Title	Testing for double threshold autoregressive conditional heteroscedastic model
Author(s)	Wong, CS; Li, WK
Citation	Statistica Sinica, 2000, v. 10 n. 1, p. 173-189
Issued Date	2000
URL	http://hdl.handle.net/10722/45348
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Statistica Sinica **10**(2000), 173-189

TESTING FOR DOUBLE THRESHOLD AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTIC MODEL

C. S. Wong and W. K. Li

The University of Hong Kong

Abstract: The testing problem for the hypothesis of linearity against the double threshold autoregressive conditional heteroscedastic model is addressed. The problem is nonstandard as the threshold parameter is a nuisance parameter which is absent under the null hypothesis. We will show that the asymptotic null distribution of the Lagrange-multiplier test statistic is a functional of a zero-mean Gaussian process. In some cases, we give the upper percentage points of the test statistic. The performance of the test statistic is illustrated by extensive simulation experiments and an example.

Key words and phrases: Conditional heteroscedasticity, Gaussian process, Lagrange-multiplier test, threshold time series model.

1. Introduction

The threshold principle was first introduced by Tong (1978) to generalize the linear autoregressions (AR models) to the self-exciting threshold autoregressive (SETAR) models. Tong and Lim (1980) show that these models are capable of capturing various nonlinear phenomena, such as asymmetric cycles, jump resonance and amplitude-frequency dependence. Tong (1990) gave a comprehensive review of these models. One of the most popular models for time series with changing conditional variance is Engle's (1982) autoregressive conditional heteroscedastic (ARCH) model. Ignoring the ARCH effect in time series can lead to inefficient estimates and suboptimal statistical inferences, see Bollerslev, Chou and Kroner (1992).

Recently, the two ideas have been used in combination to form new classes of nonlinear time series models. Li and Lam (1995) and Wong and Li (1997) considered the threshold ARCH (SETAR-ARCH) models consisting of a piecewise conditional mean and an ARCH type conditional variance. Li and Li (1996) and Liu, Li and Li (1997) considered the double threshold autoregressive heteroscedastic (DTARCH) time series models consisting of a piecewise conditional mean as well as a piecewise ARCH description of the changing conditional variance. Rabemananjara and Zakoian (1993) consider a piecewise changing conditional standard deviation model but their specification is different from those considered by other researchers. Despite the difference in the specifications, these models have been shown to be useful in the modelling of financial time series.

It can be expected that the model identification, estimation, diagnostics checking and testing are much more complicated for these classes of nonlinear time series models. Li and Lam (1995), Li and Li (1996) and Liu, Li and Li (1997) tried to test the hypothesis of linearity. In these papers, model specification testing was performed as if the threshold parameter was a known constant. Clearly, this may not be a realistic assumption. However the nonlinearity test without assuming known threshold parameter is a nonstandard one, as the threshold parameter is a nuisance parameter which is absent under the null hypothesis of linearity (Davis (1977, 1987)).

Testing a SETAR-ARCH model against the null of an AR-ARCH model, without assuming known threshold parameter, was considered by Wong and Li (1997). They used a Lagrange-multiplier test approach to solve the problem. In this paper, we will extend the Lagrange-multiplier approach to the testing of linearity for the DTARCH models. We will show that the asymptotic null distribution of the Lagrange-multiplier test statistic is a functional of a Gaussian process. The main advantage of the Lagrange-multiplier statistic is that only the AR-ARCH model, which is the model under the null hypothesis of linearity, has to be estimated. Hence this approach greatly reduces the computation burden as compared to the likelihood ratio test approach, which requires the estimation of the full models with each possible value of the threshold parameter.

The organization of this paper is as follows. Section 2 introduces the set-up of the test, states the main result, and discusses the method used to tabulate the asymptotic null distribution. Section 3 gives the approximate upper percentage points of the test statistic under two different cases. Section 4 reports the results of some extensive simulation studies and an example of applying the Lagrange-multiplier test. The outline of a proof of the main result is given in Section 5. For the complete proof, see Wong (1998). Section 6 gives the proofs and justifications of the propositions in Section 4.

2. Lagrange-Multiplier Test

Denote the indicator function by $I(\cdot)$ and $I(X_t \le r)$ by $I_r(X_t)$. Let $\mathcal{F}_t$ be the information set up to time t. The double-threshold autoregressive conditional heteroscedastic model under consideration may be defined by

$$X_t-\theta_0-\theta_1X_{t-1}-\cdots-\theta_{p_1}X_{t-p_1}-I(X_{t-d}\le r)(\phi_0+\phi_1X_{t-1}+\cdots+\phi_{p_2}X_{t-p_2})=\varepsilon_t, \tag{2.1}$$

$$\varepsilon_t \sim N(0,h_t), \tag{2.2}$$

$$h_t=E(\varepsilon_t^2|\mathcal{F}_{t-1})=\alpha_0+\sum_{i=1}^{q_1}\alpha_i\varepsilon_{t-i}^2+I(X_{t-d}\le r)\left(\beta_0+\sum_{j=1}^{q_2}\beta_j\varepsilon_{t-j}^2\right), \tag{2.3}$$

where r is the threshold parameter and d is the delay parameter. The non-negative integers p_1, p_2, q_1, q_2 and d are assumed known, and such that $0 \le p_2 \le p_1$, $0 \le q_2 \le q_1$ and $1 \le d$. It is also assumed that the threshold parameter r belongs to a known bounded subset $\tilde{R}$ of R. In general, $\tilde{R}$ is a finite interval. The roots of the characteristic equation $x^{p_1} - \theta_1 x^{p_1 - 1} - \cdots - \theta_{p_1} = 0$ are assumed to lie inside the unit circle. Also, we have $\infty > \alpha_0 > 0$; $\infty > \alpha_0 + \beta_0 > 0$; $\alpha_i \ge 0$, $i = 1, \ldots, q_1$; $\alpha_i + \beta_i \ge 0$, $i = 1, \ldots, q_2$; $\alpha_1 + \cdots + \alpha_{q_1} + \beta_1 + \cdots + \beta_{q_2} < 1$. We further assume that $E(\varepsilon_t^2) < \infty$ and $E(\varepsilon_t^4) < \infty$. The process $\{X_t\}$ is assumed to be α-mixing with exponentially decreasing rate, i.e., there is a sequence of positive numbers $\{a(m)\}$, convergent to zero, such that, for any two bounded mappings $A : R^{k+1} \to R$ and $M : R^{l+1} \to R$, $|E\{A(X_t, \ldots, X_{t-k}) M(X_{t+m}, \ldots, X_{t+m+l})\} - E\{A(\cdot)\}E\{M(\cdot)\}| \le a(m)$, see Davidson and MacKinnon (1993).

Given observations $X_1, \ldots, X_N$, consider testing the null hypothesis H_0 : $\phi_0 = \phi_1 = \cdots = \phi_{p_2} = 0$ and $\beta_0 = \beta_1 = \cdots = \beta_{q_2} = 0$. Under H_0, the nuisance parameter r is absent. The conditional log likelihood is

$$l = \sum_t l_t = \sum_t \left(-\frac{1}{2} \log h_t - \frac{1}{2} \frac{\varepsilon_t^2}{h_t} \right)$$

where the summation is from $n_0 = \max\{p_1, q_1, d\} + 1$ to N. Let $n = N - n_0 + 1$, $\theta = (\theta_0, \ldots, \theta_{p_1})'$, $\phi = (\phi_0, \ldots, \phi_{p_2})'$, $\alpha = (\alpha_0, \ldots, \alpha_{q_1})'$ and $\beta = (\beta_0, \ldots, \beta_{q_2})'$, where $'$ denotes the transpose of a vector or matrix. It is easy to obtain the score functions and the expectation of the second derivatives of l with respect to θ, ϕ, α and β, and hence they are omitted. By Theorem 4 of Engle (1982), we have $E\{\partial^2 l/(\partial\theta\, \partial\alpha')\} = E\{\partial^2 l/(\partial\theta\, \partial\beta')\} = E\{\partial^2 l/(\partial\phi\, \partial\alpha')\} = E\{\partial^2 l/(\partial\phi\, \partial\beta')\} = 0$. This means that the information matrix of the score function is block diagonal. Hence following Davies (1977, 1987), the Lagrange-Multiplier test statistic under consideration is

$$S = \sup_{r \in \tilde{R}} \left\{ T_{1r}'(C_{1r} - L_{1r}' C_1^{-1} L_{1r})^{-1} T_{1r} + T_{2r}'(C_{2r} - L_{2r}' C_2^{-1} L_{2r})^{-1} T_{2r} \right\}, \quad (2.4)$$

where

$$T_{1r} = n^{-1/2} \left. \frac{\partial l}{\partial \phi} \right|_{\hat{\theta}, \hat{\alpha}, \phi=0, \beta=0}, \qquad C_1 = -\frac{1}{n} E \left\{ \frac{\partial^2 l}{\partial\theta\, \partial\theta'} \right\}_{\hat{\theta}, \hat{\alpha}, \phi=0, \beta=0},$$

$$C_{1r} = -\frac{1}{n} E \left\{ \frac{\partial^2 l}{\partial\phi\, \partial\phi'} \right\}_{\hat{\theta}, \hat{\alpha}, \phi=0, \beta=0}, \qquad L_{1r} = -\frac{1}{n} E \left\{ \frac{\partial^2 l}{\partial\theta\, \partial\phi'} \right\}_{\hat{\theta}, \hat{\alpha}, \phi=0, \beta=0},$$

$$T_{2r} = n^{-1/2} \left. \frac{\partial l}{\partial \beta} \right|_{\hat{\theta}, \hat{\alpha}, \phi=0, \beta=0}, \qquad C_2 = -\frac{1}{n} E \left\{ \frac{\partial^2 l}{\partial\alpha\, \partial\alpha'} \right\}_{\hat{\theta}, \hat{\alpha}, \phi=0, \beta=0},$$

$$C_{2r} = -\frac{1}{n} E \left\{ \frac{\partial^2 l}{\partial\beta\, \partial\beta'} \right\}_{\hat{\theta}, \hat{\alpha}, \phi=0, \beta=0}, \qquad L_{2r} = -\frac{1}{n} E \left\{ \frac{\partial^2 l}{\partial\alpha\, \partial\beta'} \right\}_{\hat{\theta}, \hat{\alpha}, \phi=0, \beta=0}.$$

Here $\hat{\theta}$ and $\hat{\alpha}$ are the maximum likelihood estimates under the null hypothesis, obtained for example, by the Newton-Raphson method. In practice, C_1, C_{1r}, L_{1r}, C_2, C_{2r} and L_{2r} have to be estimated by their sample counterparts, i.e., replacing the expectations by sample averages.

Remark 2.1. If we just consider the SETAR-ARCH model given by (2.1), (2.2) and $h_t = E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \alpha_0 + \sum_{i=1}^{q_1} \alpha_i \varepsilon_{t-i}^2$, the test statistic (2.4) reduces to that considered by Wong and Li (1997).

Remark 2.2. It is also possible to test the AR-TARCH model given by (2.2), (2.3) and $X_t - \theta_0 - \theta_1 X_{t-1} - \cdots - \theta_{p_1} X_{t-p_1} = \varepsilon_t$. The test statistic S in (2.4) reduces to $\sup_{r\in\tilde{R}} T_{2r}'(C_{2r} - L_{2r}'C_2^{-1}L_{2r})^{-1}T_{2r}$.

Theorem 2.1. *The limiting distributions of $\{T_{1r}\}$ and $\{T_{2r}\}$ are, respectively, that of a (p_2+1)-dimensional Gaussian process $\{\xi_{1r}\}$ and a (q_2+1)-dimensional Gaussian process $\{\xi_{2r}\}$, where $\{\xi_{ir}\}$, $i = 1, 2$, are indexed by the threshold parameter $r \in R$, and $\{\xi_{1r}\}$, $\{\xi_{2r}\}$ are independent processes. For each $r \in R$, $\xi_{1r} \sim N_{p_2+1}(0, C_{1r} - L_{1r}'C_1^{-1}L_{1r})$, $\xi_{2r} \sim N_{q_2+1}(0, C_{2r} - L_{2r}'C_2^{-1}L_{2r})$ and, for $r \neq s$, $\operatorname{cov}(\xi_{ir}, \xi_{is}) = C_{i,\min(r,s)} - L_{ir}'C_i^{-1}L_{is}$, $i = 1, 2$. The asymptotic null distribution of the Lagrange-Multiplier test statistic S in (2.4) is the same as the distribution of*

$$\sup_{r\in\tilde{R}} \left\{ \xi_{1r}'(C_{1r} - L_{1r}'C_1^{-1}L_{1r})^{-1}\xi_{1r} + \xi_{2r}'(C_{2r} - L_{2r}'C_2^{-1}L_{2r})^{-1}\xi_{2r} \right\}.$$

Corollary. *By Theorem 2.1 for each fixed r, $\xi_{1r}'(C_{1r} - L_{1r}'C_1^{-1}L_{1r})^{-1}\xi_{1r} + \xi_{2r}'(C_{2r} - L_{2r}'C_2^{-1}L_{2r})^{-1}\xi_{2r}$ is asymptotically distributed as $\chi^2_{p_2+q_2+2}$.*

Lemma 2.1. *If $p_1 = p_2 = p$ in (2.1) and $q_1 = q_2 = q$ in (2.3),*

$$\sup_{r\in\tilde{R}} \sum_{i=1}^{2} \xi_{ir}'(C_{ir} - L_{ir}'C_i^{-1}L_{ir})^{-1}\xi_{ir} = \sup_{r\in\tilde{R}} \left\{ \sum_{j=1}^{p+1} \frac{B_{1j}^2(r)}{\lambda_{1j}(r) - \lambda_{1j}^2(r)} + \sum_{j=1}^{q+1} \frac{B_{2j}^2(r)}{\lambda_{2j}(r) - \lambda_{2j}^2(r)} \right\},$$

where $B_{ij}(r)$'s are independent Gaussian processes with mean zero and $\operatorname{cov}\{B_{ij}(r), B_{ij}(s)\} = \lambda_{ij}\{\min(r,s)\} - \lambda_{ij}(r)\lambda_{ij}(s)$.

Proof. We have $C_{1r} = L_{1r}$ and $C_{2r} = L_{2r}$. Since for $i = 1, 2$, C_{ir} and $C_i - C_{ir}$ are positive definite, there exist invertible matrices Q_i and diagonal matrices D_i such that $Q_iC_iQ_i'$ are identity matrices and $Q_iC_{ir}Q_i' = D_i$ with all diagonal entries of D_i being strictly between 1 and 0. Let the diagonal elements of D_1 be $\lambda_{11}(r), \ldots, \lambda_{1,p+1}(r)$ and the diagonal elements of D_2 be $\lambda_{21}(r), \ldots, \lambda_{2,q+1}(r)$; $Q_1\xi_{1r} = \{B_{11}(r), \ldots, B_{1,p+1}(r)\}'$ and $Q_2\xi_{2r} = \{B_{21}(r), \ldots, B_{2,q+1}(r)\}'$. The result follows from similar arguments to those found in Wong and Li (1997).

Note that we can choose the $\lambda_{ij}(r)$'s to be continuous functions of r. It remains to find the matrices Q_i's or the $\lambda_{ij}(r)$'s, but we defer this discussion to the next section. We can make use of the following lemma to tabulate the asymptotic null distribution of S.

Lemma 2.2. *If the conditions in Lemma* 2.1 *hold, then for large* y,

$$pr\left\{\sup_{r\in\tilde{R}}\sum_{i=1}^{2}\xi'_{ir}(C_{ir}-L'_{ir}C_i^{-1}L_{ir})^{-1}\xi_{ir}\le y\right\}$$

$$\sim\exp\left\{-2\,\chi^2_{p+q+2}\left(\frac{y}{p+q+2}-1\right)\left(\sum_{j=1}^{p+1}\int_{\tilde{R}}\frac{d\,t_{1j}}{d\,r}d\,r+\sum_{j=1}^{q+1}\int_{\tilde{R}}\frac{d\,t_{2j}}{d\,r}d\,r\right)\right\},\qquad(2.5)$$

where $\chi^2_h(\cdot)$ *denotes the probability density function of the Chi-square distribution with* h *degrees of freedom and* $t_{ij}=\frac{1}{2}\log\left[\lambda_{ij}(r)/\left\{1-\lambda_{ij}(r)\right\}\right]$.

Proof. We use the Poisson clumping heuristic developed by Aldous (1989). The proof is a modified version of the proof of Theorem 1 in Chan (1991) and hence is omitted.

Remark 2.3. For some special cases where there is only one independent Gaussian process, i.e., only one $B_{ij}(r)$, a better approximation of the tail probability is given by (2.5) in Wong and Li (1997). The testing of SETAR-ARCH model with $p = 0$ and the testing of AR-TARCH model with $q = 0$ are two of these special cases.

3. Tabulation of the Asymptotic Null Distribution of S

In this section, we discuss the tabulation of the asymptotic null distribution of S. We separate the tabulation into two cases: the model with intercept term in the conditional mean, and the model without the intercept term in the conditional mean. Note that both cases are important in applications.

So far, we have not discussed the choice of $\tilde{R}$. In this paper, we choose the interval between the 10th and 90th percentiles of X_t as $\tilde{R}$. Practically, we use the empirical percentiles. Other choices of $\tilde{R}$ are possible, such as those used in Chan and Tong (1990).

To make use of formula (2.5), we must first determine the $\lambda_{ij}(r)$'s. To do this we make use of some propositions from Wong and Li (1997). In Section 6, we give some proofs and justifications of the propositions.

We assume the white noise model $X_t=\varepsilon_t,\quad \varepsilon_t\sim N(0,\sigma^2)$, in the tabulation of the asymptotic null distribution after obtaining the $\lambda_{ij}(r)$'s, since some expectations are very difficult to evaluate under the general case. In general, the asymptotic null distribution depends on p, q and the values of parameters through the $\lambda_{ij}(r)$'s. However, some extensive simulation experiments demonstrate that the dependence of the asymptotic null distribution on the parameters is very weak. We will present one such simulation result in this paper.

Table 1. Upper 10% points for the asymptotic null distribution of S for Case A.

		q							
		–	0	1	2	3	4	5	6
	–		7.75	9.76	12.34	14.57	16.60	18.52	20.34
	0	7.75	10.31	12.34	14.57	16.60	18.52	20.34	22.11
	1	11.05	13.26	15.04	17.00	18.86	20.65	22.38	24.06
p	2	13.26	15.30	17.00	18.86	20.65	22.38	24.06	25.71
	3	15.30	17.22	18.86	20.65	22.38	24.06	25.71	27.32
	4	17.22	19.05	20.65	22.38	24.06	25.71	27.32	28.91
	5	19.05	20.82	22.38	24.06	25.71	27.32	28.91	30.47
	6	20.82	22.53	24.06	25.71	27.32	28.91	30.47	32.01

Table 2. Upper 5% points for the asymptotic null distribution of S for Case A.

		q							
		–	0	1	2	3	4	5	6
	–		9.33	11.63	14.31	16.63	18.75	20.74	22.63
	0	9.21	12.15	14.31	16.63	18.75	20.74	22.63	24.46
	1	12.85	15.18	17.07	19.11	21.06	22.91	24.72	26.47
p	2	15.18	17.31	19.11	21.06	22.91	24.72	26.47	28.17
	3	17.31	19.32	21.06	22.91	24.72	26.47	28.17	29.85
	4	19.32	21.23	22.91	24.72	26.47	28.17	29.85	31.49
	5	21.23	23.07	24.72	26.47	28.17	29.85	31.49	33.10
	6	23.07	24.86	26.47	28.17	29.85	31.49	33.10	34.70

Table 3. Upper 1% points for the asymptotic null distribution of S for Case A.

		q							
		–	0	1	2	3	4	5	6
	–		12.87	15.58	18.44	20.94	23.21	25.34	27.38
	0	12.80	16.06	18.44	20.94	23.21	25.34	27.38	29.33
	1	16.72	19.26	21.35	23.55	25.63	27.63	29.56	31.42
p	2	19.25	21.57	23.55	25.63	27.63	29.56	31.42	33.25
	3	21.57	23.73	25.63	27.63	29.56	31.42	33.25	35.03
	4	23.73	25.79	27.63	29.56	31.42	33.25	35.03	36.78
	5	25.79	27.77	29.56	31.42	33.25	35.03	36.78	38.50
	6	27.77	29.68	31.42	33.25	35.03	36.78	38.50	40.19

Case A. The general model is given by (2.1), (2.2) and (2.3), with $p = p_1 = p_2$ and $q = q_1 = q_2$. The null hypothesis is $H_0 : \phi_i = 0$, $i = 0, \ldots, p$, and $\beta_i = 0$, $i = 0, \ldots, q$.

Case B. The general model is given by (2.1) with $\theta_0 = \phi_0 = 0$, $p = p_1 = p_2$ and (2.2), (2.3) with $q = q_1 = q_2$. The null hypothesis is $H_0 : \phi_i = 0$, $i = 1, \ldots, p$, and $\beta_i = 0$, $i = 0, \ldots, q$.

Table 4. Upper 10% points for the asymptotic null distribution of S for Case B.

		q							
		–	0	1	2	3	4	5	6
	1	5.81	9.21	11.49	13.96	16.12	18.11	19.99	21.79
	2	9.21	12.00	13.96	16.12	18.11	19.99	21.79	23.53
p	3	12.00	14.31	16.12	18.11	19.99	21.79	23.53	25.22
	4	14.31	16.40	18.11	19.99	21.79	23.53	25.22	26.87
	5	16.40	18.34	19.99	21.79	23.53	25.22	26.87	28.49
	6	18.34	20.19	21.79	23.53	25.22	26.87	28.49	30.08

Proposition 3.1. *For case* A *and $p > 1$, there exists a nonsingular matrix Q_1 such that $Q_1C_1Q_1'$ is an identity matrix and $Q_1C_{1r}Q_1'$ is a block diagonal matrix with two blocks. The first block is given by the matrix*

$$\begin{pmatrix} s_1(r) & s_3(r) \\ s_3(r) & s_2(r) \end{pmatrix} \tag{3.1}$$

and the second block is diag$\{s_1(r), \ldots, s_1(r)\}$ where $s_1(r) = a_{11r}/a_{11}$, $s_2(r) = a_{22r}/a_{22}$, $s_3(r) = a_{12r}/(a_{11}a_{22})^{1/2}$ and

$$a_{11} = \frac{2}{n}\sum E\left\{\frac{1}{h_t^2}\left(\sum_{i=1}^{q}\alpha_i^2\varepsilon_{t-i}^2\right)\right\} + \frac{1}{n}\sum E\left(\frac{1}{h_t}\right), \tag{3.2}$$

$$a_{22} = \frac{2}{n}\sum E\left\{\frac{1}{h_t^2}\left(\sum_{i=1}^{q}\alpha_i^2\varepsilon_{t-i}^2X_{t-d-i}^2\right)\right\} + \frac{1}{n}\sum E\left(\frac{X_{t-d}^2}{h_t}\right), \tag{3.3}$$

$$a_{11r} = \frac{2}{n}\sum E\left[\frac{1}{h_t^2}\left\{\sum_{i=1}^{q}\alpha_i^2\varepsilon_{t-i}^2I_r(X_{t-d-i})\right\}\right] + \frac{1}{n}\sum E\left\{\frac{1}{h_t}I_r(X_{t-d})\right\}, \tag{3.4}$$

$$a_{12r} = \frac{2}{n}\sum E\left[\frac{1}{h_t^2}\left\{\sum_{i=1}^{q}\alpha_i^2\varepsilon_{t-i}^2X_{t-d-i}I_r(X_{t-d-i})\right\}\right] + \frac{1}{n}\sum E\left\{\frac{X_{t-d}}{h_t}I_r(X_{t-d})\right\}, \tag{3.5}$$

$$a_{22r} = \frac{2}{n}\sum E\left[\frac{1}{h_t^2}\left\{\sum_{i=1}^{q}\alpha_i^2\varepsilon_{t-i}^2X_{t-d-i}^2I_r(X_{t-d-i})\right\}\right] + \frac{1}{n}\sum E\left\{\frac{X_{t-d}^2}{h_t}I_r(X_{t-d})\right\}. \tag{3.6}$$

For $p = 0$, we have $s_1(r)$ only. For $p = 1$, we have matrix (3.1) *with some modifications: replace X_{t-d-i} and X_{t-d} by X_{t-1-i} and X_{t-1} in* (3.3), (3.5) *and* (3.6)*, except for those arguments in the indicator function.*

Proposition 3.2. *For case B and $p > 1$, there exists a nonsingular matrix Q_1 such that $Q_1C_1Q_1'$ is an identity matrix and $Q_1C_{1r}Q_1' =$ diag$\{s_2(r), s_1(r), \ldots,$*

$s_1(r)\}$, where $s_1(r)$ and $s_2(r)$ are defined in Proposition 3.1. For $p = 1$ we have $s_2(r)$ only, with a similar modification as in the case of $p = 1$ in Proposition 3.1.

Table 5. Upper 5% points for the asymptotic null distribution of S for Case B.

		q							
		–	0	1	2	3	4	5	6
	1	7.33	11.13	13.52	16.06	18.29	20.36	22.31	24.17
	2	11.13	13.99	16.06	18.29	20.36	22.31	24.17	25.97
p	3	13.99	16.39	18.29	20.36	22.31	24.17	25.97	27.72
	4	16.39	18.56	20.36	22.31	24.17	25.97	27.72	29.43
	5	18.56	20.57	22.31	24.17	25.97	27.72	29.43	31.10
	6	20.57	22.49	24.17	25.97	27.72	29.43	31.10	32.74

Table 6. Upper 1% points for the asymptotic null distribution of S for Case B.

		q							
		–	0	1	2	3	4	5	6
	1	10.81	15.11	17.73	20.42	22.79	25.00	27.08	29.07
	2	15.11	18.15	20.42	22.79	25.00	27.08	29.07	30.98
p	3	18.15	20.72	22.79	25.00	27.08	29.07	30.98	32.85
	4	20.72	23.03	25.00	27.08	29.07	30.98	32.85	34.66
	5	23.03	25.20	27.08	29.07	30.98	32.85	34.66	36.43
	6	25.20	27.24	29.07	30.98	32.85	34.66	36.43	38.17

Proposition 3.3. *For cases A and B with $q > 1$, there exists a nonsingular matrix Q_2 such that $Q_2C_2Q_2'$ is an identity matrix and $Q_2C_{2r}Q_2' = diag\{\tilde{s}_1(r), \tilde{s}_2(r), \tilde{s}_3(r), \ldots, \tilde{s}_3(r)\}$, where $\tilde{s}_1(r)$, $\tilde{s}_2(r)$ are the eigenvalues of $\tilde{C}_2^{-1/2}\tilde{C}_{2r}\tilde{C}_2^{-1/2}$ with*

$$\tilde{C}_2 = \begin{pmatrix} \sum E(1/h_t^2) & \sum E(\varepsilon_{t-d}^2/h_t^2) \\ \sum E(\varepsilon_{t-d}^2/h_t^2) & \sum E(\varepsilon_{t-d}^4/h_t^2) \end{pmatrix}, \tag{3.7}$$

$$\tilde{C}_{2r} = \begin{pmatrix} \sum E\{I_r(X_{t-d})/h_t^2\} & \sum E\{\varepsilon_{t-d}^2 I_r(X_{t-d})/h_t^2\} \\ \sum E\{\varepsilon_{t-d}^2 I_r(X_{t-d})/h_t^2\} & \sum E\{\varepsilon_{t-d}^4 I_r(X_{t-d})/h_t^2\} \end{pmatrix}, \tag{3.8}$$

and $\tilde{s}_3(r) = \sum E\{I_r(X_{t-d})/h_t^2\}/\sum E(1/h_t^2)$. For $q = 0$ we have simply $E\{I_r(X_{t-d})\}$ and, for $q = 1$, we have $diag\{\tilde{s}_1(r), \tilde{s}_2(r)\}$ with ε_{t-d} replaced by ε_{t-1} in (3.7) *and* (3.8).

As discussed earlier, we make use of a white noise model in the tabulation of the asymptotic null distribution for the more general models. Then X_t and ε_t are normal distributed unconditionally and the $s_i(r)$'s and $\tilde{s}_i(r)$'s can be evaluated by direct integration. Tables 1 to 3 give the upper 10%, 5% and 1% points for Case

A, while Tables 4 to 6 give the upper percentage points for Case B. In the tables, an unlabelled column gives the percentage points for SETAR-ARCH models, while an unlabelled row has percentage points for the AR-TARCH models.

Table 7. Simulation results for comparison of the sample values of $s_2(r)$, $\tilde{s}_1(r)$ and $\tilde{s}_2(r)$, for some specific value of r.

Value of r as the percentiles of X_t											
10%			30%			70%			90%		
$s_2(r)$	$\tilde{s}_1(r)$	$\tilde{s}_2(r)$	$s_2(r)$	$\tilde{s}_1(r)$	$\tilde{s}_2(r)$	$s_2(r)$	$\tilde{s}_1(r)$	$\tilde{s}_2(r)$	$s_2(r)$	$\tilde{s}_1(r)$	$\tilde{s}_2(r)$
(1) True value under $\theta_1 = 0.0$ and $\alpha_1 = 0.0$											
.3249	.4663	.0310	.4822	.4994	.2167	.5178	.7833	.5006	.6751	.9690	.5337
(2) $\theta_1 = 0.0$ and $\alpha_1 = 0.0$											
.3217	.4493	.0291	.4804	.4884	.2128	.5168	.7823	.4942	.6747	.9692	.5302
(3) $\theta_1 = 0.0$ and $\alpha_1 = 0.1$											
.2945	.4324	.0164	.4771	.4948	.1808	.5219	.8165	.4988	.7035	.9831	.5591
(4) $\theta_1 = 0.5$ and $\alpha_1 = 0.1$											
.3062	.3484	.0399	.4788	.4696	.2310	.5204	.7677	.5240	.6916	.9597	.6444
(5) $\theta_1 = 0.9$ and $\alpha_1 = 0.1$											
.3203	.1714	.0861	.4853	.3545	.2878	.5204	.7121	.6418	.6799	.9140	.8260
(6) $\theta_1 = 0.0$ and $\alpha_1 = 0.4$											
.2662	.3696	.0053	.4694	.4963	.1215	.5294	.8769	.5032	.7314	.9947	.6280

In order to justify the tabulation using a white noise model, we simulated the model $X_t = \theta_1 X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0, h_t)$, $h_t = E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = 1.0 + \alpha_1 \varepsilon_{t-1}^2$. For each replication, we computed the sample values of $s_2(r)$, $\tilde{s}_1(r)$ and $\tilde{s}_2(r)$ using Propositions 3.2 and 3.3 for some specific values of r. Averages over 100 replications are shown in Table 7. Each replication has sample size 10000. Except for the cases $\theta_1 = 0.9$ or $\alpha_1 = 0.4$, the sample values are close to the true values with the white noise model. Hence the dependence of the values of $s_2(r)$, $\tilde{s}_1(r)$ and $\tilde{s}_2(r)$ on the values of parameters are actually very weak. Note that we should have $\alpha_1 < 0.5773$ for the existence of the fourth moment. As the values of parameters (θ_1 or α_1) are close to the boundaries of the moment conditions, approximation of the asymptotic null distribution to S became poorer, as expected. Hence it seems justified to use the white noise model in the tabulation of the asymptotic null distribution of S. The approximation is further assessed by simulation experiments in the next section.

4. Simulations and An Example

We carried out extensive simulation experiments to assess the empirical properties of the test. The results of 15 experiments are shown in Table 8. We list the empirical sizes or powers at the nominal upper 10%, 5% and 1% points. For each experiment, the number of replications is 10000 and the sample size is 500. Note that a sample size of 500 is quite common for financial data.

Table 8. Results of the simulation experiments for assessing the approximation of the test.

Experiment	Case	p	q	d	Empirical sizes/powers 10%	5%	1%
(1) Model 1 with $\theta_1 = \theta_2 = \alpha_1 = \alpha_2 = 0.0$	A	0	0	1	9.72	4.62	1.03
(2) Model 1 with $\theta_1 = \theta_2 = \alpha_1 = \alpha_2 = 0.0$	A	1	0	1	8.38	4.16	0.97
(3) Model 1 with $\alpha_1 = 0.4$, $\theta_1 = \theta_2 = \alpha_2 = 0.0$	A	1	1	1	14.19	8.60	3.99
(4) Model 1 with $\alpha_1 = \alpha_2 = 0.1$, $\theta_1 = \theta_2 = 0.0$	A	0	2	1	12.71	7.39	3.02
(5) Model 1 with $\theta_1 = 0.5$, $\alpha_1 = 0.4$, $\theta_2 = \alpha_2 = 0.0$	A	1	1	1	11.14	6.34	2.48
(6) Model 1 with $\theta_1 = 0.9, \alpha_1 = 0.4, \theta_2 = \alpha_2 = 0.0$	A	1	1	1	9.88	5.45	1.58
(7) Model 1 with $\theta_1 = \theta_2 = \alpha_1 = \alpha_2 = 0.0$	B	1	0	1	10.07	4.82	1.06
(8) Model 1 with $\theta_1 = 0.5$, $\alpha_1 = 0.4$, $\theta_2 = \alpha_2 = 0.0$	B	1	1	1	13.50	7.68	2.76
(9) Model 1 with $\theta_1 = 0.9, \alpha_1 = 0.4, \theta_2 = \alpha_2 = 0.0$	B	1	1	1	11.16	5.83	1.65
(10) Model 2 with $\theta_1 = 0.0$	A	1	3	1	14.29	10.37	6.40
(11) Model 2 with $\theta_1 = 0.5$	A	1	3	1	11.12	7.25	3.78
(12) Model 3	A	1	1	1	98.88	97.50	91.19
(13) Model 4	A	1	1	1	48.19	35.41	15.70
(14) Model 5	A	1	1	1	41.62	29.23	11.83
(15) Model 6	A	1	1	1	99.59	98.98	97.12

The first nine experiments assessed the approximation of the test. Model 1 can be represented by the following general form: $X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \varepsilon_t, \varepsilon_t \sim N(0, h_t), h_t = 1.0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2$. It can be seen that the approximation in general is quite good. However, a few observations can be made from these experiments. First, we see that the approximation decays if the values of the parameters of the true processes are close to the region with non-existing fourth moment, in Experiment (3) for example. Secondly, the approximation is acceptable even when the models are close to non-stationary, Experiments (6) and (9) for example. Lastly, the approximation of the 10% and 5% points of the asymptotic null distribution of S seems much better than that of the 1% points, probably due to the fact that the extreme tail is more difficult to approximate. We also did some simulation experiments with smaller sample sizes, 250 was one case. The approximation of the test is acceptable, but some caution is in order when using the test with small samples.

A referee suggested a refined approach to estimating the upper percentage points with $s(r)$'s and $\tilde{s}(r)$'s determined via simulation (similar to those in Table 7) from the fitted model under the null hypothesis. This approach may in general give more precise upper percentage points when the ARCH's parameters are close to the region of non-existing fourth moment. However, the improvement is negligible if the ARCH parameters are well within the region of existence of the fourth moment.

Another interesting question concerns robustness of the test to misspecification of the conditional variance function. Other models in the literature for modelling the conditional variance function include the GARCH model (Bollerslev (1986)), the stochastic volatility model (Taylor (1986)) and the CHARMA model (Tsay (1987)). We looked at robustness of the test in two simulation experiments. Model 2, simulated in Experiments (10) and (11) of Table 8, is the AR-GARCH model represented by the following: $X_t = \theta_1 X_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, h_t), h_t = 1.0 + 0.1\varepsilon_{t-1}^2 + 0.8h_{t-1}$. We set $q = 3$ in the simulation experiments in order to capture the dependence of conditional variance due to the GARCH model. It can be seen that the empirical sizes are larger than the nominal sizes. However, these empirical sizes are quite reasonable as the conditional variance function is actually misspecified.

We also checked the power of the test in detecting threshold structure in the models. The true models, simulated in Experiments (12) to (15) of Table 8, are Model 3: $X_t = 0.2X_{t-1} - 0.4\, I(X_{t-1} \le 0)X_{t-1} + \varepsilon_t, h_t = 1.0 + 0.1\varepsilon_{t-1}^2 + I(X_{t-1} \le 0)(0.5 + 0.3\varepsilon_{t-1}^2)$; Model 4: $X_t = 0.2X_{t-1} - 0.4\, I(X_{t-1} \le 0)X_{t-1} + \varepsilon_t, h_t = 1.0 + 0.1\varepsilon_{t-1}^2$; Model 5: $X_t = 0.5X_{t-1} + \varepsilon_t, h_t = 1.0 + 0.1\varepsilon_{t-1}^2 + 0.3\, I(X_{t-1} \le 0)\varepsilon_{t-1}^2$; Model 6: $X_t = 0.9X_{t-1} - 0.5I(X_{t-1} \le 0)X_{t-1} + \varepsilon_t, h_t = 1.0 + 0.1\varepsilon_{t-1}^2 + I(X_{t-1} \le 0)(0.5 + 0.3\varepsilon_{t-1}^2)$. The results demonstrate the power of the test in detecting the threshold structure in the model.

Table 9. Results of applying the test to the Hong Kong Hang Seng Index 1980-91.

				Test statistics S against		
Period	p	q	d	DTARCH model	SETAR-ARCH model	AR-TARCH model
1980-81	1	6	1	24.59 ‡	7.40 ‡	21.61 †
1982-83	1	3	1	24.52 ‡	6.25 †	24.28 ‡
1984-85	1	5	1	27.52 ‡	11.78 ‡	21.37 ‡
1986-87	1	4	1	64.33 ‡	16.93 ‡	56.57 ‡
1988-89	3	2	1	27.54 ‡	17.65 ‡	10.98
1990-91	1	3	1	23.37 ‡	4.28	21.82 ‡

* ‡ significant at 5% level, † significant at 10% level.

As a concrete example, we applied the Lagrange-multiplier test to the daily return series of the closing Hong Kong Hang Seng Index from 1980 to 1991, see Liu, Li and Li (1997). The return series is defined as the difference of the log index. The dataset is divided into six non-overlapping two year periods: 1980-81, 1982-83, 1984-85, 1986-87, 1988-89 and 1990-91. The results are reported in Table 9. We also report the results of the sub-tests against SETAR-ARCH and

AR-TARCH models. There are about 500 observations for each two year period. The delay parameter d is set to one for all periods, p and q for each periods are the same as those used in Liu, Li and Li (1997), although their specification of the models is slightly different from ours. Here we only consider Case B. For the periods 1986-87 and 1988-89, the data are trimmed to within plus/minus three sample standard deviations of the respective period. The null hypothesis of linearity against the DTARCH models is rejected in all periods. Note that it is the first time that nonlinearity in the period 1990-91 is detected (Li and Lam (1995), Wong and Li (1997) and Liu, Li and Li (1997)). Furthermore, from the results of the sub-tests, nonlinearity is observed in both the conditional means and variances in the first four periods. In the 1988-89 period, nonlinearity is detected only in the conditional mean. On the contrary, nonlinearity is observed only in the conditional variance for the period 1990-91.

5. Proof of Theorem 2.1

In this section, we give an outline of the proof of Theorem 2.1. For a complete proof, see Wong (1998).

We consider the limiting distribution of $\{T_{1r}\}$. Let $w_{1r} = (T'_{1\infty}, T'_{1r})'$, where $T_{1\infty} = n^{-1/2}\partial l/\partial\theta$. Also, let $\psi' = (\psi_{10}, \ldots, \psi_{1p_1}, \psi_{20}, \ldots, \psi_{2p_2})$ where $\psi_{ij} \in R$. Consider

$$\psi' w_{1r} = n^{-1/2}\left(\sum_{j=0}^{p_1}\psi_{1j}\frac{\partial l}{\partial\theta_j} + \sum_{j=0}^{p_2}\psi_{2j}\frac{\partial l}{\partial\phi_j}\right) = n^{-1/2}\sum_{t=1}^{n}u_{1t}.$$

After some manipulation, $E(u_{1t}|\mathcal{F}_{t-1}) = 0$ and $E(u_{1t}^2) < \infty$ as the time series is stationary, ergodic and the fourth moment exists. By the Martingale Central Limit Theorem (see, for example, Theorem 23.1 in Billingsley (1968)), $\psi' w_{1r}$ is normally distributed with mean 0 and variance $\psi' E(w_{1r}w'_{1r})\psi$. By the Cramer-Wold device (see, for example, Billingsley (1968, p.48) or Brockwell and Davis (1987)),

$$w_{1r} \sim> N\left(0, \begin{pmatrix} C_1 & L_{1r} \\ L'_{1r} & C_{1r}\end{pmatrix}\right),$$

where $\sim>$ denotes convergence in distribution. Hence for each r, conditional on $T_{1\infty} = 0$, we obtain $T_{1r} \sim> N(0, (C_{1r} - L'_{1r}C_1^{-1}L_{1r}))$. Similarly, for $r \neq s$, the conditional distribution of (T_{2r}, T_{2s}) under the null hypothesis is

$$\begin{pmatrix} T_{1r} \\ T_{1s}\end{pmatrix} \sim N\left(0, \begin{pmatrix} C_{1r} & C_{1,\min(r,s)} \\ C'_{1,\min(r,s)} & C_{1s}\end{pmatrix} - \begin{pmatrix} L'_{1r} \\ L'_{1s}\end{pmatrix} C_1^{-1}\begin{pmatrix} L_{1r} & L_{1s}\end{pmatrix}\right).$$

Hence (T_{1r}, T_{1s}) has a joint normal distribution with covariance $= C_{1,\min(r,s)} - L'_{1r}C_1^{-1}L_{1s}$.

Now we consider the limiting distribution of $\{T_{2r}\}$. Let $w_{2r} = (T'_{2\infty}, T'_{2r})'$, where $T_{2\infty} = n^{-1/2}\partial l/\partial\alpha$. Also, let $\varphi' = (\varphi_{10}, \ldots, \varphi_{1q_1}, \varphi_{20}, \ldots, \varphi_{2q_2})$ where $\varphi_{ij} \in R$. Consider

$$\varphi' w_{2r} = n^{-1/2}\left(\sum_{j=0}^{q_1}\varphi_{1j}\frac{\partial l}{\partial\alpha_j} + \sum_{j=0}^{q_2}\varphi_{2j}\frac{\partial l}{\partial\beta_j}\right) = n^{-1/2}\sum_{t=1}^{n} u_{2t}.$$

It is easy to show that $E(u_{2t}|\mathcal{F}_{t-1}) = 0$ and $E(u_{2t}^2) < \infty$. By the Martingale Central Limit Theorem and the Cramer-Wold device, we have

$$w_{2r} \sim> N\left(0, \begin{pmatrix} C_2 & L_{2r} \\ L'_{2r} & C_{2r}\end{pmatrix}\right).$$

Hence for each r, conditional on $T_{2\infty} = 0$, we obtain $T_{2r} \sim> N(0, (C_{2r} - L'_{2r}C_2^{-1}L_{2r}))$. Similarly, it can be shown that (T_{2r}, T_{2s}), $r \neq s$, has a joint normal distribution with covariance $= C_{2,\min(r,s)} - L'_{2r}C_2^{-1}L_{2s}$.

It remains to consider the tightness of the distribution and the topology of the function space. Consider the spaces of functions that map $(-\infty,\infty)$ ($[-b,b], b > 0$) into R^k, right continuous with left-hand limits. Denote these spaces by $D_k(-\infty,\infty)$ and $D_k[-b,b]$, respectively. 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 continuous functions. See, for example, Pollard (1984) for more details on these spaces. Now $\{T_{1r}, -\infty < r < \infty\}$ lives on $D_{p_2+1}(-\infty,\infty)$ and $\{T_{2r}, -\infty < r < \infty\}$ lives on $D_{q_2+1}(-\infty,\infty)$.

Suppose that H_0 and our assumptions hold. We are going to show that (i) $\{T_{1r}\}$ converges weakly to $\{\xi_{1r}\}$ in $D_{p_2+1}(-\infty,\infty)$ and each realization of $\{\xi_{1r}\}$ belongs to $C_{p_2+1}(-\infty,\infty)$ a.s., and (ii) $\{T_{2r}\}$ converges weakly to $\{\xi_{2r}\}$ in $D_{q_2+1}(-\infty,\infty)$ and each realization of $\{\xi_{2r}\}$ belongs to $C_{q_2+1}(-\infty,\infty)$ a.s. Note that it suffices to verify the tightness of $\{T_{1r}, -b \le r \le b\}$ and $\{T_{2r}, -b \le r \le b\}$ componentwise.

For the tightness of $\{T_{1r}\}$, we only consider the case where $h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2$. Note that under H_0 the β_j's are all zero. The proof is easily extended to more general cases. Without loss of generality, consider the last component of $\{T_{1r}, -b \le r \le b\}$. It is tight if and only if

$$g_{1n}(r) = n^{-1/2}\sum\left\{\frac{\varepsilon_t}{h_t}X_{t-p_2}I_r(X_{t-d}) + \alpha_1\frac{\varepsilon_{t-1}}{h_t}X_{t-p_2-1}\left(1 - \frac{\varepsilon_t^2}{h_t}\right)I_r(X_{t-d-1})\right\}$$

is tight. Let $-b \le s \le r \le b$ be two arbitrary numbers, M_i, $i = 1,2,3$, be constants independent of n, and $g_{1n}(r) - g_{1n}(s) = n^{-1/2}\sum\kappa_{1t}(r,s)$, where

$$\kappa_{1t}(r,s) = \frac{\varepsilon_t}{h_t}X_{t-p_2}I_{s,r}(X_{t-d}) + \alpha_1\frac{\varepsilon_{t-1}}{h_t}X_{t-p_2-1}\left(1 - \frac{\varepsilon_t^2}{h_t}\right)I_{s,r}(X_{t-d-1})$$

and $I_{s,r}(X_t) = I(s < X_t \le r)$. It can be shown that for $i = 1, \ldots, p_2$,

$$E\left\{\left|\frac{\varepsilon_t}{h_t}X_{t-i}I(s < X_{t-d} \le r)\right|^2\right\} \le M_1(r-s)$$

and

$$E\left\{\left|\frac{\varepsilon_{t-1}}{h_t}X_{t-i-1}\left(1-\frac{\varepsilon_t^2}{h_t}\right)I(s < X_{t-d-1} \le r)\right|^2\right\} \le M_2(r-s).$$

After some manipulation, we have

$$E\left\{\sup_{|r-s|<\tau}|\kappa_{1t}(r,s)|^2\right\} \le M_3\tau.$$

By (2.1) in Andrews and Pollard (1984), we have the required geometric bound for Theorem 2.2 of Andrews and Pollard (1984) to hold. Hence, we have for each $\epsilon > 0$, $\eta > 0$, there exists a δ, $\delta > 0$, such that

$$pr\left\{\sup_{-b\le s<r\le b,|r-s|<\delta}|g_{1n}(r)-g_{1n}(s)| \ge \epsilon\right\} \le \eta$$

for large n. This gives the tightness of $\{T_{1r}\}$.

For the tightness of $\{T_{2r}\}$, without loss of generality, we consider the last component of $\{T_{2r}, -b \le r \le b\}$. It is tight if and only if

$$g_{2n}(r) = n^{-1/2}\sum\left\{\frac{1}{h_t}\varepsilon_{t-q_2}^2\left(\frac{\varepsilon_t^2}{h_t}-1\right)I_r(X_{t-d})\right\}$$

is tight.

Let $-b \le s \le r \le b$ be two arbitrary numbers, M_4 and M_5 be constants independent of n, and $g_{2n}(r) - g_{2n}(s) = n^{-1/2}\sum\kappa_{2t}(r,s)$, where

$$\kappa_{2t}(r,s) = \frac{1}{h_t}\varepsilon_{t-q_2}^2\left(\frac{\varepsilon_t^2}{h_t}-1\right)I_{s,r}(X_{t-d}).$$

It can be shown that for $i = 1, \ldots, q_2$,

$$E\left\{\left|\frac{1}{h_t}\varepsilon_{t-i}^2\left(\frac{\varepsilon_t^2}{h_t}-1\right)I(s < X_{t-d} \le r)\right|^2\right\} \le M_4(r-s),$$

and hence after some manipulation,

$$E\left\{\sup_{|r-s|<\tau}|\kappa_{2t}(r,s)|^2\right\} \le M_5\tau.$$

So by Theorem 2.2 of Andrews and Pollard (1994) with the required geometric bound, we have for each $\epsilon > 0$, $\eta > 0$, there exists a δ, $\delta > 0$, such that

$$pr\left\{\sup_{-b\le s\le r\le b, |r-s|\le\delta} |g_{2n}(r) - g_{2n}(s)| \ge \epsilon\right\} \le \eta$$

for large n. The tightness of $\{T_{2r}\}$ follows.

From the above discussion, $\{T_{1r}\}$ is asymptotically a $(p_2 + 1)$-dimensional Gaussian process and $\{T_{2r}\}$ is asymptotically a $(q_2 + 1)$-dimensional Gaussian process, both indexed by the threshold parameter $r \in R$. Denote the limiting Gaussian processes by $\{\xi_{1r}\}$ and $\{\xi_{2r}\}$ and we have completed our proof.

6. Proofs and Justifications of the Propositions

It can be observed that Proposition 3.3 is trivial if $q = 0$ or 1. The special cases $p = 0$ and $p = 1$ in Proposition 3.1 and $p = 1$ in Proposition 3.2 can be proved by utilizing the following result from Engle (1982). Let u and v be any two random variables; the expectation $E\{g(u, v)|v\}$ is an anti-symmetric function of v if g is anti-symmetric in v, the conditional density of $u|v$ is symmetric in v, and the expectation exists.

Let W_t be a constant with respect to $\mathcal{F}_{t-1}$. Using the result above, we can show that $E(\varepsilon_{t-i}W_{t-i}/h_t^2) = 0$, for $1 \le i \le m$, if these expectations exist. Clearly $\varepsilon_{t-i}W_{t-i}/h_t^2$ is anti-symmetric in ε_{t-i}, which is part of the information set $\mathcal{F}_{t-i}$. Because h_t is symmetric, the conditional density must be symmetric in ε_{t-i} and, using Engle's Lemma, $g(\varepsilon_{t-i}) = E\left\{(\varepsilon_{t-i}W_{t-i}/h_t^2)\middle|\,\mathcal{F}_{t-i}\right\}$ is anti-symmetric, as the density of ε_{t-i} conditional on the information set $\mathcal{F}_{t-i-1}$ is a symmetric (normal) density. Hence $E\{g(\varepsilon_{t-i})|\mathcal{F}_{t-i-1}\} = 0$.

Now we can apply the above argument to obtain the special cases in Propositions 3.1 and 3.2. As an illustration, we obtain $s_2(r)$ for $p = 1$ in Proposition 3.2. We have

$$\begin{aligned} C_{1r} &= \frac{1}{2n}\sum E\left[\frac{1}{h_t^2}\left\{2\sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}X_{t-1-i}I_r(X_{t-d-i})\right\}^2\right] + \frac{1}{n}\sum E\left\{\frac{1}{h_t}X_{t-1}^2 I_r(X_{t-d})\right\} \\ &= \frac{2}{n}\sum E\left[\frac{1}{h_t^2}\left\{\sum_{i=1}^{q}\alpha_i^2\varepsilon_{t-i}^2 X_{t-1-i}^2 I_r(X_{t-d-i})\right\}\right] + \frac{1}{n}\sum E\left\{\frac{1}{h_t}X_{t-1}^2 I_r(X_{t-d})\right\}, \end{aligned}$$

as all the crossproduct terms in the first summation are zero, by the above argument. Similarly, we get

$$C_1 = \frac{2}{n}\sum E\left[\frac{1}{h_t^2}\left\{\sum_{i=1}^{q}\alpha_i^2\varepsilon_{t-i}^2 X_{t-1-i}^2\right\}\right] + \frac{1}{n}\sum E\left\{\frac{1}{h_t}X_{t-1}^2\right\}.$$

Hence the result.

Table 10. Results of simulation experiments to justify the use of the Propositions to compute $\lambda_{ij}(r)$'s.

Value of r as the percentiles of X_t	$\lambda_{11}(r)$	$\lambda_{12}(r)$	$\lambda_{13}(r)$	$\lambda_{14}(r)$	$\lambda_{21}(r)$	$\lambda_{22}(r)$	$\lambda_{23}(r)$	$\lambda_{24}(r)$
Method 1: average eigenvalues of $C_1^{-1/2}C_{1r}C_1^{-1/2}$ and $C_2^{-1/2}C_{2r}C_2^{-1/2}$								
10%	.4204	.1239	.1070	.0046	.3665	.2013	.1544	.0454
30%	.7387	.3210	.3016	.0346	.4627	.3719	.3293	.2393
50%	.8860	.5073	.4894	.1045	.5241	.4995	.4770	.4491
70%	.9599	.6949	.6749	.2480	.7535	.6546	.6035	.4907
90%	.9937	.8910	.8731	.5607	.9508	.8370	.7763	.5812
Method 2: average eigenvalues from Propositions 3.1 and 3.3								
10%	.4200	.1009	.1009	.0047	.3621	.0995	.0995	.0535
30%	.7384	.3008	.3008	.0348	.4525	.2999	.2999	.2491
50%	.8857	.4996	.4996	.1047	.5072	.4997	.4997	.4633
70%	.9597	.6986	.6986	.2482	.7461	.7000	.7000	.4953
90%	.9936	.8984	.8984	.5613	.9440	.8999	.8999	.5834

For the general cases of the propositions, we justify them through extensive simulation experiments. In Table 10, we report some of the results. The model simulated is $X_t = 0.5X_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim N(0, h_t)$, $h_t = E(\varepsilon_t^2|\mathcal{F}_{t-1}) = 1.0 + 0.1\varepsilon_{t-1}^2 + 0.1\varepsilon_{t-2}^2 + 0.1\varepsilon_{t-3}^2$. For each replication, we compute the sample values of $\lambda_{ij}(r)$'s with $p = 3$ and $q = 3$ in Case A. Each replication has sample size 5000 and we perform 100 replications. The first method directly computes the eigenvalues of $C_1^{-1/2}C_{1r}C_1^{-1/2}$ and $C_2^{-1/2}C_{2r}C_2^{-1/2}$. The second method uses Propositions 3.1 and 3.3 to compute the $\lambda_{ij}(r)$. It can be observed that the $\lambda_{ij}(r)$ computed by the two methods are close to each other and hence it seems justified to use the Propositions to compute them.

Acknowledgement

The authors would like to thank the referee and the Editor for comments that led to improvement of the paper. W. K. Li's research is supported by the Hong Kong Research Grant Council and C. S. Wong's research is supported by the Committee on Research and Conference Grants, The University of Hong Kong.

References

Aldous, D. J. (1989). *Probability approximation via the Poisson clumping heuristic.* Appl. Math. Sci. 77. Springer-Verlag, New York.

Andrews, D. W. K. and Pollard, D. (1994). An Introduction to functional central limit theorems for dependent stochastic processes. *Internat. Statist. Rev.* **62**, 119-132.

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

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **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. Econometrics* **52**, 5-59.

Brockwell, P. J. and Davis, R. A. (1987). *Time Series: Theory and Methods*. Springer, New York.

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. (1990). On likelihood ratio tests for threshold autoregression. *J. Roy. Statist. Soc. Ser. B* **52**, 469-476.

Davidson, J. and MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics*. Oxford University Press, Oxford.

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). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987-1007.

Li, C. W. and Li, W. K. (1996). On a double-threshold autoregressive heteroscedastic time series model. *J. Appl. Econometrics* **11**, 253-274.

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

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.

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

Rabemananjara, R. and Zakoian, J. M. (1993). Threshold ARCH models and asymmetries in volatility. In *Nonlinear Dynamics Chaos and Econometrics* (Edited by M. H. Pesaran and S. M. Potter), 179-197. Wiley, New York.

Taylor, S. J. (1986). *Modelling Financial Time Series*. John Wiley, Chichester.

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 (with discussion). *J. Roy. Statist. Soc. Ser. B* **42**, 245-292.

Tsay, R. S. (1987). Conditional heteroscedastic time series models. *J. Amer. Statist. Assoc.* **82**, 590-604.

Wong, C. S. (1998). *Statistical Inference for Some Nonlinear Time Series Models*. Ph.D. Thesis, The University of Hong Kong.

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

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

E-mail: cswonga@hkusua.hku.hk

E-mail: hrntlwk@hku.hk

(Received August 1997; accepted March 1999)