



## Testing for Nonlinearity with Partially Observed Time Series

Henghsiu Tsai; K. S. Chan

*Biometrika*, Vol. 87, No. 4. (Dec., 2000), pp. 805-821.

Stable URL:

<http://links.jstor.org/sici?sici=0006-3444%28200012%2987%3A4%3C805%3ATFNWPO%3E2.0.CO%3B2-5>

*Biometrika* is currently published by Biometrika Trust.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/bio.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

# Testing for nonlinearity with partially observed time series

BY HENGHSIU TSAI

*Department of Statistics, Tunghai University, Taichung, Taiwan 407, R.O.C.*  
htsai@mail.thu.edu.tw

AND K. S. CHAN

*Department of Statistics & Actuarial Science, University of Iowa, Iowa City,  
Iowa 52242, U.S.A.*  
kchan@stat.uiowa.edu

## SUMMARY

We have implemented a Lagrange multiplier test specifically for the alternative of a nonlinear continuous-time autoregressive model with the instantaneous mean having one degree of nonlinearity. The test is then extended to testing for the alternative of general nonlinear continuous-time autoregressive models with multiple degrees of nonlinearity. The performance of the test in the finite-sample case is compared with several existing tests for nonlinearity including Keenan's (1985) test, Petruccielli & Davies' (1986) test and Tsay's (1986, 1989) tests. The comparison is based on simulated data from some linear autoregressive models, self-exciting threshold autoregressive models, bilinear models and the nonlinear continuous-time autoregressive models which the Lagrange multiplier test is designed to detect. The Lagrange multiplier test outperforms the other tests in detecting the model for which it is designed. Compared with the other tests, the test has excellent power in detecting bilinear models, but seems less powerful in detecting self-exciting threshold autoregressive nonlinearity. The test is further illustrated with the Hong Kong beach water quality data.

*Some key words:* Euler scheme; Irregularity sample data; Kalman filter; Lagrange multiplier test; Stochastic differential equation.

## 1. INTRODUCTION

Let  $\{Y_{t_i}\}_{i=0,\dots,N}$  be a time series sampled with possibly unequal time intervals. We are interested in testing the data for nonlinearity. For irregularly sampled data, several tests for nonlinearity are available; see Petruccielli & Davies (1986), Tsay (1989) and Tong & Yeung (1991). These tests are, however, implicitly targeted at the alternative of threshold models and none of these tests is likelihood based. Here, we develop Lagrange multiplier score tests for the alternative of nonlinear continuous-time autoregressive models with general nonlinearity in the instantaneous mean function.

Continuous-time processes are often modelled by stochastic differential equations. For rigorous accounts of stochastic differential equations, see Kunita (1990), Lipster & Shiriyayev (1977), Øksendal (1995) and Wong & Hajek (1985). The loglikelihood function of discrete-time data sampled from a nonlinear continuous-time process is generally

intractable. However, for sufficiently small sampling intervals, a stochastic differential equation may be adequately approximated by a nonlinear difference equation which yields a tractable but approximate likelihood function. In practice, the sampling interval of the observed data may not be small enough to justify the preceding approximation. Our basic idea is to embed the observed data in a sequence of time series with a sufficiently small sampling interval. In the terminology of the incomplete-data literature, the observed data are called the incomplete data whereas the finely sampled ambient series is called the complete data; see Dempster, Laird & Rubin (1977) and Little & Rubin (1987). The latter will generally contain extensive missing data, the number of which depends on the specified sampling interval. Using this incomplete data framework, we develop formulae for approximating the Lagrange multiplier test statistics. The score is computed by Kalman filtering whereas the information matrix is computed by Kalman filtering and Monte Carlo.

In this paper, we implement the Lagrange multiplier test for a specific class of nonlinear continuous-time autoregressive models. In § 2, we briefly review nonlinear continuous-time autoregressive modelling, and in § 3 we develop the Lagrange multiplier test statistic for nonlinearity. In § 4, properties of the test are discussed, and empirical sizes of the test in the finite-sample case are reported in § 5. In § 6 the power of the test is studied via simulation, and in § 7 the test is illustrated with the Hong Kong beach water quality data. Some extensions are given in § 8.

## 2. NONLINEAR CONTINUOUS-TIME AUTOREGRESSIVE MODELLING

Many time series data,  $\{Y_t\}_{t=0,\dots,N}$ , say, are sampled with unequal time intervals, often coming from irregularly sampling an underlying continuous-time process. That is, there exists a continuous-time process  $\{X_t, t \in \mathfrak{R}\}$  such that  $Y_t = X_{t_i}$ . More generally,  $Y_t$  can be some functional of the underlying continuous-time process, measured perhaps with observation error. The underlying continuous-time process is often modelled by linear stochastic differential equations, namely a continuous-time autoregressive moving average model. This linear specification results in a tractable likelihood for the observed discrete-time data, and hence this method has been routinely used in analysing discrete-time sampled time series; see Harvey (1989), Bergstrom (1990), Tong (1990) and Jones (1993). In some cases, the underlying continuous-time process is not real, but is merely a device for providing a convenient analysis. In other cases, the continuous-time process may be the object of the study (Bergstrom, 1990). For continuous time series modelling in economics, many interesting dynamical processes are nonlinear. There is much literature on regularly sampled nonlinear time series; see Tong (1990, 1995), Granger & Teräsvirta (1993) and Cox (1997).

In order to analyse irregularly sampled nonlinear time series data, it is natural to consider nonlinear continuous-time modelling. Here, we shall mainly consider continuous-time nonlinear autoregressive models with additive noise of constant instantaneous variance, defined below. This flexible class of models will be called the NL CAR( $p$ ) models. A NL CAR( $p$ ) process is defined to be a solution of the  $p$ th-order differential equation

$$dY_t^{(p-1)} = g(Y_t, Y_t^{(1)}, \dots, Y_t^{(p-1)}; \theta) dt + \sigma dW_t, \quad (1)$$

where  $g$  is a measurable function, the superscript  $(j)$  denotes  $j$ -fold differentiation with respect to  $t$ ,  $\{W_t, t \geq 0\}$  is standard Brownian motion, and  $\theta$  and  $\sigma$  are parameters. We assume that  $\sigma > 0$ . Brownian motion is nowhere differentiable, so the solution of

equation (1) is interpreted as satisfying the integral equation

$$Y_t^{(p-1)} - Y_0^{(p-1)} = \int_0^t g(Y_s, Y_s^{(1)}, \dots, Y_s^{(p-1)}; \theta) ds + \sigma W_t.$$

The term  $g(Y_t, Y_t^{(1)}, \dots, Y_t^{(p-1)}; \theta)$  is referred to as the instantaneous mean of the process, and  $\sigma$  as the instantaneous standard deviation. For a Lipschitz continuous function  $g$ , the solution of equation (1) exists and is unique; see Theorem 5.2.9 of Karatzas & Shreve (1991). More generally,  $\sigma$  can be state dependent, but this generalisation will not be pursued here.

A NL $\text{CAR}(p)$  model becomes a continuous-time autoregressive  $\text{CAR}(p)$  model when the instantaneous mean is linear. Linearity can be checked by enlarging the  $\text{CAR}(p)$  model to a larger model in some directions of departure from the linear mean, followed by testing for the need for the enlargement. For example, we add an exponential term to the linear instantaneous mean to enlarge a  $\text{CAR}(p)$  model to a NL $\text{CAR}(p)$  model. Specifically, let  $h(X_t; \theta)$ , also simply written as  $h(X_t)$ , be a nonlinear function, where  $X_t = [X_t^{(0)}, \dots, X_t^{(p-1)}]'$ , the prime ' denotes the transpose of a vector and  $X_t^{(0)} = Y_t$ . Let  $\theta = (\lambda, \alpha_0, \dots, \alpha_p)$ . Consider the model

$$dX_t^{(p-1)} = (\alpha_0 + a'X_t + e^{\lambda h(X_t)} - 1) dt + \sigma dW_t, \tag{2}$$

where  $\alpha = (\alpha_1, \dots, \alpha_p)'$  and  $dX_t^{(j-1)} = X_t^{(j)} dt$  ( $j = 1, \dots, p - 1$ ).

The linear  $\text{CAR}(p)$  model is obtained by setting  $\lambda = 0$  in equation (2). Equation (2) can be equivalently cast in terms of the following observation and state equations (Brockwell, 1993):

$$Y_t = H'X_t \quad (t \geq 0),$$

$$dX_t = \{AX_t + (\alpha_0 + e^{\lambda h(X_t)} - 1)\delta_p\} dt + \sigma\delta_p dW_t, \tag{3}$$

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 & \dots & \alpha_p \end{bmatrix}, \quad X_t = \begin{bmatrix} X_t^{(0)} \\ X_t^{(1)} \\ \vdots \\ X_t^{(p-2)} \\ X_t^{(p-1)} \end{bmatrix}, \quad \delta_p = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \quad H = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}.$$

Equation (3) is an Itô differential equation for the state vector  $X_t$ . We assume that  $X_0$  is independent of  $\{W_t, t \geq 0\}$  and  $X_0$  is determined by initial conditions that could be random or deterministic. We also assume that  $\alpha_1 \neq 0$ , so that the inverse of  $A$  exists.

The problem of testing for nonlinearity in the instantaneous mean function becomes that of testing  $H_0: \lambda = 0$  against  $H_1: \lambda \neq 0$ . Note that  $e^{\lambda h(\cdot)} - 1 = \lambda h(\cdot) + o(\lambda)$ . For small  $\lambda$ , therefore, the alternatives are in the form of curvature as specified by  $h$  in the instantaneous mean function. This corresponds to the test of Keenan (1985) in the case of regular sampling if we set  $h(X_t) = (\alpha_0 + \alpha'X_t)^2$ . Moreover, the framework can be extended to the general case of several directions of departure from linearity, in which case

$$dX_t^{(p-1)} = (\alpha_0 + \alpha'X_t + e^{\sum_i \lambda_i h_i(X_t)} - 1) dt + \sigma dW_t,$$

where the  $h_i$ 's are suitable functions of the state variables. Setting the  $h_i$ 's to be interaction terms such as  $X_t^{(j)} X_t^{(k)}$  yields an analogue to the test of Tsay (1986).

## 3. THE TEST STATISTIC

## 3.1. Definition of the test statistic

Let the parameter space be denoted by  $\Omega$  and an arbitrary element of  $\Omega$  be denoted by  $\theta = (\lambda, \nu)$ , where  $\lambda$  is the parameter of interest and  $\nu = (\alpha_0, \dots, \alpha_p)$  are nuisance parameters. Under the null hypothesis  $H_0: \lambda = 0$ , let the constrained maximum likelihood estimate of  $\nu$  be denoted by  $\hat{\nu}_0$ , and let the vector of score statistics for  $\lambda$  evaluated at  $\lambda = 0$  be denoted by  $S(0, \hat{\nu}_0)$ . The information matrix  $I(\lambda, \nu)$  may be partitioned according to the partition of the vector  $(\lambda, \nu)$ :

$$I(\lambda, \nu) = \begin{bmatrix} I_{11} & I_{12} \\ I'_{12} & I_{22} \end{bmatrix}.$$

The Lagrange multiplier statistic for  $H_0$  is given by

$$\text{LM} = S(0, \hat{\nu}_0)' I_{11.2}^{-1}(0, \hat{\nu}_0) S(0, \hat{\nu}_0) = \left( \frac{\partial l_Y}{\partial \lambda} \Big|_{\lambda=0, \hat{\nu}_0} \right)^2 I_{11.2}^{-1}(0, \hat{\nu}_0), \quad (4)$$

where  $I_{11.2} = I_{11} - I_{22}^{-1} I'_{12}$ , and  $l_Y$  is the loglikelihood function of the observed data; see Cox & Hinkley (1974, p. 324) for further discussion. Alternatively,  $I(0, \hat{\nu}_0)$  may be replaced by its expected value. In practice,  $\sigma^2$  is a nuisance parameter which is needed in computing (4), and is replaced by the constrained maximum likelihood estimate of  $\sigma^2$  under  $H_0$ .

Under  $H_0$ , the NL CAR( $p$ ) model (2) is a linear CAR( $p$ ) model, and the likelihood function of a CAR( $p$ ) model can be computed via Kalman filters. The constrained maximum likelihood estimate  $\hat{\nu}_0$  is then obtained by means of a nonlinear optimisation algorithm; see § 3.2. The score vector is also computed via Kalman filters, whereas the observed and the expected Fisher information are computed by a Monte Carlo method; see §§ 3.3–3.5.

## 3.2. Constrained maximum likelihood estimation

When  $\lambda = 0$  in equation (2), the model becomes the linear CAR( $p$ ) model

$$dX_t^{(p-1)} = (\alpha_0 + \alpha' X_t) dt + \sigma dW_t,$$

and equation (3) becomes

$$dX_t = (AX_t + \alpha_0 \delta_p) dt + \sigma \delta_p dW_t. \quad (5)$$

The solution of (5) can be written as

$$X_t = e^{At} X_0 + \alpha_0 \int_0^t e^{A(t-u)} \delta_p du + \sigma \int_0^t e^{A(t-u)} \delta_p dW_u,$$

where  $e^{At} = I + \sum_{n=1}^{\infty} \{(At)^n (n!)^{-1}\}$ , and  $I$  is the identity matrix.

Let the mean vector of  $\{X_t\}$  be denoted by  $\mu_t$ , which satisfies the equation

$$\mu_t = \frac{\alpha_0}{\alpha_1} (e^{At} - I) H + e^{At} \mu_0.$$

The covariance kernel of  $\{X_t\}$ , denoted by  $\gamma_{s,t}$ , equals

$$\begin{aligned} \gamma_{s,t} &= E\{(X_s - \mu_s)(X_t - \mu_t)'\} = e^{As} V_0 e^{A't} + \sigma^2 \int_0^{t \wedge s} e^{A(s-u)} \delta_p \delta_p' e^{A'(t-u)} du \\ &= \begin{cases} e^{A(s-t)} V_t & (0 \leq t \leq s < \infty), \\ V_s e^{A'(t-s)} & (0 \leq s \leq t < \infty), \end{cases} \end{aligned}$$

where  $t \wedge s = \min(s, t)$  and

$$V_t = \gamma_{t,t} = e^{At} V_0 e^{A't} + \sigma^2 \int_0^t e^{A(t-u)} \delta_p \delta_p' e^{A'(t-u)} du.$$

It follows from the above equations for  $\mu_t$  and  $V_t$  that the states and the observations,  $X_{t_i}$  and  $Y_{t_i}$ , at the sampling times  $t_0, t_1, \dots$ , satisfy the discrete-time state and observation equations

$$X_{t_{i+1}} = \mu_{t_{i+1}} + e^{A(t_{i+1}-t_i)}(X_{t_i} - \mu_{t_i}) + Z_{t_i} \quad (i = 0, 1, \dots), \quad (6)$$

$$Y_{t_i} = H' X_{t_i} \quad (i = 0, 1, \dots), \quad (7)$$

where  $Z_{t_i}$  is independent of  $X_{t_i}$ , and  $\{Z_{t_i}, i = 0, 1, \dots\}$  is an independent sequence of Gaussian random vectors with zero mean and covariance matrices

$$\Sigma_i = E(Z_{t_i} Z_{t_i}') = \sigma^2 \int_{t_i}^{t_{i+1}} e^{A(t_{i+1}-u)} \delta_p \delta_p' e^{A'(t_{i+1}-u)} du.$$

These equations are needed for applications of the Kalman recursions; see for example Brockwell & Davis (1991, Ch. 12). Let capital letters be used for random vectors and corresponding lower case letters for observed vectors. Let  $Y = \{Y_{t_i}\}_{i=0, \dots, N}$  and define  $\hat{X}_{t_i|s}$  as the conditional expectation of  $X_{t_i}$  based on the observations up to time  $s$  and  $P_{t_i|s}$  the corresponding covariance matrix; that is

$$\hat{X}_{t_i|s} = E(X_{t_i} | y_j, j \leq s) = (\hat{X}_{t_i|s}^{(0)}, \dots, \hat{X}_{t_i|s}^{(p-1)}), \quad P_{t_i|s} = \text{var}(X_{t_i} | y_j, j \leq s).$$

Then we can compute recursively  $\tilde{y}_{t_i} = y_{t_i} - \hat{X}_{t_i|t_{i-1}}^{(0)}$ , the predictive residuals, and  $p_{t_i|t_{i-1}}^{(1,1)}$ , the (1, 1) element of  $P_{t_i|t_{i-1}}$  ( $i \geq 1$ ), which are required for computing minus twice the loglikelihood function:

$$-2l_Y(\theta, \sigma^2) = \sum_{i=0}^N \left( \frac{\tilde{y}_{t_i}^2}{p_{t_i|t_{i-1}}^{(1,1)}} + \log p_{t_i|t_{i-1}}^{(1,1)} \right) + (N + 1) \log(2\pi).$$

We start with a diffuse initial condition as we do not assume stationarity; that is we let  $\hat{X}_{t_{-1}|t_{-1}} = [\bar{y}, 0, \dots, 0]'$  and  $P_{t_{-1}|t_{-1}} = \delta s_y^2 I$ , where  $t_{-1} < t_0$  is some arbitrarily chosen time point,  $\delta$  is some positive number, and  $\bar{y}$  and  $s_y^2$  are the sample mean and sample variance of  $y$ , respectively. A reasonable choice for  $\delta$  is 5.

A nonlinear optimisation algorithm can then be used in conjunction with the expression for  $-2l_Y(\theta, \sigma^2)$  to find the maximum likelihood estimate of  $(\theta, \sigma^2)$ . The calculation of  $e^{At}$  is most readily performed by first block-diagonalising  $A$  and then applying a Padé approximation to each block (Ward, 1977). For computing of  $\Sigma_i$ , see Shoji & Ozaki (1998) and Tsai & Chan (2000), or an unpublished University of Iowa technical report by H. Tsai and K. S. Chan, that uses an EM approach to obtain an approximate but more efficient constrained maximum likelihood estimator.

### 3.3. The score vector

Let  $Y = y$  be the observed incomplete data, and  $X = \{X_0, X_{1/m}, \dots, X_{k_N/m}\}$  be the unobserved complete data of which  $Y$  is a measurable function, where  $Y = \{Y_{t_j}\}_{0, \dots, N}$  and  $m$  is chosen to be some moderately large integer such that, for each  $0 \leq j \leq N$ ,  $t_j = k_j/m$  for

some integer  $k_j$ . Henceforth in this section, we write  $X_k$  for  $X_{k/m}$ ,  $Y_{k_j}$  for  $Y_{t_j}$  and  $q$  for  $k_N$ . Also,  $E_\theta(\cdot|y)$ ,  $\text{var}_\theta(\cdot|y)$  and  $\text{cov}_\theta(\cdot|y)$  denote the conditional expectation, variance and covariance of the enclosed expression given by  $Y=y$ , respectively, where  $\theta$  is the true parameter. To simplify notation, the parameter  $\theta$  is omitted if no confusion would occur.

Since the likelihood function for nonlinear continuous-time models is intractable, the stochastic differential equation

$$dX_t^{(p-1)} = (\alpha_0 + \alpha'X_t + e^{\lambda h(X_t)} - 1) dt + \sigma dW_t$$

using Euler's method (Milstein, 1995, p. 18) is approximated by the difference equation

$$\Delta X_t^{(p-1)} = (\alpha_0 + \alpha'X_t + e^{\lambda h(X_t)} - 1) \Delta t + \sigma \Delta W_t, \quad (8)$$

where  $\Delta X_t^{(p-1)} = X_{t+\Delta t}^{(p-1)} - X_t^{(p-1)}$ ,  $\Delta W_t = W_{t+\Delta t} - W_t$  and  $\Delta t = 1/m$ .

Conditional on  $X_0$ , equation (8) can be used to get  $l_X(\theta, \sigma^2)$ , the loglikelihood function of  $X$ :

$$l_X(\theta, \sigma^2) = C - \frac{m}{2\sigma^2} \sum_{j=1}^q \left[ X_j^{(p-1)} - X_{j-1}^{(p-1)} - \frac{1}{m} \{(\alpha_0 + \alpha'X_{j-1}) + e^{\lambda h(X_{j-1})} - 1\} \right]^2, \quad (9)$$

where  $C = -(q/2) \log(2\pi/m) - (q/2) \log \sigma^2$ . Let  $D$  be the differential operator with respect to  $\theta$ . Louis (1982) has shown that

$$Dl_Y(\theta, \sigma^2) = E_{\theta, \sigma^2} \{Dl_X(\theta, \sigma^2)|y\}, \quad (10)$$

$$\begin{aligned} -D^2l_Y(\theta, \sigma^2) &= E_{\theta, \sigma^2} \{-D^2l_X(\theta, \sigma^2)|y\} - E_{\theta, \sigma^2} \{Dl_X(\theta, \sigma^2)D'l_X(\theta, \sigma^2)|y\} \\ &\quad + Dl_Y(\theta, \sigma^2)D'l_Y(\theta, \sigma^2). \end{aligned} \quad (11)$$

In § 3.4, equation (10) is used to estimate the score vector via Kalman filters, and equation (11) is used to estimate the observed Fisher information via Kalman filters and the Monte Carlo method.

Under the null hypothesis, the conditional distribution of  $X$  given  $Y=y$  is Gaussian. For  $0 \leq k \leq q$ , the conditional mean and conditional variance of  $X_k$  given  $Y=y$  can be computed by a forward Kalman filtering sequence, followed by backward iteration (Anderson & Moore, 1979, p. 189).

We now outline the Kalman computation. For  $0 \leq k \leq q$ ,  $\hat{X}_{k|k}$ ,  $\hat{X}_{k+1|k}$ ,  $P_{k|k}$  and  $P_{k+1|k}$  can be computed via a forward Kalman filter as follows. First, let  $\hat{X}_{-1|-1} = [\bar{y}, 0, \dots, 0]'$  and  $P_{-1|-1} = \delta s_y^2 I$ , as in § 3.2. Then, for  $0 \leq k \leq q$ , compute  $P_{k|k-1}$  and  $P_{k|k}$  recursively as follows:

$$P_{k|k-1} = e^{A/m} P_{k-1|k-1} e^{A'/m} + \Sigma, \quad (12)$$

$$P_{k|k} = \begin{cases} P_{k|k-1} - \frac{1}{p_{k|k-1}^{(1,1)}} P_{k|k-1} H H' P_{k|k-1} & \text{if } k \in \{k_0, \dots, k_N\}, \\ P_{k|k-1} & \text{if } k \notin \{k_0, \dots, k_N\}, \end{cases} \quad (13)$$

where  $p_{k|k-1}^{(i,j)}$  is the  $(i, j)$  element of  $P_{k|k-1}$ ,  $\Sigma = V - e^{A/m} V e^{A'/m}$  and  $V$  is the solution of the matrix equation  $AV + VA' = -\sigma^2 \delta_p \delta_p'$ . The preceding result about  $\Sigma$  is well known for the stationary case but it also holds for the nonstationary case; see Tsai & Chan (2000) for a proof.

For  $0 \leq k \leq q$ , compute  $\hat{X}_{k|k-1}$  and  $\hat{X}_{k|k}$  recursively through

$$\hat{X}_{k|k-1} = \mu + e^{A/m} (\hat{X}_{k-1|k-1} - \mu), \quad (14)$$

$$\hat{X}_{k|k} = \begin{cases} \hat{X}_{k|k-1} + \frac{1}{P_{k|k-1}^{(1,1)}} P_{k|k-1} H(y_k - \hat{X}_{k|k-1}) & \text{if } k \in \{k_0, \dots, k_N\}, \\ \hat{X}_{k|k-1}, & \text{if } k \notin \{k_0, \dots, k_N\}, \end{cases} \quad (15)$$

where  $\mu = -\alpha_0 H/\alpha_1$ . Next, using the backwards Kalman filter, we can compute the conditional means and the variances of the  $X_k$ 's given all observed data:

$$\hat{X}_{k|q} = \hat{X}_{k|k} + B_k(\hat{X}_{k+1|q} - \hat{X}_{k+1|k}), \quad (16)$$

$$P_{k|q} = P_{k|k} + B_k(P_{k+1|q} - P_{k+1|k})B'_k, \quad (17)$$

where  $B_k = P_{k|k} e^{A/m} P_{k+1|k}^{-1}$ , for  $k = q - 1, \dots, 0$ .

For  $0 \leq k \leq q$ , the conditional covariance of  $X_k$  and  $X_{k-1}$  given  $Y = y$  will also be needed, and is stated in the following theorem from the University of Iowa technical report by H. Tsai and K. S. Chan.

THEOREM 1. (a) Given  $Y_l = y_l$  ( $l \leq k$ ) and  $X_{k+1} = x_{k+1}$ ,

$$X_k \sim N(\hat{X}_{k|k} + B_k(X_{k+1} - \hat{X}_{k+1|k}), P_{k|k} - B_k P_{k+1|k} B'_k),$$

where  $B_k = P_{k|k} e^{A/m} P_{k+1|k}^{-1}$ , for  $0 \leq k \leq q - 1$ .

(b) For  $1 \leq j \leq q - 1$ ,

$$E(X_j X'_{j-1} | y) = \hat{X}_{j|q} \hat{X}'_{j-1|q} + P_{j|q} B'_{j-1}. \quad (18)$$

To obtain the expressions for the score and the observed and expected information we need to specify the form of the function  $h(X_t)$ . Here, we follow Keenan's (1985) approach by taking  $h(X_t) = (\alpha_0 + \alpha' X_t)^2$ . Note that, under  $H_0: \lambda = 0$ ,

$$\begin{aligned} \frac{\partial l_Y}{\partial \lambda} \Big|_{\lambda=0} &= \frac{1}{\sigma^2} \sum_{j=1}^q E \left[ \left\{ \delta'_p(X_j - X_{j-1}) - \frac{1}{m} (\alpha_0 + \alpha' X_{j-1}) \right\} (\alpha_0 + \alpha' X_{j-1})^2 \Big| y \right] \\ &= \frac{1}{\sigma^2} \sum_{j=1}^q \left[ 2\delta'_p(P_{j|q} B'_{j-1} - P_{j-1|q}) \alpha (\alpha_0 + \alpha' X_{j-1|q}) \right. \\ &\quad \left. + \delta'_p(X_{j|q} - X_{j-1|q}) \{ (\alpha' P_{j-1|q} \alpha) + (\alpha_0 + \alpha' X_{j-1|q})^2 \} \right. \\ &\quad \left. - \frac{1}{m} \{ 3(\alpha' P_{j-1|q} \alpha) (\alpha_0 + \alpha' X_{j-1|q}) + (\alpha_0 + \alpha' X_{j-1|q})^3 \} \right], \quad (19) \end{aligned}$$

where we have used the fact that  $\text{cov}(X_j, X_{j-1} | y) = P_{j|q} B'_{j-1}$ , which follows from equation (18).

Note that applying Euler's method in approximating the likelihood function of the observations is equivalent to using step functions to approximate the integrals involved in the score vector. For a representation of the integral form of the score and further discussion, see the technical report by Tsai and Chan.

### 3.4. The observed information matrix

The partial derivatives, up to order 2, of the loglikelihood function of  $X$  evaluated under  $H_0$  are given as follows:

$$\begin{aligned} \left. \frac{\partial l_X}{\partial \lambda} \right|_{\lambda=0} &= \frac{1}{\sigma^2} \sum_{j=1}^q \hat{\varepsilon}_j (\alpha_0 + \alpha' X_{j-1})^2, \\ \left. \frac{\partial l_X}{\partial \alpha_0} \right|_{\lambda=0} &= \frac{1}{\sigma^2} \sum_{j=1}^q \hat{\varepsilon}_j, \\ \left. \frac{\partial l_X}{\partial \alpha_r} \right|_{\lambda=0} &= \frac{1}{\sigma^2} \sum_{j=1}^q \hat{\varepsilon}_j X'_{j-1} \delta_r \quad (r = 1, \dots, p), \\ \left. \frac{\partial^2 l_X}{\partial \lambda^2} \right|_{\lambda=0} &= \frac{1}{\sigma^2} \sum_{j=1}^q \left\{ \left( \hat{\varepsilon}_j - \frac{1}{m} \right) (\alpha_0 + \alpha' X_{j-1})^4 \right\}, \\ \left. \frac{\partial^2 l_X}{\partial \lambda \partial \alpha_0} \right|_{\lambda=0} &= \frac{1}{\sigma^2} \sum_{j=1}^q \left[ \left\{ 2\hat{\varepsilon}_j - \frac{1}{m} (\alpha_0 + \alpha' X_{j-1}) \right\} (\alpha_0 + \alpha' X_{j-1}) \right], \\ \left. \frac{\partial^2 l_X}{\partial \lambda \partial \alpha_r} \right|_{\lambda=0} &= \frac{1}{\sigma^2} \sum_{j=1}^q \left[ \left\{ 2\hat{\varepsilon}_j - \frac{1}{m} (\alpha_0 + \alpha' X_{j-1}) \right\} (\alpha_0 + \alpha' X_{j-1}) X'_{j-1} \delta_r \right] \quad (r = 1, \dots, p), \\ \left. \frac{\partial^2 l_X}{\partial \alpha_0^2} \right|_{\lambda=0} &= -\frac{N}{\sigma^2}, \\ \left. \frac{\partial^2 l_X}{\partial \alpha_0 \partial \alpha_r} \right|_{\lambda=0} &= -\frac{1}{m\sigma^2} \sum_{j=1}^q X'_{j-1} \delta_r \quad (r = 1, \dots, p), \\ \left. \frac{\partial^2 l_X}{\partial \alpha_s \partial \alpha_t} \right|_{\lambda=0} &= -\frac{1}{m\sigma^2} \sum_{j=1}^q \delta'_s X_{j-1} X'_{j-1} \delta_t \quad (1 \leq r, s \leq p), \end{aligned}$$

where, for  $r = 1, \dots, p$ ,  $\delta_r = [\delta_{j,r}]_{p \times 1}$  is a  $p \times 1$  vector with  $\delta_{j,r} = 1$ , if  $j = r$ , and  $\delta_{j,r} = 0$  if  $j \neq r$ , and  $\hat{\varepsilon}_j = \delta'_p (X_j - X_{j-1}) - (\alpha_0 + \alpha' X_{j-1})/m$ .

Using the above equations and equations (10) and (11), we can obtain a closed-form expression for the observed information, but this is very complicated and here we prefer to use Monte Carlo methods to compute the observed information.

We need to simulate  $X_0, \dots, X_q$  given  $Y = y$  under the null hypothesis. Let  $[W|z]$  denote the conditional density of  $W$  given  $Z = z$ . Then

$$\begin{aligned} [X_0, X_1, \dots, X_q | y] &= \left( \prod_{k=0}^{q-1} [X_k | X_{k+1}, \dots, X_q, y] \right) [X_q | y] \\ &= \left( \prod_{k=0}^{q-1} [X_k | X_{k+1}, y_l, l \leq k] \right) [X_q | y], \end{aligned}$$

where the second equality follows from the Markov property. As a result of Gaussianity,  $\hat{X}_{q|q}$  and  $P_{q|q}$  allow us to simulate  $X_q$  given  $Y = y$ . For  $0 \leq k \leq q-1$ ,  $X_k$  given  $X_{k+1}$  and  $y_l$  ( $l \leq k$ ) is a Normal random vector with its first two moments given in Theorem 1(a).

Given  $y_l$  ( $l \leq q$ ),  $X_q \sim N(\hat{X}_{q|q}, P_{q|q})$ , and Theorem 1(a) shows that the other  $X_k$ 's satisfy the backward equation

$$X_k = \hat{X}_{k|k} + B_k(X_{k+1} - \hat{X}_{k+1|k}) + F_k, \quad (20)$$

where  $F_k \sim N(0, \rho_k)$ ,  $\rho_k = P_{k|k} - B_k P_{k+1|k} B'_k$ , for  $0 \leq k \leq q-1$ . Equation (20) enables us to simulate the  $X_j$ 's given  $Y = y$ . Apply forward Kalman filtering to obtain  $\hat{X}_{k|k}$ ,  $\hat{X}_{k+1|k}$ ,  $B_k$

and  $\rho_k$ , for  $k = 0, \dots, q - 1$ , simulate  $X_q$  from  $N(\hat{X}_{q|q}, P_{q|q})$ , and then use (20) to get the  $X_j$ 's in a backwards manner.

### 3.5. The expected information matrix

An alternative to the observed Fisher information is the expected information

$$-E_{\theta, \sigma^2} \{D^2 l_Y(\theta, \sigma^2)\} = E_{\theta, \sigma^2} \{Dl_Y(\theta, \sigma^2) D'l_Y(\theta, \sigma^2)\}.$$

Note that, under  $H_0: \lambda = 0$ ,  $Dl_Y(\theta, \sigma^2)$  is a  $(p + 2) \times 1$  vector with elements given by (19) along with

$$\begin{aligned} \frac{\partial l_Y}{\partial \alpha_0} &= \frac{1}{\sigma^2} \sum_{j=1}^q E \left\{ \delta'_p(X_j - X_{j-1}) - \frac{1}{m} (\alpha_0 + \alpha' X_{j-1}) \middle| y \right\} \\ &= \frac{1}{\sigma^2} \left\{ \delta'_p(\hat{X}_{q|q} - \hat{X}_{0|q}) - \frac{q}{m} \alpha_0 - \frac{\alpha'}{m} \sum_{j=1}^q \hat{X}_{j-1|q} \right\}, \end{aligned} \tag{21}$$

$$\begin{aligned} \frac{\partial l_Y}{\partial \alpha_r} &= \frac{1}{\sigma^2} \sum_{j=1}^q E \left[ \left\{ \delta'_p(X_j - X_{j-1}) - \frac{1}{m} (\alpha_0 + \alpha' X_{j-1}) \right\} X'_{j-1} \delta_r \middle| y \right] \\ &= \frac{1}{\sigma^2} \left[ \sum_{j=1}^q E \{ \delta'_p(X_j - X_{j-1}) X'_{j-1} \delta_r | y \} - \frac{1}{m} \alpha_0 \sum_{j=1}^q \hat{X}'_{j-1|q} \delta_r \right. \\ &\quad \left. - \frac{1}{m} \alpha' \sum_{j=1}^q (P_{j-1|q} + \hat{X}_{j-1|q} \hat{X}'_{j-1|q}) \delta_r \right] \quad (r = 1, \dots, p). \end{aligned} \tag{22}$$

Equations (19), (21) and (22) can be used to compute  $Dl_Y(\theta, \sigma^2)$ , and  $I(\theta, \sigma^2)$  can be estimated by Monte Carlo as follows. First, do (i) to (iii)  $L$  times recursively, where  $L$  is a prescribed positive integer:

- (i) generate random variables  $\{Y_{t_j}^*\}_{j=0,1,\dots,N}$  from the stationary CAR( $p$ ) model with parameter  $(\theta, \sigma^2)$ , by the discrete-time state and observation equations (6) and (7);
- (ii) compute  $Dl_{Y^*}(\theta, \sigma^2)$  by equations (19), (21) and (22);
- (iii) compute  $Dl_{Y^*}(\theta, \sigma^2) D'l_{Y^*}(\theta, \sigma^2)$ .

Then the average of the  $L$  matrices computed from (iii) is an estimate of  $I(\theta, \sigma^2)$ .

## 4. PROPERTIES OF THE LAGRANGE MULTIPLIER TEST

**THEOREM 2** (*Asymptotic null distribution of the Lagrange multiplier statistic*). Under the null hypothesis, the statistic defined by equation (4) asymptotically has a  $\chi^2$  distribution with one degree of freedom, if  $O/N$ , where  $O$  is the observed information matrix, converges to a positive definite matrix under the null hypothesis.

The proof of Theorem 2 follows from Theorems 1 and 2 and Corollary 1 of Sweeting (1980). Next, we show that the test statistic is invariant with respect to linear transformations. Consequently, the choice of  $m$  in the computation of the Lagrange multiplier test statistic is independent of the scale of the series.

**THEOREM 3.** Consider model (2) with  $h(X_t) = (\alpha_0 + \alpha' X_t)^2$ :

$$dX_t^{(p-1)} = (\alpha_0 + \alpha' X_t + e^{\lambda(\alpha_0 + \alpha' X_t)^2} - 1) dt + \sigma dW_t.$$

Let  $\{Y_{t_j}\}_{j=0,\dots,N}$  and  $\{Y_{t_j}^*\}_{j=0,\dots,N}$  be two series of discrete-time data, where  $Y_{t_j}^* = aY_{t_j} + b$ ,

for all  $j$ , and  $a \neq 0$ . Let LM be the Lagrange multiplier statistic of  $\{Y_{t_j}\}_{j=0,\dots,N}$  defined by equation (4), that is

$$LM = \left( \frac{\partial l_Y}{\partial \lambda} \Big|_{\hat{\theta}_0, \hat{\sigma}_0^2} \right)^2 I_{11 \cdot 2}^{-1}(\hat{\theta}_0, \hat{\sigma}_0^2), \tag{23}$$

where  $\hat{\theta}_0 = (0, \hat{\alpha}_0, \dots, \hat{\alpha}_p)$  and  $\hat{\sigma}_0^2$  are the restricted maximum likelihood estimators of  $\theta = (\lambda, \alpha_0, \dots, \alpha_p)$  and  $\sigma^2$  based on  $\{Y_{t_j}\}$  under  $H_0: \lambda = 0$ . Here,  $I_{11 \cdot 2}$  is computed based on the expected information. Similarly, let LM\* denote the Lagrange multiplier statistic of  $\{Y_{t_j}^*\}$  defined in equation (23) with  $\{Y_{t_j}\}$  replaced by  $\{Y_{t_j}^*\}$  and  $(\hat{\theta}_0, \hat{\sigma}_0^2)$  replaced by  $(\hat{\theta}_0^*, \hat{\sigma}_0^{*2})$ , the restricted maximum likelihood estimators of the linear CAR( $p$ ) model based on  $\{Y_{t_j}^*\}$ . Then  $LM^* = LM$ .

### 5. EMPIRICAL SIZES OF THE LAGRANGE MULTIPLIER TEST

In this section we use Monte Carlo to study the empirical size of the Lagrange multiplier test with nominal significance level equal to 0.10, 0.05 and 0.01. Regularly spaced time series data,  $Y_{t_i} = X_{t_i}^{(0)}$  ( $i = 0, 1, 2, \dots, N$ ) are simulated from two stationary continuous-time autoregressive processes:

Model 1,  $dX_{t_i}^{(0)} = -0.25X_{t_i}^{(0)} dt + dW_{t_i}$ ,

Model 2,  $dX_{t_i}^{(1)} = (-0.3X_{t_i}^{(0)} - 0.2X_{t_i}^{(1)}) dt + dW_{t_i}$ .

The empirical rejection frequencies of the Lagrange multiplier statistic based on the observed information and based on 1000 replications are given in Table 1. In Table 1,  $m - 1$  is the number of imputations between every two adjacent observations, and  $L$  is the number of replicates used in the Monte Carlo simulation for computing the observed and the expected information matrices.

Table 1. *The empirical frequencies of rejecting a linear model based on 1000 replications generated from Models 1 and 2*

	N	m	L	Nominal levels		
				0.100	0.050	0.010
Observed information						
Model 1	100	20	100	0.241	0.209	0.180
	100	20	2000	0.230	0.211	0.179
	100	100	100	0.221	0.192	0.159
	200	20	100	0.201	0.162	0.125
	400	20	100	0.154	0.112	0.063
	1000	20	100	0.122	0.073	0.037
Expected information						
Model 1	50	20	100	0.044	0.017	0.004
	50	40	100	0.045	0.019	0.003
	100	20	100	0.070	0.033	0.008
	100	40	100	0.063	0.029	0.008
	200	20	100	0.073	0.036	0.012
	200	40	100	0.078	0.041	0.011
	400	20	100	0.098	0.051	0.018
Model 2	100	20	100	0.074	0.032	0.007

Table 1 shows that, if the test is based on the observed information matrices, the empirical rejection frequencies of the test tend to be much larger than the nominal frequencies, especially for sample size not larger than 400. This is because, at the constrained maximum likelihood estimator under  $H_0$ , the observed information need not be positive definite. If we use the expected information, the test statistics converge to the  $\chi^2$  distribution much faster. The empirical sizes of the test are much closer to the nominal sizes, although they seem to be conservative, especially for sample size not larger than 100. Table 1 also shows that the empirical frequencies of the test seem robust to the values of  $m$  and  $L$  whenever  $m \geq 20$  and  $L \geq 100$ .

## 6. POWER OF THE TEST AND COMPARISON

In this section we use Monte Carlo to compare the Lagrange multiplier test with Keenan's (1985) test, Petrucci & Davies' (1986) test and Tsay's (1986, 1989) tests. The computation of the Lagrange multiplier test statistic is based on the expected information. In practice, the choice of  $m$  is important. One way to choose  $m$  is to set  $m$  initially at 20, say, and then to increase  $m$  in a systematic way until the test statistics converge. Here, for the simulation, we simply choose  $m$  to be 20. The values of  $p$  and  $L$  are chosen to be 1 and 100, respectively. According to the suggestions of Keenan (1985) and Tsay (1989), we choose an autoregressive approximation of order  $M = 4$  to calculate Keenan's test and Tsay's (1986) test, and use

$$r_{\min} = (N/10) + p,$$

with  $N$  the sample size and  $p$  the fitted autoregressive order, to compute the Petrucci & Davies test and Tsay's (1989) test. Here  $r_{\min}$  is defined in Petrucci & Davies (1986). The values of  $p$  and  $d$  in the Petrucci & Davies test and Tsay's (1989) test are both chosen to be 1.

For simplicity, only regularly spaced time series data are simulated. All of the empirical frequencies of rejecting a linear process are based on 1000 realisations with critical value equal to 5%, and the sample size is always 101. Three underlying models were studied.

*Nonlinear continuous-time autoregressive model.* This is the model that the Lagrange multiplier test is specifically designed to detect:

$$dX_t = (\alpha_0 + \alpha_1 X_t + e^{\lambda(\alpha_0 + \alpha_1 X_t)^2} - 1) dt + \sigma dW_t, \quad (24)$$

with  $(\alpha_0, \alpha_1) = (0.0, -0.25)$  and  $\lambda$  varies from  $-3.0$  to  $+1.0$ . The data are simulated using the local linearisation scheme of Ozaki (1985); see also Ozaki (1992), Shoji (1998) and Shoji & Ozaki (1997, 1998). First, we simulate a discretised process  $\{X_{t_i}, i = 1, \dots, 12\,000\}$  with step size  $h = 0.0125$ . The regularly spaced time series data  $\{Y_i = X_{4000+80i}, i = 0, 1, \dots, 100\}$  are then used for testing for nonlinearity. The empirical frequencies of rejecting a linear process are given in Table 2. It is clear that, not surprisingly, the Lagrange multiplier test is generally more powerful than the other tests. Also note that, for the linear model, that is  $\lambda = 0$ , the Lagrange multiplier test results in relatively smaller type I errors than the other tests. For  $\lambda = -2$ , the empirical frequency of the Lagrange multiplier test with  $(m, h) = (40, 0.0125)$  is 0.960, which is 0.014 larger than for the test with  $(m, h) = (20, 0.0125)$ . Note that the NLAR(1) process defined by equation (24) with  $\alpha_1 < 0$  is asymptotically stationary if  $\lambda \leq 0$  and is nonstationary if  $\lambda > 0$ . The model becomes linear when  $\lambda \rightarrow -\infty$ ; hence the decrease in the power of the Lagrange multiplier test for  $\lambda < -2$ .

Table 2. *The empirical frequencies of rejecting a linear model based on 1000 replications generated from the model in (24). The nominal size of the test is 5%*

Test	$\lambda$										
	-3	-2.5	-2	-1.5	-1	-0.5	0	0.25	0.5	0.75	1
LM	0.592	0.810	0.946	0.703	0.209	0.058	0.035	0.042	0.101	0.271	0.591
Tsay (1989)	0.227	0.153	0.092	0.058	0.045	0.042	0.059	0.087	0.139	0.182	0.290
PD	0.215	0.232	0.111	0.036	0.035	0.044	0.046	0.059	0.086	0.106	0.111
Tsay (1986)	0.550	0.613	0.511	0.331	0.161	0.071	0.044	0.050	0.078	0.180	0.272
Keenan (1985)	0.251	0.266	0.208	0.170	0.139	0.065	0.033	0.036	0.082	0.155	0.189

LM, Lagrange multiplier; PD, Petrucci & Davies (1986).

*Self-exciting threshold autoregressive model.* We used the model used in the power study of Petrucci & Davies (1986):

$$Y_t = \begin{cases} \phi_1 Y_{t-1} + a_t & (Y_t \geq 0), \\ \phi_2 Y_{t-1} + a_t & (Y_t < 0), \end{cases} \tag{25}$$

where  $a_t \sim N(0, 1)$ ,  $\phi_1 = -0.5, 0.0, 0.5$  and  $\phi_2 = -2.0, 0.0$ . The stationary region of (25) was shown to be  $\phi_1 < 1$ ,  $\phi_2 < 1$  and  $\phi_1 \phi_2 < 1$  by Petrucci & Woolford (1984). Thus  $(\phi_1, \phi_2) = (-0.5, -2)$  is outside the stationary region. To mitigate the transient effect, we generated 1100 observations from model (25) with  $Y_0 = 0$  and discarded the first 1000 values. With 1000 realisations, the empirical frequencies of rejecting a linear process are given in Table 3. The Lagrange multiplier test is less powerful than the other four tests. For the linear model  $(\phi_1, \phi_2) = (0, 0)$ , the Lagrange multiplier test results in smaller Type I error.

Table 3. *The empirical frequencies of rejecting a linear model based on 1000 replications generated from the model in (25). The nominal size of the test is 5%*

Test	$\phi_1 = -0.5$		$\phi_1 = 0$		$\phi_1 = 0.5$	
	$\phi_2 = -2$	$\phi_2 = 0$	$\phi_2 = -2$	$\phi_2 = 0$	$\phi_2 = -2$	$\phi_2 = 0$
LM	0.050	0.018	0.757	0.017	0.375	0.054
Tsay (1989)	0.951	0.195	1.000	0.045	0.997	0.214
PD	0.933	0.116	0.987	0.043	0.990	0.148
Tsay (1986)	1.000	0.117	0.974	0.055	1.000	0.124
Keenan (1985)	0.973	0.139	0.998	0.056	0.410	0.167

LM, Lagrange multiplier; PD, Petrucci & Davies (1986).

One of the reasons for the poor performance of the Lagrange multiplier test with  $p = 1$  is that the autocorrelations of a CAR(1) model are always positive, since  $e^{\alpha_1 t} > 0$  always. However, some of the autocorrelations of model (25) may be negative. This restriction on the correlation pattern can be lifted by considering higher-order continuous-time autoregressive models. When we try the Lagrange multiplier test for the higher orders  $p = 2, 3$  and  $4$ , the empirical frequencies become  $0.911, 0.881$  and  $0.402$ , respectively, in testing the case with  $(\phi_1, \phi_2) = (-0.5, -2)$ . For the same case, the empirical frequency of the Lagrange multiplier test is  $0.064$  with  $(p, m) = (1, 40)$ , which is only  $0.014$  larger than with  $(p, m) = (1, 20)$ . The low power of the Lagrange multiplier test may also be caused by the slow convergence rate of Euler’s method; we shall study the use of other approximation methods

in the future. Another way to increase the power of the Lagrange multiplier tests is to consider more general alternative hypotheses; see § 8.

*Bilinear model.* We used the bilinear model also considered by Petrucci & Davies (1986):

$$Y_t = (\phi + \beta a_t)Y_{t-1} + a_t, \quad (26)$$

where  $a_t \sim N(0, 1)$ ,  $\beta = 1.0$  and  $\phi = -0.9, -0.5, 0.5, 0.9$ . Results from the same simulation method as used above are given in Table 4. In general, the Lagrange multiplier test is less powerful than Tsay's (1986) test, but is more powerful than the others.

Table 4. *The empirical frequencies of rejecting a linear model based on 1000 replications generated from the model in (26). The nominal size of the test is 5%*

Test	$\beta = 1$ $\phi = -0.9$	$\beta = 1$ $\phi = -0.5$	$\beta = 1$ $\phi = 0.5$	$\beta = 1$ $\phi = 0.9$
LM	0.765	0.649	0.809	0.878
Tsay (1989)	0.853	0.626	0.504	0.714
PD	0.562	0.498	0.428	0.641
Tsay (1986)	0.946	0.808	0.810	0.930
Keenan (1985)	0.701	0.497	0.447	0.540

LM, Lagrange multiplier; PD, Petrucci & Davies (1986).

## 7. APPLICATION

*Example: Hong Kong beach water quality data (Tong, 1990).* Over the years of 1980–85, irregularly sampled water quality data were taken one to three times a month from four Hong Kong beaches, namely Butterfly, Anglers, Repulse Bay and Shek O. On each occasion, the dissolved oxygen content, which acts as an indicator of bacterial activities, was measured in  $10^{-5}$  mg per litre. The unit of time is taken as 10 days. The time series plots of the log-transformed data are displayed in Fig. 1. Petrucci & Davies' (1986) test and Tsay's (1989) test were applied to the log-transformed data with  $p = 1$  and  $r_{\min} = 11$ . The results are reported in Tong & Yeung (1990) and Tong (1990, p. 279), with the tentative conclusion based on 5% significance level that the dissolved oxygen series of Butterfly, Anglers and Repulse Bay are linear, whereas the Shek O series is nonlinear. Tong (1990) noted that the Repulse Bay series is marginally nonlinear as one of his tests yields a  $p$ -value of 0.115.

We tested the four log-transformed series for nonlinearity using the Lagrange multiplier test statistic after removal of the same outliers as were deleted by Tong & Yeung (1990). The  $p$ -values are shown in Table 5. For each model, we also report the Akaike information criterion, defined by  $AIC = -2\{l_Y(\hat{\theta}) - r\}$ , where  $r$  is the number of parameters in the model, and  $\hat{\theta}$  is the restricted maximum likelihood estimate under the null hypothesis.

Based on AIC, order  $p = 1$  is selected for all series, so the tentative conclusions drawn below are mainly based on the tests with order equal to one. Furthermore, for the case of unit autoregressive order, the tests were replicated with  $m = 20$  and  $m = 40$ , and yielded similar results. The  $p$ -values for Butterfly, Repulse Bay and Shek O data shown in Table 5 essentially suggest linearity, whereas those of Anglers beach suggest nonlinearity. The

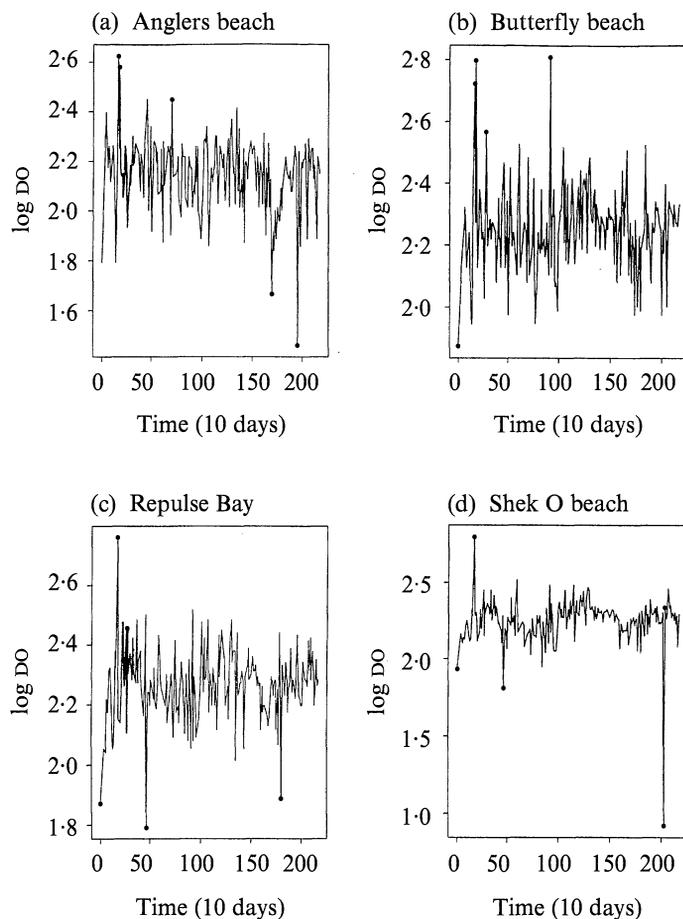


Fig. 1. Time series plots of the log transformed dissolved oxygen, DO, content in four beaches in Hong Kong; outliers marked as solid circles were deleted from the analysis.

Table 5. Results of the Lagrange multiplier test applied to log transformed dissolved oxygen content of Hong Kong beaches after outlier deletion

$p$	$m$	Sample size	$p$ -value	AIC	Sample size	$p$ -value	AIC
				Anglers beach			
1	20	193	0.0081	-240.96	192	0.1775	-266.41
1	40	193	0.0041	-242.40	192	0.1510	-268.05
2	20	193	0.1568	-210.56	192	0.2476	-260.80
3	20	193	0.0314	-175.20	192	0.3052	-245.97
4	20	193	0.0026	-88.06	192	0.3640	-220.30
				Repulse Bay			
1	20	188	0.6992	-313.90	192	0.4311	-337.23
1	40	188	0.6868	-316.50	192	0.3638	-339.21
2	20	188	0.9760	-304.52	192	0.5224	-302.41
3	20	188	0.4782	-285.73	192	0.5825	-277.77
4	20	188	0.0510	-250.25	192	0.0000	-205.86
				Shek O beach			

discrepancy between Tong & Yeung’s (1990) results and ours is intriguing. As shown in § 6, the Lagrange multiplier test is quite powerful in detecting smooth nonlinearity but less powerful in detecting threshold nonlinearity. We conjecture that the discrepancy may have arisen because the dissolved oxygen process at Anglers beach and Shek O beach may be of different nonlinear types. The modelling of these two nonlinear processes is clearly an interesting future research problem.

8. SOME EXTENSIONS

The Lagrange multiplier approach we have developed for (2) can be extended to models of the general form

$$dX_t^{(p-1)} = \{\alpha_0 + \alpha'X_t + g(X_t; \theta, \lambda)\} dt + \sigma dW_t, \tag{27}$$

where  $\theta = (\alpha_0, \dots, \alpha_p)$ ,  $\lambda = (\lambda_1, \dots, \lambda_l)$ , for some positive integer  $l$ , and  $g(X_t; \theta, \lambda = 0) = 0$ . Then testing for linearity is equivalent to testing  $H_0: \lambda_j = 0$ , for all  $j$ , versus  $H_1: \lambda_j \neq 0$ , for some  $j$ . The stochastic differential equation (27) is approximated by the difference equation

$$\Delta X_t^{(p-1)} = \{\alpha_0 + \alpha'X_t + g(X_t)\} \Delta t + \sigma \Delta W_t, \tag{28}$$

where we write  $g(X_t) = g(X_t; \theta, \lambda)$  for brevity,

$$\Delta X_t^{(p-1)} = X_{t+\Delta t}^{(p-1)} - X_t^{(p-1)}, \quad \Delta W_t = W_{t+\Delta t} - W_t, \quad \Delta t = 1/m.$$

With notation defined in §§ 3·3–3·5, equation (28) allows us to write  $l_X(\theta, \lambda, \sigma^2)$ , the log-likelihood function of  $X$  conditional on  $X_0$ , as

$$l_X(\theta, \lambda, \sigma^2) = C - \frac{m}{2\sigma^2} \sum_{j=1}^q \left[ X_j^{(p-1)} - X_{j-1}^{(p-1)} - \frac{1}{m} \{(\alpha_0 + \alpha'X_{j-1}) + g(X_{j-1})\} \right]^2, \tag{29}$$

where  $C = -(q/2) \log(2\pi/m) - (q/2) \log \sigma^2$ . It is straightforward to replicate the methodology of §§ 3·1–3·5 and to implement the test. As an example, consider the following model that is analogous to the alternative model used in Tsay’s (1986) discrete-time work:

$$dX_t^{(p-1)} = \left\{ \alpha_0 + \alpha'X_t + \exp \left( \sum_{r=1}^p \sum_{s=r}^p \lambda_{r,s} X_t^{(r-1)} X_t^{(s-1)} \right) - 1 \right\} dt + \sigma dW_t.$$

The elements of the score vector are given as follows. For  $1 \leq r \leq s \leq p$ ,

$$\begin{aligned} & \left. \frac{\partial l_Y}{\partial \lambda_{r,s}} \right|_{\lambda_{r,s}=0, \text{ all } r,s} \\ &= \frac{1}{\sigma^2} \sum_{j=1}^q E \left[ \left\{ \delta'_p(X_j - X_{j-1}) - \frac{1}{m} (\alpha_0 + \alpha'X_{j-1}) \right\} (\delta'_r X_{j-1})(\delta'_s X_{j-1}) \middle| y \right] \\ &= \frac{1}{\sigma^2} \sum_{j=1}^q \left[ \delta'_p(X_{j|q} - X_{j-1|q})(\delta'_r X_{j-1|q})(\delta'_s X_{j-1|q}) + \delta'_p(X_{j|q} - X_{j-1|q})(\delta'_r P_{j-1|q} \delta_s) \right. \\ & \quad + (\delta'_r X_{j-1|q}) \delta'_p(P_{j|q} B'_{j-1} - P_{j-1|q}) \delta_s + (\delta'_s X_{j-1|q}) \delta'_p(P_{j|q} B'_{j-1} - P_{j-1|q}) \delta_r \\ & \quad - \frac{1}{m} \{(\alpha_0 + \alpha'X_{j-1|q})(\delta'_r X_{j-1|q})(\delta'_s X_{j-1|q}) + (\alpha_0 + \alpha'X_{j-1|q})(\delta'_r P_{j-1|q} \delta_s) \\ & \quad \left. + (\delta'_r X_{j-1|q})(\alpha' P_{j-1|q} \delta_s) + (\delta'_s X_{j-1|q})(\alpha' P_{j-1|q} \delta_r) \right\} \end{aligned}$$

We leave for future research the problem of the order determination for nonlinear continuous-time autoregressive models.

#### ACKNOWLEDGEMENT

Part of this work is based on Tsai's Ph.D. dissertation. We would like to thank the University of Iowa for partial support. We thank the editor, an associate editor and a referee for constructive comments leading to a streamlined version of the paper.

#### APPENDIX

##### *Proof of Theorem 3*

From equations (21), (22) and (19) and some algebra, we have the following relationship between the first derivatives of the loglikelihood functions of  $\{Y_{t_j}^*\}$  and  $\{Y_{t_j}\}$ :

$$\begin{aligned}\frac{\partial l_{Y^*}}{\partial \alpha_0^*} \Big|_{\hat{\theta}_0^*, \hat{\sigma}_0^{*2}} &= \frac{1}{a} \frac{\partial l_Y}{\partial \alpha_0} \Big|_{\hat{\theta}_0, \hat{\sigma}_0^2}, \\ \frac{\partial l_{Y^*}}{\partial \alpha_r^*} \Big|_{\hat{\theta}_0^*, \hat{\sigma}_0^{*2}} &= \frac{\partial l_Y}{\partial \alpha_r} \Big|_{\hat{\theta}_0, \hat{\sigma}_0^2} + \frac{b}{a} \frac{\partial l_Y}{\partial \alpha_0} \Big|_{\hat{\theta}_0, \hat{\sigma}_0^2} H' \delta_r \quad (r = 1, \dots, p), \\ \frac{\partial l_{Y^*}}{\partial \lambda^*} \Big|_{\hat{\theta}_0^*, \hat{\sigma}_0^{*2}} &= a \frac{\partial l_Y}{\partial \lambda} \Big|_{\hat{\theta}_0, \hat{\sigma}_0^2},\end{aligned}$$

where  $H$  and  $\delta_r$  are defined as before. Let

$$D(\hat{\theta}_0, \hat{\sigma}_0^2) = \left( \frac{\partial l_Y}{\partial \lambda} \Big|_{\hat{\theta}_0, \hat{\sigma}_0^2}, \frac{\partial l_Y}{\partial \alpha_0} \Big|_{\hat{\theta}_0, \hat{\sigma}_0^2}, \dots, \frac{\partial l_Y}{\partial \alpha_p} \Big|_{\hat{\theta}_0, \hat{\sigma}_0^2} \right)'$$

denote the vector of first derivatives evaluated at the restricted maximum likelihood estimates under  $H_0: \lambda = 0$ . Then, from the above equations, we have  $D^*(\hat{\theta}_0^*, \hat{\sigma}_0^{*2}) = QD(\hat{\theta}_0, \hat{\sigma}_0^2)$ , where

$$Q = \begin{bmatrix} a & 0 & 0 & 0 & \dots & 0 \\ 0 & 1/a & 0 & 0 & \dots & 0 \\ 0 & b/a & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{bmatrix} = \begin{bmatrix} a & O' \\ O & B \end{bmatrix},$$

in which  $O = [0]_{(p+1) \times 1}$ . The expected information for  $\{Y_{t_j}^*\}_{j=0,1,\dots,N}$ , evaluated at the restricted maximum likelihood estimate, is

$$I^*(\hat{\theta}_0^*, \hat{\sigma}_0^{*2}) = QE_{\hat{\theta}_0, \hat{\sigma}_0^2} [D(\hat{\theta}_0, \hat{\sigma}_0^2) D(\hat{\theta}_0, \hat{\sigma}_0^2)'] Q' = \begin{bmatrix} a^2 I_{11}(\hat{\theta}_0, \hat{\sigma}_0^2) & a I_{12}(\hat{\theta}_0, \hat{\sigma}_0^2) B' \\ a B I_{21}(\hat{\theta}_0, \hat{\sigma}_0^2) & B I_{22}(\hat{\theta}_0, \hat{\sigma}_0^2) B' \end{bmatrix}.$$

Thus

$$\begin{aligned}I_{11 \cdot 2}^*(\hat{\theta}_0^*, \hat{\sigma}_0^{*2}) &= a^2 I_{11}(\hat{\theta}_0, \hat{\sigma}_0^2) - a I_{12}(\hat{\theta}_0, \hat{\sigma}_0^2) B' \{B I_{22}(\hat{\theta}_0, \hat{\sigma}_0^2) B'\}^{-1} \{a B I_{21}(\hat{\theta}_0, \hat{\sigma}_0^2)\} \\ &= a^2 I_{11 \cdot 2}(\hat{\theta}_0, \hat{\sigma}_0^2).\end{aligned}$$

These results follow from the expression of the Langrange multiplier statistic given in (23).

#### REFERENCES

- ANDERSON, B. D. O. & MOORE, J. (1979). *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall.  
BERGSTROM, A. R. (1990). *Continuous Time Econometric Modeling*. New York: Oxford University Press.

- BROCKWELL, P. J. (1993). Threshold ARMA processes in continuous time. In *Dimension Estimation and Models*, Ed. H. Tong, pp. 170–90. River Edge, NJ: World Scientific.
- BROCKWELL, P. J. & DAVIS, R. A. (1991). *Time Series: Theory and Methods*. New York: Springer-Verlag.
- COX, D. R. (1997). The current position of statistics: a personal view (with Discussion). *Int. Statist. Rev.* **65**, 261–90.
- COX, D. R. & HINKLEY, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall.
- DEMPSTER, A. P., LAIRD, N. M. & RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with Discussion). *J. R. Statist. Soc. B* **39**, 1–38.
- GRANGER, C. W. J. & TERÄSVIRTA, T. (1993). *Modelling Nonlinear Economic Relationships*. New York: Oxford University Press.
- HARVEY, A. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. New York: Cambridge University Press.
- JONES, R. H. (1993). *Longitudinal Data with Serial Correlation: A State-space Approach*. London: Chapman and Hall.
- KARATZAS, I. & SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*. New York: Springer-Verlag.
- KEENAN, D. M. (1985). A Tukey nonadditivity-type test for time series nonlinearity. *Biometrika* **72**, 39–44.
- KUNITA, H. (1990). *Stochastic Flows and Stochastic Differential Equations*. Cambridge: Cambridge University Press.
- LIPSTER, R. S. & SHIRYAYEV, A. N. (1977). *Statistics of Random Processes I: General Theory*. New York: Springer-Verlag.
- LITTLE, R. J. A. & RUBIN, D. B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.
- LOUIS, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *J. R. Statist. Soc. B* **44**, 226–33.
- MILSTEIN, G. N. (1995). *Numerical Integration of Stochastic Differential Equations*. Boston: Kluwer Academic Publishers.
- ØKSENDAL, B. K. (1995). *Stochastic Differential Equations: An Introduction with Applications*. New York: Springer.
- OZAKI, T. (1985). Non-linear time series models and dynamical systems. In *Handbook of Statistics, 5: Time Series in the Time Domain*, Ed. E. J. Hannan, P. R. Krishnaiah and M. M. Rao, pp. 25–83. New York: Elsevier Science.
- OZAKI, T. (1992). A bridge between nonlinear time series models and nonlinear stochastic dynamical systems: a local linearization approach. *Statist. Sinica* **2**, 113–35.
- PETRUCCELLI, J. D. & DAVIES, N. (1986). A portmanteau test for self-exciting threshold autoregressive-type nonlinearity in time series. *Biometrika* **73**, 687–94.
- PETRUCCELLI, J. D. & WOOLFORD, S. W. (1984). A threshold AR(1) model. *J. Appl. Prob.* **21**, 270–86.
- SHOJI, I. (1998). Approximation of continuous time stochastic processes by a local linearisation method. *Math. Comp.* **67**, 287–98.
- SHOJI, I. & OZAKI, T. (1997). Comparative study of estimation methods for continuous time stochastic processes. *J. Time Ser. Anal.* **18**, 485–506.
- SHOJI, I. & OZAKI, T. (1998). A statistical method of estimation and simulation for systems of stochastic differential equations. *Biometrika* **85**, 240–3.
- SWEETING, T. J. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *Ann. Statist.* **8**, 1375–81.
- TONG, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.
- TONG, H. (1995). A personal overview of non-linear time series analysis from a chaos perspective (with Discussion). *Scand J. Statist.* **22**, 399–445.
- TONG, H. & YEUNG, I. (1990). On tests for threshold type nonlinearity in irregularly spaced time series. *J. Statist. Comp. Simul.* **34**, 177–94.
- TONG, H. & YEUNG, I. (1991). On tests for SETAR-type nonlinearity in partially observed time series. *Appl. Statist.* **40**, 43–62.
- TSAI, H. & CHAN, K. S. (2000). A note on the covariance structure of a continuous-time ARMA process. *Statist. Sinica* **10**, 989–98.
- TSAY, R. S. (1986). Nonlinearity tests for time series. *Biometrika* **73**, 461–6.
- TSAY, R. S. (1989). Testing and modeling threshold autoregressive process. *J. Am. Statist. Assoc.* **84**, 231–40.
- WARD, R. C. (1977). Numerical computation of the matrix exponential with accuracy estimate. *SIAM J. Numer. Anal.* **14**, 600–10.
- WONG, E. & HAJEK, B. (1985). *Stochastic Processes in Engineering Systems*. Berlin: Springer-Verlag.

[Received March 1999. Revised May 2000]