Statistica Sinica 3(1993), 245-254

A SIMPLE ONE DEGREE OF FREEDOM TEST FOR NON-LINEAR TIME SERIES MODEL DISCRIMINATION

W. K. Li

University of Hong Kong

Abstract: A test procedure for discriminating among different nonlinear time series models is proposed. This test procedure is seen to perform reasonably well in simulation experiments. One advantage of the test procedure is its simplicity. A real example based on the annual sunspot series is also given.

Key words and phrases: Bilinear models, Lagrange multiplier statistics, non-nested alternatives, threshold models.

1. Introduction

In recent years there has been rapid growth in the literature on nonlinear time series models. Many different types of models have been suggested. Two major classes are the threshold models (Tong (1978), Tong and Lim (1980)) and the bilinear models (Granger and Andersen (1978), Subba Rao (1981)). The recent book by Tong (1990) contains a comprehensive summary of most of the proposed nonlinear models. A natural and important problem is to develop tests to discriminate among the various models. Many tests have been proposed for testing different nonlinear models against linear (ARMA) models but not among nonlinear models. Saikkonen and Luukkonen (1988) give a summary review of the former procedures. For the latter, various informal arguments have been suggested. For example, it has been argued that threshold models can mimic limit cycle behaviour but bilinear models cannot (Tong and Lim (1980)). Consequently, one should consider threshold models for data that appear to have a limit cycle. Another common approach is to compare the post sample forecast ability of the different models (Ghaddar and Tong (1981)) or the residual sum of squares (Gabr and Subba Rao (1981)). Other arguments include parsimony in terms of model parameters and whiteness of residuals. Although these arguments are valid and important it may still be beneficial if formal tests can be developed for distinguishing between different nonlinear models. Clearly, the problem is more difficult than testing nonlinearity versus linearity since different types of nonlinear models in general cannot be nested within one another. Under the assumption of Gaussian innovations and nested models, comparing residual sums of squares is equivalent to the likelihood ratio test which is, in general, asymptotically chi-squared distributed under the null hypothesis. However, for non-nested models the likelihood ratio statistic will not normally have an asymptotic chi-squared distribution and thus the comparison of residual variances does not usually fit into the hypothesis testing framework. A possible approach is to consider a Cox test for separate families of hypotheses (Cox (1962)). This, however, requires evaluating the expectation and variance of the log likelihood ratio under the null hypothesis. For nonlinear time series this is a difficult task. Li (1989) proposed a bootstrap procedure to overcome this difficulty. However, such an approach is not too convenient to use and could encounter numerical problems. In this paper a simple one degree of freedom test for discriminating among nonlinear models is developed. This new test supercedes the bootstrapped Cox test in that it is easy to compute and that it avoids the conceptual problem that faces the bootstrap. More importantly, simulation results suggest that the test statistic has satisfactory power and approximately the correct sizes in large samples. It will also be shown that the test statistics are in some way related to the comparison of residual variances. Hence the proposed methodology may be regarded as a formalization of the latter procedure. The test is derived in Section 2; some simulation results and a real example based on Wolf's annual sunspot data are given in Section 3.

2. The Test Procedure

For simplicity we consider only two possible hypotheses. Generalization to the more general case is direct. Denote the time series process by $\{y_t\}$. It is assumed that $\{y_t\}$ is stationary with at least finite second order moments. Let F_t be the σ-field generated by $\{y_t, y_{t-1}, \ldots\}$, and $\{a_{it}\}$, $i = 1, 2$, be Gaussian white noise processes with means zero and variances σ_i^2, $i = 1, 2$. The null and alternative hypotheses are respectively

$$H_0: \quad y_t = f(F_{t-1}; \underset{\sim}{\gamma}) + a_{1t}$$

and

$$H_1: \quad y_t = g(F_{t-1}; \underset{\sim}{\beta}) + a_{2t},$$

where the forms of f and g are known and both have continuous second order derivatives with respect to $\underset{\sim}{\gamma}$ and $\underset{\sim}{\beta}$. Here $\underset{\sim}{\gamma}$ and $\underset{\sim}{\beta}$ are $p_i \times 1$ vectors of unknown parameters, $i = 1, 2$. To avoid the possibility of unidentifiability it is further assumed that the two families of models $\{f(F_{t-1}, \underset{\sim}{\gamma})\}$ and $\{g(F_{t-1}, \underset{\sim}{\beta})\}$ are nonoverlapping. That is, $\{f(F_{t-1}, \underset{\sim}{\gamma})\} \bigcap \{g(F_{t-1}, \underset{\sim}{\beta})\} = \phi$. In the case of bilinear and threshold models this would mean that the possibility of a linear model is excluded. In practice, tests such as those in Saikkonen and Luukkonen (1988) can be employed to see if linear models are adequate. Note that in Vuong (1989) a variance test is suggested in the independent case to check if two families of models can be considered as equivalent. Vuong proposes that if such is the case then no more testing will be needed. Extension of his result to the time series situation is certainly relevant and important but is clearly too involved to be included in the present paper. Denote maximum likelihood estimators of $\underset{\sim}{\gamma}$ and $\underset{\sim}{\beta}$ by $\hat{\underset{\sim}{\gamma}}$ and $\hat{\underset{\sim}{\beta}}$. Denote the corresponding residuals by $\hat{a}_{it}$, $i = 1, 2$ and let $\tilde{y}_t = g(F_{t-1}; \hat{\underset{\sim}{\beta}})$, the prediction of y_t under the alternative model. Consider the model

$$y_t = f(F_{t-1}; \underset{\sim}{\gamma}) + \lambda g(F_{t-1}; \hat{\underset{\sim}{\beta}}) + a_t, \qquad (1)$$

where $\{a_t\}$ are zero mean Gaussian white noise with variance σ^2. A test of H_o against the alternative can be based on testing $H_0' : \lambda = 0$. This test may be interpreted as a test of the adequacy of the null model versus a possible deviation in the direction of the alternative. Note that McAleer et al. (1988) adopted a similar approach for testing a pure moving average model against a pure autoregressive model. The test of H_o' can be based on the Lagrange multiplier approach (White (1984, p.72)). Let $S = \sum a_t^2/2\sigma^2$ and $\underset{\sim}{\theta} = (\underset{\sim}{\gamma}', \lambda)'$. Then the Lagrange multiplier test for $\lambda = 0$ is given by

$$T = \left(\frac{\partial S}{\partial \underset{\sim}{\theta}}\right)' \left[E\left(\frac{\partial S}{\partial \underset{\sim}{\theta}} \frac{\partial S'}{\partial \underset{\sim}{\theta}}\right)\right]^{-1} \left(\frac{\partial S}{\partial \underset{\sim}{\theta}}\right)$$

where the expectation is evaluated under the null hypothesis. Under the null hypothesis T would be asymptotically chi-squared distributed with one degree of freedom. For simplicity, let n be the same as the effective sample size in estimating $\hat{\underset{\sim}{\gamma}}$. Since $\partial S/\partial \underset{\sim}{\theta} = \sigma^{-2} \sum a_t \partial a_t/\partial \underset{\sim}{\theta}$, the statistic T can be rewritten as

$$T = \sigma_1^{-2} \sum \hat{a}_t \left(\frac{\partial a_t}{\partial \underset{\sim}{\gamma}'}, \tilde{y}_t\right) \left[E\left(\sum \frac{\partial a_t}{\partial \underset{\sim}{\theta}} \sum \frac{\partial a_t}{\partial \underset{\sim}{\theta}'}\right)\right]^{-1} \sum \left(\frac{\partial a_t}{\partial \underset{\sim}{\gamma}'}, \tilde{y}_t\right)' \hat{a}_t$$

where $\tilde{y}_t = g(F_{t-1}; \hat{\underset{\sim}{\beta}})$ and $\partial a_t/\partial x$ is evaluated under H_0, and $\hat{a}_t = \hat{a}_{1t}$. For n large enough we may drop the expectation operator and rewrite T as

$$T' = na'W'(WW')^{-1}Wa \Big/ \sum \hat{a}_t^2$$

where W' is the $n \times (p_1 + 1)$ matrix of regressors formed by stacking $(\partial a_t/\partial \underset{\sim}{\gamma}', \tilde{y}_t)$ and $a' = (\hat{a}_1, \ldots, \hat{a}_n)$. The statistic T' will have the same asymptotic distribution as T under H_0. Thus, as in Godfrey (1979), the T' statistic can be interpreted as n times the coefficient of determination of the regression of $\hat{a}_{1t}$ on $\partial a_t/\partial \underset{\sim}{\gamma} | \hat{\underset{\sim}{\gamma}}$ and $\tilde{y}_t$. In other words, the Lagrange multiplier statistic for testing $\lambda = 0$ can be easily obtained from an auxiliary ordinary regression. It is desirable in non-nested testing to intercharge the role of the null and the alternative (Cox (1962)). There is, of course, the possibility of having both hypotheses rejected. Although the interpretation problem can be difficult, such a result is still informative in the sense that it may lead us to a better model different from the existing possibilities. Clearly, generalization of the above procedure to the case of more than one alternative is direct. In the next section the empirical size and power of T' in discriminating among different nonlinear time series models are considered using simulation.

We now show that the T' statistic is related to the method of comparing residual variances. Consider the regressions

$$\hat{a}_{1t} = \tau \tilde{y}_t + \varepsilon_t, \tag{2}$$

and

$$\tilde{y}_t = \frac{\partial f(F_{t-1}; \underset{\sim}{\gamma})}{\partial \underset{\sim}{\gamma}} \cdot \underset{\sim}{K} + V_t, \tag{3}$$

where ε_t, V_t are independent zero mean normal random variates; τ and $\underset{\sim}{K}$ are the respective regression parameters. For simplicity, let $\sigma_1^2 = 1$. Then under H_0 the score vector $\partial S / \partial \underset{\sim}{\theta} = -(\underset{\sim}{0}, \sum \hat{a}_{1t} \tilde{y}_t)$ and the observed Fisher information matrix,

$$I = \begin{bmatrix} \sum \frac{\partial f_t}{\partial \underset{\sim}{\gamma}} \frac{\partial f_t}{\partial \underset{\sim}{\gamma}^T} & \sum \tilde{y}_t \frac{\partial f_t}{\partial \underset{\sim}{\gamma}} \\ \sum \tilde{y}_t \frac{\partial f_t}{\partial \underset{\sim}{\gamma}^T} & \sum \tilde{y}_t^2 \end{bmatrix},$$

where $f_t = f(F_{t-1}; \underset{\sim}{\gamma})$. Hence, the statistic T' can be written as

$$\begin{aligned} T' &= \left(\sum \hat{a}_{1t} \tilde{y}_t\right)^2 \left[\sum \tilde{y}_t^2 - \left(\sum \tilde{y}_t \frac{\partial f_t}{\partial \underset{\sim}{\gamma}^T}\right)\left(\sum \frac{\partial f_t}{\partial \underset{\sim}{\gamma}} \cdot \frac{\partial f_t}{\partial \underset{\sim}{\gamma}^T}\right)^{-1}\left(\sum \tilde{y}_t \frac{\partial f_t}{\partial \underset{\sim}{\gamma}}\right)\right]^{-1} \\ &= \left[\frac{\sum \hat{a}_{1t} \tilde{y}_t}{\sum \tilde{y}_t^2}\right]^2 \cdot \frac{\sum \tilde{y}_t^2}{[1 - r^2]}, \end{aligned}$$

where

$$r^2 = \frac{\left(\sum \tilde{y}_t \frac{\partial f_t}{\partial \underset{\sim}{\gamma}^T}\right)\left(\sum \frac{\partial f_t}{\partial \underset{\sim}{\gamma}} \cdot \frac{\partial f_t}{\partial \underset{\sim}{\gamma}^T}\right)^{-1}\left(\sum \tilde{y}_t \frac{\partial f_t}{\partial \underset{\sim}{\gamma}}\right)}{\sum \tilde{y}_t^2}.$$

The quantity r^2 is the coefficient of determination for the auxiliary regression (3). Note that $\sum \hat{a}_{1t} \tilde{y}_t / \sum \tilde{y}_t^2 = \hat{\tau}$, the least squares estimate of τ in (2). Hence, using standard regression results

$$\begin{aligned} T' &= \frac{\sum \hat{\tau}^2 \tilde{y}_t^2}{1-r^2} \\ &= \frac{\sum \hat{a}_{1t}^2 - \sum \hat{\varepsilon}_t^2}{1-r^2}. \end{aligned} \tag{4}$$

We observe from (4) that if H_0 is the true model then $\sum \hat{a}_{1t}^2$ should be small and $\sum \hat{\varepsilon}_t^2$ should be close to $\sum \hat{a}_{1t}^2$. However if H_1 is true then $\sum \hat{a}_{1t}^2$ should be large while $\sum \hat{\varepsilon}_t^2$ should be small. A similar result holds when we interchange the hypotheses. Thus the testing procedure can be interpreted as a way to compare residual variances after adjusting them by auxiliary regressions (2) and (3). One advantage of the approach is, clearly, that the statistic T' has a known asymptotic distribution under the null hypothesis and therefore we can have meaningful discussions on sizes and power at least asymptotically. The parameter r^2 can be interpreted as a measure of the similarity between $g(F_{t-1};\beta)$ and $f(F_{t-1};\gamma)$ since, in the special case, where $g(F_{t-1};\beta) = \beta g(F_{t-1})$ and $f(F_{t-1};\gamma) = \gamma f(F_{t-1})$ then $r^2 = 1$ if $cf = g$ for some constant c. Note also that since $0 < r^2 < 1$, the test statistic can be much larger than its numerator and hence the procedure can be more sensitive in detecting significant differences of the models than the method of comparing residual variances. We shall give more discussions on this aspect in the next section.

3. Some Empirical Results

Two major classes of models in nonlinear time series analysis are the threshold models (Tong and Lim (1980)) and the bilinear models (Subba Rao (1981)). Simulation experiments were performed to study the effectiveness of the T' statistic in discriminating these models. In the first experiment the true model was the simple SETAR (2;1,1) model

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + a_t, \quad \text{if } X_{t-1} \geq 0, \\ X_t &= \phi_1' X_{t-1} + a_t, \quad \text{otherwise}, \end{aligned} \tag{5}$$

where a_t is Gaussian with mean 0 and variance 1. Two T' statistics T_1' and T_2' were computed. The statistic T_1' had the true model as the null hypothesis and the simple bilinear model

$$X_t = CX_{t-1} + bX_{t-k}e_{t-\ell} + e_t, \tag{6}$$

as alternative. The statistic T_2' had the simple bilinear model as null and the true threshold model as the alternative. The parameters for the SETAR model were $(\phi_1, \phi_1') = (0.5, -0.5)$ and $(0.8, 0.3)$. The values of (k, ℓ) in the bilinear model were (1,1), (1,2) and (2,1). There were 200 independent realizations each of length 100 for each combination of (ϕ_1, ϕ_1') and (k, ℓ). The derivatives of the bilinear model were obtained recursively by setting initial values at zero as in Gabr and Subba Rao (1981). The estimation of bilinear models was based on the Newton-Raphson procedure. IMSL subroutines were used to generate the series and to estimate the model parameters. The empirical mean, variance and the upper 10% and 5% significance levels of T_1' and T_2' are reported in Table 1. It can be seen that the chi-square distribution with one degree of freedom gave very reasonable approximation to the distribution of T_1'. The power of the T_2' statistic was also very impressive. In most cases the frequency of rejecting the bilinear null hypothesis in favour of the threshold model was well over 50%.

In the second experiment the true model was the simple bilinear model

$$X_t = CX_{t-1} + bX_{t-k}e_{t-\ell} + e_t$$

with $(C, b) = (0.5, 0.2)$, $(k, \ell) = (1,1)$, $(1,2)$ and $(2,1)$ and e_t normal with mean 0 and variance 1. The alternative models were the SETAR (2;1,1) model and the SETAR (2;2,2) model

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t, \quad \text{if } X_{t-2} \geq 0 \\ X_t &= \phi_1' X_{t-1} + \phi_2' X_{t-2} + a_t, \quad \text{otherwise.} \end{aligned}$$

Again we computed two statistics T_1'' and T_2'' where T_1'' took the true model as the null hypothesis and T_2'' took the threshold model as the null hypothesis. For cases involving the SETAR (2;1,1) model the length of each realization was again 100. For cases involving the SETAR (2;2,2) model the series lengths were 100 and 200. Again for each combination of hypotheses there were 200 independent replications. The empirical means, variances, and the upper 10% and 5% significance levels of T_1'' and T_2'' with respect to the chi-square distribution of one degree of freedom are reported in Table 2. From Table 2, the chi-square distribution gave a very reasonable approximation to the distribution of T_1'' in cases involving the SETAR (2;1,1) model. The power of T_2'' was also very good and the two statistics could be considered to be doing their job well. When the SETAR (2;2,2) model was considered and when $n = 100$ the chi-square distribution was somewhat too sensitive for T_1''. The power of T_2'' remained reasonably good and was much greater than the nominal sizes. However, there was considerable improvement when $n = 200$; and both statistics repeated the strong showing given by T_1' and T_2' in Table 1. These experiments suggested that the simple Lagrange multiplier statistic T_1' as proposed in Section 2 can be used to discriminate between nonlinear time series models with reasonable results.

As suggested in the previous section the proposed statistics could be more sensitive in detecting model differences. In Table 3 the mean absolute difference (DRES) of the residual variances under the two hypotheses and the mean absolute difference (DTEST) of the T' statistics were computed for the first three cases of Tables 1 and 2. It can be seen that on the average the differences between residual variances are rather small. In the absence of a reference distribution it could be difficult to judge their significance. However, the differences between the T' statistics were on the average much larger than the upper 5% value of the χ_1^2 distribution. This suggested that differences between models might be less susceptible to detection using residual variances than by using the proposed tests.

It may be of interest to investigate the power of the tests when both hypotheses are false. In Table 4, the true model is given by

$$X_t = \phi X_{t-1} + d_1 X_{t-k} a_{t-\ell} + a_t, \qquad \text{if } X_{t-1} \geq 0$$

and

$$X_t = d_2 X_{t-k} a_{t-\ell} + a_t, \qquad \text{if } X_{t-1} < 0$$

where a_t are normal with mean 0 and variance 1.

We may call these models "threshold bilinear" models. Investigation on the properties of this new class of nonlinear models is beyond the scope of this paper. However, it is interesting to see how the T' statistics work in this situation. Denote by T_1^* the T' statistic corresponding to the null hypothesis given by (5) and the alternative given by (6). Denote by T_2^* the T' statistic when the roles of (5) and (6) are reversed. The power of T_1^* and T_2^* are recorded in Table 4. The series length is 200 and the number of replications in each case is 100. We use the same (k, ℓ) in simulation and estimation. It can be seen that for the cases considered both the T_1^* and T_2^* statistics appear to have reasonable power when the true model is in fact not covered by both hypotheses. This result suggests further evidence on the usefulness of the proposed approach.

As a real example we considered the annual Wolf sunspot numbers (1700-1921). Tong (1990) considered a SETAR (2;3,11) model and Gabr and Subba Rao (1981) suggested that a subset bilinear model may gave a better fit. These nonlinear models were refitted by considering the first 11 observation as fixed and two T_1' statistics $\tilde{T}_1$ and $\tilde{T}_2$ were computed. The $\tilde{T}_1$ statistic had the threshold model as the null and the subset bilinear model as the alternative and the $\tilde{T}_2$ statistic had the hypotheses the other way round. The refitted models and the $\tilde{T}_i$ statistics are as follows. For the threshold model we had

$$X_t = \begin{cases} 10.7678 + 1.7344X_{t-1} - 1.2957X_{t-2} + 0.4740X_{t-3} + \varepsilon_t, \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{if } X_{t-3} \le 36.6, \\ 7.5791 + 0.7332X_{t-1} - 0.0403X_{t-2} - 0.1971X_{t-3} \\ + 0.1597X_{t-4} - 0.2204X_{t-5} + 0.0220X_{t-6} + 0.1491X_{t-7} \\ - 0.2403X_{t-8} + 0.3121X_{t-9} - 0.3691X_{t-10} + 0.3881X_{t-11} \\ + \varepsilon_t, \qquad\qquad\qquad\qquad\qquad \text{if } X_{t-3} > 36.6 \end{cases}$$

and $\tilde{T}_1 = 51.84$. Note that here the residuals for both branches of the model were taken to have same variance. For the subset bilinear model we had

$$\begin{aligned} X_t &= 6.8922 + 1.5012X_{t-1} - 0.7671X_{t-2} + 0.1152X_{t-9} \\ &\quad -0.0146X_{t-2}e_{t-1} + 0.0063X_{t-8}e_{t-1} \\ &\quad -0.0072X_{t-1}e_{t-3} + 0.0068X_{t-4}e_{t-3} \\ &\quad +0.0036X_{t-1}e_{t-6} + 0.0043X_{t-2}e_{t-4} \\ &\quad +0.0018X_{t-3}e_{t-2} + e_t \end{aligned}$$

and $\tilde{T}_2 = 0.0268$. Hence, the $\tilde{T}_1$ statistic rejected the threshold null while the $\tilde{T}_2$ statistic accepted the bilinear null. Thus the approach here favoured the bilinear model to the threshold model. The residual variances for the bilinear and threshold models were respectively 124.92 and 149.71. Note that the value of 0.0268, although small, was still greater than the lower 10% critical value of a chi-square distribution with one degree of freedom. However, Tong (1990, p.443) suggested that the bilinear model could be noninvertible.

Acknowledgement

I am grateful to professor Howell Tong, an associate editor, two referees and Dr. A. Y. C. Kuk for comments that led to improvements in the quality of the paper. Thanks are also due to the Edinburgh Nonlinear Time Series Workshop for partial support.

Table 1. Empirical means, variances and significance levels for T_1', T_2'.

ϕ_1, ϕ_1'	k	ℓ	T_1'				T_2'			
			90%	95%	$\bar{T}_1'$	$V(T_1')$	90%	95%	$\bar{T}_2'$	$V(T_2')$
$(0.5, -0.5)$	1	1	0.110	0.065	1.11	2.41	0.715	0.640	6.03	21.30
	1	2	0.110	0.060	1.03	2.24	0.995	0.995	16.85	30.43
	2	1	0.080	0.050	0.95	1.58	1.000	0.995	16.10	38.34
$(0.8, 0.3)$	1	1	0.110	0.060	1.03	1.94	0.510	0.365	3.76	13.97
	1	2	0.120	0.070	1.25	3.32	0.730	0.555	5.25	15.76
	2	1	0.110	0.055	1.17	3.20	0.790	0.700	6.11	16.02

Table 2. Empirical means, variances and significance levels for T_1'', T_2''.

	T_1''				T_2''			
(k,ℓ)	90%	95%	$\bar{T}_1''$	$V(T_1'')$	90%	95%	$\bar{T}_2''$	$V(T_2'')$
SETAR (2;1,1)								
$n=100$								
1 1	0.10	0.065	1.01	1.93	0.695	0.555	5.60	22.74
1 2	0.08	0.055	0.98	1.67	0.690	0.565	5.47	16.80
2 1	0.08	0.045	0.98	1.31	0.620	0.505	5.05	19.49
SETAR (2;2,2)								
$n=100$								
1, 1	0.170	0.100	1.38	2.89	0.850	0.820	10.25	43.72
1, 2	0.160	0.105	1.29	2.70	0.420	0.265	2.86	7.85
2, 1	0.165	0.090	1.32	3.30	0.565	0.435	4.25	16.38
$n=200$								
1, 1	0.115	0.045	1.07	1.88	0.990	0.980	19.18	75.90
1, 2	0.130	0.065	1.21	2.38	0.605	0.480	4.84	17.81
2, 1	0.095	0.050	1.00	1.83	0.850	0.765	8.42	31.21

Table 3. Mean absolute differences between residual variances (DRES) and Tests (DTEST).

H_0 (True model)	H_1	DRES	DTEST
SETAR (5)	Bilinear (6)		
(ϕ_1, ϕ_1')	(k,ℓ)		
(0.5, −0.5)	(1, 1)	0.06	5.03
	(1, 2)	0.18	15.13
	(2, 1)	0.18	15.12
Bilinear (6)	SETAR (5)		
(1, 1)		0.05	4.93
(1, 2)		0.05	5.32
(2, 1)		0.05	5.12

Table 4. Power of T_1^* and T_2^* at nominal sizes 10% and 5%. Entries are number of rejections in 100 replications.

ϕ_1	d_1	d_2	k	ℓ	T_1^* 10%	T_1^* 5%	T_2^* 10%	T_2^* 5%
0.0	0.5	−0.5	1	1	51	45	45	37
0.5	0.5	−0.5	3	2	65	55	98	95
0.5	0.5	−0.5	2	3	76	72	61	54

References

Cox, D. R. (1962). Further results on tests of separate families of hypotheses. *J. Roy. Statist. Soc. Ser.B* **24**, 406–424.

Gabr, M. M. and Subba Rao, T. (1981). The estimation and prediction of subset bilinear time series models with applications. *J. Time Ser. Anal.* **2**, 155–171.

Ghaddar, D. K. and Tong, H. (1981). Data transformation and self-exciting threshold autoregression. *J. Roy. Statist. Soc. Ser.C* **30**, 238–248.

Godfrey, L. G. (1979). Testing the adequacy of a time series model. *Biometrika* **66**, 67–72.

Granger, C. W. J. and Andersen, A. P. (1978). *Introduction to Bilinear Time Series Models.* Vandenhoeck and Ruprecht, Göttingen.

Li, W. K. (1989). Discriminating nonlinear time series models using a bootstrapped Cox test. Technical Report, University of Hong Kong.

McAleer, M., McKenzie, C. R. and Hall, A. D. (1988). Testing separate time series models. *J. Time Ser. Anal.* **9**, 169–189.

Saikkonen, P. and Luukkonen, R. (1988). Lagrange multiplier tests for testing nonlinearities in time series models. *Scand. J. Statist.* **15**, 55–68.

Subba Rao, T. (1981). On the theory of bilinear time series models. *J. Roy. Statist. Soc. Ser.B* **43**, 244–255.

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. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion), 575–586. *J. Roy. Statist. Soc. Ser.B* **42**, 245–292.

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

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrika* **57**, 307–333.

White, H. (1984). *Asymptotic Theory for Econometricians.* Academic Press, Orlando.

Department of Statistics, University of Hong Kong, Hong Kong.

(Received November 1990; accepted May 1992)