambda_2) = 2 \sum_{t=1}^{n} [\log g_t(\lambda_2) - \log g_t(\lambda_1)],$$

where $g_t(\lambda) = \sqrt{f(\varepsilon_t(\lambda, r_0))}/\sqrt[4]{h_t(\lambda, r_0)}$, see Ling and McAleer (2003).

To find the asymptotic distribution of LR_n under H_{1n}, we need the LAN property of $\Lambda_n(\lambda_0, \lambda_0 + v_n/\sqrt{n})$ and the contiguity of $P_{\lambda_0}^n$ and $P_{\lambda_0 + v_n/\sqrt{n}}^n$, where v_n is a bounded constant sequence in $R^{p+q+r+s+1}$.

Assumption 3.1. *The density f of ε_t is absolutely continuous with a.e.-derivative f' and*

$$I_1(f) = \int \xi_1^2(x) f(x) dx < \infty, \qquad I_2(f) = \int \xi_2^2(x) f(x) dx < \infty,$$

where $\xi_1(x) = f'(x)/f(x)$ and $\xi_2 = 1 + x\xi_1(x)$.

Denote $\zeta_{1t} = \xi_1(\varepsilon_t)$, $\zeta_{2t} = \xi_2(\varepsilon_t)$, $I_3(f) = E(\zeta_{1t}\zeta_{2t})$, $I_4(f) = I_2(f) - 2I_1(f)$, and $\Gamma_{rl}^* = 0.5 I_4(f) \Gamma_{rl}^{(hh)} - I_3(f)(\Gamma_{rl}^{(eh)} + [\Gamma_{lr}^{(eh)}]')$, where $r, l \in R$. Applying Theorem 2.1 and Remark 2.1 in Ling and McAleer (2003), we have the following theorem.

Theorem 3.1. *If Assumptions* 2.1–2.3 *and* 3.1 *hold and* $\lambda_0 = (\gamma_0', 0)'$, *then*

$$(a) \quad \Lambda_n(\lambda_0, \lambda_0 + \frac{v_n}{\sqrt{n}}) = \frac{v_n'}{\sqrt{n}} \sum_{t=1}^{n} [\frac{\zeta_{1t}}{\sqrt{h_t}} \frac{\partial e_t(\lambda_0, r_0)}{\partial \lambda} - \frac{\zeta_{2t}}{2h_t} \frac{\partial h_t(\lambda_0, r_0)}{\partial \lambda}]$$

$$-\frac{v_n'}{4}[I_1(f)\Omega_{r_0} + \Gamma^*_{r_0 r_0}]v_n \text{ under } P^n_{\lambda_0},$$

(*b*) $P^n_{\lambda_0}$ *and* $P^n_{\lambda_0+v_n/\sqrt{n}}$ *are contiguous,*

where the matrix Ω_r *is defined as in the previous section.*

Furthermore, we can show the following theorem by the Central Limit Theorem, the Continuous Mapping Theorem, Theorems 2.1 and 3.1.

Theorem 3.2. *If Assumption* 3.1 *and the assumptions of Theorem* 2.1 *hold, then under* H_{1n}*,*

(*a*) $T_n(r) \Rightarrow \mu(r) + \sqrt{2}G_q(r)$ *in* $D^q[R]$,

(*b*) $LR_n \to_{\mathcal{L}} \frac{1}{2}\sup_{r\in[a,b]}\{[\mu(r) + \sqrt{2}G_q(r)]'K_{rr}^{-1}[\mu(r) + \sqrt{2}G_q(r)]\}$,

where $\mu(r) = K^*_{rr_0}h$ *and* $G_q(r)$ *is a Gaussian process defined in Theorem* 2.1*.*

The above theorem shows that the likelihood ratio test LR_n has non-trivial asymptotic power under local alternatives H_{1n}.

4. Simulation and Empirical Results

4.1. Two simulation experiments

We first performed two simulation experiments to demonstrate the usefulness of the test in Section 2. The first experiment is used to demonstrate the sample sizes and replications needed when simulation experiments were used to find out critical values of test LR_n. The following generating process was involved,

$$y_t = 0.5e_{t-1} + e_t,\ e_t = \varepsilon_t h_t^{\frac{1}{2}} \text{ and } h_t = 1.0 + 0.3e^2_{t-1} + 0.3h_{t-1},$$

where $\{\varepsilon_t\}$ is *i.i.d.* with the standard normal distribution $N(0,1)$. The sample size n was set to be 200, 500 or $5,000$, and there were 10,000 replications for each sample size. For the parameters in the alternative hypothesis, we let $q = d = 1$, $a = -1.28\sqrt{\text{var}(y_t)}$ and $b = 1.28\sqrt{\text{var}(y_t)}$, where 1.28 is the 0.9-quantile of the standard normal distribution. The Newton-Raphson algorithm was used to search for the quasi-maximum likelihood estimators for MA(1)-GARCH(1,1) and TMA(1,1,1)-GARCH(1,1) models and this algorithm was also used in the second experiment and the two data examples below. Figure 4.1 displays the tail behaviors of empirical distributions of the statistic LR_n for three different sample sizes. We may take the empirical distribution for sample size 5,000 to be the 'true' distribution. It can be seen that the empirical distribution matches the 'true' one satisfactorily for n as small as 200 and rather well for $n = 500$.

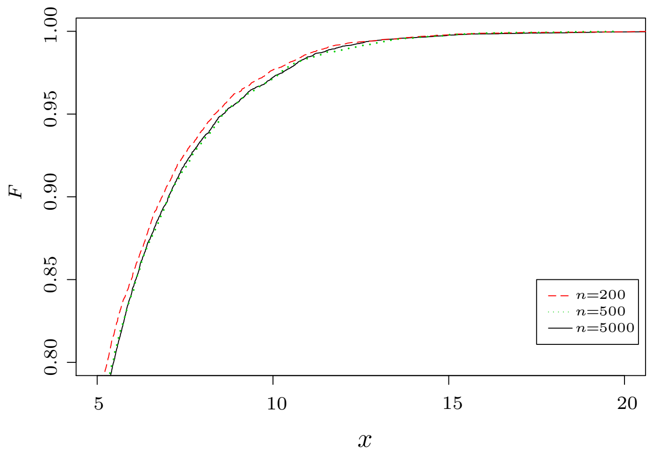

Figure 4.1. The simulated empirical distributions of the statistic LR_n with $n = 200$, 500 and 5,000, based on 10,000 replications for each sample size.

The second experiment was performed to illustrate that ignoring the effect of conditional heteroscedasticity could lead to misleading conclusions. In this experiment, we considered the generating process

$$y_t = 0.5e_{t-1} + e_t,\ e_t = \varepsilon_t h_t^{\frac{1}{2}} \text{ and } h_t = 1.0 + 0.5e_{t-1}^2,$$

where $\{\varepsilon_t\}$ is *i.i.d.* with $N(0, 1)$, and the test of Ling and Tong (2005) was used to check whether or not there exists a threshold structure in the generated sequence. We tried three different models with $(p, q, d) = (1, 1, 2)$, $(1, 1, 3)$ or $(2, 2, 3)$ and the sample size was set to be 400. There were 1,000 replications for each model and the quantities a and b were respectively selected to be 0.1- and 0.9-quantiles of the empirical distribution based on the sample $\{y_1, \ldots, y_{400}\}$. The empirical sizes for some upper percentage points are presented in Table 4.1, where the respective critical values are given by Andrews (1993). The empirical sizes in Table 4.1 are much greater than the nominal values. Hence, the test for threshold moving average with *i.i.d.* errors is sensitive to the presence of conditional heteroscedasticity.

4.2. Two data examples

We first consider the weekly closing price of S&P 500. The time range is from Jan.1, 1996 to Dec.31, 2005 and there are 521 observations in total. The modeling results for its centered log return sequence (as percentage) are summarized in Table 4.2, and eight models, MA(p), TMA(p,d,p), MA(p)-GARCH(1,1) and

Table 4.1. Empirical sizes for the theoretical upper percentage points using the quasi-likelihood ratio test in Ling and Tong (2005) for time series with conditional heteroscedasticity.

Models		Empirical sizes			Critical values		
p or q	d	10%	5%	1%	10%	5%	1%
1	2	37.8	29.2	16.9	7.63	9.31	12.69
1	3	44.3	31.3	17.7	7.63	9.31	12.69
2	3	41.9	32.2	16.5	10.50	12.27	16.04

TMA(p,d,p)-GARCH(1,1) with $p = 1$ and 2, are involved. The quasi-maximum likelihood method is employed and the delay parameter d and the threshold parameter r are estimated by

$$(d_0, r_0) = \underset{r \in [a,b], 1 \le d \le 16}{\operatorname{argmin}} \tilde{L}_{1n}(\tilde{\lambda}(r), r),$$

where $(a, b) = (-1.1892, 1.2442)$ are, respectively, the empirical 0.1- and 0.9-quantiles of the centered log return sequence. In Table 2, the BICs of models with GARCH(1,1) errors are less than those of models with *i.i.d.* errors and the estimated values of a_1 and b_1 are all significant. We may claim that there exists the phenomenon of time-varying conditional variance in this time series. Furthermore, the BICs of threshold models are all greater than those of the corresponding models with no threshold and this can be considered as evidence of no threshold in the centered log return sequence of S&P 500 weekly closing price. To draw a conclusion formally, Ling and Tong's test (LR_n^{LT}) and the test in this article (LR_n^{LL}) are considered. The results are presented in Table 4.3, and the critical values for the test, MA *versus* TMA, are from Andrews (1993). The critical values for the test, MA-GARCH(1,1) *versus* TMA-GARCH(1,1), are obtained via simulation. In the simulation experiments, the sample sizes are set to be 1,000 and the obtained critical values are based on 10,000 replications. The test LR_n^{LL} suggests that there is no threshold in the sequence at all three significance levels. Note that the test LR_n^{LT} rejects the hypothesis of no threshold at the significance level of 0.05 and this may be because of the presence of conditional heteroscedasticity as in the second simulation experiment in the previous subsection.

We next consider the centered log return (as percentage) of weekly exchange rate of Japanese Yen against USA dollar from Jan. 1, 1994 to Dec. 31, 2003 with 521 observations. The same modeling process as the above example is considered again with $(a, b) = (-0.6992, 0.6049)$ the 0.1- and 0.9-quantiles of the empirical distribution. For the parameters in the GARCH part, we considered the sparse model, $h_t = a_0 + a_5 e_{t-5}^2 + b_1 h_{t-1}$, instead of the full model, $h_t = a_0 + \sum_{i=1}^5 a_i e_{t-i}^2$

Table 4.2. Modeling results for the centered log return of S&P 500 weekly closing price(1996-2005).

Models	ψ_1	ψ_2	ϕ_1	ϕ_2	r_0	a_0	a_1	b_1	BIC
Models with *i.i.d.* errors									
MA(1)			-0.0657						528.98
TMA(1,4,1)	0.3181		-0.1680		-0.7488				537.11
MA(2)			-0.0642	0.0476					531.69
TMA(2,6,2)	0.3862	-0.3282	-0.1552	0.1283	-1.0672				542.54
Models with GARCH(1,1) errors									
MA(1)			-0.0820			0.2113	0.1826	0.6313	526.54
TMA(1,9,1)	-0.2929		-0.0072		-0.5838	0.2106	0.1802	0.6303	528.26
MA(2)			-0.0778	0.0624		0.2156	0.1798	0.6285	527.94
TMA(2,7,2)	0.3096	-0.0586	-0.1963	0.1200	-0.2318	0.2121	0.1725	0.6335	529.50

Table 4.3. Testing results for the centered log return of S&P 500 weekly closing price(1996-2005).

	The test statistics		Critical values		
	LR_n^{LT}	LR_n^{LL}	10%	5%	1%
MA *versus* TMA					
$p=q=1, d=4$	12.4992		7.63	9.31	12.69
$p=q=2, d=6$	18.9044		10.50	12.27	16.04
MA-GARCH(1,1) *versus* TMA-GARCH(1,1)					
$p=q=1, d=9$		6.4307	7.03	8.54	12.07
$p=q=2, d=7$		9.3016	10.02	11.72	14.43

Table 4.4. Modeling and testing results for the centered log return of weekly exchange rate of Japanese Yen against USA dollar (1994-2003).

Models	ψ_1	ϕ_1	r_0	a_0	a_5	b_1	BIC	LR_n^{LT}	LR_n^{LL}
Models with *i.i.d.* errors									
MA(1)		0.263					676.22		
TMA(1,10,1)	-0.367	0.373	-0.108				676.08	18.74	
Models with GARCH(5,1) errors									
MA(1)		0.258		0.157	0.231	0.237	282.40		
TMA(1,10,1)	-0.282	0.338	-0.097	0.174	0.219	0.179	277.72		12.82

Critical values at the upper 5% for LR_n^{LT} and LR_n^{LL} are, respectively, 9.31 and 8.69.

$+b_1h_{t-1}$, as the fitted parameters $\widehat{a}_i, i=1,\ldots,4$ are insignificant in the process of estimation. Hence, in this example, we consider four models, MA(1), TMA(1,d,1), MA(1)-GARCH(5,1) and TMA(1,d,1)-GARCH(5,1), and the modeling results are presented in Table 4.4. Here the BICs strongly prefer the conditional heteroscedastic models and also suggest the existence of the threshold structure in the time series. Our test LR_n^{LL} rejects the null hypothesis of no threshold in the centered log return sequence at the significance level of 0.05 and so does the test LR_n^{LT}. Furthermore, these two threshold models share the same value of d and the estimated threshold parameters are approximately the same. Therefore, it seems that the threshold MA structure with time varying conditional variance really exists in this time series.

Acknowledgement

W. K. Li thanks the Hong Kong Research Grant Council grant HKU 7036/06P for partial support.

Appendix: Proof of Theorem 2.1

We state the proofs of Theorem 2.1 and two key lemmas in this appendix. For convenience, all parameters here are evaluated at the true ones unless otherwise specified.

Let

$$T_{0n}(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\varepsilon_t}{\sqrt{h_t}} \Big[\sum_{i=0}^{\infty} u_p' \Phi^i u_p Z_{t-1-i} I(y_{t-d-i} \le r) \Big],$$

$$T_{jn}(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{(1-\varepsilon_t^2)}{h_t} \times \Big\{ \sum_{i=0}^{\infty} u_s' B^i u_s e_{t-i-j} \Big[\sum_{l=0}^{\infty} u_p' \Phi^l u_p Z_{t-1-i-j-l} I(y_{t-d-i-j-l} \le r) \Big] \Big\},$$

where $j = 1, \ldots, m$, $Z_t = (e_t, \ldots, e_{t-q+1})'$, $u_p = (1, 0, \ldots, 0)'_{1\times p}$, $u_s = (1, 0, \ldots, 0)'_{1\times s}$, matrices Φ and B are defined as in Lemma A.1 and evaluated at the true parameter vector λ_0. Following Stute (1997) and Ling and Tong (2005), we call $\{T_{jn}(r), r \in R\}, j = 0, 1, \ldots, m$ the marked empirical processes. It can be verified that

$$T_n(r) = \frac{\Delta_r}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial l_t(\lambda_0, r)}{\partial \lambda} = -\frac{\Sigma_{1r}' \Sigma^{-1}}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial l_t(\gamma_0)}{\partial \gamma} + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial l_t(\lambda_0, r)}{\partial \psi},$$

where the first item at the right hand side of $T_n(r)$ is independent of the threshold parameter r, and the second item can be rewritten as,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial l_t(\lambda_0, r)}{\partial \psi} = -2[T_{0n}(r) + \sum_{j=1}^{m} T_{jn}(r)].$$

Hence, it is sufficient to show the tightness of the marked empirical processes $\{T_{jn}(r), r \in R_\gamma\}, j = 0, 1, \ldots, m$.

The tightness of $T_{0n}(r)$. Note that, for any $\kappa > 0$,

$$\begin{aligned} T_{0n}(r' + \kappa) - T_{0n}(r') &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\varepsilon_t}{\sqrt{h_t}} \Big[\sum_{i=0}^{\infty} u_p' \Phi^i u_p Z_{t-1-i} I(r' < y_{t-d-i} \le r' + \kappa) \Big], \\ \sup_{r' < r \le r' + \kappa} \|T_{0n}(r) - T_{0n}(r')\| &\le \frac{1}{a_0\sqrt{n}} \sum_{t=1}^{n} |\varepsilon_t| \Big[\sum_{i=0}^{\infty} \|\Phi^i\| \|Z_{t-1-i}\| I(r' < y_{t-d-i} \le r' + \kappa) \Big]. \end{aligned}$$

Hence, by Burkholder's inequality (Hall and Heyde (1980, p.23)), Holder's inequality, Minkowskii's inequality, Lemmas A.1 and A.2, it can be shown that

$$E\|T_{0n}(r' + \kappa) - T_{0n}(r')\|^4 \le C[\frac{\kappa}{n} + \kappa^2], \tag{A.1}$$

$$E\{ \sup_{r' < r \le r' + \kappa} \|T_{0n}(r) - T_{0n}(r')\|\}^4 \le C[\frac{\kappa}{n} + \kappa^2 + n\kappa^3 + n^2\kappa^4], \tag{A.2}$$

where C is a constant independent of r'.

For any $\delta < 1$, we choose n such that $n^{-1} \le \delta$, and then select an integer K such that $\delta n/2 \le K \le n\delta$. Let $r_{k+1} = r_k + \delta/K$, where $r_1 = r'$ and $k = 1, \ldots, K$. Thus

$$\begin{aligned} &\sup_{r' < r \le r' + \delta} \|T_{0n}(r) - T_{0n}(r')\| \\ &\quad \le \max_{1 \le k \le K} \|T_{0n}(r_k) - T_{0n}(r')\| + \max_{1 \le k \le K} \sup_{r_k < r \le r_k + \frac{\delta}{K}} \|T_{0n}(r) - T_{0n}(r_k)\|. \end{aligned} \tag{A.3}$$

Note that $T_{0n}(r_i) - T_{0n}(r_j) = \sum_{k=i+1}^{j} [T_{0n}(r_k) - T_{0n}(r_{k-1})]$ and $n^{-1} \le \delta/K$. Then, by (A.1), it follows that

$$E\|T_{0n}(r_i) - T_{0n}(r_j)\|^4 \le C\Big\{ \frac{(j-i)\delta}{nK} + \Big[\frac{(j-i)\delta}{K}\Big]^2 \Big\} \le C\Big(\sum_{k=i+1}^{j} \frac{\delta}{K} \Big)^2$$

for any $1 \le i < j \le K$. Hence, by Theorem 12.2 in Billingsley (1968, p.94) there exists a constant C_{11} independent of K, δ, r' and n such that

$$P\Big(\max_{1 \le k \le K} \|T_{0n}(r_k) - T_{0n}(r')\| > \frac{\eta}{2} \Big) \le \frac{C_{11}C}{\eta^4} \Big(\sum_{k=1}^{K} \frac{\delta}{K} \Big)^2 = \frac{C_{11}C}{\eta^4} \delta^2. \tag{A.4}$$

Furthermore, by (A.2), we can show that

$$P\Big(\max_{1\le k\le K}\sup_{r_k<r\le r_k+\delta/K}\|T_{0n}(r)-T_{0n}(r_k)\|>\frac{\eta}{2}\Big)\le\frac{32C}{\eta^4}\delta^2. \tag{A.5}$$

Given $\varepsilon>0$ and $\eta>0$, let $\delta=\min\{\varepsilon\eta^4/(CC_{11}+32C),0.5\}$. We select N to be $[\delta^{-1}]$, the largest integer less than δ^{-1}. Thus, by (A.3)-(A.5),

$$P\Big(\sup_{r'<r\le r'+\delta}\|T_{0n}(r)-T_{0n}(r')\|>\eta\Big)\le\frac{C(C_{11}+32)}{\eta^4}\delta^2\le\delta\varepsilon$$

for any $r'\in R_\gamma$ and $n>N$. By Theorem 15.5 in Billingsley (1968) and the proof of his Theorem 16.1, it can be shown that $\{T_{0n}(r),r\in R_\gamma\}$ is tight.

The tightness of $T_{jn}(r)$ *with* $j=1,\ldots,m$. For $T_{jn}(r)$, $j=1,\ldots,m$, it holds that

$$\sqrt{u_s'B^iu_s}\le M_1\rho^{\frac{i}{2}}\quad\text{and}\quad\frac{\sqrt{u_s'B^iu_s}e_{t-i-j}}{h_t}\le M_3\text{ a.s.},$$

where $u_s=(1,0,\ldots,0)'_{1\times s}$, $i\ge0$, $j=1,\ldots,m$, ρ is defined in Lemma A.1, and M_1 and M_3 are constants. Then, by arguments similar used for $\{T_{0n}(r),r\in R_\gamma\}$, we can claim that $\{T_{jn}(r),r\in R_\gamma\},j=1,\ldots,m$ are tight under the finite eighth moment of ε_t.

Lemma A.1. *If Assumptions* 2.1 *and* 2.2 *hold, then it holds that*

$$(a)\quad\sup_{\lambda\in\Theta}\sup_{r\in[a,b]}\Big\|\prod_{i=1}^{j}[\Phi+\Psi I(y_{t-i}\le r)]\Big\|=O(\rho^j)$$

$$(b)\quad\sup_{\lambda\in\Theta}\|B^j\|=O(\rho^j),$$

where $0<\rho<1$,

$$\Phi=\begin{pmatrix}-\phi_1 & \cdots & -\phi_p\\ & I_{p-1} & \boldsymbol{0}_{(p-1)\times1}\end{pmatrix},\Psi=\begin{pmatrix}-\psi_1 & \cdots & -\psi_p\\ & \boldsymbol{0}_{(p-1)\times p} & \end{pmatrix},$$

$$B=\begin{pmatrix}b_1 & \cdots & b_s\\ & I_{s-1} & \boldsymbol{0}_{(s-1)\times1}\end{pmatrix},$$

with I_k *being the* $k\times k$ *identity matrix and* $\boldsymbol{0}_{k\times l}$ *the* $k\times l$ *zero matrix.*

Proof. In fact, (a) is just Theorem A.1 of Ling and Tong (2005). By a similar method, we can show that (b) also holds.

Lemma A.2. *If Assumptions* 2.1 *and* 2.3 *hold then, under* H_0, *it follows that*

$$(a)\quad E\{|e_{t-j}|^kI(r'<y_{t-d}\le r)\}\le C(r-r'),\ \text{as } k=0,2,4,$$

(b) $\quad E\{\|Z_{t-1}\|^k I(r' < y_{t-d} \le r) \le C(r-r'),$ *as* $k = 2, 4,$

(c) $\quad E\{|e_{t-j}|^4 I(r' < y_{t-d} \le r) \prod_{i=1}^{l} I(r' < y_{t_i-d} \le r)\} \le C(r-r')^{l+1},$

(d) $\quad E\{\|Z_{t-1}\|^4 I(r' < y_{t-d} \le r) \prod_{i=1}^{l} I(r' < y_{t_i-d} \le r)\} \le C(r-r')^{l+1},$

where $Z_t = (e_t, \dots, e_{t-q+1})'$, $l = 1, 2, 3$, $j \ge 1$, t, t_1, t_2, t_3 *are different from each other,* $r' < r$, $r, r' \in R_\gamma$, R_γ *is defined as before and* C *is a constant independent of* r *and* r'.

Proof. Under Assumption 2.3, it holds that $\sup_{x\in R} |x|^k f(x) \le M$ for $k = 0, 2$ and 4, where M is a constant. Let $g_t = \sum_{i=1}^{p} \phi_i e_{t-i}$ and find that

$$\begin{aligned} E\{|e_t|^k I(r' < y_t \le r)|\mathcal{F}_{t-1}\} &= E\{|\varepsilon_t|^k h_t^{\frac{k}{2}} I(r' < \varepsilon_t h_t^{\frac{1}{2}} + g_t \le r)|\mathcal{F}_{t-1}\} \\ &= h_t^{\frac{k}{2}} \int_{h_t^{-\frac{1}{2}}(r'-g_t)}^{h_t^{-\frac{1}{2}}(r-g_t)} |x|^k f(x) dx \\ &\le M h_t^{\frac{k-1}{2}} (r-r') \\ &\le C_1 h_t^{\frac{k}{2}} (r-r') \quad \text{a.s.,} \end{aligned} \tag{A.6}$$

where $k = 0, 2, 4$ and $C_1 = M/\sqrt{a_0}$. Note that $E[I(r' < y_t \le r)|\mathcal{F}_{t-1}] \le C_1(r-r')$ a.s.. Hence, it can be shown that (a) holds for the case $t - j \le t - d$.

We now consider the case $t - j > t - d$ for (a). Without loss of generality, the notations $t - j$ and $t - d$ can be replaced respectively by $t + L$ and t with $L > 0$. Let

$$H_t = (e_t^2, \dots, e_{t-m+1}^2, h_t, \dots, h_{t-s+1})'_{1\times(m+s)}, \quad u = (1, 0, \dots, 0)'_{1\times(m+s)},$$

$$\pi_t = (a_0\varepsilon_t^2, 0, \dots, 0, a_0, 0, \dots, 0)'_{1\times(m+s)},$$

$$A_t = \left(\begin{array}{ccc|ccc} a_1\varepsilon_t^2 & \cdots & a_m\varepsilon_t^2 & b_1\varepsilon_t^2 & \cdots & b_s\varepsilon_t^2 \\ & I_{m-1} & \mathbf{0}_{(m-1)\times 1} & & \mathbf{0}_{(m-1)\times s} & \\ \hline a_1 & \cdots & a_m & b_1 & \cdots & b_s \\ & \mathbf{0}_{(s-1)\times m} & & & I_{s-1} & \mathbf{0}_{(s-1)\times 1} \end{array}\right)_{(m+s)\times(m+s)},$$

where I_m is the $m \times m$ identity matrix and $\mathbf{0}_{m\times s}$ is the $m \times s$ zero matrix. From (2.1), we can rewrite e_{t+L}^2 as

$$e_{t+L}^2 = a_0\varepsilon_{t+L}^2 + \sum_{j=0}^{L-2} u' \prod_{i=0}^{j} A_{t+L-i}\pi_{t+L-j-1} + u' \prod_{j=0}^{L-1} A_{t+L-j} H_t$$

$$= \tilde{A}^{(1)}_{t+1,t+L} + \tilde{A}^{(2)}_{t+1,t+L} H_t \quad \text{a.s..},$$

where $\tilde{A}^{(1)}_{t+1,t+L} = a_0\varepsilon^2_{t+L} + \sum_{j=0}^{L-2} u' \prod_{i=0}^{j} A_{t+L-i}\pi_{t+L-j-1}$ and $\tilde{A}^{(2)}_{t+1,t+L} = u' \prod_{j=0}^{L-1} A_{t+L-j}$ are functions of $\{\varepsilon^2_{t+1}, \ldots, \varepsilon^2_{t+L}\}$, hence independent of the σ-field $\mathcal{F}_t$. Note that $\tilde{A}^{(1)}_{t+1,t+L}$ is a random variable, $\tilde{A}^{(2)}_{t+1,t+L}$ is a random vector and

$$E[H_t I(r' < y_t \le r)|\mathcal{F}_{t-1}] \le C_1(r - r')\tilde{H}_t = C_1(r - r')E[H_t|\mathcal{F}_{t-1}] \quad \text{a.s.},$$

where $\tilde{H}_t = (h_t, e^2_{t-1} ..., e^2_{t-m+1}, h_t, \ldots, h_{t-s+1})'$ is just the vector H_t with the first element replaced by h_t. Hence,

$$\begin{aligned}
&E\{e^2_{t+L} I(r' < y_t \le r)|\mathcal{F}_{t-1}\} \\
&= E(\tilde{A}^{(1)}_{t+1,t+L})E[I(r' < y_t \le r)|\mathcal{F}_{t-1}] + E(\tilde{A}^{(2)}_{t+1,t+L})E[H_t I(r' < y_t \le r)|\mathcal{F}_{t-1}] \\
&\le C_1(r - r')E(\tilde{A}^{(1)}_{t+1,t+L}) + C_1(r - r')E(\tilde{A}^{(2)}_{t+1,t+L})E[H_t|\mathcal{F}_{t-1}] \\
&= C_1(r - r')E(e^2_{t+L}|\mathcal{F}_{t-1}) \quad \text{a.s.,} \qquad \text{(A.7)}
\end{aligned}$$

where C_1 is the same as (A.6). Denote $\tilde{A}^{(2)}_{t+1,t+L}$ by $c = (c_1, \ldots, c_{m+s})'$, and then

$$\begin{aligned}
&E[(c'H_t)^2 I(r' < y_t \le r)|\mathcal{F}_{t-1}] \\
&\le (m + s)E[(\sum_{i=1}^{m} c_i^2 e^4_{t+1-i} + \sum_{i=1}^{s} c^2_{m+i} h^2_{t+1-i})I(r' < y_t \le r)|\mathcal{F}_{t-1}] \\
&\le (m + s)C_1(r - r')(E[c_1^2]h_t^2 + \sum_{i=2}^{m} E[c_i^2]e^4_{t+1-i} + \sum_{i=1}^{s} E[c^2_{m+i}]h^2_{t+1-i}) \\
&\le (m + s)C_1C_2(r - r')[E[c_1^2](E\varepsilon_t^4)h_t^2 + \sum_{i=2}^{m} E[c_i^2]e^4_{t+1-i} + \sum_{i=1}^{s} E[c^2_{m+i}]h^2_{t+1-i}] \\
&= C_3(r - r')E[\sum_{i=1}^{m} c_i^2 e^4_{t+1-i} + \sum_{i=1}^{s} c^2_{m+i} h^2_{t+1-i}|\mathcal{F}_{t-1}] \\
&\le C_3(r - r')E[(c'H_t)^2|\mathcal{F}_{t-1}] \quad \text{a.s.,}
\end{aligned}$$

where $C_3 = (m + s)C_1C_2$, $C_2 = \max\{(E\varepsilon_t^4)^{-1}, 1\}$. Thus

$$\begin{aligned}
&E\{e^4_{t+L} I(r' < y_t \le r)|\mathcal{F}_{t-1}\} \\
&\quad \le 2E[\tilde{A}^{(1)}_{t+1,t+L}]^2 E[I(r' < y_t \le r)|\mathcal{F}_{t-1}] \\
&\qquad + 2E\{[\tilde{A}^{(2)}_{t+1,t+L} H_t]^2 I(r' < y_t \le r)|\mathcal{F}_{t-1}\} \\
&\quad \le 2C_1(r - r')E[\tilde{A}^{(1)}_{t+1,t+L}]^2 + 2C_3(r - r')E\{[\tilde{A}^{(2)}_{t+1,t+L} H_t]^2|\mathcal{F}_{t-1}\} \\
&\quad \le C_4(r - r')E\{[\tilde{A}^{(1)}_{t+1,t+L}]^2 + [\tilde{A}^{(2)}_{t+1,t+L} H_t]^2|\mathcal{F}_{t-1}\}
\end{aligned}$$

$$\leq C_4(r - r')E(e_{t+L}^4 | \mathcal{F}_{t-1}) \quad \text{a.s.,} \tag{A.8}$$

where $C_4 = 2\max\{C_1, C_3\}$. Following (A.7) and (A.8), we can show that (a) also holds for the case $t - j > t - d$.

Applying (A.6) and (A.8) repeatedly, it is readily verified that (c) holds. From (a) and (c), we get (b) and (d) immediately.

References

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61**, 821-55.

Berkes, I. and Horvath, L. (2004). The efficiency of the estimators of the parameters in GARCH processes. *Ann. Statist.* **32**, 633-55.

Berkes, I., Horvath, L. and Kokoszka, P. (2003). GARCH processes: structure and estimation. *Bernoulli* **9**, 201-27.

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Bollerslev, T. (1986). Generalized autoregression conditional heteroscedasticity. *J. Economet.* **31**, 307-327.

Bollerslev, T., Chou, R. Y. and Kroner, K. F. (1992). ARCH model in finance: A review of the theory and empirical evidence. *J. Economet.* **52**, 5-59.

Caner, M. and Hansen, B.E. (2001). Threshold autoregression with a unit root. *Econometrica* **69**, 1555-1596.

Chan, K. S. (1990). Testing for threshold autoregression. *Ann. Statist.* **18**, 1886-1894.

Chan, K. S. (1991). Percentage points of likelihood ratio tests for threshold autoregression. *J. Roy. Statist. Soc. Ser. B* **53**, 691-696.

Chan, K. S. and Tong, H. (1985). On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. Appl. Probab.* **17**, 666-678.

Chan, K. S. and Tong, H. (1990). On likelihood ratio tests for threshold autoregression. *J. Ror. Statist. Soc. Ser. B* **52**, 469-476.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64**, 247-254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74**, 33-43.

Engle, R. F. (1982). Autoregression conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50**, 987-1008.

Hall, P. and Heyde, C. C. (1980) *Martingale Limit Theory and its Application.* Academic Press, New York.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* **64**, 413-430.

Li, W. K. and Lam, K. (1995). Modelling the asymmetry in stock returns using threshold ARCH model. *Statistician* **44**, 333-341.

Ling, S. (1999). On probability properties of a double threshold ARMA conditional heteroskedasticity model. *J. Appl. Probab.* **36**, 688-705.

Ling, S. and McAleer, M. (2003). On adaptive estimation in nonstationary ARMA models with GARCH errors. *Ann. Statist.* **31**, 642-676.

Ling, S. and Tong, H. (2005). Testing for a linear MA model against threshold MA models. *Ann. Statist.* **33**, 2529-2552.

Liu, J., Li, W. K. and Li C. W. (1997). On a threshold autoregression with conditional heteroscedastic variances. *J. Statist. Plann. Inference* **62**, 279-300.

Peng, L. and Yao, Q. (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* **90**, 967-975.

Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25**, 613-641.

Tong, H. (1978). On a threshold model. In *Pattern Recognition and Signal Processing*, (Edited by C. H. Chen), 575-586. Sijthoff and Noordhoff, Amsterdam.

Tong, H. (1990). *Nonlinear Time Series: A Dynamical System Approach.* Oxford University Press, Oxford.

Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *J. Roy. Statist. Soc. Ser.B* **42**, 245-292.

Tsay, R. S. (1987). Conditional heteroscedastic time series models. *J. Amer. Statist. Assoc.* **82**, 590-605.

Wong, C. S. and Li, W. K. (1997). Testing for threshold autoregression with conditional heteroscedasticity. *Biometrika* **84**, 407-418.

Wong, C. S. and Li, W. K. (2000). Testing for double threshold autoregressive conditional heteroscedastic model. *Statist. Sinica* **10**, 173-189.

Division of Mathematical Sciences, School of Physcial and Mathematical Sciences, Nanyang Technological University, Block 5, Level 3, Singapore 637616.

E-mail: gdli@ntu.edu.sg

Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, Hong Kong.

E-mail: hrntlwk@hku.hk

(Received June 2006; accepted November 2006)