Studies in Mathematical Sciences **ISSN 1923-8444 [Print]**
Vol. 2, No. 1, 2011, pp. 1-20 **ISSN 1923-8452 [Online]**
www.cscanada.org **www.cscanada.net**

Testing for Multivariate Threshold Autoregression

Shu-Ing LIU[1,*]

Abstract: In this article we propose a testing procedure for multivariate threshold autoregression with the disturbances following conditional homoscedastic martingale difference sequences. A right-tailed asymptotic distribution of the proposed test statistics is derived and the accuracy is investigated by simulations. The numerical simulations however show a remarkable robustness to a miss-specification of the order of the AR model. This encourages one to apply the asymptotic results, which will make the computation more convenient in actual applications. Furthermore, some numerical simulations indicate that the proposed test is more powerful than the test in [12].

Key Words: Eigenvalues; Lagrange-multiplier test; Likelihood ratio test; Martingale differences; Threshold autoregression

1. INTRODUCTION

The self-exciting threshold autoregressive model is one of the nonlinear time series models first proposed by Tong[8] and later discussed in detail by Tong and Lim[10]. Since then, these models have become popular. They have a piecewise linear conditional mean in the threshold space, which is able to describe many phenomena such as jumps, limit cycles and financial data. A comprehensive review of these models can be found in Tong[9]. Tsay[11] has used the arranged autoregression technique to test the existence of threshold nonlinearity. Chan[3], Chan and Tong[4], on the other hand, have proposed the likelihood ratio approach as a test for threshold autoregression with normally distributed noise. In some special cases, the problem may be reduced to a first-passage problem associated with a Gaussian process. Chan[5] has derived an asymptotic tailed distribution of the test statistic for the general case when the threshold effect is absent.

Later, Wong and Li[14] presented an asymptotic null distribution of the Lagrange-multiplier test statistic for threshold autoregression with conditional heteroscedasticity. Some numerical results indicate that the likelihood ratio test may have a higher Type I error probability than the specified one when conditional heteroscedasticity exists. However, the Lagrange-multiplier test will be more robust when hetero-scedasticity exists. Tsay[12] extended his previous work to multivariate data, giving descriptions of some multivariate threshold models, and proposed testing procedures, to detect the existence of threshold nonlinearity. In this paper, a unified testing procedure for a multivariate threshold model is investigated. Essentially, it is an extension of Chan[3], Chan and Tong's work[4], and a partial extension of Wong and Li's work[14], but with martingale difference noise sequences under some regular conditions.

The paper is organized as follows: In Section 2, the discussed testing procedure is introduced and an approximate tailed distribution of the test statistic is discussed. Illustrations of the accuracy of the approximation, the testing power investigated by simulations are included in Section 3. Finally, Section 4 gives a brief conclusion.

[1]Department of Finance, Shih Hsin University, No. 111, Mu-Cha Road, Sec 1, Taipei, 11604, Taiwan.
E-mail adress: siliu@cc.shu.edu.tw.
*Corresponding author.
[†]Received 26 December 2010; accepted 6 February 2011.

2. TEST FOR MULTIVARIATE THRESHOLD MODELS

Chan and Tong[4] applied the likelihood ratio statistic to test threshold nonlinearity for univariate data. Later, Wong and Li[14] derived a Lagrange-multiplier test statistic for threshold nonlinearity with conditional heteroscedasticity. By using the properties of the mosaic process[1], an approximation of the asymptotic tailed distribution of the test statistic was obtained. In the next subsection, a unified approach for producing test statistic for multivariate data will be investigated.

2.1 The Model

Consider a multivariate time series $\{X_t\}$, where X_t is a k-dimension random vector, denoted by $X_t = (X_{1,t}, \cdots, X_{k,t})^T$. A simple multivariate threshold model with an autoregression order p, a lag parameter d, $1 \le d \le p$, and a threshold parameter r is defined as follows:

$$X_t - \Theta_0 - \sum_{j=1}^{p} \Theta_j X_{t-j} - I(f(X_{t-d}) \le r)(\Phi_0 + \sum_{j=1}^{p} \Phi_j X_{t-j}) = a_{t,r}, \tag{1}$$

for $t = p+1, \cdots, n$. Here, $I(\cdot)$ is an indicator function, and for simplicity, function f is a projection function defined by $f(X_t) = X_{i,t}$ for some i-th component of X_t, with $1 \le i \le k$. Θ_0 and Φ_0 are $k \times 1$ parameter vectors, and the Θ_i's and Φ_j's, for $1 \le i, j \le p$ are $k \times k$ parameter matrices, satisfying some stationary conditions: All the roots of the equation, $\det\left(Z^p - \Theta_1 Z^{p-1} - \cdots - \Theta_p\right) = 0$, lie strictly inside the unit circle. The threshold parameter r is supposed to belong to an finite interval of R, say (b, c). In the following analysis, for simplicity, the lag parameter d and the order of autoregression p are assumed to be known in advance. Moreover, the $\{a_{t,r}\}$ is a sequence of martingale differences satisfying $E(a_{t,r} \,|\, F_{t-1,r}) = 0$, $Cov(a_{t,r}, a_{t,r} \,|\, F_{t-1,r}) = \Omega_{1,r}$ and $Cov(a_{t,r}, a_{s,r} \,|\, F_{t \wedge s-1,r}) = 0$, for $t \ne s$. $t \wedge s = \min(t, s)$, $F_{t-1,r}$ is the σ-field generated by $\{X_{t-i}, a_{t-i,r} \,|\, i = 1, 2, \cdots\}$, and $\Omega_{1,r}$ is a symmetric positive definite matrix. Also, X_t and $a_{t+i,r}$ are uncorrelated for $i > 0$, and for all r.

Before proposing the test statistics, model (1) is re-represented in a matrix form. Define

$$Y = \begin{pmatrix} X_{p+1}^T \\ X_{p+2}^T \\ \vdots \\ X_n^T \end{pmatrix}, \quad X = \begin{pmatrix} 1 & X_p^T & \cdots & X_1^T \\ 1 & X_{p+1}^T & \cdots & X_2^T \\ \vdots & & \vdots & \\ 1 & X_{n-1}^T & \cdots & X_{n-p}^T \end{pmatrix}, \quad \varepsilon_{1,r} = \begin{pmatrix} a_{p+1,r}^T \\ a_{p+2,r}^T \\ \vdots \\ a_{n,r}^T \end{pmatrix},$$

$$Y_r = \begin{pmatrix} I(f(X_{p+1-d}) \le r) & I(f(X_{p+1-d}) \le r) X_p^T & \cdots & I(f(X_{p+1-d}) \le r) X_1^T \\ I(f(X_{p+2-d}) \le r) & I(f(X_{p+2-d}) \le r) X_{p+1}^T & \cdots & I(f(X_{p+2-d}) \le r) X_2^T \\ \vdots & & \vdots & \\ I(f(X_{n-d}) \le r) & I(f(X_{n-d}) \le r) X_{n-1}^T & \cdots & I(f(X_{n-d}) \le r) X_{n-p}^T \end{pmatrix},$$

$$\Theta_{(F)}^T = (\Theta_0, \Theta_1, \cdots, \Theta_p) \text{ and } \Phi_{(F)}^T = (\Phi_0, \Phi_1, \cdots, \Phi_p).$$

Then, model (1) is re-written as

$$Y = X\Theta_{(F)} + Y_r \Phi_{(F)} + \varepsilon_{1,r}. \tag{2}$$

For convenience, model (2) will be re-written in a vector form. If A is an $m \times \ell$ matrix, let A^V denote the $m\ell$-vector obtained by vectorizing A; that is, by stacking the column of A on top of one another.

Finally, model (2) is re-expressed as

$$Y^V = (I_k \otimes X)\Theta^V_{(F)} + (I_k \otimes Y_r)\Phi^V_{(F)} + \varepsilon^V_{1,r}, \tag{3}$$

where I_ℓ is an $\ell \times \ell$ identity matrix. Now, our null hypothesis is formulated as

$$\mathrm{H}_0 : \Phi^V_{(F)} = 0 \text{ against } \mathrm{H}_1 : \Phi^V_{(F)} \neq 0 .$$

Under the null hypothesis, model (3) is reduced to

$$Y^V = (I_k \otimes X)\Theta^V_{(R)} + \varepsilon^V_2 .$$

Here $\Theta^V_{(R)}$ denotes the parameter matrix corresponding to $\Theta^V_{(F)}$, under H_0. ε^V_2 has the same structure as $\varepsilon^V_{1,r}$ with Ω_2 as the covariance matrix corresponding to $\Omega_{1,r}$, under H_0.

The least squares estimates of each parameter are respectively expressed as follows:

$$\hat{\Theta}^V_{(R)} = \left\{ I_k \otimes \left(X^T X \right)^{-1} X^T \right\} Y^V ,$$

$$\hat{\Theta}^V_{(F)} = \hat{\Theta}^V_{(R)} - \left\{ I_k \otimes \left(X^T X \right)^{-1} X^T Y_r \left(Y_r^T G Y \right)^{-1} Y_r^T G \right\} Y^V ,$$

$$\hat{\Phi}^V_{(F)} = \left\{ I_k \otimes \left(Y_r^T G Y_r \right)^{-1} Y_r^T G \right\} Y^V ,$$

$$\hat{\Omega}_{1,r} = \hat{\varepsilon}^T_{1,r} \hat{\varepsilon}_{1,r} / (n-p) \text{ and } \hat{\Omega}_2 = \hat{\varepsilon}^T_2 \hat{\varepsilon}_2 / (n-p),$$

where $G = I_{n-p} - X(X^T X)^{-1} X^T$, $\hat{\varepsilon}^V_{1,r} = Y^V - (I_k \otimes X)\hat{\Theta}^V_{(F)} - (I_k \otimes Y_r)\hat{\Phi}^V_{(F)}$ and $\hat{\varepsilon}^V_2 = Y^V - (I_k \otimes X)\hat{\Theta}^V_{(R)}$. It is worthy to note that the relation $Y_r^T X = Y_r^T Y_r = X^T Y_r$ is true under model (1).

2.2 The Proposed Test Statistics

In this subsection, some convenient test statistics for testing H_0 against H_1 are defined in terms of the residual sum of the squares. First, define

$$\lambda_r(\Omega) = \Lambda_2(\Omega)^T \Lambda_2(\Omega) - \Lambda_{1,r}(\Omega)^T \Lambda_{1,r}(\Omega),$$

where

$$\Lambda_2(\Omega) = \left(\Omega^{-1/2} \otimes I_{n-p} \right) \hat{\varepsilon}^V_2 \text{ and } \Lambda_{1,r}(\Omega) = \left(\Omega^{-1/2} \otimes I_{n-p} \right) \hat{\varepsilon}^V_{1,r} .$$

$\lambda_r(\Omega)$ can be used to measure the difference in the weighted residual sum of the squares between the reduced model and the full model. After algebra, $\lambda_r(\Omega)$ can be re-expressed as

$$\lambda_r(\Omega) = (Y^V)^T \left(\Omega^{-1/2} \otimes G Y_r \right) \left\{ I_k \otimes \left(Y_r^T G Y_r \right)^{-1} \right\} \left(\Omega^{-1/2} \otimes Y_r^T G \right) Y^V .$$

Now, a different estimate of Ω will produce a different test statistic. Here, two test statistics are investigated. They are $\lambda_{(1)} = \sup_{b \le r \le c} \lambda_r(\hat{\Omega}_{1,r})$ and $\lambda_{(2)} = \sup_{b \le r \le c} \lambda_r(\hat{\Omega}_2)$.

Suppose the normality assumptions for the disturbances are imposed in the following discussions: For a fixed r and $k = 1$, $\lambda_r(\hat{\Omega}_{1,r})$ asymptotically reduces to the likelihood ratio test statistic proposed by Chan and Tong[4]; $\lambda_r(\hat{\Omega}_2)$ asymptotically reduces to the Lagrange-multiplier test statistic discussed by Wong and Li[14]. For $k \ge 1$, it can be analytically proven that $\lambda_r(\hat{\Omega}_2)$ is asymptotically equivalent to the

Lagrange-multiplier test statistic (see Appendix A), and that $\lambda_r(\hat{\Omega}_{1,r})$ is asymptotically equivalent to the Wald test statistic[13]. Moreover, it can be shown that the difference between the two covariance estimators, $\hat{\Omega}_2$ and $\hat{\Omega}_{1,r}$, is

$$\hat{\Omega}_2 - \hat{\Omega}_{1,r} = Y^T G Y_r \left(Y_r^T G Y_r\right)^{-1} Y_r^T GY/(n-p).$$

Under hypothesis H_0, the two covariance matrices, $\hat{\Omega}_{1,r}$ and $\hat{\Omega}_2$, will converge to the same covariance matrix when the sample size is sufficiently large. Therefore, the asymptotical distribution of the two test statistics will be the same.

Suppose the following assumption holds,

$$\frac{1}{n}\begin{pmatrix} X^T X & X^T Y_r \\ Y_r^T X & Y_r^T Y_r \end{pmatrix} \xrightarrow[n\to\infty]{a.s.} \begin{pmatrix} \sum & \sum_r \\ \sum_r & \sum_r \end{pmatrix},$$

then $n^{-1}Y_r^T G Y_r \xrightarrow[n\to\infty]{a.s.} (\sum_r - \sum_r \sum^{-1} \sum_r)$. Furthermore, by applying the results of Lai and Wei[7] (detailed proof referred to Appendix B), $\hat{\Phi}_{(F)}^V$ is asymptotically normally distributed. Therefore, under H_0,

$$n^{-1/2}\left(\hat{\Omega}_2^{-1/2} \otimes Y_r^T G\right) Y^V \xrightarrow[n\to\infty]{d} \xi_r,$$

where the random vector ξ_r is $\mathrm{N}(0, I_k \otimes (\sum_r - \sum_r \sum^{-1} \sum_r))$ distributed and Cov $(\xi_g, \xi_h) = I_k \otimes (\sum_{g\wedge h} - \sum_g \sum^{-1} \sum_h)$. A straightly forward proof of the covariance stricture is given is Appendix C. Finally, under H_0,

$$\lambda_{(2)} \xrightarrow[n\to\infty]{d} \sup_{b\le r\le c} \eta_r,$$

where $\eta_r = \xi_r^T [I_k \otimes (\sum_r - \sum_r \sum^{-1} \sum_r)^{-1}] \xi_r$ and η_r is Chi-square distributed with $k(pk+1)$ degrees of freedom. Actually, the aforementioned result can be found in Chan[3], Theorem 2.2 and 2.3. The remaining problem is to find a suitable approximation of the asymptotic distribution of $\sup_r \eta_r$. Moreover, since the distribution of η_r is parameter-free, only depending upon the dimension of data random vector X_t and the order p of the fitted model. Thus, we could infer that under hypothesis H_0, the distribution of $\lambda_{(2)}$ or $\lambda_{(1)}$ will be parameter-free. Later some simulation results will provide partial evidences to the stated conclusion.

When y is sufficiently large, by using the mosaic process property[1], the following approximation can be obtained:

$$\Pr(\sup_{b\le r\le c} \eta_r \le y) \doteq \exp\left[-2\left(\frac{y}{pk+1} - k\right)\chi^2_{k(pk+1)}(y)\left\{\sum_{i=1}^{pk+1}\left(t_i(c) - t_i(b)\right)\right\}\right], \tag{4}$$

where $t_i(r) = 0.5\ln\{\delta_i(r)/[1-\delta_i(r)]\}$, $i = 1, \cdots, pk+1$. $\delta_i(r)$'s are eigenvalues of $\sum^{-1/2} \sum_r \sum^{-1/2}$, and $\chi^2_{(\ell)}(\cdot)$ is the density function of the Chi-square random variable with ℓ degrees of freedom. In fact, when $k = 1$, formula (4) is reduced to the same results as those given by Chan[5], or Wong and Li[14]. Using their approach, a heuristic proof of the approximation (4) is given in Appendix D. Finally, when H_0 is true, an approximate tailed probability of $\lambda_{(j)}$, $j = 1, 2$, can be expressed as

$$\Pr(\lambda_{(j)} \le y) \doteq \exp\left[-2\left(\frac{y}{pk+1}-k\right)\chi^2_{k(pk+1)}(y)\left\{\sum_{i=1}^{pk+1}(t_i(c)-t_i(b))\right\}\right], \tag{5}$$

when y is large enough. Furthermore, without the intercept term, that is, when $\Theta_0 = \Phi_0 = 0$, an approximated tailed probability is reduced to

$$\Pr(\lambda_{(j)} \le y) \doteq \exp\left[-2\left(\frac{y}{pk}-k\right)\chi^2_{pk^2}(y)\left\{\sum_{i=1}^{pk}(t_i(c)-t_i(b))\right\}\right]. \tag{6}$$

It is worthy to note that the approximate probabilities in formulae (5)-(6) both formally depend upon the order of the model from which the data originated.

3. NUMERICAL ILLUSTRATIONS

In this section, the accuracy of formulae (5)-(6), the power of the discussed test statistics, $\lambda_{(i)}$, $i = 1, 2$, and Tsay's test will be examined using simulations.

3.1 Accuracy of the Approximate Tailed Probability

Both formulae (5)-(6) depend upon the eigenvalues of the corresponding matrix $\sum^{-1/2}\sum_r\sum^{-1/2}$. In real applications, this is replaced by $\hat{\sum}^{-1/2}\hat{\sum}_r\hat{\sum}^{-1/2}$, where $\hat{\sum} = X^T X/n$ and $\hat{\sum}_r = Y_r^T Y_r/n$. In order to prevent the initial value from influencing data generated in the following simulations, whenever a data set with n observations is needed, then $n+100$ data are generated, and the last n data are analyzed.

3.1.1 Univariate Case

Now, consider the testing of the univariate threshold nonlinearity. Data are generated using some special autoregressive processes, with or without an intercept term. First, consider cases where the data are generated from a model with an intercept term:

(u1). $(1-\psi_1 B)X_t = \psi_0 + a_t$,

(u2). $\Pi_{i=1}^{2}(1-\psi_i B)X_t = \psi_0 + a_t$,

(u3). $\Pi_{i=1}^{3}(1-\psi_i B)X_t = \psi_0 + a_t$,

(u4). $\Pi_{i=1}^{4}(1-\psi_i B)X_t = \psi_0 + a_t$,

(u5). $\Pi_{i=1}^{5}(1-\psi_i B)X_t = \psi_0 + a_t$.

Here, $a_t \overset{iid}{\sim} \mathrm{N}(0,1)$.

When data are generated from model (uj), the hypothesis testing is formulated as

$$\mathrm{H}_0 : X_t - \theta_0 - \sum_{j=1}^{p}\theta_j X_{t-j} = a_t,$$

as against

$$\mathrm{H}_1 : X_t - \theta_0 - \sum_{j=1}^{p}\theta_j X_{t-j} - I(X_{t-1} \le r)(\varphi_0 + \sum_{j=1}^{p}\varphi_j X_{t-j}) = a_{t,r},$$

where $1 \le p \le 5$. Therefore, the sensitivity of the miss-specification of the order of the autoregressive process is simultaneously investigated. The values of the parameters are set as follows: $\psi_0 = -0.3$ and $(\psi_1^{-1}, \psi_2^{-1}, \psi_3^{-1}, \psi_4^{-1}, \psi_5^{-1}) = (-2.5, +3.8, +1.4, +4.9, -2.0)$. After some sets of data with a sample size n are generated from (uj), an upper percentile point computed from formula (5) is obtained, under a specified order " p " in H_0 and H_1.

For a given upper percentile α, the corresponding percentile point, denoted by $\hat{y}_\alpha$, can be approximately obtained by solving the following equation:

$$\exp\left[-2\left(\frac{\hat{y}_\alpha}{pk+1} - k\right)\chi^2_{k(pk+1)}(\hat{y}_\alpha)\left\{\sum_{i=1}^{pk+1}\left(\hat{t}_i(c) - \hat{t}_i(b)\right)\right\}\right] = 1-\alpha, \tag{7}$$

where b and c are respectively defined as the 10% and 90% data points. That is, $b = X_{(n_1)}$ and $c = X_{(n_2)}$, with $n_1 = 0.1n$ and $n_2 = 0.9n$, and where $\{X_{(t)}\}_{t=1}^n$ are the order statistics for the data set $\{X_t\}_{t=1}^n$. $\hat{t}_i(r) = 0.5\ln\left\{\hat{\delta}_i(r)/[1-\hat{\delta}_i(r)]\right\}$ and $\hat{\delta}_i(r)$'s are the eigenvalues for the matrix $\hat{\Sigma}^{-1/2}\hat{\Sigma}_r\hat{\Sigma}^{-1/2}$. In order to obtain a more accurate percentile point, the average solutions of *M* data sets from equation (7) are used. That is, the α percentile point is estimated by

$$\hat{y}_\alpha = \sum_{m=1}^{\mathrm{M}} \hat{y}_\alpha^{(m)} / \mathrm{M}, \tag{8}$$

where $\hat{y}_\alpha^{(m)}$ is solved from (7) by using the m-th data set.

To numerically verify the accuracy of the percentile points obtained from (8), the empirical distribution of $\lambda_{(1)}$ is estimated, by computing

$$\lambda_{(1)} = \sup_{X_{n_1} \le r \le X_{n_2}} \lambda_r(\hat{\Omega}_{1,r}).$$

Suppose that L sets of data are independently generated from the model (uj); then we obtain $\{\hat{\lambda}_{(1)}^{(\ell)}\}_{\ell=1}^L$. This set can be used to construct an empirical distribution of $\lambda_{(1)}$. Presumably, the empirical results can be regarded as the true one when the sample size is sufficiently large. Simulations are conducted for $n = 500$, $\alpha = 5\%$, 2.5% or 1%, $\mathrm{M} = 2{,}000$ and $L = 2{,}000$. The simulation results expressed in Table 1 indicate that the approximated results are rather close to the empirical results.

To investigate the accuracy in greater detail, the percentile error of the size is computed. Let α_* denote the proportion of cases $\{\hat{\lambda}_{(1)} > \hat{y}_\alpha\}$ among L replications. Then, comparisons between α_* and α, for example when the relative percentage error defined by, $\left(|\alpha_* - \alpha|/\alpha\right)\times 100\%$, will be used to measure the accuracy of formula (5). In summary, the relative percentage errors for $\alpha = 5\%$ are less than 5%. Actually, when the α probability decreases, the percentage errors also decrease. To save time, the percentile points for a case without an intercept term are only investigated for an AR(1) model: Data are generated from $X_t = 0.4X_{t-1} + a_t$, where $a_t \overset{iid}{\sim} \mathrm{N}(0,1)$ and the testing hypothesis is set as $\mathrm{H}_0: X_t - \theta X_{t-1} = a_t$, as against $\mathrm{H}_1: X_t - \theta X_{t-1} - I(X_{t-1} \le r)(\varphi X_{t-1}) = a_{t,r}$.

Both the true model and the hypothetical model are without an intercept term. A procedure similar to the aforementioned one is performed. This time an approximate upper percentile point may be obtained by solving formula (6). Similar conclusions to the ones mentioned above are indicated by the simulation results shown in Table 2. By comparing the percentile points for cases with or without an intercept term,

Table 1: Percentile points of the proposed test statistics

	True model				
p	(u1)	(u2)	(u3)	(u4)	(u5)
			$\alpha = 5\%$		
1	12.89†	12.89	12.92	12.94	12.91
	(12.52)‡	(12.59)	(12.36)	(12.77)	(12.93)
2	15.24	15.24	15.27	15.28	15.27
	(14.43)	(15.11)	(14.95)	(14.84)	(15.31)
3	17.40	17.40	17.43	17.44	17.38
	(16.90)	(16.68)	(17.49)	(16.79)	(17.03)
4	19.44	19.44	19.46	19.47	19.40
	(19.88)	(19.26)	(18.70)	(19.07)	(19.12)
5	21.38	21.38	21.40	21.41	21.32
	(21.51)	(21.07)	(21.49)	(20.86)	(21.24)
			$\alpha = 2.5\%$		
1	14.59†	14.59	14.63	14.64	14.59
	(14.24)‡	(14.45)	(14.03)	(14.41)	(14.35)
2	17.04	17.04	17.07	17.08	17.03
	(16.22)	(16.93)	(16.59)	(16.47)	(16.94)
3	19.28	19.28	19.31	19.32	19.26
	(18.91)	(18.84)	(19.13)	(19.04)	(18.74)
4	21.39	21.39	21.41	21.42	21.35
	(22.37)	(21.47)	(20.69)	(21.37)	(21.24)
5	23.40	23.40	23.42	23.43	23.35
	(23.77)	(23.23)	(23.91)	(22.89)	(23.17)
			$\alpha = 1\%$		
1	16.76†	16.76	16.79	16.81	16.75
	(16.55)‡	(16.92)	(16.50)	(16.32)	(16.79)
2	19.31	19.31	19.34	19.35	19.30
	(18.61)	(18.90)	(18.95)	(18.93)	(18.83)
3	21.65	21.65	21.67	21.68	21.63
	(21.32)	(21.40)	(22.09)	(21.50)	(21.26)
4	23.85	23.85	23.86	23.87	23.80
	(24.05)	(23.61)	(22.98)	(23.11)	(23.39)
5	25.93	25.93	25.95	25.95	25.88
	(26.00)	(25.67)	(25.50)	(25.82)	(25.46)

An AR(p) model is specified in H_0 ; †computed from formula (5); ‡computed from the empirical distribution of $\hat{\lambda}_{(1)}$

the presence of the intercept term seems to cause the tailed distribution of the test statistic $\lambda_{(1)}$ to be slightly heavier. The reason for this is that more parameters lead to a greater uncertainty in the test statistic. It is worthy of note that it is more convenient and saves computing time to use formulae (5)-(6).

Furthermore, by inspecting the numerical results in Tables 1-2, it is quite interesting to find that, for fixed *b* and *c*, when data are generated from an AR(q) model, but with $p < q$ specified in the null hypothesis, then the percentile points will be automatically adjusted. Therefore, the numerical simulations show a remarkable robustness to the miss-specification of the model order. Indeed, Chan[5] indicated similar results that the LRT percentiles depend only on the fitted autoregression order. Therefore, the miss-specification of the order, in particular its under-specification, does not cause difficulty in obtaining critical points. It is thus more convenient in practical applications.

Table 2: Percentile points of test statistics, data generated from $X_t = 0.4X_{t-1} + a_t$

		α	
p	5%	2.5%	1%
1	6.39†	8.12†	10.26†
	(6.77)‡	(8.19)‡	(10.23)‡
2	11.16	12.93	15.14
	(10.88)	(12.54)	(15.16)
3	14.05	15.89	18.20
	(14.02)	(16.09)	(17.91)
4	16.46	18.38	20.78
	(16.80)	(18.38)	(20.91)
5	18.64	20.63	23.11
	(18.49)	(20.85)	(23.76)

An AR(p) model is specified in H_0; †computed from formula (6);

‡computed from the empirical distribution of $\hat{\lambda}_{(1)}$

3.1.2 Multivariate Case

For the multivariate model, for simplicity, a two-dimensional autoregressive threshold model is discussed. Data are generated respectively from the following four models; the first two models have an intercept term while the last two models have no intercept term:

(m1). $X_t = \Psi_0 + \Psi_1 X_{t-1} + a_t$,

(m2). $X_t = \Psi_0 + \Psi_1 X_{t-1} + \Psi_2 X_{t-2} + a_t$,

(m3). $X_t = \Psi_1 X_{t-1} + a_t$,

(m4). $X_t = \Psi_1 X_{t-1} + \Psi_2 X_{t-2} + a_t$,

$$\Psi_0 = \begin{pmatrix} 1.2 \\ -0.5 \end{pmatrix}, \quad \Psi_1 = \begin{pmatrix} 0.86 & -0.80 \\ 0.90 & -1.10 \end{pmatrix} \text{ and } \Psi_2 = \begin{pmatrix} -0.40 & -0.30 \\ -0.53 & -0.21 \end{pmatrix}.$$

Here, $a_t \overset{iid}{\sim} \mathrm{N}(0, \Omega)$, with $\Omega = \begin{pmatrix} 1.20 & 0.72 \\ 0.72 & 1.20 \end{pmatrix}$.

Table 3: Percentile points of the proposed test statistics

	True model			
	with intercept		without intercept	
p	(m1)	(m2)	(m3)	(m4)
		$\alpha = 5\%$		
1	21.54†	21.54	15.82*	15.81
	(20.99)‡	(21.87)	(15.52)*	(16.12)
2	28.55	28.55	24.12	24.12
	(28.60)	(28.61)	(24.19)	(23.93)
		$\alpha = 2.5\%$		
1	23.55†	23.55	17.77*	17.76
	(23.45)‡	(23.72)	(17.11)*	(17.95)
2	30.80	30.81	26.31	26.31
	(30.54)	(30.91)	(26.25)	(26.51)
		$\alpha = 1\%$		
1	26.07†	26.07	20.20*	20.19
	(25.96)‡	(26.60)	(20.10)*	(20.77)
2	33.60	33.60	29.02	29.02
	(33.88)	(33.22)	(29.16)	(28.78)

A VAR(p) model is specified in H_0;
†computed from formula (5); *computed from formula (6)
‡the empirical distribution of $\hat{\lambda}_{(1)}$, with an intercept term
*the empirical distribution of $\hat{\lambda}_{(1)}$, without an intercept term

When data are generated from model (mj), $j = 1, 2$, with an intercept term, the testing hypothesis is formulated as

$$\mathrm{H}_0 : X_t - \Theta_0 - \sum_{j=1}^{p} \Theta_j X_{t-j} = a_t ,$$

as against

$$\mathrm{H}_1 : X_t - \Theta_0 - \sum_{j=1}^{p} \Theta_j X_{t-j} - I(X_{1,t-1} \le r)(\Phi_0 + \sum_{j=1}^{p} \Phi_j X_{t-j}) = a_{t,r} . \quad (9)$$

On the other hand, in cases without an intercept term, the tested hypothesis is restricted by setting $\Theta_0 = \Phi_0 = 0$. To save time, the values of p are investigated for $p = 1, 2$.

Adopting the previous setup, $n = 500$, $\alpha = 5\%$, 2.5% or 1%, $\mathrm{M} = 2,000$ and $L = 2,000$ are assumed. We let $b = X_{1,(n_1)}$ and $c = X_{1,(n_2)}$, again $n_1 = 0.1n$ and $n_2 = 0.9n$, where $\{X_{1,(t)}\}_{t=1}^{n}$ are the order statistics of the first component data set $\{X_{1,t}\}_{t=1}^{n}$. Then, the tailed probability of some specified Type I error probability α is computed; when the fitted model has an intercept term, formula (5) is applied; without an intercept term formula (6) is used. Meanwhile, the desired percentage point, from the empirical distribution of $\lambda_{(1)}$, is estimated by

$$\tilde{\lambda}_{(1)} = \sup_{X_{1,(n_1)} \le r \le X_{1,(n_2)}} \lambda_r(\hat{\Omega}_{1,r}).$$

The results, which are shown in Table 3, indicate that both approaches, the approximated and the empirical, are rather close. The relative percentage errors for each specified α -level are all less than 3% for the intercept term case; without the intercept term, the percentile point values tend to be smaller, with the corresponding relative percentage errors being less than 4%.

Some of the aforementioned numerical studies (Tables 1-2) indicate that in the univariate case, the miss-specification of the order does not cause difficulty in computing critical points. To partially clarify this point in the multivariate case, some simulations are conducted. Data are generated respectively from the five two-dimensional vector autoregressive models denoted by VAR(j), say, $X_t = \Psi_0 + \Psi_1 X_{t-1} + \cdots + \Psi_j X_{t-j} + a_t$, where $j = 1, 2, \cdots, 5$. The definitions of Ψ_0, Ψ_1 and Ψ_2 are the same as above, the remaining Ψ_i's are defined as follows:

$$\Psi_3 = \begin{pmatrix} 0.25 & -0.40 \\ 0.62 & -0.76 \end{pmatrix}, \quad \Psi_4 = \begin{pmatrix} -0.60 & 0.08 \\ -0.05 & -0.40 \end{pmatrix}, \quad \Psi_5 = \begin{pmatrix} 0.50 & 0.10 \\ 0.25 & -0.20 \end{pmatrix}.$$

Again, $a_t \overset{iid}{\sim} \mathrm{N}(0, \Omega)$, with $\Omega = \begin{pmatrix} 1.20 & 0.72 \\ 0.72 & 1.20 \end{pmatrix}$ and the testing hypothesis is formulated as before, (9).

Table 4: Examinations of the percentile points in the multivariate cases

	True model				
p	VAR(1)	VAR(2)	VAR(3)	VAR(4)	VAR(5)
			$\alpha = 5\%$		
1	21.54	21.54	21.55	2159	21.58
2	28.57	28.57	28.60	28.70	28.62
3	35.15	35.14	35.15	35.21	35.13
4	41.36	41.36	41.36	41.41	41.31
5	47.35	47.35	47.36	47.35	47.27
			$\alpha = 2.5\%$		
1	23.55	23.55	23.59	23.56	23.58
2	30.80	30.81	30.85	30.81	30.82
3	37.58	37.58	37.58	37.68	37.57
4	43.95	43.95	43.96	44.00	43.91
5	50.10	50.10	50.11	50.09	50.08
			$\alpha = 1\%$		
1	26.07	26.08	26.11	26.21	26.12
2	33.61	33.60	33.61	33.72	33.61
3	40.59	40.59	40.60	40.67	40.58
4	47.16	47.16	47.17	47.21	47.13
5	53.48	53.48	53.49	53.47	53.48

A VAR(p) model is specified in H_0; Each entry is computed from formula (5)

To investigate which percentile points are independent of the order of the vector autoregressive model generated by the data, the fitted order *p* is considered for $p = 1, 2, \cdots, 5$. The remaining conditions are the same as those that produced Table 3. By inspecting the numerical results listed in Table 4, we may conclude that the percentile points are parameter-free and only depend upon the order of the fitted model in the discussed multivariate case again. This advantage is helpful in practical applications.

3.2 Power Comparisons

In this subsection, the power of the two defined test procedures, $\lambda_{(1)}$ and $\lambda_{(2)}$, and of Tsay's test procedure, will be compared by simulations, for both univariate and multivariate data.

3.2.1 Univariate Case

Adopting Tsay's[11] numerical framework, consider the following univariate threshold model,

$$X_t = \begin{cases} \mu + 0.5X_{t-1} + a_t, & if \quad X_{t-1} < r, \\ \mu + \varphi X_{t-1} + a_t, & if \quad X_{t-1} \geq r, \end{cases} \tag{10}$$

where the a_t's are iid N(0,1) distributed and *n*=50 or 100 are set. Different values of φ are set, say -2, -1, -0.5, 0 and 0.5. The testing power for $\varphi = 0.5$ is then just the Type I error probability, which can be used to verify the applicability of the parameter-free conjecture, and the accuracy of formulae (5)-(6), in particular, in situations with a small sample size. To investigate the impact of the intercept term, the values of (μ, r) are set as (1.0, 1.0) or (0.0, 0.0).

For simplicity, the testing hypothesis is formulated by setting $p = 1$ in (9). Moreover, the presence of an intercept term in the testing hypothesis depends upon the value of μ. For each specified parameter value, 10,000 sets of data are generated from model (10). For each data set, the value of the test statistic $\hat{\lambda}_{(i)}$ is computed, to decide whether the null hypothesis should be rejected. $\hat{\lambda}_{(1)}$ has been defined in subsection 3.1.1 and

$$\hat{\lambda}_{(2)} = \sup_{X_{(n_1)} \leq r \leq X_{(n_2)}} \lambda_r(\hat{\Omega}_2).$$

The critical values may be obtained from Tables 1-2. They are computed from (5) or (6); say 12.89 for $\alpha = 5\%$ and 16.76 for $\alpha = 1\%$, when the intercept term exists in the fitted model; without the intercept term, 6.39 and 10.26 are used respectively as the corresponding critical point.

Let us first consider a model with an intercept, $(\mu, r) = (1.0, 1.0)$: When $\varphi = 0.5$, the powers of $\hat{\lambda}_{(1)}$ are quite close to the specified Type I error probability. This means that the approximation formula (5) still works well for small sample sizes. However, the other two test statistics, $\hat{\lambda}_{(2)}$, or Tsay's test statistic, are both less than the specified Type I error probability. This will cause the test statistic to be slightly conservative. They will therefore be less powerful than Chan and Tong's test when the true model deviates gradually from the null hypothesis.

Next, we consider a case without an intercept term, $(\mu, r) = (0.0, 0.0)$: The powers of Tsay's test, at $\varphi = 0.5$, will be larger than the specified probability, especially for a small error probability, say 1%. Under the alternative hypothesis, this will lead to an over-estimation of the power. Moreover, as the alternative hypothesis is far from the null hypothesis, especially, in a non-stationary situation, the power of Tsay's test is rather low. The powers of $\hat{\lambda}_{(1)}$, at $\varphi = 0.5$, is rather close to the desired one, but is sometimes over, therefore, it may be unfair to discuss the testing power further. However, the powers of $\hat{\lambda}_{(2)}$, at $\varphi = 0.5$, is rather conservative compared to that of $\hat{\lambda}_{(1)}$, sometimes less than the specified one.

Nevertheless, even in the case when less than the given Type I error probability, the power is almost higher than that of Tsay's test. From the numerical results presented in Table 5, in summary, if there is an intercept term, the $\hat{\lambda}_{(1)}$ test statistic is suggested, while the $\hat{\lambda}_{(2)}$ test statistic is suggested when the intercept term is absent.

Table 5: Probability (%) of rejecting an AR model when the data follow model (10)

	$(\mu, r) = (1.0, 1.0)$					$(\mu, r) = (0.0, 0.0)$				
	φ					φ				
n	-2	-1	-0.5	0	0.5	-2	-1	-0.5	0	0.5
					$\alpha = 5\%$					
50	100.0*	98.4	68.9	14.7	4.6	99.9*	98.3	84.0	34.9	6.9
	100.0*	94.7	45.9	4.3	1.0	99.9*	97.2	79.1	27.6	4.5
	(100.0)†	(99.0)	(67.2)	(12.0)	(3.8)	(61.2)†	(53.9)	(42.6)	(20.7)	(5.9)
100	100.0	100.0	95.3	25.3	4.2	100.0	100.0	99.3	64.4	6.4
	100.0	100.0	91.6	16.6	1.9	100.0	100.0	99.1	60.9	5.2
	(100.0)	(100.0)	(94.3)	(23.7)	(2.6)	(85.6)	(80.4)	(70.1)	(37.4)	(5.9)
					$\alpha = 1\%$					
50	100.0*	95.6	51.1	5.8	1.1	99.7*	92.7	61.8	13.5	1.1
	100.0*	81.8	19.6	0.5	0.0	99.1*	86.1	45.5	6.5	0.3
	(100.0)†	(92.0)	(41.6)	(3.5)	(0.9)	(41.8)†	(35.7)	(25.1)	(9.3)	(1.8)
100	100.0	100.0	87.5	11.6	0.9	100.0	99.9	95.3	35.7	1.3
	100.0	99.9	77.2	5.3	0.2	100.0	99.9	92.7	29.2	0.7
	(100.0)	(100.0)	(83.6)	(8.7)	(0.6)	(74.0)	(67.2)	(53.5)	(22.9)	(1.9)

* $\hat{\lambda}_{(1)}$ test statistic, using 10,000 replications; * $\hat{\lambda}_{(2)}$ test statistic, using 10,000 replications;
†coded from Tsay[11], using 1,000 replications

3.2.2 Multivariate Case

To adopt the discussion to a univariate case, the same framework will be utilized for multivariate data. First, we consider the following two-dimensional threshold model with an intercept term:

$$X_t = \begin{cases} \Psi_0 + \Psi_1 X_{t-1} + a_t, & if \quad X_{1,t-1} < 0.2, \\ \Psi_0 + \Gamma_i X_{t-1} + a_t, & if \quad X_{1,t-1} \geq 0.2. \end{cases} \tag{11}$$

The Ψ_i's are defined in subsection 3.1.2.

$$\Gamma_1 = \begin{pmatrix} 3.86 & -1.10 \\ 13.25 & -4.00 \end{pmatrix}, \quad \Gamma_2 = \begin{pmatrix} 2.86 & -0.90 \\ 9.42 & -3.10 \end{pmatrix} \text{ and } \Gamma_3 = \Psi_1.$$

The eigenvalues of the Γ_i matrices are, respectively, (-0.94, 0.80) for Γ_1, (-0.75, 0.51) for Γ_2, and (-0.61, 0.37) for Γ_3. Therefore, the Γ_2 matrix is closer to the Γ_3 matrix than the Γ_1 matrix.

Wong and Li[14] have derived Lagrange-multiplier test statistic for a threshold autoregression with conditional heteroscedasticity. Owing to the complexity of the distribution of the test statistic, they suggested that the percentile points derived under homoscedasticity be used when conditional

heteroscedasticity exists. The applicability of the percentile points listed in Table 3 will be examined when the a_t's follow the ARCH(1) model. Here, the distribution of the a_t's is assumed to follow $\mathrm{N}(0, \mathrm{H}_t)$. To save time, only the iid normality and the ARCH(1) are investigated. For the latter, $\mathrm{H}_t = \mathrm{A} + \mathrm{B}^T a_{t-1} a_{t-1}^T \mathrm{B}$, where

$$\mathrm{A} = \begin{pmatrix} 1.20 & 0.72 \\ 0.72 & 1.20 \end{pmatrix}, \qquad \mathrm{B} = \begin{pmatrix} 0.10 & 0.20 \\ -0.20 & 0.30 \end{pmatrix},$$

while for the former, $\mathrm{H}_t = \mathrm{A}$ is assumed. The sample sizes are set as $n = 150$ or 300. Again, for each specified parameter value, 10,000 sets of data are generated from model (11). The critical values are obtained from Table 3, computed with (5), say 21.54 for $\alpha = 5\%$ and 26.07 for $\alpha = 1\%$.

Intuitively, since the Γ_1 matrix is farthest from the Ψ_1 matrix, it should be best able to detect the existence of the threshold nonlinearity. The power of the Γ_3 matrix should be quite close to the specified Type I error probability. The numerical results exhibited in Table 6 really do reflect this phenomenon. Again, in the normally distributed case, the power of $\tilde{\lambda}_{(1)}$ under Γ_3 is closer to the specified Type I error probability. On the other hand, in the case of ARCH(1), the power of $\tilde{\lambda}_{(2)}$, where $\tilde{\lambda}_{(2)} = \sup_{X_{1,(n_1)} \le r \le X_{1,(n_2)}} \lambda_r(\hat{\Omega}_2)$, under Γ_3, is closer to the specified significant level. Similarly, we would suggest that $\tilde{\lambda}_{(1)}$ is suitable for cases with normally distributed noise; while $\tilde{\lambda}_{(2)}$ is suitable for cases of heteroscedasticity.

Table 6: Power(%) of detecting the threshold effect under model (11)

	Normal			ARCH(1)		
n	Γ_1	Γ_2	Γ_3	Γ_1	Γ_2	Γ_3
			$\alpha = 5\%$			
150	95.6*	81.3	5.1	98.8*	91.8	8.2
	95.1†	78.6	2.5	98.7†	90.8	4.0
300	99.5	94.6	4.6	99.9	99.0	7.2
	99.4	94.2	3.1	99.9	99.0	5.0
			$\alpha = 1\%$			
150	94.9*	77.9	1.1	98.6*	90.3	2.4
	94.4†	75.3	0.4	98.5†	89.4	0.8
300	99.4	93.3	1.0	99.9	98.8	2.1
	99.3	92.8	0.6	99.9	98.7	1.2

α denotes the Type I error probability; * $\tilde{\lambda}_{(1)}$ test statistic; † $\tilde{\lambda}_{(2)}$ test statistic

Tsay[12] considered the following two-dimensional vector model without an intercept term:

$$X_t = \begin{cases} \Delta_1 X_{t-1} + a_t^{(1)}, & \text{if} \quad X_{1,t-1} < 0, \\ \Delta_2 X_{t-1} + a_t^{(2)}, & \text{if} \quad X_{1,t-1} \ge 0, \end{cases} \tag{12}$$

where $a_t^{(i)} \sim \mathrm{N}(0, \Sigma_i)$,

$$\Delta_1 = \begin{pmatrix} 0.7 & 0.0 \\ 0.3 & 0.7 \end{pmatrix}, \ \Sigma_1 = \begin{pmatrix} 1.0 & 0.2 \\ 0.2 & 1.0 \end{pmatrix}, \ \Delta_2 = \begin{pmatrix} -0.7 & 0.0 \\ -0.3 & -0.7 \end{pmatrix} \text{ and } \Sigma_2 = \begin{pmatrix} 1.0 & -0.3 \\ -0.3 & 1.0 \end{pmatrix}.$$

Our testing hypothesis is set as:

$$\mathrm{H}_0 : X_t - \Theta X_{t-1} = a_t ,$$

as against

$$\mathrm{H}_1 : X_t - \Theta X_{t-1} - I(X_{1,t-d} \le r)(\Phi X_{t-1}) = a_{t,r} . \tag{13}$$

In the aforementioned discussion, the lag parameter d is treated as known. However, in practical usage, it may be unknown. To examine how the power is influenced when the value of d is miss-specified, different values of d will be examined, say $d = 1, 2, 3, 4$. According to Tsay's demonstrations, the sample size is set to be $n = 150$ and $n = 300$. To save time, only $\alpha = 5\%$ is used, with the corresponding critical point, 15.82, being obtained from Table 3, via formula (6). 10,000 data sets are generated with model (12). The results, exhibited in Table 7, again show that the proposed test statistics for either $\tilde{\lambda}_{(1)}$ or $\tilde{\lambda}_{(2)}$ are more powerful than Tsay's test statistic, whether the lag parameter d is correctly specified or not.

Table 7: Power (%) of detecting the threshold effect, α =5%

	Model (12)				Model (14)			
	True *d*	Misspecified *d*			True *d*	Misspecified *d*		
n	1	2	3	4	1	2	3	4
	100.0*	67.9	33.0	24.8	100.0*	96.2	65.9	42.4
150	100.0†	60.8	26.8	19.1	100.0†	94.5	59.3	35.1
	(99.4)‡	(46.2)	(23.2)	(16.9)	(87.7)‡	(34.3)	(25.4)	(18.5)
	100.0	94.4	52.0	32.6	100.0	100.0	86.1	54.7
300	100.0	92.7	47.4	28.9	100.0	99.9	84.4	52.4
	(100.0)	(80.3)	(43.9)	(25.6)	(97.9)	(66.8)	(43.3)	(29.6)

* $\tilde{\lambda}_{(1)}$ test statistic; † $\tilde{\lambda}_{(2)}$ test statistic; ‡coded from Tsay[12]

Finally, data generated from the following three regime vector TAR models are investigated:

$$X_t = \begin{cases} \Pi_1 X_{t-1} + a_t, & \text{if} \quad X_{1,t-1} < -3.3, \\ \Pi_2 X_{t-1} + a_t, & \text{if} \quad -3.3 \le X_{1,t-1} < 3.3, \\ \Pi_3 X_{t-1} + a_t, & \text{if} \quad X_{1,t-1} \ge 3.3, \end{cases} \tag{14}$$

here $a_t \sim \mathrm{N}(0, I_2)$, $\Pi_1 = \begin{pmatrix} -0.9 & 0.0 \\ 0.2 & -0.9 \end{pmatrix}$, $\Pi_2 = \begin{pmatrix} 1.2 & 0.0 \\ 0.0 & 0.6 \end{pmatrix}$ and $\Pi_3 = \begin{pmatrix} -0.8 & 0.0 \\ 0.2 & 0.8 \end{pmatrix}$.

Again, the hypothesis is formulated as (13). Results exhibited in Table 7 indicate that both the $\tilde{\lambda}_{(1)}$ and $\tilde{\lambda}_{(2)}$ test statistics perform significantly better than does Tsay's test.

4. CONCLUSIONS

This article has proposed some test procedures, an extension of Chan[3], Chan and Tong[4], and Wong and Li's[14] works, for multivariate threshold autoregression with conditional homoscedastic martingale difference noise sequences. An approximate asymptotic tailed distribution of the proposed test is derived. Some numerical illustrations, which include an examination of the accuracy of the approximate tailed distribution and the power of the suggested testing procedure are also investigated. The simulation results

suggest that the approximate tailed distribution is rather accurate and parameter-free on the original model. Though the results depend on the order of the fitted autoregression model, however, a remarkable robustness to a miss-specification of the model order is shown. These results have also been pointed out by Chan[5] in a univariate study. Some simulations indicate that this property still holds to the discussed multivariate models. This makes the proposed testing procedure simpler and more convenient in real applications.

Furthermore, based on the simulated data, we suggest that when the noise is normally distributed, the test statistic $\lambda_{(1)}$ be used; while under a conditional heteroscedasticity situation, the test statistic $\lambda_{(2)}$ may be safer to use to guarantee that the Type I error probability will be closer to the specified one. Moreover, for small or moderate sample sizes, the simulated results show some evidence that the proposed test may be more powerful than Tsay's test[11,12], for the discussed cases. Tsay's test essentially treating the arranged autoregression model as a linear regression model, the core test statistic of the threshold effect is the simple F -statistic. Since the dependence between response variables and explanatory variables are ignored, this may cause the test to be less informative, and thus less powerful than the discussed test in this paper.

In this paper, the discussed threshold model is defined by a simple threshold structure, an indicator function that is only dependent upon one component. A further theoretical study on how to relax the threshold restriction would be worthy of investigation.

ACKNOWLEDGEMENTS

This research was supported by the National Science Council, the Republic of China under contract #NSC90-2121-M-008-020. The author wishes to thank the anonymous referee for valuable and constructive suggestions and comments which lead to a considerable improvement in this article, and Mr. Tsay M. C. for his assistance in carrying out part of the computing work.

REFERENCES

[1] Aldous, D. (1989). *Probability approximation via the Poisson clumping heuristic*. New York: Springer-Verlag.

[2] Basilevsky, A. (1983). *Applied matrix algebra in the statistical sciences*. Amsterdam: North-Holland.

[3] Chan, K. S. (1990). Testing for threshold autoregression. *The Annals of Statistics, 18*(4), 1886-1894.

[4] Chan, K. S., & Tong, H. (1990). On likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society, Series B, 52*(3), 469-476.

[5] Chan, K. S. (1991). Percentage points of likelihood ratio tests for autoregression. *Journal of the Royal Statistical Society, Series B,53*(3), 691-696.

[6] Karatzas, S. E. S. (1991). *Brownian motion and stochastic calculus* (2nd edition). New York: Springer-Verlag.

[7] Lai, T. L., & Wei, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *The Annals of Statistics, 10*(1), 154-166.

[8] Tong, H. (1978). *On a threshold model in pattern recognition and signal processing* (ed. by C. H. Chen). Amsterdam: Sijhoff & Noordhoff.

[9] Tong, H. (1983). *Threshold nodels in nonlinear time series analysis*. Springer Lecture Notes in Statistics, 21. New York: Springer-Verlag.

[10] Tong, H., & Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion). *Journal of the Royal Statistical Society, Series B, 42*(3), 245-292.

[11] Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association, 84*(405), 231-240.

[12] Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association, 93*(443), 1188-1202.

[13] Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematics Society, 54*(3), 426-482.

[14] Wong, C. S., & Li, W. K. (1997). Testing for threshold autoregression with conditional heteroscedasticity. *Biometrika, 84*(2), 407-418.

Appendix A

Proof that $\lambda_r(\hat{\Omega}_2)$ is asymptotically the Lagrange-multiplier test.

Under a normality assumption, the log-likelihood function of model (3) is

$$l \propto -\frac{1}{2}\ln\left|I_{n-p}\otimes\Omega_{1,r}\right| - \frac{1}{2}(\varepsilon_{1,r}^V)^T(I_{n-p}\otimes\Omega_{1,r})^{-1}(\varepsilon_{1,r}^V).$$

Let $\beta = \begin{pmatrix}\Theta_{(F)}^V \\ \Phi_{(F)}^V \\ \Omega_{1,r}^V\end{pmatrix}$ and $\hat{\beta} = \begin{pmatrix}\hat{\Theta}_{(R)}^V \\ 0 \\ \hat{\Omega}_2^V\end{pmatrix}$ be the MLE under H_0. Then the Lagrange-multiplier test statistic λ_{LM} is

$$\lambda_{LM} = -\left(\frac{\partial l}{\partial\beta}|_{\hat{\beta}}\right)^T\left[E\left(\frac{\partial^2 l}{\partial\beta\partial\beta^T}\right)|_{\hat{\beta}}\right]^{-1}\left(\frac{\partial l}{\partial\beta}|_{\hat{\beta}}\right).$$

After algebra, it can be shown that

$$\frac{\partial l}{\partial\Theta_{(F)}^V}|_{\hat{\beta}} = 0,\ \frac{\partial l}{\partial\Phi_{(F)}^V}|_{\hat{\beta}} = (\hat{\Omega}_2^{-1}\otimes Y_r^T G)Y^V,\ \frac{\partial l}{\partial\left(\Omega_{1,r}^{-1}\right)^V}|_{\hat{\beta}} = 0,$$

$$E\left(\frac{\partial^2 l}{\partial(\Omega_{1,r}^{-1})^V\partial(\Theta_{(F)}^V)^T}\right) = E\left(\frac{\partial^2 l}{\partial(\Omega_{1,r}^{-1})^V\partial(\Phi_{(F)}^V)^T}\right) = 0,$$

$$\frac{\partial^2 l}{\partial\Phi_{(F)}^V\partial(\Theta_{(F)}^V)^T}|_{\hat{\beta}} = -(\hat{\Omega}_2^{-1}\otimes Y_r^T X),$$

$$\frac{\partial^2 l}{\partial\Theta_{(F)}^V\partial(\Theta_{(F)}^V)^T}|_{\hat{\beta}} = -(\hat{\Omega}_2^{-1}\otimes X^T X) \text{ and } \frac{\partial^2 l}{\partial\Phi_{(F)}^V\partial(\Phi_{(F)}^V)^T}|_{\hat{\beta}} = -(\hat{\Omega}_2^{-1}\otimes Y_r^T Y_r).$$

Therefore, the Lagrange-multiplier test statistic is asymptotically equivalent to

$$\lambda_r(\hat{\Omega}_2) = \omega_r^T\{I_k\otimes(Y_r^T G Y_r)^{-1}\}\omega_r,$$

where $\omega_r = (\hat{\Omega}_2^{-1/2}\otimes Y_r^T G)Y^V$.

Appendix B

Proof that $\hat{\Phi}_{(F)}^V$ is asymptotically normally distributed.

Model (2) could be re-written as

$$Y^V = \begin{pmatrix}I_k\otimes X & I_k\otimes Y_r\end{pmatrix}\begin{pmatrix}\Theta_{(F)}^V \\ \Phi_{(F)}^V\end{pmatrix} + \varepsilon_{1,r}^V = Z\begin{pmatrix}\Theta_{(F)}^V \\ \Phi_{(F)}^V\end{pmatrix} + \varepsilon_{1,r}^V,$$

where $Z = (I_k\otimes X \quad I_k\otimes Y_r)$. By assumption, $\frac{1}{n}Z^T Z$ converges almost surely. Define $q = 2k^2(p+1)$, $m = k(n-p)$, z_i^T the i-th row vector of Z, say, $z_i^T = \begin{pmatrix}z_{i1} & z_{i2} & \cdots & z_{iq}\end{pmatrix}$, and $Z^T = \begin{pmatrix}z_1 & z_2 & \cdots & z_m\end{pmatrix}$, then

$$\frac{1}{n}Z^T Z = \frac{1}{n}\sum_{i=1}^{m} z_i z_i^T = \frac{1}{n}\begin{pmatrix} \sum_{i=1}^{m} z_{i1}z_{i1} & \sum_{i=1}^{m} z_{i1}z_{i2} & \cdots & \sum_{i=1}^{m} z_{i1}z_{iq} \\ \sum_{i=1}^{m} z_{i2}z_{i1} & \sum_{i=1}^{m} z_{i2}z_{i2} & \cdots & \sum_{i=1}^{m} z_{i2}z_{iq} \\ & \vdots & & \\ \sum_{i=1}^{m} z_{iq}z_{i1} & \sum_{i=1}^{m} z_{iq}z_{i2} & \cdots & \sum_{i=1}^{m} z_{iq}z_{iq} \end{pmatrix}.$$

Let $S_m = \sum_{i=1}^{m}\sum_{j=1}^{q} z_{ij}z_{ij} = \sum_{i=1}^{m}\|z_i\|^2$, then $\frac{S_m}{m} = \frac{S_m}{n}\cdot\frac{n}{m} \xrightarrow{a.s.} \gamma$, and $\frac{S_m}{m} - \frac{S_{m-1}}{m} \xrightarrow{a.s.} 0$.

Thus $\frac{\|z_m\|}{\sqrt{m}} \xrightarrow{a.s.} 0$; that is $\forall \varepsilon > 0$, $\exists M$, such that $\frac{\|z_m\|}{\sqrt{m}} < \varepsilon$, as $m \geq M$. Moreover,

$\max_{1\leq i\leq m}\frac{\|z_i\|}{\sqrt{m}} \leq \max_{1\leq i\leq M}\frac{\|z_i\|}{\sqrt{m}} + \max_{M\leq i\leq m}\frac{\|z_i\|}{\sqrt{m}} \leq c^*/\sqrt{m} + \varepsilon \underset{m\to\infty}{\to} 0$, where $c^* = \max_{1\leq i\leq M}\|z_i\| < \infty$.

Therefore, applying result of Theorem 3 of Lai and Wei (page 164), and taking $B_n = \sqrt{k(n-p)}$, then the asymptotically normally distributed of $\hat{\Phi}^V_{(F)}$ is established.

Appendix C

Prove that under H_0,

$Cov(n^{-1/2}(\hat{\Omega}_2^{-1/2}\otimes Y_g^T G)Y^V, n^{-1/2}(\hat{\Omega}_2^{-1/2}\otimes Y_h^T G)Y^V) \xrightarrow[n\to\infty]{} I_k \otimes (\Sigma_{g\wedge h} - \Sigma_g \Sigma^{-1}\Sigma_h)$ pf): By assumption, under H_0,

$$n^{-1/2}\left\{\left(\hat{\Omega}_2^{-1/2}\otimes Y_g^T G\right)Y^V - \left(\Omega_2^{-1/2}\otimes W_g^T\right)\varepsilon_2^V\right\} \xrightarrow[n\to\infty]{a.s.} 0,$$

where $W_g = Y_g - X\Sigma^{-1}\Sigma_g$. Let $(W_g)_{ij}$ be the (i,j)-th entry of matrix W_g, therefore by definition $(W_g)_{ij}$ is a function of $\{X_\ell : \ell \leq p+i-1\}$. Let $\varepsilon_2^V = (e_1^T, e_2^T, \cdots, e_k^T)^T$ and $e_i = (e_{p+1,i}, e_{p+2,i}, \cdots, e_{n,i})^T$, then under H_0,

$$Cov(e_{i\tau}, e_{j\nu}\,|\,F_{i\wedge j-1}) = \begin{cases} 0 & , \text{ if } i \neq j, \\ (\Omega_2)_{\tau\nu} & , \text{otherwise.} \end{cases}$$

Define $Z_g = (\Omega_2^{-1/2}\otimes W_g^T)$, $\varepsilon_2^V = (Z_{1g}^T, Z_{2g}^T, \cdots, Z_{kg}^T)^T$ and $Z_{ig} = \sum_{\nu=1}^{k}(\Omega_2^{-1/2})_{i\nu} W_g^T e_\nu$. The s-th element of Z_{ig} is

$$(Z_{ig})_s = \sum_{\tau=1}^{k}(\Omega_2^{-1/2})_{i\tau}\{\sum_{\nu=1}^{n-p}(W_g)_{\nu s} e_{p+\nu,\tau}\} = \sum_{\nu=1}^{n-p}(W_g)_{\nu s} u_{p+\nu,i},$$

where $u_{vi} = \sum_{T=1}^{k} e_{vT}(\Omega_2^{-1/2})_{iT}$. Therefore, under H_0

$$Cov(u_{iq}, u_{j\ell}\,|\,F_{i\wedge j-1}) = \begin{cases} 0 & \text{if } i \neq j, \\ \delta_{q\ell} & \text{otherwise} \end{cases}.$$

Here, $\delta_{q\ell}=1$, if $q=\ell;=0$, otherwise. After algebra, $E(Z_{ig}Z_{jh}^T)=\delta_{ij}E(W_g^TW_h)$. Since $n^{-1}W_g^TW_h \xrightarrow[n\to\infty]{a.s.} \sum_{g\wedge h}-\sum_g\sum^{-1}\sum_h$, thus $n^{-1}E(Z_{i,g}Z_{i,h}^T)\xrightarrow[n\to\infty]{}\delta_{ij}(\sum_{g\wedge h}-\sum_g\sum^{-1}\sum_h)$. Finally, we have $n^{-1}Cov(Z_g,Z_h)\xrightarrow[n\to\infty]{}I_k\otimes(\sum_{g\wedge h}-\sum_g\sum^{-1}\sum_h)$.

Appendix D

Derivation of the approximate tailed distribution (4), under hypothesis H_0.

Let $\zeta_r=(I_k\otimes\sum^{-1/2})\xi_r$ and $\Gamma_r=\sum^{-1/2}\sum_r\sum^{-1/2}$, then covariance between ζ_r, is $Cov(\zeta_g,\zeta_h)=I_k\otimes(\Gamma_{g\wedge h}-\Gamma_g\Gamma_h)$. Moreover, since Γ_r is symmetric, by using Basilevsky's result[2], there exists an orthogonal matrix Q_r, such that $Q_r\Gamma_rQ_r^T=D_r$, where $D_r=diag(\delta_1(r),\cdots,\delta_{pk+1}(r))$ and $\delta_1(r)\ge\delta_2(r)\ge\cdots\ge\delta_{pk+1}(r)$, for each r. Then after algebra,

$$\begin{aligned}\eta_r&=\left[(I_k\otimes Q_r)\zeta_r\right]^T\left\{I_k\otimes Q_r(\Gamma_r-\Gamma_r\Gamma_r)Q_r^T\right\}^{-1}\left[(I_k\otimes Q_r)\zeta_r\right]\\&=\left[(I_k\otimes Q_r)\zeta_r\right]^T\left\{I_k\otimes(D_r-D_rD_r)\right\}^{-1}\left[(I_k\otimes Q_r)\zeta_r\right]\\&=\sum_{j=1}^{k}\sum_{i=1}^{pk+1}\frac{(\mathrm{B}_{\delta_i(r)}^{(j)})^2}{\delta_i(r)-\delta_i^2(r)},\end{aligned}$$

where $(I_k\otimes Q_r)\zeta_r=(\mathrm{B}_{\delta_1(r)}^{(1)},\cdots,\mathrm{B}_{\delta_{pk+1}(r)}^{(1)},\mathrm{B}_{\delta_1(r)}^{(2)},\cdots,\mathrm{B}_{\delta_{pk+1}(r)}^{(2)},\cdots,\mathrm{B}_{\delta_1(r)}^{(k)},\cdots,\mathrm{B}_{\delta_{pk+1}(r)}^{(k)})^T$.

For a fixed j, $\{\mathrm{B}_{\delta_i(r)}^{(j)}\}_{\delta_i(r)}$ is a Gaussian process with mean zero and

$$Cov(\mathrm{B}_{\delta_u(g)}^{(e)},\mathrm{B}_{\delta_v(h)}^{(f)})=0,$$

for $e\ne f$ or $u\ne v$. Let $\zeta_g=\left(\zeta_{1,g}^T,\zeta_{2,g}^T,\cdots,\zeta_{k,g}^T\right)^T$, then $Q_g\zeta_{j,g}=(\mathrm{B}_{\delta_1(g)}^{(j)},\cdots,\mathrm{B}_{\delta_{pk+1}(g)}^{(j)})^T$. Since $Cov(\zeta_g,\zeta_h)=I_k\otimes\left(\Gamma_{g\wedge h}-\Gamma_g\Gamma_h\right)$, therefore, $Cov\left(\zeta_{i,g},\zeta_{i,h}\right)=\left(\Gamma_{g\wedge h}-\Gamma_g\Gamma_h\right)$, $Cov\left(Q_g\zeta_{i,g},Q_h\zeta_{i,h}\right)=Q_g(\Gamma_{g\wedge h}-\Gamma_g\Gamma_h)Q_h^T$, and

$$Cov(\mathrm{B}_{\delta_i(g)}^{(j)},\mathrm{B}_{\delta_i(h)}^{(j)})=\{\delta_i(g\wedge h)-\delta_i(g)\delta_i(h)\}Q_{i,g}^TQ_{i,h},$$

where $Q_{i,g}^T$ is the i-th row of the matrix Q_g.

Since $Q_{i,g}^TQ_{i,h}=\|Q_{i,g}\|\|Q_{i,h}\|Cos\theta_{i,g,h}=Cos\theta_{i,g,h}$, where $\theta_{i,g,h}$ is an angle between the two vectors $Q_{i,g}$ and $Q_{i,h}$, by a first order Taylor expansion of $\theta_{i,g,h}$ at $\theta_{i,g,h}=0$, then $Q_{i,g}^TQ_{i,h}\doteq1$. The monotone property of $\delta_i(r)$ in i will make the true $\theta_{i,g,h}$ more close to 0. Therefore, in the following discussion, we treat

$$Cov(\mathrm{B}_{\delta_i(g)}^{(j)},\mathrm{B}_{\delta_i(h)}^{(j)})=\delta_i(g\wedge h)-\delta_i(g)\delta_i(h).$$

Define $t_i(r)=0.5\ln\left(\kappa_i(r)\right)$, $\kappa_i(r)=\delta_i(r)/\left(1-\delta_i(r)\right)$, and $X_{\kappa_i(r)}^{(j)}=\mathrm{B}_{\delta_i(r)}^{(j)}/\left(1-\delta_i(r)\right)$, then $\kappa_i(r)=\exp\left(2t_i(r)\right)$, $\delta_i(r)=\kappa_i(r)/\left(1+\kappa_i(r)\right)$, and $\left\{X_{\kappa_i(r)}^{(j)}\right\}$ is a Gaussian process, such that $Cov\left(X_{\kappa_i(r)}^{(j)},X_{\kappa_i(s)}^{(j)}\right)=\kappa_i(r\wedge s)$. That is $\left\{X_{\kappa_i(r)}^{(j)}\right\}_{\kappa_i(r)}$ is a Brownian motion.

Define

$$U_{j,i}(t_i(r)) = \frac{\mathrm{B}^{(j)}_{\delta_i(r)}}{\sqrt{\delta_i(r) - \delta_i^2(r)}} = X^{(j)}_{\exp(2t_i(r))} \cdot \exp\left(-t_i(r)\right),$$

then by results of Karatzas[6], $\{U_{j,i}(t_i(r))\}_{t_i(r)}$ becomes an independent Ornstein-Uhlenbeck process with covariance, $Cov[U_{j,i}(t_i(r+s)),\ U_{j,i}(t_i(s))] = \exp\left(-t_i(r+s) + t_i(s)\right)$. By using Aldous's result[1], $U_{j,i}(t_i(r))$ satisfies the following stochastic differential equation

$$dU_{j,i}(t_i(r)) = -U_{j,i}(t_i(r))dt_i(r) + \sqrt{2}dW_{j,i}(t_i(r)),$$

where $\left\{W_{j,i}(t_i(r))\right\}_{t_i(r)}$ is the Brownian motion. After algebra, it turns out that

$$\eta_r = \sum_{j=1}^{k}\sum_{i=1}^{pk+1} U^2_{j,i}(t_i(r)),$$

and by using Ito's lemma

$$d\eta_r = -2\sum_{j=1}^{k}\sum_{i=1}^{pk+1}[U^2_{j,i}(t_i(r)) - 1]dt_i(r) + 2\sqrt{2}\sum_{j=1}^{k}\sum_{i=1}^{pk+1} U^2_{j,i}(t_i(r))dW_{j,i}(t_i(r)).$$

If $r \in (b,c)$, then by using the properties of the mosaic process[1], the following approximation can be obtained:

$$\Pr(\sup_{b \le r \le c} \eta_r \le y) \doteq \exp\left[-\int_b^c \alpha(r)dr\right].$$

Here the clump rate $\alpha(r)$ is defined as

$$\begin{aligned}\alpha(r) &= E\left\{\max\left(0, 2\sum_{j=1}^{k}\sum_{i=1}^{pk+1}[U^2_{j,i}(t_i(r)) - 1]\frac{dt_i(r)}{dr}\right) \mid \eta_r = y\right\} f_{\eta_r}(y) \\ &= \max\left\{0, 2\sum_{i=1}^{pk+1}\sum_{j=1}^{k}\frac{dt_i(r)}{dr}E\left[(U^2_{j,i}(t_i(r)) - 1) \mid \eta_r = y\right]\right\}\chi^2_{k(pk+1)}(y).\end{aligned}$$

By the independently and identically distributed property of $\left\{U_{j,i}(t_i(r))\right\}$, $\alpha(r)$ is simplified as

$$\begin{aligned}\alpha(r) &= \max\left\{0, 2\sum_{i=1}^{pk+1}\sum_{j=1}^{k}\frac{dt_i(r)}{dr}\left(\frac{y}{k(pk+1)} - 1\right)\right\}\chi^2_{k(pk+1)}(y) \\ &= \max\left\{0, 2k\left(\frac{y}{k(pk+1)} - 1\right)\sum_{i=1}^{pk+1}\frac{dt_i(r)}{dr}\right\}\chi^2_{k(pk+1)}(y).\end{aligned}$$

Therefore, when *y* is large enough, $\alpha(r)$ is approximated by

$$\alpha(r) \doteq 2\left(\frac{y}{pk+1} - k\right)\chi^2_{k(pk+1)}(y)\left\{\sum_{i=1}^{pk+1}\frac{dt_i(r)}{dr}\right\},$$

and finally $\Pr(\sup_{b \le r \le c} \eta_r \le y) \doteq \exp\left[-2\left(\frac{y}{pk+1} - k\right)\chi^2_{k(pk+1)}(y)\left\{\sum_{i=1}^{pk+1}\left(t_i(c) - t_i(b)\right)\right\}\right]$.