Statistica Sinica 2(1992), 157-169

POWER OF THE SCORE TEST AGAINST BILINEAR TIME SERIES MODELS

D. Guegan and T. D. Pham

University of Paris XIII and Laboratory of Modelling and Computation

Abstract: We investigate local power of the Lagrange Multiplier test against Bilinear alternatives which are contiguous to the null hypothesis. An empirical study is made, through simulations, of the power of this test as a function of the values of the bilinear parameters, both in the neighbourhood of the null hypothesis and far away from it. The theoretical comparison of the local power has been compared with simulations.

Key words and phrases: Bilinear model, Lagrange multiplier test, contiguity, local power.

1. Introduction

Recently there has been considerable interest in nonlinear time series modelling. An important problem in this area is the detection of non-linearity. Various tests for linearity have been proposed in the literature. See, for example, Subba Rao and Gabr (1980), Hinich (1982), MacLeod and Li (1983), Guegan (1984), Keenan (1985), Tsay (1986), Petruccelli and Davies (1986), Chan and Tong (1986), Luukkonen, Saikkonen and Terasvirta (1988), Saikkonen and Luukkonen (1988) and Chan and Tong (1990). A general class of tests for non-linearity can be obtained from the Lagrange Multipliers (L.M.) method. This test has been considered by Saikkonen and Luukkonen (1988), who have derived an explicit form in the bilinear case, amongst others, and have compared the performance of this test with some other tests. Until now, however, there has not been a systematic study of the power of the L.M. test, in particular, for the case of bilinear alternatives. Theoretical computation of the power of the test is in fact very difficult, except when the alternatives are restricted to being contiguous. In this paper we do a theoretical study of the local power of the L.M. test, and we also examine empirically, through simulations, the power of this test as a function of the values of the bilinear parameters, both in the neighbourhood of the null hypothesis and far from it. The theoretical computation of the local power has been compared with that based on simulations, and there appears to be good agreement.

2. Theoretical Study of the Local Power of the Test

In this section, we deal with the power of the L.M. test under a sequence of alternatives, contiguous to the null hypothesis. We begin by specifying the notations and recalling the basic results of the test in a general setup.

Consider a statistical model specified by a vector parameter θ belonging to some open set $\Theta \subset \mathbf{R}^{m+k}$. We are interested in the null hypothesis H_0 defined by $\theta_2 = 0$ when θ_2 is formed by the last k components of θ. Let us denote by θ_1, the vector formed by the first m components of θ, so that $\theta = (\theta_1, \theta_2)$.

Let ℓ denote the log-likelihood function of the sample. The score vector is defined by the first derivative of ℓ, denoted by $\partial\ell/\partial\theta = [(\partial\ell/\partial\theta_1)', (\partial\ell/\partial\theta_2)']'$ where prime represents the transpose, with m components in $\frac{\partial\ell}{\partial\theta_1}$ and k components in $\frac{\partial\ell}{\partial\theta_2}$.

The Fisher information matrix J is the negative of the mathematical expectation of the second derivatives of the log-likelihood ℓ. This matrix can be partitioned, in an obvious way, into four blocks. $J_{ij} = -E\frac{\partial^2\ell}{\partial\theta_i\partial\theta_j}$, $i, j = 1, 2$.

Let $\hat{\theta} = (\hat{\theta}_1, 0)$ be the maximum likelihood estimate of θ under H_0. Then the L.M. (or score) test is defined by:

$$T_n = \frac{1}{n}\left(\frac{\partial\ell}{\partial\theta_2}\right)'_{\hat{\theta}} \cdot \hat{J}_{2,1}^{-1} \cdot \left(\frac{\partial\ell}{\partial\theta_2}\right)_{\hat{\theta}} \tag{2.1}$$

where $\hat{J}_{2,1}$ is a consistent estimate of $J_{2,1} = J_{22} - J_{21} \cdot J_{11}^{-1} \cdot J_{12}$. In general, one estimates J by

$$-\left(\frac{1}{n}\frac{\partial^2\ell}{\partial\theta^2}\right)_{\hat{\theta}},$$

and thus can take:

$$\hat{J}_{2,1} = -\frac{1}{n}\left[\frac{\partial^2\ell}{\partial\theta_2^2} - \frac{\partial^2\ell}{\partial\theta_1\partial\theta_2} \cdot \left(\frac{\partial^2\ell}{\partial\theta_1^2}\right)^{-1} \cdot \frac{\partial^2\ell}{\partial\theta_2\partial\theta_1}\right]_{\hat{\theta}}.$$

From the theory of score tests, (cf. Moran (1970)), it is known that under adequate regularity assumptions the test statistic T_n converges in distribution, under the null hypothesis H_0, to the χ^2 random variable with k degrees of freedom.

We consider now a general stationary diagonal bilinear model defined by:

$$X(t) = \sum_{i=1}^{m} a_i X(t-i) + \epsilon(t) + \sum_{j=1}^{q} c_j \epsilon(t-j) + \sum_{k=1}^{p} b_{kk} X(t-k)\epsilon(t-k) \tag{2.2}$$

where $\epsilon(t)$ is a sequence of independent identically distributed Gaussian random variables with mean zero and variance σ^2. The Gaussian assumption is actually needed only to write down the likelihood; but the asymptotic results below hold without this assumption. For simplicity we shall confine attention to the case where there are no moving average terms, i.e., $q = 0$. Then the parameter is $\theta = \{a_1, \dots, a_m, b_{11}, \dots, b_{pp}\}$ and the null hypothesis is specified by $b_{jj} = 0$, $j = 1, \dots, p$. The log-likelihood function of the model, based on n observations $X_1, \dots, X_n$ is approximately

$$\ell = -\frac{1}{2\sigma^2} \sum_{t=1}^{n} \epsilon_\theta^2(t)$$

where $\epsilon_\theta(t)$ is defined recursively from the model (2.2), i.e.,

$$\epsilon_\theta(t) = -\sum_{k=1}^{p} b_{kk} X(t-k)\epsilon_\theta(t-k) + X(t) - \sum_{j=1}^{m} a_j X(t-j) \tag{2.3}$$

with starting value $\epsilon_\theta(t) = 0$ for $t < 1$. The above approximation of the likelihood is justified by the fact that $\epsilon_\theta(t) - \epsilon(t) \to 0$ almost surely, uniformly in θ, as $t \to \infty$, provided that the model is invertible. Note that the model (2.2) is invertible if the "nonlinear coefficients" $b_{11}, \dots, b_{pp}$ are small enough. The special case where they vanish corresponds to an AR model which is always invertible. Since we are interested only in a small neighbourhood of the null hypothesis, we need not bother about invertibility. Based on the above (approximate) log-likelihood, the score vector can be computed as:

$$\frac{\partial \ell}{\partial a_i} = -\frac{1}{\sigma^2} \sum_{t=1}^{n} \epsilon_\theta(t) u(t-i), \tag{2.4}$$

$$\frac{\partial \ell}{\partial b_{jj}} = -\frac{1}{\sigma^2} \sum_{t=1}^{n} \epsilon_\theta(t) v_\theta(t-j), \tag{2.5}$$

where $u(t) = -X(t)$ and $v_\theta(t)$ is the solution of the recurrence equation

$$v_\theta(t) = -\sum_{k=1}^{p} b_{kk} X(t-k) v_\theta(t-k) - X(t)\epsilon_\theta(t) \tag{2.6}$$

with initial condition $v_\theta(t) = 0$ for $t < 1$. The second derivatives of ℓ are:

$$\frac{\partial^2 \ell}{\partial a_i \partial a_j} = -\sum_{t=1}^{n} u(t-i)u(t-j)$$

$$\frac{\partial^2 \ell}{\partial a_i \partial b_{jj}} = -\sum_{t=1}^{n} u(t-i)v_\theta(t-j)$$

$$\frac{\partial^2 \ell}{\partial b_{ii}\partial b_{jj}} = -\sum_{t=1}^{n} v_\theta(t-i)v_\theta(t-j) - \sum_{t=1}^{n} \epsilon_\theta(t)\frac{\partial}{\partial b_{jj}}v_\theta(t-j).$$

The last term in the above right hand side may be neglected since $\epsilon_\theta(t)$ is close to $\epsilon(t)$ and $E[\epsilon(t)\partial v_\theta(t-j)/\partial b_{jj}] = 0$. Thus, one may take as an estimate of J:

$$\hat{J} = \frac{1}{n}\sum_{t=1}^{n} \begin{bmatrix} u(t-1) \\ \vdots \\ u(t-m) \\ v_{\hat{\theta}}(t-1) \\ \vdots \\ v_{\hat{\theta}}(t-p) \end{bmatrix} [u(t-1)\cdots u(t-m)v_{\hat{\theta}}(t-1)\dots v_{\hat{\theta}}(t-p)]. \tag{2.7}$$

The test statistic T_n can then be constructed according to (2.1), (2.4), (2.5) and (2.7).

We now study the local power of the above L.M. test. Consider a sequence of alternatives specified by a vector parameter of the form $(\theta_1^*, \varphi_2/\sqrt{n})$, where $\theta_1^* = (a_1^*, \dots, a_m^*)$ and $\varphi_2 = (\beta_{11}, \beta_{22}, \dots, \beta_{pp})$, and n denotes the sample size. If this sequence of alternatives can be proved to be contiguous to the null hypothesis (the one specified by $(\theta_1^*, 0)$) then one may apply LeCam's third lemma (Hajeck and Sidak (1967), p.208) to obtain the power of the test under this sequence of alternatives, provided the Central Limit Theorem (C.L.T.) holds under the null hypothesis, for the pair of random variables formed by the test statistic and the log-likelihood ratio. However, this approach encounters technical difficulties due mainly to the fact that the log-likelihood cannot be obtained exactly but only approximately by ignoring end-effects. Therefore we shall adopt an heuristic approach which yields, in a very simple way, an expression for the local power, which is our main interest, since it serves as a basis for comparison with the empirical studies in next section. Note that the main point is the lack of a rigorous proof for the contiguity of the sequence of alternatives specified by $(\theta_1^*, \varphi_2/\sqrt{n})$ to the null $(\theta_1^*, 0)$ hypothesis. The validity of the C.L.T. is not a problem since, under contiguity one need only work under the null hypothesis, for which the observations follow an autoregressive process.

Before proceeding further, we define some notation. Let $\theta^* = (\theta_1^*, 0)$ and denote by $\boldsymbol{X}(t, m)$, $\boldsymbol{v}_{\theta^*}(t, p)$ and $\boldsymbol{v}_0(t, p)$ the random vectors with components $X(t-1), \dots, X(t-m), v_{\theta^*}(t-1), \dots, v_{\theta^*}(t-p)$ and $-\epsilon(t)X(t-1), \dots, -\epsilon(t)X(t-p)$, respectively.

In order to study the behaviour of the statistic (2.1) under the alternative, we first look at that of the score vector $\frac{\partial \ell}{\partial \theta_2}$ evaluated at $\hat{\theta}$. Using a Taylor's expansion, one may write

$$\left(\frac{\partial \ell}{\partial \theta_2}\right)_{(\hat{\theta}_1,0)} = \left(\frac{\partial \ell}{\partial \theta_2}\right)_{\theta^*} + \left(\frac{\partial^2 \ell}{\partial \theta_2 \partial \theta_1}\right)(\hat{\theta}_1 - \theta_1^*) + O_p(\sqrt{n}), \tag{2.8}$$

where $O_p(n^\alpha)$ represents a random variable such that $n^{-\alpha}O_p(n^\alpha)$ tends to zero in probability as n goes to infinity. The above expansion relies on the fact that $\hat{\theta}_1 - \theta_1^*$ is of order $1/\sqrt{n}$ in probability (under the alternatives). This can be seen from the expression of $\hat{\theta}_1 - \theta^*$ derived below. Note that the log-likelihood ℓ is quadratic in θ_1 (for fixed θ_2). Thus,

$$0 = \left(\frac{\partial \ell}{\partial \theta_1}\right)_{(\hat{\theta}_1,0)} = \left(\frac{\partial \ell}{\partial \theta_1}\right)_{\theta^*} + \left(\frac{\partial^2 \ell}{\partial \theta_1^2}\right)_{\theta^*}(\hat{\theta}_1 - \theta^*)$$

giving

$$\hat{\theta}_1 - \theta_1^* = -\left(\frac{\partial^2 \ell}{\partial \theta_1^2}\right)^{-1}\left(\frac{\partial \ell}{\partial \theta_1}\right)_{\theta^*} \tag{2.9}$$

The score vector in the above right hand side is explicitly given by

$$\left(\frac{\partial \ell}{\partial \theta_1}\right)_{\theta^*} = -\frac{1}{\sigma^2}\sum_{t=1}^{n} \epsilon_{\theta^*}(t)\boldsymbol{X}(t-m)$$

where

$$\begin{aligned}\epsilon_{\theta^*}(t) &= X(t) - \sum_{i=1}^{m} a_i^* X(t-i) \\ &= \epsilon(t) + \sum_{j=1}^{p} \frac{\beta_{jj}}{\sqrt{n}} X(t-j)\epsilon(t-j).\end{aligned} \tag{2.10}$$

From the above results, one can see that $\hat{\theta}_1 - \theta_1^*$ is of order $1/\sqrt{n}$. Note also that $\hat{\theta}_1$, as an estimator of θ_1^*, has, under the alternative, a bias of order $1/\sqrt{n}$, which comes from the last term in the above right hand side.

Let us return to the score vector in the right hand side of (2.8). We have:

$$\left(\frac{\partial \ell}{\partial \theta_2}\right)_{\theta^*} = -\frac{1}{\sigma^2}\sum_{t=1}^{n} \epsilon_{\theta^*}(t)\boldsymbol{v}_{\theta^*}(t,p). \tag{2.11}$$

Clearly $v_{\theta^*}(t) = -X(t)\epsilon_{\theta^*}(t)$. Hence from (2.8),

$$v_{\theta^*}(t) = v_0(t) - X(t)\sum_{\ell=1}^{p} \frac{\beta_{\ell\ell}}{\sqrt{n}} X(t-\ell)\epsilon(t-\ell) \tag{2.12}$$

where $v_\theta(t) = -X(t)\epsilon(t)$.

From the above results, it can be shown (see details in the Appendix) that under the alternative, the random vector $n^{-1/2}\frac{\partial \ell}{\partial \theta_2}$ evaluated at $(\hat{\theta}_1, 0)$ converges in probability as $n \to \infty$ to a Gaussian vector with mean c and covariance matrix $J_{1,2}$. The elements of c and $J_{1,2}$ are given by (A.4) and (A.5). Hence the test statistic T_n converges under the alternative to a noncentral χ^2 variate with p degrees of freedom and noncentral parameter $c' J_{1,2}^{-1} c$.

Remark. (a) If one considers a fixed alternative, then $(1/n)(\partial \ell / \partial \theta_2)$ evaluated at $(\hat{\theta}_1, 0)$ would converge to a finite (nonzero) limit. Hence the test statistic would converge to infinity, meaning that the test is consistent. The power of the test against contiguous alternatives, however converges to a limit less than 1, as n goes to infinity.

(b) The above result does not require the assumption that the $\epsilon(t)$ are Gaussian but only that their joint 3rd and 4th cumulants are zero. Moreover these conditions are needed only to get the simple expressions (A.4) and (A.5) for c and $J_{1,2}$.

Example. Consider the first order bilinear model

$$X(t) = aX(t-1) + bX(t-1)\epsilon(t-1) + \epsilon(t). \tag{2.13}$$

This model reduces to an AR(1) model when $b = 0$. Then $E[X^2(t)] = \sigma^2/(1-a^2)$, which yields

$$c = \beta\sigma^2\Big(2 + \frac{1}{1-a^2}\Big) \quad (\beta = b/\sqrt{n})$$

and

$$J_{2,1} = \sigma^2\Big(2 + \frac{1}{1-a^2}\Big).$$

Hence the noncentral parameter is:

$$\left(|\beta|\sigma\sqrt{2 + \frac{1}{1-a^2}}\right)^2. \tag{2.14}$$

3. Empirical Study of the Power

In this section we study empirically the power of the L.M. test of an AR(1) model against the diagonal BL(1,0,1,1) alternative, denoted by DBL(1,0,1,1). However, to test against nonlinearity, the usual approach would be to fit first a linear model and then test its adequacy against possible nonlinearities. In the L.M. approach, one would apply a standard identification procedure, resulting in

Table 1. Percentage of rejections of the two L.M. tests at level 5% for (1) testing AR(1) against DBL(1, 0, 1, 1) and (2) for testing ARMA(1,1) against DBL(1, 1, 1, 1); 500 replications – sample size 200.

a_1 / b_1	$-.8$		$-.5$		$-.2$		$.2$		$.5$		$.8$	
Models	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
$-.9$					70.8	56	57.6	52.6				
$-.8$			82.2	78.8	84.2	58.4	65	62.2	56.8	67.6		
$-.7$			92.2	88.8	92.8	57	79.6	75.2	59.2	73.6		
$-.6$	84	86.2	97.2	96	98.4	57	88	86.4	64.6	83.8	56.2	88
$-.5$	89.2	89.2	99.2	99	98.8	63.6	97.2	93.8	82	93.2	56.8	90.6
$-.4$	95.6	96	99.6	99.6	100	75.6	99.4	96.8	91.4	99	62.2	89.8
$-.3$	99.4	99.4	100	100	100	88.6	100	97	99.6	100	88.2	94.8
$-.2$	99.8	99.8	99.6	99.8	99.2	95.4	98.8	96.4	99	99.2	99.8	99.6
$-.1$	83.2	83.6	72.4	71.8	69.6	62.6	67.8	58.2	77.2	68.4	81.6	83.4
$-.05$	33	34.6	24.8	26	22.6	20.4	21.4	18.8	22.4	23	27	33.8
0	06.2	06.8	05.4	05.2	05.4	5.8	06	05	06	06	04.4	08.4
.05	31.6	31.4	22.8	22.8	20	20.8	20.8	17.8	22.6	23.8	26.2	34.4
.1	82	82	68.2	67.6	65.2	59.8	64	57.8	67.2	67.4	79.2	83.2
.2	100	100	99.8	99.8	99.4	95.4	97.2	95.6	99.4	99.8	99.2	99.8
.3	99.4	99.6	100	100	100	90.6	100	98	100	99.8	88.6	95
.4	95.8	96.4	99.8	99.8	100	76	100	96.8	91.4	98.4	63.8	87.8
.5	90.2	90.6	99.2	98.8	99.8	64.6	98	93.6	81.6	91.6	50.2	88.6
.6	85.2	85.2	96.2	96.2	98.4	59.2	89.2	83.2	64.6	82	53.6	87
.7			91.4	89.8	92.4	58.8	75.2	68.8	52.4	71.6		
.8			81.2	80	83	57.6	64.4	57	47.8	60.6		
.9					70.6	56.4	58.6	49.8				

an ARMA(p, q) model, say, using the L.M. test to test it against the bilinear extension of this model. If the data actually obey an BL$(1, 0, 1, 1)$ model, the identification procedure would usually result in an ARMA(1,1) model since this model has the same covariance structure as the diagonal BL$(1, 0, 1, 1)$ model; and hence one would end up testing the ARMA(1,1) against the BL$(1, 1, 1, 1)$ alternative. Therefore in this study we shall consider, in addition to the L.M. test for AR(1) against DBL$(1, 0, 1, 1)$, the L.M. test for ARMA(1,1) against DBL$(1, 1, 1, 1)$ as well. In both cases, the data actually follow the DBL$(1, 0, 1, 1)$ given by (2.13). We generate 500 series of length 200 from this model. The parameter a takes the values -0.8; -0.5; -0.2; 0.2; 0.5; 0.8, and the parameter σb varies between -0.9 to 0.9 with a step 0.2 or 0.1. (Note that the distribution of the test statistic depends only on a and σb since the process $X(t)$ has the same distribution as $\sigma Y(t)$, where $Y(t)$ obeys a DBL$(1, 0, 1, 1)$ model with parameters a, σb and noise variance 1). The results of the L.M. test (in the percentage of rejections) are listed in Table 1. Some entries in the table are empty because the bilinear model $X(t)$ (2.13) defined by the corresponding values of a and σb is not stationary.

The row $b = 0$ corresponds to the null hypothesis and one may see that the percentage of rejections correspond well to the 5% level that has been chosen. Note that the results concerning the two tests are very similar.

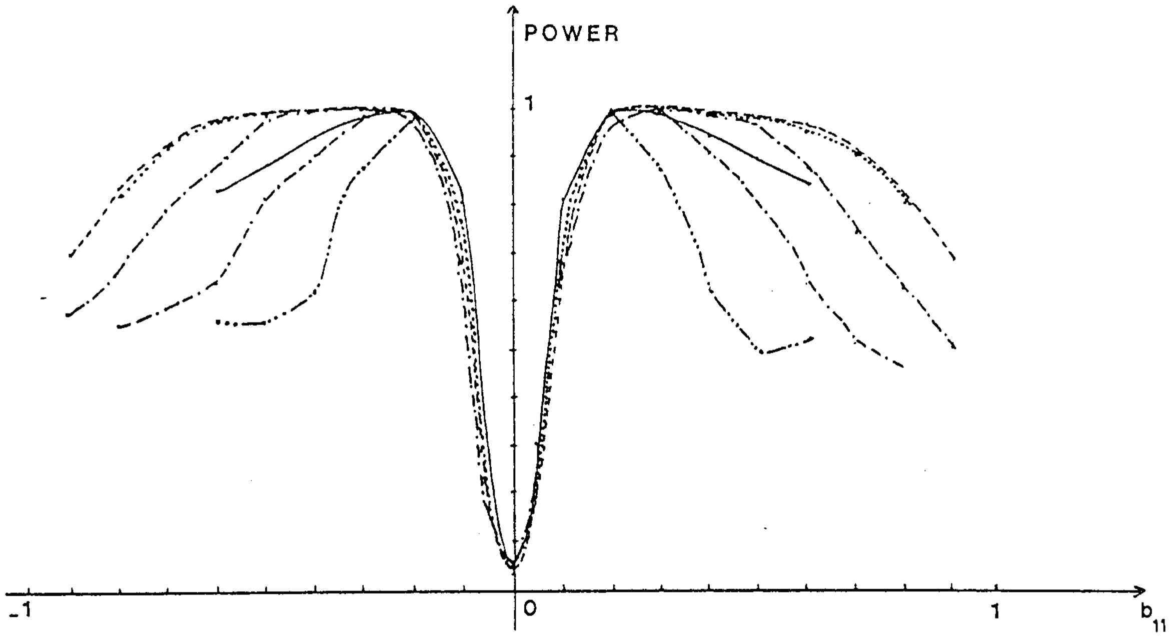

Figure 1. Power of the L.M. test at level 5% for testing model AR(1) against DBL(1, 0, 1, 1): $a = -0.8$ ——, $a = -0.5$ · — — ·, $a = -0.2$ — — — —, $a = 0.2$ · — · —, $a = 0.5$ · · · · ·, $a = 0.8$ · · · — · · · .

To have a better idea of the variation of the power of the L.M. test as a function of the parameters, we have plotted its graph in Figure 1. The different curves correspond to the different values of a. One sees that the power grows quickly in the neighbourhood of the null hypothesis when b becomes larger (in absolute value), but from a certain threshold, the power becomes worse. This phenomenon is more important when the parameter a is large. Thus the L.M. test is more efficient for testing a local bilinear alternative than testing an alternative far from the null hypothesis. This is not surprising because the L.M. test has been developed in a local context. We may remark that the empirical power function is more or less symmetric with respect to the bilinear parameter b. In the neighbourhood of the null hypothesis, this function is indeed symmetric as can be seen from the computations of the previous paragraph. Nevertheless we cannot establish its symmetry for the whole range of b.

To see if the theoretical power of the L.M. test agrees with that based

Table 2. Theoretical (column T) and empirical (column E) percentages of rejections of the L.M. test at 5% level for testing AR(1) against DBL(1, 0, 1, 1).

$a\backslash b$	−.8		−.5		−.2		.2		.5		.8	
Power	E	T	E	T	E	T	E	T	E	T	E	T
--.2	99.8	100	99.6	100	99.2	99.8	99.8	99.8	99	100	99.8	100
−.18	99.6	100	98.6	99.6	98.4	99.3	98.2	99.3	98.2	99.6	98.8	100
−.16	99	99.8	97	98.5	95.8	97.6	95.6	97.6	95.4	98.5	97.6	99.8
−.14	96.6	99	93	95.0	91.6	93.1	90.8	93.1	91.4	95.0	96.2	99
−.12	93.6	95.9	87.4	87	84	83.8	83.6	83.8	84.4	87	90.8	95.9
−.1	83.2	87	72.4	73	69.6	69.1	67.8	69.1	77.2	73	81.7	87
−.09	77	79.3	63.4	64	58.4	59.8	58.2	59.8	61.8	64	72.8	79.3
−.08	66	69.4	51.4	54	46.6	50.3	46	50.3	48.8	54	63	69.4
−.07	55	57.9	41.4	43.7	38.2	40.6	38	40.6	39.4	43.7	51.6	57.9
−.06	43.6	45.7	33.4	33.8	29.2	31.3	28.8	31.3	30.6	33.8	38	45.7
−.05	33	33.7	24.8	25.2	22.6	23.3	21.4	23.3	22.4	25.2	27	33.7
−.04	22.6	23.3	18.2	17.9	16.4	16.6	17.4	16.6	15.4	17.9	20	23.3
−.03	17.8	15.3	13.4	12.2	13	11.4	11.2	11.4	12	12.2	13.6	15.3
−.02	10.6	9.5	10.4	8.3	10.2	7.9	8.8	7.9	8	8.3	8.8	9.5
−.01	6.6	6.1	7	5.9	7	5.7	7	5.7	6.6	5.9	5	6.1
0	6.2	5	5.4	5	5.4	5	6	5	6	5	4.4	5
.01	6	6.1	6.6	5.9	5.8	5.7	5.2	5.7	5.6	5.9	5.8	6.1
.02	8.6	9.5	8.2	8.3	8.2	7.9	7	79	6.4	8.3	8.2	9.5
.03	15.8	15.3	11.6	12.2	11.4	11.4	10.6	11.4	09.2	12.2	11.4	15.3
.04	23.6	23.3	16	17.9	16.6	16.6	14.6	16.6	13.6	17.9	18	23.3
.05	31.6	33.7	22.8	25.2	20	23.3	20.8	23.3	22.6	25.2	26.2	33.7
.06	42	45.7	31	33.8	29.2	31.3	27.4	31.3	30	33.8	37.2	45.7
.07	51.2	57.9	40.4	43.7	36.6	40.6	36.8	40.6	38	43.7	50	57.9
.08	64.2	69.4	49.2	54	45.4	50.6	45.4	50.6	48.6	54	62.2	69.4
.09	73.2	79.3	59.2	64	54	59.8	55.6	59.8	57.6	64	71.4	79.3
.1	82	87	68.2	73	65.2	69.1	64	69.1	67.2	73	79.2	87
.12	91.2	95.9	82.6	87	78.8	83.6	79.2	83.8	80.2	87	88.8	95.9
.14	97.2	.99	91.4	95.0	89.2	93.1	88.4	93.1	90.8	95.0	95.4	99
.16	99.6	99.8	96.6	98.5	94.8	97.6	94.2	97.6	94.8	98.5	98	99.8
.18	100	100	98.8	99.6	97.4	99.3	97.8	99.3	98.2	99.6	98.8	100
.2	100	100	99.8	100	99.4	99.8	99.2	99.8	99.4	100	99.2	100

E: empirical
T: theoretical

on simulation, we have put in the same figure the graph obtained empirically and the graph of the theoretical power (the parameter σb being restricted to the range, $(-.2, .2)$. There are six graphs corresponding to the different values of a, displayed in Figure 2 (a,b,c,d,e,f). The empirical results and theoretical values are listed in Table 2. The theoretical power has been computed from the noncentral χ^2 distribution with noncentral parameter obtained from the formula (2.14), given in the previous section.

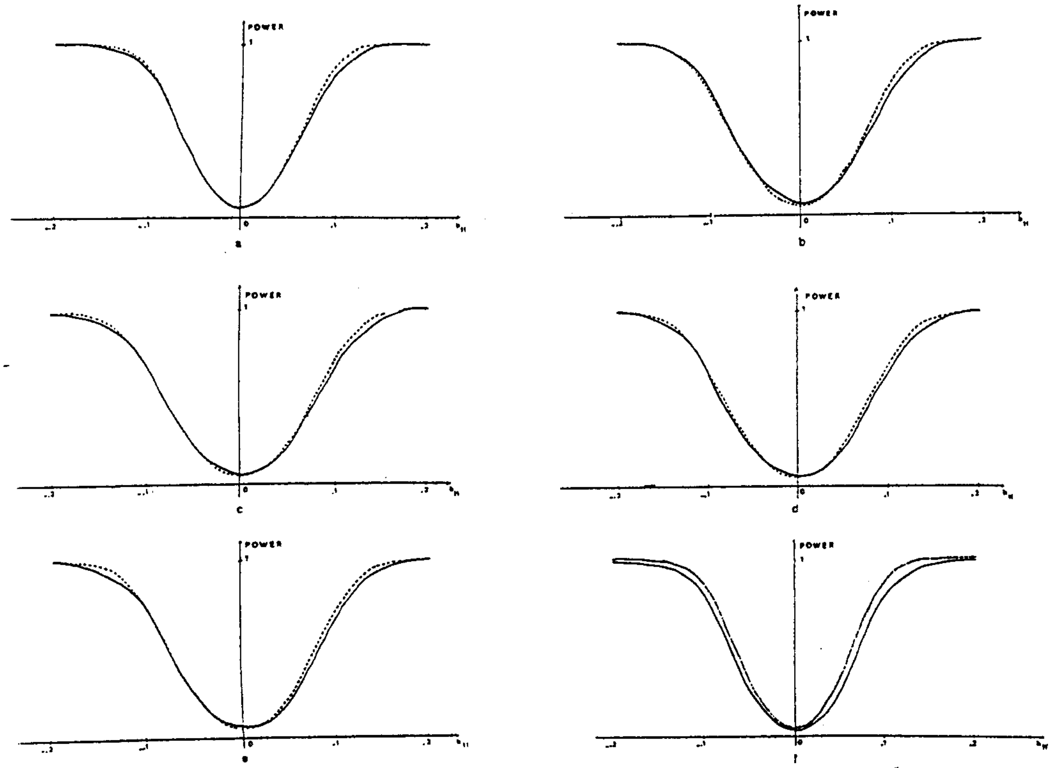

Figure 2. Theoretical and empirical powers of the L.M. test at level 5% for testing the model AR(1) against DBL(1, 0, 1, 1): empirical power ——, theoretical power $\cdots$; (a) $a = -0.8$, (b) $a = -0.5$, (c) $a = -0.2$, (d) $a = 0.2$, (e) $a = 0.5$, (f) $a = 0.8$.

Appendix: Behaviour of the score vector $\left(\frac{\partial \ell}{\partial \theta_2}\right)$ at $(\theta_1^*, 0)$ under the alternative

From the results (2.8)–(2.12) in Section 2, one may write,

$$\frac{1}{\sqrt{n}}\left(\frac{\partial \ell}{\partial \theta_2}\right)_{(\hat{\theta}_1,0)} = -\frac{1}{\sigma^2\sqrt{n}}\sum_{t=1}^{n}\epsilon_{\theta^*}(t)v_{\theta^*}(t,p)$$
$$-\frac{J_{21}J_{11}^{-1}}{\sigma^2\sqrt{n}}\sum_{t=1}^{n}\epsilon_{\theta^*}(t) + \boldsymbol{X}(t-m) + O_p(1).$$

Replacing $\epsilon_{\theta^*}(t)$ by its expression (2.10), yields:

$$\begin{aligned}\frac{1}{\sqrt{n}}\left(\frac{\partial \ell}{\partial \theta_2}\right)_{(\hat{\theta}_1,0)} &= \frac{1}{\sigma^2\sqrt{n}}\sum_{t=1}^{n}\epsilon(t)\boldsymbol{v}_{\theta^*}(t,p) - \frac{J_{21}J_{11}^{-1}}{\sigma^2\sqrt{n}}\sum_{t=1}^{n}\epsilon(t)\boldsymbol{X}(t,m)\\ &\quad - \frac{1}{\sigma^2\sqrt{n}}\sum_{t=1}^{n}\sum_{j=1}^{p}\frac{\beta_{jj}}{\sqrt{n}}X(t-j)\epsilon(t-j)\boldsymbol{v}_{\theta^*}(t,p)\\ &\quad - \frac{1}{\sigma^2\sqrt{n}}J_{21}\cdot J_{11}^{-1}\sum_{t=1}^{n}\sum_{j=1}^{p}\frac{\beta_{jj}}{\sqrt{n}}X(t-j)\epsilon(t-j)\boldsymbol{X}(t,m)\\ &\quad + O_p(1). \end{aligned} \tag{A.1}$$

Replacing $v_{\theta^*}(t)$ by (2.12) and using (2.8), the first term in the right hand side of (A.1) can be rewritten as:

$$\frac{1}{\sigma^2\sqrt{n}}\sum_{t=1}^{n}\epsilon(t)\boldsymbol{v}_0(t,p) - \frac{1}{n\sigma^2}\sum_{t=1}^{n}\Big(\sum_{j=1}^{p}\beta_{jj}X(t-j)\epsilon(t-j)\Big)\boldsymbol{X}(t,m)\epsilon(t) \tag{A.2}$$

and the third term as:

$$-\frac{1}{n\sigma^2}\sum_{t=1}^{n}\Big(\sum_{j=1}^{p}\beta_{jj}X(t-j)\epsilon(t-j)\Big)\boldsymbol{v}_0(t,p) + O_p(1). \tag{A.3}$$

Then using (A.2) and (A.3), Equation (A.1) becomes:

$$\frac{1}{\sqrt{n}}\left(\frac{\partial \ell}{\partial \theta_2}\right)_{(\hat{\theta}_1,0)} = A + B + C + D$$

where

$$\begin{aligned} A &= \frac{1}{\sigma^2\sqrt{n}}\sum_{t=1}^{n}\epsilon(t)\big[\boldsymbol{v}_\theta(t,p) - J_{21}J_{11}^{-1}\boldsymbol{X}(t,m)\big]\\ B &= -\frac{1}{\sigma^2 n}\sum_{t=1}^{n}\Big(\sum_{j=1}^{p}\beta_{jj}X(t-j)\epsilon(t-j)\Big)\boldsymbol{X}(t,m)\epsilon(t)\\ C &= -\frac{1}{\sigma^2 n}\sum_{t=1}^{n}\Big(\sum_{j=1}^{p}\beta_{jj}X(t-j)\epsilon(t-j)\Big)\boldsymbol{v}_0(t,p)\\ D &= -\frac{1}{\sigma^2 n}J_{21}J_{11}^{-1}\sum_{t=1}^{n}\Big(\sum_{j=1}^{p}\beta_{jj}X(t-j)\epsilon(t-j)\Big)\boldsymbol{X}(t,m) + O_p(1). \end{aligned}$$

Using the C.L.T., A converges in distribution, as $n \to \infty$, to the Gaussian distribution with mean zero and covariance matrix $J_{2,1} = J_{22} - J_{21}J_{11}^{-1}J_{12}$.

As for the random variables $(B)-(D)$, one may invoke the Law of Large Numbers (ergodic theorem) to find that they differ from their expectation by a term tending to zero, as n goes to infinity. But the alternative also "tends" to the null hypothesis with n, therefore one may evaluate the expectation at the null hypothesis. Thus:

$$E[(\epsilon(t)X(t-\ell)\epsilon(t-\ell)X(t-j)\epsilon(t-j)] = 0 \quad \ell = 1,\dots,p \text{ and } j = 1,\dots,m,$$
$$E[X(t-j)\epsilon(t-j)X(t-\ell)] = 0 \quad j = 1,\dots,p \text{ and } \ell = 1,\dots,m.$$

The last equality requires that the $\epsilon(t)$ have zero third order moment, but they need not be Gaussian. It follows that (B) and (D) converge to zero in probability as $n \to \infty$. As for (C), it converges to the constant vector

$$c = -\frac{1}{\sigma^2}\sum_{j=1}^{p}\beta_{jj}E[X(t-j)\epsilon(t-j)v_0(t,p)].$$

The rth element of c is

$$c_r = \frac{1}{\sigma^2}\sum_{j=1}^{p}\beta_{jj}E[X(t-j)\epsilon(t-j)X(t-r)\epsilon(t-r)].$$

Finally the random vector $n^{-1/2}(\partial\ell/\partial\theta_2)$ evaluated at $(\hat{\theta}_1, 0)$ converges in distribution to a Gaussian vector with mean c and covariance matrix $J_{1,2}$. The elements of the vector c are given as above. As for that of $J_{1,2}$, one has

$$(J_{2,1})_{rs} = \frac{1}{\sigma^2}E[X(t-r)\epsilon(t-r)X(t-s)\epsilon(t-s)],$$

since, under the null hypothesis all joint third order moments of $X(t)$ vanish, which implies that $J_{21} = 0$. If we assume that the fourth cumulant of $\epsilon(t)$ vanish (which is true if the $\epsilon(t)$ are Gaussian), then, under the null hypothesis, all joint fourth order cumulants of the process $X(t)$ vanish. Hence

$$\begin{aligned}
E[X(t-j)&\epsilon(t-j)X(t-r)\epsilon(t-r)] \\
&= \operatorname{cov}\{X(t-j),\epsilon(t-j)\}\operatorname{cov}\{X(t-r),\epsilon(t-r)\} \\
&\quad + \operatorname{cov}\{X(t-j),\epsilon(t-r)\}\operatorname{cov}\{X(t-r),\epsilon(t-j)\} \\
&\quad + \operatorname{cov}\{X(t-j),X(t-r)\}\operatorname{cov}\{\epsilon(t-j),\epsilon(t-r)\}.
\end{aligned}$$

The last two terms vanish unless $j = r$. Thus

$$E[X(t-j)\epsilon(t-j)X(t-r)\epsilon(t-r)] = \begin{cases} \sigma^4 & \text{if } j \neq r \\ 2\sigma^4 + \sigma^2 E[X^2(t)] & \text{if } j = r \end{cases}.$$

This yields

$$c_r = \sigma^2\Big(\sum_{j=1}^{k}\beta_{jj}\Big) + \beta_{rr}\{\sigma^2 + E[X^2(t)]\} \tag{A.4}$$

$$J_{2,1} = \sigma^2 + E[X^2(t)]I + \sigma^2\mathbf{1} \tag{A.5}$$

where I is the identity matrix and $\mathbf{1}$ is the matrix with all elements equal to one.

References

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

Chan, W. S. and Tong, H. (1986). On tests for non-linearity in time series analysis. *J. Forecast.* **5**, 217–228.

Guegan, D. (1984). Tests de modèles non lineaires. *Proceedings of the 3rd Franco-Belgian meeting of statisticians.* Nov. 82. Bruxelles. U.L.B. 45–65.

Hajek, J. and Sidak, Z. S. (1967). *Theory of Rank Tests.* Academic Press.

Hinich, M. J. (1982). Testing for gaussianity and linearity of a stationary time series. *J. Time Ser. Anal.* **3**, 169–176.

Keenan, D. M. (1985). A Tukey nonadditivity-type test for time series nonlinearity. *Biometrika* **72**, 39–44.

McLeod, A. I. and Li, W. K. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *J. Time Ser. Anal.* **4**, 269–273.

Luukkonen, R., Saikkonen, P. and Terasvirta, T. (1988). Testing linearity in univariate time series models. *Scand. J. Statist.* **15**, 161–175.

Moran, P. A. P. (1970). On asymptotically optimal tests of composite hypotheses. *Biometrika* **57**, 47–55.

Petruccelli, J. and Davies, N. (1986). A portmanteau test for self-exciting threshold autoregressive-type linearity in time series. *Biometrika* **73**, 687–694.

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

Subba Rao, T. and Gabr, M. M. (1980). A test for linearity of stationary time series. *J. Time Ser. Anal.* **1**, 145–158.

Tsay, R. S. (1986). Nonlinearity tests for time series. *Biometrika* **73**, 461-466.

Department of Mathematics, University of Paris XIII, CSP, Avenue J. B. Clément, 93430 Villetaneuse, France.

Laboratory of Modelling and Computation, CNRS, Grenoble, B. P. 53x, 38041 Grenoble Cedex, France.

(Received December 1989; accepted August 1991)