Biometrika (2011), **98**, 1, *pp.* 243–250
doi: 10.1093/biomet/asq074

Printed in Great Britain
Advance Access publication 8 February 2011

Testing a linear time series model against its threshold extension

BY GUODONG LI AND WAI KEUNG LI

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

gdli@hku.hk hrntlwk@hku.hk

SUMMARY

This paper derives the asymptotic null distribution of a quasilikelihood ratio test statistic for an autoregressive moving average model against its threshold extension. The null hypothesis is that of no threshold, and the error term could be dependent. The asymptotic distribution is rather complicated, and all existing methods for approximating a distribution in the related literature fail to work. Hence, a novel bootstrap approximation based on stochastic permutation is proposed in this paper. Besides being robust to the assumptions on the error term, our method enjoys more flexibility and needs less computation when compared with methods currently used in the literature. Monte Carlo experiments give further support to the new approach, and an illustration is reported.

Some key words: Autoregressive moving average model; Bootstrap method; Quasilikelihood ratio test; Threshold model.

1. INTRODUCTION

Threshold models were first proposed by Tong & Lim (1980), and have since become a standard class of nonlinear time series models; see Tong (1990). Due to the complexity of nonlinear models, there arises the important problem of testing whether a threshold model can provide a better fit to the data than a linear one. This problem has attracted a lot of attention, e.g., a simple portmanteau test was considered in Petruccelli & Davies (1986) and Tsay (1998); the quasilikelihood ratio test in Chan (1990), Chan & Tong (1990) and Ling & Tong (2005); and the Wald test in Hansen (1996) and Caner & Hansen (2001). All the existing references in this literature only consider either a pure autoregressive or a pure moving average model under the null hypothesis. However, it is well known that the autoregressive moving average model is superior to the pure autoregressive model from the viewpoint of parsimony, and superior to the pure moving average model from the viewpoint of interpretation. We propose a quasilikelihood ratio test for the autoregressive moving average model against its threshold extension.

In these tests, the threshold parameter is usually assumed to be unknown under the alternative hypothesis, and is absent under the null hypothesis. As a result, the null distributions of likelihood-based tests take on a very complicated form even for some simple cases, see Davies (1977, 1987). Methods based on the Poisson clumping heuristic or a stationary Ornstein–Uhlenbeck process had been developed to approximate the tail of the null distributions, and some commonly used critical values were tabulated for use in applications, see Chan & Tong (1990) and Chan (1991). These can be considered as trade-offs between flexibility and computation time. Contemporary computing power gives us a chance to provide a more flexible solution. Bootstrap approximations are generally powerful, and can maintain significance levels at the same time. Ling & Tong (2005) suggested a simulation method to calculate the p-values, of the quasilikelihood ratio test for the threshold moving average model, which is actually a parametric bootstrap method. A huge amount of computation may be involved in this approach, when a complicated model is considered, due to the optimization step for each bootstrapped sample. The classical block-wise bootstrap method

(Chan et al., 2009) or the wild bootstrap method (Gonzalo & Wolf, 2005) encounter the same problem. Hansen (1996) considered a bootstrap method based on stochastic permutation, to approximate the p-values of a Wald test. There is no optimization involved in the bootstrapped samples in this method, and this is important especially for our cases since a large amount of computation is required for the optimization of threshold autoregressive moving average models. However, the test statistic of this method is required to have a quadratic form, see also Zhu & Ng (2003). In this paper, we derive an asymptotically equivalent quadratic form for the quasilikelihood ratio test statistic, and then propose a stochastic permutation-based bootstrap method to approximate the null distribution.

It is usually necessary to assume that the error term is identically and independently distributed in tests for the threshold structure, see the simulation experiments in Wong & Li (1997) and Li & Li (2008). However this condition is too strong in practice. For example, many time series in finance and economics may have a time varying conditional variance (Engle, 1982; Bollerslev, 1986). In order to handle this case, a generalized autoregressive conditional heteroscedastic structure is usually assumed for the errors (Wong & Li, 1997; Li & Li, 2008), and these extra parameters may burden the testing procedure. This paper derives the null distribution of the test for threshold autoregressive moving average models without assuming identically and independently distributed errors. The proposed bootstrap method is robust with respect to this assumption by persevering the unknown structure of errors in the process of approximating the null distribution. Hence, the possible dependence structure in the errors becomes irrelevant in establishing the p-value of the test statistic.

2. QUASILIKELIHOOD RATIO TEST

Let $\{y_t\}$ be a strictly stationary and ergodic time series generated by the ARMA(p, q) model,

$$y_t = \mu + \sum_{i=1}^{p} \phi_{1i} y_{t-i} + \sum_{j=1}^{q} \phi_{2j} e_{t-j} + e_t, \tag{1}$$

where p and q are known positive integers, and $\{e_t\}$ is an uncorrelated error sequence with zero mean. It is of interest to check whether or not the model (1) can provide an adequate fit for the real data with a threshold model as the alternative, i.e. whether the sequence $\{y_t\}$ is generated by

$$y_t = \mu + \sum_{i=1}^{p} \phi_{1i} y_{t-i} + \sum_{j=1}^{q} \phi_{2j} e_{t-j} + \left(\mu_1 + \sum_{i=1}^{p} \psi_{1i} y_{t-i} + \sum_{j=1}^{q} \psi_{2j} e_{t-j} \right) I(y_{t-d} \leqslant r) + e_t, \tag{2}$$

where the delay parameter d is a known positive integer, $I(\cdot)$ is the indicator function, and $r \in \mathbb{R}$ is the unknown threshold parameter.

Let $\phi = (\mu, \phi_{11}, \ldots, \phi_{1p}, \phi_{21}, \ldots, \phi_{2q})'$, $\psi = (\mu_1, \psi_{11}, \ldots, \psi_{1p}, \psi_{21}, \ldots, \psi_{2q})'$, and $\lambda = (\phi', \psi')'$, where λ is the parameter vector of model (2). Denote the parameter space by $\Theta = \Theta_\phi \times \Theta_\psi$, where Θ_ϕ and Θ_ψ are compact subsets of $\mathbb{R}^{p+q+1}$. Suppose the true parameter vector $\lambda_0 = (\phi_0', \psi_0')'$ is an interior point of the parameter space Θ.

Given observations $y_1, \ldots, y_n$, we consider the following hypotheses,

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

By temporarily assuming normality for e_t, we have the loglikelihood functions, conditional on $y_0, y_{-1}, \ldots$, respectively, under H_0 and H_1 as follows,

$$L_{0n}(\phi) = \sum_{t=1}^{n} \{e_t(\phi)\}^2, \quad L_{1n}(\lambda, r) = \sum_{t=1}^{n} \{e_t(\lambda, r)\}^2, \tag{3}$$

where $e_t(\phi) = e_t(\lambda, -\infty)$ and $e_t(\lambda, r)$ is defined based on the iterative equation (2). The likelihood functions (3) are dependent on past observations infinitely far away, and then initial values are needed. For simplicity, we assume that $y_i = 0$, $i \leqslant 0$ and these functions evaluated at these initial values can be denoted, respectively, by $\widetilde{e}_t(\phi)$, $\widetilde{e}_t(\lambda, r)$, $\widetilde{L}_{0n}(\phi)$ and $\widetilde{L}_{1n}(\lambda, r)$.

For a given r, let $\widetilde{\phi}_n = \text{argmin}_{\phi\in\Theta_\phi} \widetilde{L}_{0n}(\phi)$ and $\widetilde{\lambda}_n(r) = \text{argmin}_{\lambda\in\Theta} \widetilde{L}_{1n}(\lambda, r)$. The quasilikelihood ratio test statistic is then $\widetilde{L}R_n(r) = \widetilde{L}_{0n}(\widetilde{\phi}_n) - \widetilde{L}_{1n}(\widetilde{\lambda}_n(r), r)$. Since r is unknown and the quantity $\sup_{r\in\mathbb{R}} \widetilde{L}R_n(r)$ will diverge to infinity in probability as $n \to \infty$ (Andrews, 1993), the quasilikelihood ratio test statistic in this paper is defined as

$$LR_n = \frac{1}{\widetilde{\sigma}_e^2} \sup_{r\in[a,b]} \widetilde{L}R_n(r),$$

where $\widetilde{\sigma}_e^2 = \widetilde{L}_{0n}(\widetilde{\phi}_n)/n$, and $[a, b]$ is a predetermined interval.

Denote $\sigma_e^2 = E(e_t^2)$,

$$K_{rs} = E\left\{e_t^2 \frac{\partial e_t(\lambda_0, r)}{\partial\lambda} \frac{\partial e_t(\lambda_0, s)}{\partial\lambda'}\right\}, \quad \Omega_r = \begin{pmatrix} \Sigma & \Sigma_{1r} \\ \Sigma'_{1r} & \Sigma_{rr} \end{pmatrix} = E\left\{\frac{\partial e_t(\lambda_0, r)}{\partial\lambda} \frac{\partial e_t(\lambda_0, r)}{\partial\lambda'}\right\},$$

and $\Omega_{1r} = \Omega_r^{-1} - \text{diag}(\Sigma^{-1}, 0) = (-\Sigma'_{1r}\Sigma^{-1}, \text{I})'(\Sigma_{rr} - \Sigma'_{1r}\Sigma^{-1}\Sigma_{1r})^{-1}(-\Sigma'_{1r}\Sigma^{-1}, \text{I})$, where Σ, Σ_{1r}, Σ_{rr}, 0 and I are $(p + q + 1) \times (p + q + 1)$ matrices, 0 is a zero matrix and I is an identity matrix. Let $\{G_{2(p+q+1)}(r), r \in \mathbb{R}\}$ be a $2(p + q + 1)$-dimensional vector Gaussian process with zero mean and covariance kernel K_{rs}; almost all its paths are continuous. To investigate the asymptotic power of LR_n, we also consider the following local alternatives, $H_{1n} : \psi_0 = n^{-1/2}h$ for a constant vector $h \in \mathbb{R}^{p+q+1}$ and $r = r_0 \in \mathbb{R}$ is a fixed value.

THEOREM 1. *Under H_0, if Assumptions* A1–A4 *in the Appendix hold and $n \to \infty$, then in distribution,*

$$LR_n \longrightarrow \frac{1}{\sigma_e^2} \sup_{r\in[a,b]} \{G'_{2(p+q+1)}(r)\Omega_{1r}G_{2(p+q+1)}(r)\}.$$

THEOREM 2. *Under H_{1n}, if Assumptions* A1–A5 *in the Appendix hold and $n \to \infty$, then in distribution,*

$$LR_n \longrightarrow \frac{1}{\sigma_e^2} \sup_{r\in[a,b]} [\{G_{2(p+q+1)}(r) + \mu(r)\}'\Omega_{1r}\{G_{2(p+q+1)}(r) + \mu(r)\}],$$

where $\mu(r) = K_{rr_0}h$.

The proofs of the above theorems are similar to those of Theorems 2.2 and 3.2 in Li & Li (2008), and are omitted.

The values of the matrix K_{rs} depend on the possible dependence structure of e_t, and this is why the existing approximation methods fail to work when the independence assumption is broken. It is therefore necessary to develop a new approximation method, and hence the bootstrap method in §3 is proposed.

3. BOOTSTRAP APPROXIMATION BY STOCHASTIC PERMUTATION

Let $T_n(r) = n^{-1/2} \sum_{t=1}^n e_t \partial e_t(\lambda_0, r)/\partial\lambda$. Under H_0 and Assumptions A1–A3 in the Appendix, we can show that

$$\sup_{r\in[a,b]} |\widetilde{L}R_n(r) - T'_n(r)\Omega_{1r}T_n(r)| = o_p(1); \tag{4}$$

see also the proof of Lemma 2.1 in Ling & Tong (2005). The quantity $T'_n(r)\Omega_{1r}T_n(r)$ is a quadratic form, so we can consider a bootstrap method to approximate it by permutating stochastically the summand in $T_n(r)$. The uniform expansion in (4) makes sure that we can equivalently handle $T'_n(r)\Omega_{1r}T_n(r)$, so we can avoid time-consuming optimization in searching for the quasimaximum likelihood estimates for each bootstrapped sample in the traditional methods.

We first consider removing any possible threshold structure in a candidate time series $\{y_t\}$ since the uniform expansion (4) and the null distribution both depend on H_0. Let $\widetilde{r}_n = \text{argmin}_{r\in[a,b]} \widetilde{L}_{1n}\{\widetilde{\lambda}_n(r), r\}$, and $\widetilde{\lambda}_n = (\hat{\phi}_n', \hat{\psi}_n')' = \widetilde{\lambda}_n(\widetilde{r}_n)$, where $\hat{\phi}_n' = (\hat{\mu}, \hat{\phi}_{11}, \ldots, \hat{\phi}_{2q})'$ and $\widetilde{\lambda}_n(r)$ is defined in §2. Let $\widetilde{e}_t = \widetilde{e}_t(\widetilde{\lambda}_n, \widetilde{r}_n)$ and $\widetilde{y}_t = \hat{\mu} + \sum_{i=1}^p \hat{\phi}_{1i}\widetilde{y}_{t-i} + \sum_{j=1}^q \hat{\phi}_{2j}\widetilde{e}_{t-j} + \widetilde{e}_t$, where $\widetilde{e}_t = 0$ for $t \leqslant 0$. Note that $\{\widetilde{y}_t\}$ is a time series generated by a linear autoregressive moving average model with parameters $\hat{\phi}_n$ and innovations $\{\widetilde{e}_t\}$.

Define the vector functions

$$\frac{\partial \widetilde{e}_t(r)}{\partial \phi} = -\widetilde{z}_t - \sum_{j=1}^q \hat{\phi}_{2j}\frac{\partial \widetilde{e}_{t-j}}{\partial \phi}, \quad \frac{\partial \widetilde{e}_t(r)}{\partial \psi} = -\widetilde{z}_t I(\widetilde{y}_{t-d} \leqslant r) - \sum_{j=1}^q \hat{\phi}_{2j}\frac{\partial \widetilde{e}_{t-j}(r)}{\partial \psi},$$

and $\partial \widetilde{e}_t(r)/\partial\lambda = (\partial \widetilde{e}_t(r)/\partial\phi', \partial \widetilde{e}_t(r)/\partial\psi')'$, where $\widetilde{z}_t = (1, \widetilde{y}_{t-1}, \ldots, \widetilde{y}_{t-p}, \widetilde{e}_{t-1}, \ldots, \widetilde{e}_{t-q})'$ and $\partial \widetilde{e}_t(r)/\partial\lambda = 0$ for $t \leqslant 0$. Let

$$\widetilde{\Omega}_r = \begin{pmatrix} \widetilde{\Sigma} & \widetilde{\Sigma}_{1r} \\ \widetilde{\Sigma}_{1r}' & \widetilde{\Sigma}_{rr} \end{pmatrix} = \frac{1}{n}\sum_{t=1}^n \frac{\partial \widetilde{e}_t(r)}{\partial\lambda}\frac{\partial \widetilde{e}_t(r)}{\partial\lambda'}, \quad \widetilde{\Omega}_{1r} = \widetilde{\Omega}_r^{-1} - \text{diag}\left\{\widetilde{\Sigma}^{-1}, 0\right\}.$$

Suppose $\{\varepsilon_t^*\}$ is an extra identically and independently distributed sequence with zero mean, variance unity and finite fourth moment. Let $T_n(\varepsilon^*, r) = n^{-1/2}\sum_{t=1}^n \varepsilon_t^* \widetilde{e}_t \partial \widetilde{e}_t(r)/\partial\lambda$, and

$$LR_n(\varepsilon^*) = \frac{1}{\hat{\sigma}_e^2} \sup_{r\in[a,b]} T_n'(\varepsilon^*, r)\widetilde{\Omega}_{1r} T_n(\varepsilon^*, r),$$

where $\hat{\sigma}_e^2 = \widetilde{L}_{1n}(\widetilde{\lambda}_n, \widetilde{r}_n)/n$.

THEOREM 3. *Under H_0 or H_1, if Assumptions* A1–A4 *in the Appendix hold, then, conditional on $y_1, \ldots, y_n$,*

$$LR_n(\varepsilon^*) \longrightarrow \mathcal{L}\frac{1}{\sigma_e^2} \sup_{r\in[a,b]} \{G_{2(p+q+1)}'(r)\Omega_{1r} G_{2(p+q+1)}(r)\}$$

in probability, where $\to \mathcal{L}$ means the convergence in distribution, and $\{G_{2(p+q+1)}(r), r \in R\}$ and Ω_{1r} are defined as in Theorem 1.

The proof of Theorem 3 is in the Appendix. The conditional asymptotic distribution in Theorem 3 is the same as the unconditional one in Theorem 1, the bootstrap method can be used to approximate the p-values of LR_n.

Note that $T_n'(\varepsilon^*, r)\widetilde{\Omega}_{1r} T_n(\varepsilon^*, r)$ is a stepwise function with possible jumps at $\widetilde{y}_1, \ldots, \widetilde{y}_n$, so the amount of computation depends only on the sample size n and the number of bootstrapped samples. This greatly reduces the computational burden for more sophisticated models in applications. In practice, we need to know the interval $[a, b]$ before performing the test, and the values of a and b can be set to empirical quantiles as in Chan (1991).

4. SIMULATION EXPERIMENTS

We conducted two simulation experiments to check the performance of the proposed testing procedure. In each experiment, the sample size is set to 200, the number of bootstrapped samples is 1000, the number of replications is 1000, and the significance level is 0·05.

The first experiment is conducted to compare the effectiveness of three different permutating distributions: (i) the Rademacher distribution, which takes values ± 1 with probability 0·5, (ii) the uniform distribution on $[-\surd 3, \surd 3]$ and (iii) the standard normal distribution. The data generating process is

$$y_t = 0{\cdot}7y_{t-1} + 0{\cdot}6e_{t-1} - \psi(0{\cdot}7y_{t-1} + 0{\cdot}6e_{t-1})I(y_{t-1} \leqslant 0) + e_t,$$

where $e_t = (0{\cdot}5 + 0{\cdot}5e_{t-1}^2)^{1/2}\varepsilon_t$, $\{\varepsilon_t\}$ are identically and independently distributed random variables with the standard normal distribution, $\psi = 0$ corresponds to the size and $\psi \neq 0$ corresponds to the power. We performed the test with $p = q = 1$, and the parameters μ and μ_1 are suppressed to reduce the computation in searching for the quasimaximum likelihood estimates of the autoregressive moving average model and

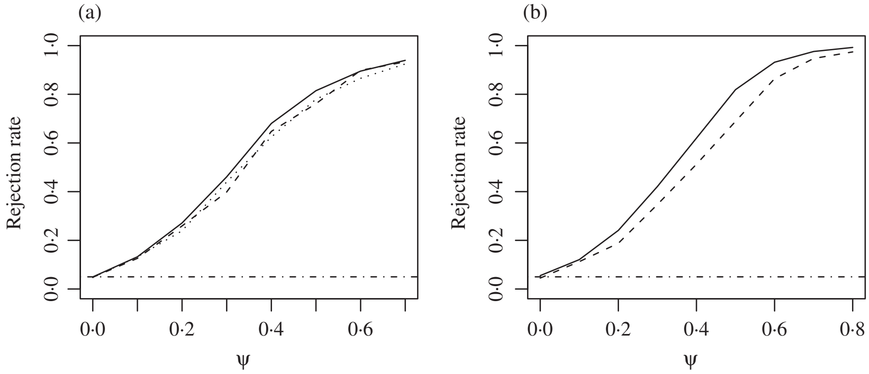

Fig. 1. The rejection rates of our test with different departures of values of ψ from zero. The horizontal dot-dashed line corresponds to the nominal rate of 0·05. (a) Rademacher (solid), uniform (dashed), normal (dotted). (b) $p = 2$ (solid), AIC (dashed).

its threshold extension. The value of a, or b, for the interval $[a, b]$ is set to be the empirical 0·2, or 0·8, quantile of each replication, and the Newton–Raphson algorithm was employed to perform all the optimizations. Fig. 1(a) presents the rejection rates of our test with different values of ψ. It can be seen that all sizes are very close to the nominal rate 0·05, and the test using the Rademacher distribution is slightly more powerful.

The second experiment is considered to assess the performance of the test when an information criterion is employed to select the orders of autoregressive moving average models. The null distribution of LR_n heavily depends on the orders, see Chan (1991). However, the bootstrap test is adaptable to the selection of orders, and is hence supposed to maintain the sizes. We generate the samples in this experiment by

$$y_t = 0{\cdot}1y_{t-1} + 0{\cdot}1y_{t-2} - \psi(y_{t-1} + y_{t-2})I(y_{t-1} \leqslant 0) + e_t,$$

where e_t and the interval $[a, b]$ are the same as those in the first experiment. The test was performed with $q = 0$, and the parameters μ and μ_1 are suppressed again. We consider two methods to select the order p: (i) the test advocated in this paper with $p = 2$, and (ii) the test with order $1 \leqslant p \leqslant 10$ selected by the AIC. Figure 1(b) presents the rejection rates of the test permuted by the Rademacher distribution. It can be seen that the sizes are both close to 0·05, and the test based on the AIC is less powerful. The results for the tests permuted by the normal distribution or the uniform distribution are similar, and are omitted to save space.

5. An example

We consider the centred log-return as a percentage of the monthly exchange rate of Japanese Yen against USA dollar from January 1972 to December 2003, and denote these observations by $y_1, \ldots, y_{384}$. A linear model is first considered to fit the data. At the 0·05 significance level, the sample autocorrelation function of $\{y_t\}$ is significant at lags 1 and 11, and the sample partial autocorrelation function is significant at lags 1, 8 and 11. Hence, we first try an ARMA (12,1) model, then remove the most insignificant coefficients one by one, and finally reach the model

$$y_t = 0{\cdot}116_{0{\cdot}051}y_{t-11} + 0{\cdot}376_{0{\cdot}047}e_{t-1} + e_t,$$

where e_t has zero mean and variance 1·297, the value of the AIC is 1159·45, and the values 0·051 and 0·047 correspond to the standard errors. The p-values of Ljung–Box test statistics Q(M) at lags M = 6, 12, and 18 are, respectively, 0·64, 0·67 and 0·51. We also tried other ARMA(p, 1) models, however, the p-values of Q(18) are all less than 0·2 as $p < 11$.

The fitted sparse ARMA (11,1) should provide a good fit from the viewpoint of linear time series modelling, and we next try its threshold version to $\{y_t\}$ with $d = 1$. The interval (a, b) is set to be $(-1{\cdot}59, 1{\cdot}36)$, the empirical 0·1 and 0·9 quantiles of $\{y_t\}$, and the fitted model is

$$y_t = 0{\cdot}158_{0{\cdot}057} y_{t-11} + 0{\cdot}293_{0{\cdot}061} e_{t-1} - (0{\cdot}263_{0{\cdot}145} y_{t-11} - 0{\cdot}333_{0{\cdot}096} e_{t-1}) I(y_{t-1} \leqslant -1{\cdot}394_{0{\cdot}606}) + e_t,$$

where e_t has zero mean and variance 1·253, and the value of the AIC is 1152·53. The threshold model is preferred by the AIC although the coefficient −0·263 is insignificant at the 0·05 significance level. We next employ our test to check whether the linear model can provide a better fit. The number of bootstrapped samples is set to 10 000, and the Rademacher distribution is employed for the stochastic permutation. The calculated p-value is 0·017, and hence the fitted threshold model should provide a better fit to $\{y_t\}$ compared with the linear one.

ACKNOWLEDGEMENT

The authors thank the editor and two referees for comments that led to the substantial improvement of this paper, and the Hong Kong Research Grant Council for partial support.

APPENDIX

Technical details

In this paper, we assume the error term $e_t = \varepsilon_t \sigma_t$, where $\{\varepsilon_t\}$ are identically and independently distributed random variables with zero mean and variance unity, and $\sigma_t > 0$ is a measurable function with respect to the information set $\mathcal{F}_t = \sigma(\varepsilon_t, \varepsilon_{t-1}, \ldots)$.

Assumption A1. All roots of the polynomials $1 - \phi_{11}x - \cdots - \phi_{1p}x^p$ and $1 + \phi_{21}x + \cdots + \phi_{2q}x^q$ are outside the unit circle, and these two polynomials are coprime. The polynomials $1 - \psi_{11}x - \cdots - \psi_{1p}x^p$ and $1 + \psi_{21}x + \cdots + \psi_{2q}x^q$ are also coprime. It holds that $\sum_{i=1}^{q} |\phi_{2i}| < 1$ and $\sum_{i=1}^{q} |\phi_{2i} + \psi_{2i}| < 1$.

Assumption A2. The random variable ε_t has a continuous and positive density function $f(\cdot)$ on $\mathbb{R}$ with $E(\varepsilon_t^4) < \infty$ and $\sup_{x\in\mathbb{R}} x^4 f(x) < \infty$.

Assumption A3. The stochastic process $\{\sigma_t^2\}$ is strictly stationary and ergodic with $E(\sigma_t^4) < \infty$, and is bounded away from zero with probability one, i.e., there exists an $\eta > 0$ such that $\text{pr}(\sigma_t > \eta) = 1$.

Assumption A4. The following inequality holds

$$E\left\{e_s^4 \prod_{i=1}^{l} I(r_1 < y_{t_i} \leqslant r)\right\} \leqslant C(r - r_1)^l,$$

where $l = 1, \ldots, 4$, and s, t_1, t_2, t_3 and t_4 are different integers, $r_1 < r$, with $r_1, r \in [a, b]$, C is a constant independent of r and r_1.

Assumption A5. Suppose the density f of ε_t is absolutely continuous with derivative f' almost everywhere and $\int \{f'(x)/f(x)\}^2 f(x)dx < \infty$.

Assumptions A3 and A4 are guaranteed by Lemma A.2 of Li & Li (2008) when $\{e_t\}$ is an identically and independently distributed process or the generalized autoregressive conditional heteroscedastic process (Bollerslev, 1986).

Proof of Theorem 3. Suppose $\{y_t\}$ is generated by model (2) with the parameter vector $\lambda_1 = (\phi_0', \psi_0')'$ and the threshold parameter r_0. Let $\lambda_0 = (\phi_0', 0')'$, where 0 is a $(p + q + 1)$-dimensional zero vector. By a method similar to that of §6 in Ling & Tong (2005), we can show that $\widetilde{\lambda}_n = \lambda_1 + o_p(1)$ and, if $\psi_0 \neq 0$, $\widetilde{r}_n = r_0 + o_p(1)$. It can be further shown that $\sup_{r\in[a,b]} |\widetilde{\Omega}_r - \Omega_r| = o_p(1)$ and $\sup_{r,s\in[a,b]} |\widetilde{K}_{rs} - K_{rs}| = o_p(1)$, where $\widetilde{K}_{rs} = n^{-1} \sum_{t=1}^{n} \widetilde{e}_t^2 (\partial \widetilde{e}_t(r)/\partial\lambda)(\partial \widetilde{e}_t(s)/\partial\lambda')$. Hence, it is sufficient to show the tightness of $T_n^*(r) = n^{-1/2} \sum_{t=1}^{n} \varepsilon_t^* \widetilde{e}_t \partial \widetilde{e}_t(r)/\partial\psi$, conditional on $y_1, \ldots, y_n$ or $\mathcal{F}_n = \sigma(\varepsilon_n, \varepsilon_{n-1}, \cdots)$.

Denote

$$\Gamma(r) = E\left\{e_t^2 \sum_{j=0}^{\infty} \rho^j z_{t-j}^2 I(y_{t-d-j} \leqslant r)\right\}, \quad \Gamma_n(r) = \frac{1}{n}\sum_{t=1}^{n} \widetilde{e}_t^2 \sum_{j=0}^{\infty} \rho^j \widetilde{z}_{t-j}^2 I(\widetilde{y}_{t-d-j} \leqslant r),$$

where $z_t = (1, y_{t-1}, \ldots, y_{t-p}, e_{t-1}, \ldots, e_{t-q})'$ and ρ is defined as that in Lemma A.1 of Li & Li (2008). For simplicity and without loss of generality, we will treat the vectors z_t and $\widetilde{z}_t$ as scalars. It holds that, for $r_1 < r$, $n^{-1}\sum_{t=1}^{n}[\widetilde{e}_t\{\partial\widetilde{e}_t(r)/\partial\psi - \partial\widetilde{e}_t(r_1)/\partial\psi\}]^2 \leqslant C_1[\Gamma_n(r) - \Gamma_n(r_1)]$, where C_1 is a constant. Then, by Burkholder's inequality (Hall & Heyde, 1980, p. 23), we can show that

$$E\left\{|T_n^*(r) - T_n^*(r_1)|^4 | \mathcal{F}_n\right\} \leqslant C_2\left\{\Gamma_n(r) - \Gamma_n(r_1)\right\}^2, \tag{A1}$$

where C_2 is a constant.

Note that $T_n^*(r)$ is a step-wise function with possible jump points at $\{\widetilde{y}_1, \ldots, \widetilde{y}_n\}$. For any interval $[r_1, r_1 + \delta] \subset [a, b]$, suppose that it can be divided into M parts by these jump points, say $r_1 = a_0 < a_1 < \cdots < a_M = r_1 + \delta$. For any $\eta > 0$, by (A1) and Theorem 10.2 of Billingsley (1999),

$$\begin{aligned} \mathrm{pr}\left\{\sup_{r_1 \leqslant r \leqslant r_1+\delta} |T_n^*(r) - T_n^*(r_1)| > \eta | \mathcal{F}_n\right\} &= \mathrm{pr}\left\{\max_{1\leqslant j\leqslant M} |\sum_{i=1}^{j} T_n^*(a_i) - T_n^*(a_{i-1})| > \eta | \mathcal{F}_n\right\} \\ &\leqslant \frac{C_3 C_2}{\eta^4}\left\{\Gamma_n(r_1+\delta) - \Gamma_n(r_1)\right\}^2, \end{aligned} \tag{A2}$$

where C_3 is a constant.

Note that $\Gamma(r)$ is a continuous function on $[a, b]$, and $\sup_{r\in[a,b]}|\Gamma_n(r) - \Gamma(r)| = o_p(1)$. Then, for any $\epsilon > 0$, there exist a $\delta_0 > 0$ such that $\sup_{r\in[a,b-\delta_0]}|\Gamma_n(r+\delta_0) - \Gamma_n(r)| < \epsilon$ holds for almost every sample $y_1, \ldots, y_n$ generated by (2), limited by convergence in probability. Let $\delta = \max\{C_3C_2\epsilon/\eta^4, \delta_0\}$. By (A2),

$$\mathrm{pr}\left\{\sup_{r_1\leqslant r\leqslant r_1+\delta} |T_n^*(r) - T_n^*(r_1)| > \eta | \mathcal{F}_n\right\} \leqslant \delta\epsilon.$$

As in Ling & Tong (2005) and Li & Li (2008), we can claim that $\{T_n^*(r), r \in \mathbb{R}_\gamma\}$ is tight, where $\mathbb{R}_\gamma = [a, b]$ is equipped with the corresponding product Shorohod topology (Billingsley, 1999). □

References

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

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

Bollerslev, T. (1986). Generalized autoregression conditional heteroscedasticity. *J. Economet.* **31**, 307–27.

Caner, M. & Hansen, B. E. (2001). Threshold autoregression with a unit root. *Econometrica* **69**, 1555–96.

Chan, K. S. (1990). Testing for threshold autoregression. *Ann. Statist.* **18**, 1886–94.

Chan, K. S. (1991). Percentage points of likelihood ratio tests for threshold autoregression. *J. R. Statist. Soc. B* **53**, 691–6.

Chan, K. S. & Tong, H. (1990). On likelihood ratio tests for threshold autoregression. *J. R. Statist. Soc. B* **52**, 469–76.

Chan, K. S., Tong, H. & Stenseth, N. C. (2009). Analyzing short time series data from periodically fluctuating rodent populations by threshold models: a nearest block bootstrap approach. *Sci. China Series A* **52**, 1085–112.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64**, 247–54.

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.

Gonzaloa, J. & Wolf, M. (2005). Subsampling inference in threshold autoregressive models. *J. Economet.* **127**, 201–24.

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

HANSEN, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* **64**, 413–30.

LI, G. & LI, W. K. (2008). Testing for threshold moving average with conditional heteroscedasticity. *Statist. Sinica* **18**, 647–65.

LING, S. & TONG, H. (2005). Testing for a linear MA model against threshold MA models. *Ann. Statist.* **33**, 2529–52.

PETRUCCELLI, J. & DAVIES, N. (1986). A portmanteau test for self-exciting threshold autoregressive-type nonlinearity in time series. *Biometrika* **73**, 687–94.

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

TONG, H. & LIM, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *J. R. Statist. Soc. B* **42**, 245–92.

TSAY, R. S. (1998). Testing and modeling multivariate threshold models. *J. Am. Statist. Assoc.* **93**, 1188–202.

WONG, C. S. & LI, W. K. (1997). Testing for threshold autoregression with conditional heteroscedasticity. *Biometrika* **84**, 407–18.

ZHU, L. X. & NG, K. W. (2003). Checking the adequacy of a partial linear model. *Statist. Sinica* **13**, 763–81.

[*Received February* 2010. *Revised September* 2010]