Biometrika (2004), **91**, 1, *pp.* 195–209

Printed in Great Britain

Generalised likelihood ratio tests for spectral density

By JIANQING FAN

Department of Operation Research and Financial Engineering, Princeton University, Princeton, New Jersey 08544, U.S.A.

jqfan@princeton.edu

AND WENYANG ZHANG

Institute of Mathematics and Statistics, University of Kent, Canterbury, Kent CT2 7NF, U.K.

w.zhang@kent.ac.uk

Summary

There are few techniques available for testing whether or not a family of parametric times series models fits a set of data reasonably well without serious restrictions on the forms of alternative models. In this paper, we consider generalised likelihood ratio tests of whether or not the spectral density function of a stationary time series admits certain parametric forms. We propose a bias correction method for the generalised likelihood ratio test of Fan et al. (2001). In particular, our methods can be applied to test whether or not a residual series is white noise. Sampling properties of the proposed tests are established. A bootstrap approach is proposed for estimating the null distribution of the test statistics. Simulation studies investigate the accuracy of the proposed bootstrap estimate and compare the power of the various ways of constructing the generalised likelihood ratio tests as well as some classic methods like the Cramér–von Mises and Ljung–Box tests. Our results favour the newly proposed bias reduction method using the local likelihood estimator.

Some key words: ARMA model; Generalised likelihood ratio test; Local least squares; Local likelihood; Periodogram; Spectral density.

1. Introduction

To verify whether or not a family of parametric models fits a given set of time series, we need to specify a class of alternative models. A common practice is to embed the family of parametric models into an even larger family of parametric models and to use this larger family of models as the alternative models. This parametric approach implicitly assumes that the larger family of parametric models contains the true model, and this assumption is not always true. Therefore, one can further enlarge the family of alternative models to include the structured nonparametric models; see Chapters 8 and 9 of Fan & Yao (2003).

The structured nonparametric models are much more flexible than parametric ones, but it is still possible that structured nonparametric models such as the functional autoregressive models (Chen & Tsay, 1993) and the additive model (Hastie & Tibshirani, 1990) do not fit the data well. One may therefore not wish to impose any structure on the alternative models. Under the stationarity assumption, a convenient technique is to test whether or not the spectral density of the data admits a certain parametric form. This is the approach that we follow.

The tests proposed in this paper are mainly based on the stationary assumption. This reduces dramatically the danger of model misspecification, but at the cost of power reduction when one can correctly specify a smaller family of alternative models. Tests based on spectral density functions verify only the autocovariance structure of an underlying process. When the null model is rejected, with high confidence, the family of models under the null hypothesis does not fit well the data. If one assumes further that the series comes from a stationary Gaussian process, then the autocovariance structure completely governs the law of the process.

Nonparametric methods (Brillinger, 1981; Brockwell & Davis, 1991; Fan & Yao, 2003) provide estimates of the spectral density under alternative models. Under the parametric family of models, one can obtain the form of spectral density and hence an estimate of the spectral density can be formed under the null hypothesis. The distance between the estimates under the null and the alternative hypotheses provides a measure of discrepancy between them.

In nonparametric regression models, Fan et al. (2001) argued that the likelihood ratio statistic is one of the most natural measures. However, for nonparametric models, the maximum likelihood ratio test usually does not exist, and even if it does exist it is not easily computed. Furthermore, they showed that the nonparametric maximum likelihood ratio test does not achieve the optimal rate of convergence for hypothesis testing in the sense of Ingster (1993a,b,c). This led them to replace the maximum likelihood estimator under the alternative model by any reasonable nonparametric estimator, resulting in a generalised likelihood ratio statistic. They demonstrated that the asymptotic null distribution of the generalised likelihood ratio statistic is asymptotically χ^2 and is independent of the nuisance parameters, and the generalised likelihood ratio test is optimal in the sense of Ingster (1993a,b,c). The first property allows one to compute the p-value easily.

The problem of testing whether or not a spectral density admits a certain parametric form is approximately the same as that of testing against a family of parametric models in the nonparametric regression setting, based on the observed periodogram. Thus, the generalised likelihood ratio test can be employed. We also introduce a technique for reducing the biases of the nonparametric estimator when the underlying distribution is indeed from a parametric family of models. This reduces the biases of the asymptotic null distribution of the generalised likelihood ratio test statistic.

There are several other approaches for validating a parametric form for the spectral density. Paparoditis (2000) constructed a testing procedure based on the Priestley–Chao estimator and an L_2-distance, and established the asymptotic normality of his test statistic. Some authors also considered constructed tests without smoothing, as with Dzhaparidze's (1986, p. 273) test statistic based on a cumulative rescaled spectral density. The asymptotic distribution of his test statistic is identical to that of the Cramér–von Mises test, considered by Anderson (1993), based on the integrated squared difference between the standardised sample spectral distribution function and the standardised spectral distribution under the null hypothesis. Anderson (1993) also considered the corresponding Kolmogorov–Smirnov test and derived the asymptotic null distribution. The null distributions of the test statistics can also be approximated via frequency domain bootstrap; see for example Paparoditis (2000) and references therein.

In §2, we introduce the generalised likelihood ratio test for the spectral density. The bias reduction techniques are given in §3. Section 4 presents the asymptotic properties of the proposed tests under the restrictions of stationarity, linearity and positivity of the spectrum. Numerical examples are given in §5 and technical proofs are deferred to the Appendix.

2. The generalised likelihood ratio test

2·1. *Introduction*

Let X_t, for $t = 0, \pm 1, \pm 2, \ldots$, be a stationary time series with mean zero and autocovariance function

$$\gamma(u) = E(X_t X_{t+u}) \quad (u = 0, \pm 1, \pm 2, \ldots).$$

Its spectral density is

$$g(x) = (2\pi)^{-1} \sum_{u=-\infty}^{\infty} \gamma(u) \exp(-iux), \quad x \in [0, \pi].$$

Let $X_1, \ldots, X_T$ be the observed time series. The raw material for estimating the spectral density is the periodogram

$$I_T(w_k) = T^{-1} \left| \sum_{t=1}^{T} X_t e^{-itw_k} \right|^2, \quad w_k = 2\pi k/T \quad (k = 1, \ldots, n,\ n = [(T-1)/2]).$$

It is well known that the periodogram is not a consistent estimator of the spectral density (Brillinger, 1981, Ch. 5; Brockwell & Davis, 1991, Ch. 10). Smoothing techniques are needed.

There are two approaches for smoothing $\{I_T(w_k)\}$. The methods can be classified by whether they are based on the spline or kernel-local linear fit, or the least-squares or the Whittle (1962) likelihood-based fit. For example, Wahba (1980) applied the least-squares method to the log-periodogram using spline approximations, while Pawitan & O'Sullivan (1994) and Kooperberg et al. (1995a,b) used the Whittle likelihood to estimate parameters in the spline models. Fan & Kreutzberger (1998) proposed automatic procedures for estimating spectral densities, using a local linear fit. Both the least-squares and the Whittle-likelihood methods are studied there, and it was shown that the Whittle-likelihood type of estimators are asymptotically more efficient. The approaches of Fan & Kreutzberger (1998) will be employed here for constructing the generalised likelihood ratio test statistics.

2·2. *Relationship to nonparametric regression*

It is known (Fan & Yao, 2003) that periodograms $I_T(w_k)$ are asymptotically exponentially distributed with mean $g(w_k)$ and asymptotically independent:

$$(2\pi)^{-1} I_T(w_k) = g(w_k) V_k + R_{n,k} \quad (k = 1, \ldots, n),$$

where V_k $(k = 1, \ldots, n)$ are independently and identically distributed with the standard exponential distribution and $R_{n,k}$ is a term that is asymptotically negligible. Furthermore, let $Y_k = \log\{I_T(w_k)/(2\pi)\}$ and $m(.) = \log g(.)$. Then we have

$$Y_k = m(w_k) + z_k + r_k \quad (k = 1, \ldots, n), \tag{2·1}$$

where

$$r_k = \log[1 + R_{n,k}/\{g(w_k) V_k\}], \quad z_k = \log V_k.$$

Thus, $\{z_k, k=1,\ldots,n\}$ are independently and identically distributed random variables with density function

$$f_z(x) = \exp\{-\exp(x) + x\},$$

and r_k is an asymptotically negligible term; see Lemma A1.

Suppose we wish to test whether or not the spectral density of an observed time series belongs to a specific parametric family $\{g_\theta(.): \theta \in \Theta\}$. The problem can be formulated as testing the hypothesis

$$H_0: g(.) = g_\theta(.) \quad \text{versus} \quad H_1: g(.) \neq g_\theta(.),$$

which is equivalent to

$$H_0: m(.) = m_\theta(.) \quad \text{versus} \quad H_1: m(.) \neq m_\theta(.), \tag{2·2}$$

where $m_\theta(.) = \log g_\theta(.)$. To test (2·2), we apply the generalised likelihood ratio test of Fan et al. (2001), pretending that the following regression model holds:

$$Y_k = m(w_k) + z_k \quad (k=1,\ldots,n), \tag{2·3}$$

which is an approximation of (2·1) obtained by ignoring r_k.

The mean of z_k in model (2·3) is the Euler constant, which is $E(z_k) = C_0 = -0{\cdot}57721$, and the variance is $\text{var}(z_k) = \pi^2/6$; see Davis & Jones (1968). Let $Y_k^* = Y_k - C_0$ and $z_k^* = z_k - C_0$. By (2·3), we have the standard nonparametric regression model

$$Y_k^* = m(w_k) + z_k^* \quad (k=1,\ldots,n).$$

Appealing to the method in Fan et al. (2001), we can easily construct the least-squares-based test statistic. Since the distribution of z_k is not normal, the test based on the least-squares estimate does not fully use the likelihood information and cannot be powerful. Furthermore, the likelihood-based approach is more appealing. Indeed, in an earlier draft of this paper, we conducted extensive simulations which show the likelihood-based approach is more powerful than the least-squares one. We omit those results to save space.

2·3 *Likelihood-based approaches*

For any given spectral density function, the loglikelihood function associated with (2·3) is

$$\sum_{k=1}^{n} [Y_k - m(w_k) - \exp\{Y_k - m(w_k)\}].$$

For any x, approximating $m(w_k)$ by the linear function $a + b(w_k - x)$ for w_k near x, we obtain the local loglikelihood function

$$\sum_{k=1}^{n} [Y_k - a - b(w_k - x) - \exp\{Y_k - a - b(w_k - x)\}] K_h(w_k - x). \tag{2·4}$$

The local maximum likelihood estimator $\hat{m}_{\text{LK}}(x)$ of $m(x)$ is $\hat{a}$ in the maximiser $(\hat{a}, \hat{b})$ of (2·4); see Fan & Kreutzberger (1998).

Under the null hypothesis, the loglikelihood function of (2·3) would be

$$\sum_{k=1}^{n} [Y_k - m_\theta(w_k) - \exp\{Y_k - m_\theta(w_k)\}].$$

Its maximiser $\hat{\theta}$ would be the maximum likelihood estimate of θ.

The generalised likelihood test statistic can be constructed as

$$T_{\mathrm{LK}} = \sum_{k=1}^{n} [\exp\{Y_k - m_{\hat{\theta}}(w_k)\} + m_{\hat{\theta}}(w_k) - \exp\{Y_k - \hat{m}_{\mathrm{LK}}(w_k)\} - \hat{m}_{\mathrm{LK}}(w_k)].$$

This would have been the maximum likelihood ratio test, had $\hat{m}_{\mathrm{LK}}$ been the nonparametric maximum likelihood estimator. If T_{LK} is bigger than a selected critical value, reject (2·2); otherwise, accept.

2·4. *Bandwidth selection*

The test statistic T_{LK} depends on the choice of the bandwidth h, and the optimal bandwidth for hypothesis testing differs from that for estimating the spectral density (Ingster, 1993a,b,c). The problem of choosing the optimal bandwidth for hypothesis testing has not been seriously studied in the literature, partly because the optimal bandwidth for estimating the underlying nonparametric function provides, intuitively, a good proxy for the optimal bandwidth for nonparametric testing. For this reason, we will use the data-driven bandwidth for estimating the spectral density function as the bandwidth for the hypothesis testing. More careful study of a good choice of the bandwidth for the generalised likelihood ratio test is beyond the scope of this study.

In spectral density estimation, Fan & Kreutzberger (1998) give a data-driven bandwidth selection rule, based on the pre-asymptotic substitution method of Fan & Gijbels (1995), for the least-squares estimate of the log-spectral density and that for the local likelihood estimator. That bandwidth selection rule will be used in our paper. We omit the details.

2·5. *Bootstrap estimate of null distribution*

To implement the generalised likelihood ratio statistics, we need to obtain their distributions under the null hypothesis. The asymptotic distributions, studied in §4, are independent of the nuisance parameters under the null hypothesis, thereby exhibiting what Fan et al. (2001) refer to as the Wilks phenomenon. We can fix these parameters at the estimated values and obtain by simulation the bootstrap estimate of the null distributions.

The schematic algorithm of the bootstrap estimate is as follows. The algorithm can be applied to both the least-squares based estimate and the maximum likelihood based estimate.

Step 1. Obtain the parametric estimate $\hat{\theta}$ and the nonparametric estimate $\hat{m}$ along with its associated bandwidth $\hat{h}$.

Step 2. Compute the generalised likelihood ratio test statistic T_{obs}.

Step 3. Generate a random sample of size n from model (2·3) with $m = m_{\hat{\theta}}$.

Step 4. Using the generated random sample with the bandwidth $\hat{h}$, obtain the generalised likelihood ratio statistic T^*.

Step 5. Repeat B times Steps 3 and 4 and obtain the bootstrap generalised likelihood ratio test statistics $T_1^*, \ldots, T_B^*$.

Step 6. The p-value of the test statistic is the percentage of the bootstrap generalised likelihood ratio test statistics $\{T_1^*, \ldots, T_B^*\}$ that exceed T_{obs}.

The null distribution of the generalised likelihood ratio test statistic is approximated by the distribution of the bootstrap generalised likelihood ratio statistics $\{T_1^*, \ldots, T_B^*\}$. When a parametric model such as an ARMA model is specified under the null hypothesis, one can sample directly from the ARMA model and then compute the log-periodogram. While this will give a more accurate estimate of the null distribution, it increases the computational demand. The bandwidth $\hat{h}$ in Step 4 can also be data-driven.

3. BIAS REDUCTION AND RESIDUAL-BASED TESTS

3·1. *Bias reduction*

The function $m_\theta(\,.\,)$ is typically nonlinear, so that the local linear estimator of the function $\hat{m}$ contains biases even under the null hypothesis, and this affects the null distribution of the generalised likelihood ratio statistic. In §4, we let the bandwidth go to zero sufficiently fast, but in practice we do not know the size of bandwidth which would make the bias negligible. This motivates us to introduce a bias correction technique.

We can correct the bias of the test under the null model as follows. Let $\hat{\theta}$ be an estimator of the parameters under the null hypothesis and let $m^*(w) = m(w) - m_{\hat{\theta}}(w)$. Then the testing problem (2·2) is equivalent to testing between

$$H_0^*: m^*(w) \equiv 0, \quad H_1^*: m^*(w) \neq 0. \tag{3·1}$$

Note that $m^*(\,.\,)$ is simply a reparametrisation of the function m. Hence, the testing problem (3·1) can be dealt with by applying the generalised likelihood ratio test based on the synthetic data $\tilde{Y}_k = Y_k - m_{\hat{\theta}}(w_k)$. With this new parametrisation, the local linear estimator of the function $m^*(\,.\,)$ does not incur serious biases under H_0^*.

The appropriate test statistic is

$$T_{\mathrm{LK,bc}} = \sum_{k=1}^{n} [\exp(\tilde{Y}_k) - \exp\{\tilde{Y}_k - \hat{m}_{\mathrm{LK}}^*(w_k)\} - \hat{m}_{\mathrm{LK}}^*(w_k)],$$

where $\hat{\theta}$ is obtained via the maximum likelihood estimator and $\hat{m}_{\mathrm{LK}}^*$ is the local likelihood estimator of m^* based on the data $\tilde{Y}_k$. Recall that we simply reparameterise the testing problem. The bandwidth selection and bootstrap estimation of the null distribution outlined in §2 still apply.

The above idea is similar to the prewhitening technique of Press & Tukey (1956) and the technique used by Härdle & Mammen (1993). It is also related to the nonparametric estimate using parametric start of Hjort & Glad (1995) and Glad (1998).

3·2. *Testing whiteness of residuals*

The scope of the tests can be significantly enhanced when we reduce a testing problem to the problem of testing whether or not the residuals under the null hypothesis are white noise. Suppose that we wish to verify whether or not a family of models fits a given time series. Let $\{\hat{\varepsilon}_t\}$ be the residuals resulting from a fit. If the null hypothesis is reasonable, then the residuals should behave like a white noise series with variance σ^2. In the spectral domain, this is equivalent to testing between

$$H_0: m_\theta(w) = \log\{\sigma^2/(2\pi)\}, \quad H_1: m_\theta(w) \neq \log\{\sigma^2/(2\pi)\}. \tag{3·2}$$

The generalised likelihood ratio test approach can be applied to (3·2). Since the local linear fit does not have any biases for the spectral density estimate under the null hypothesis, no bias correction is needed. The resulting procedure will be referred to as the residual-based test and we denote the test statistic by T_{R}.

4. Asymptotic properties of the test statistics

Let θ_0 be the true parameter under the null hypothesis. Then we can write

$$T_{\mathrm{LK}} = T_{\mathrm{LK},1} - T_{\mathrm{LK},2}, \tag{4·1}$$

where

$$T_{\mathrm{LK},1} = \sum_{k=1}^{n} [\exp\{Y_k - m_{\theta_0}(w_k)\} + m_{\theta_0}(w_k) - \exp\{Y_k - \hat{m}_{\mathrm{LK}}(w_k)\} - \hat{m}_{\mathrm{LK}}(w_k)],$$

$$T_{\mathrm{LK},2} = \sum_{k=1}^{n} [\exp\{Y_k - m_{\theta_0}(w_k)\} + m_{\theta_0}(w_k) - \exp\{Y_k - m_{\hat{\theta}}(w_k)\} - m_{\hat{\theta}}(w_k)].$$

The test statistic $T_{\mathrm{LK},1}$ is the generalised likelihood ratio test statistic for testing between

$$H_0 : m(.) = m_{\theta_0}(.), \quad H_1 : m(.) \neq m_{\theta_0}(.), \tag{4·2}$$

while $T_{\mathrm{LK},2}$ is the maximum likelihood ratio test statistic for testing between

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0 \tag{4·3}$$

in the parametric family of models $\{m_\theta(.)\}$. Under certain regularity conditions, the asymptotic null distribution of $T_{\mathrm{LK},2}$ is χ^2_p, where p is the dimension of vector θ. Hence,

$$T_{\mathrm{LK},2} = O_P(1). \tag{4·4}$$

It will be shown that $T_{\mathrm{LK},1}$ follows asymptotically a scaled χ^2 distribution with degrees of freedom tending to infinity. Thus, $T_{\mathrm{LK},2}$ is asymptotically negligible and the asymptotic distribution is completely determined by $T_{\mathrm{LK},1}$. Therefore, we need only to consider the case where the null hypothesis is simple, see (4·2), in which case T_{LK} can be simplified to

$$T_{\mathrm{LK}} = \sum_{k=1}^{n} [\exp\{Y_k - m_{\theta_0}(w_k)\} - \exp\{Y_k - \hat{m}_{\mathrm{LK}}(w_k)\} + m_{\theta_0}(w_k) - \hat{m}_{\mathrm{LK}}(w_k)].$$

In the sequel, we will drop the subscript θ_0.

We now introduce some notation to facilitate the presentation. Let

$$v_0 = \int\int t^2(s+t)^2 K(t)K(t+s)dt\, ds,$$

$$\mu_n = h^{-1}\pi\left\{K(0) - 2^{-1}\int K^2(t)dt\right\}, \quad \sigma_n^2 = 2h^{-1}\pi\int\{K(t) - 2^{-1}K*K(t)\}^2 dt,$$

$$D_K = \left\{K(0) - 2^{-1}\int K^2(t)dt\right\} \Big/ \int\{K(t) - 2^{-1}K*K(t)\}^2 dt,$$

$$d_{1ng} = 8^{-1}h^4 v_0 \sum_{k=1}^{n} m''_\theta(w_k)^2/g_\theta(w_k).$$

Theorems 1 and 2 below require the following technical conditions.

Condition 1. The kernel function $K(t)$ is a symmetric probability density function, bounded with a bounded support, and $K'(t)$ is bounded.

Condition 2. The series $\{X_t\}$ is a linear Gaussian process; that is $X_t = \sum_{j=-\infty}^{\infty} a_j \zeta_{t-j}$, ζ_j are independently and identically distributed as $N(0, \sigma^2)$, and $\sum_{j=-\infty}^{\infty} |a_j| j^2 < \infty$.

Condition 3. The spectral density $g(.)$ is positive on $[0, \pi]$.

It follows from Conditions 2 and 3 that the second derivative $m''(.)$ of $m(.)$ is continuous on $[0, \pi]$; see for example Kooperberg et al. (1995b).

In the sequel, we use the notation $Y_n \sim_a \chi^2_{a_n}$ to denote a sequence of random variables Y_n for which

$$(Y_n - a_n)/(2a_n)^{\frac{1}{2}} \to N(0, 1),$$

in distribution, as $n \to \infty$.

THEOREM 1. *If Conditions* 1–3 *are satisfied, then, under* H_0, *as* $h \to 0$, $nh^2 \to \infty$ *and* $n^{(\xi-1)/\xi} h \geqslant c_0 \log^\delta n$ *for some* $\delta > (\xi - 1)/(\xi - 2)$ *and* $\xi > 2$,

$$\sigma_n^{-1}(T_{\mathrm{LK}} - \mu_n + d_{1ng}) \to N(0, 1),$$

in distribution. Furthermore, if $nh^{9/2} \to 0$, *then*

$$D_K T_{\mathrm{LK}} \sim_a \chi^2_{D_K \mu_n}.$$

The result shows that the asymptotic null distribution is independent of nuisance parameters when the bandwidth converges to zero fast enough. Fan et al. (2001) called such a phenomenon the Wilks phenomenon. This result permits one to use parametric bootstrap to estimate the asymptotic null distribution. Recently, Zhang (2003) provided a useful method for calibrating the degrees of freedom, which makes the approximation more accurate.

After bias correction with known θ_0, there will be no bias in the test under the null hypothesis. Thus, the bias is governed by the quality of the estimator $\hat{\theta}$. It is to be expected that the d_{1ng} in Theorem 1 will disappear, and the following theorem confirms this.

THEOREM 2. *If Conditions* 1–3 *are satisfied*, $\partial m_\theta(x)/\partial\theta$ *is continuous and bounded, and* $\hat{\theta} - \theta = O_P(n^{-\frac{1}{2}} \log n)$, *then, under* H_0, *as* $h \to 0$, $nh^2 \to \infty$ *and* $n^{(\xi-1)/\xi} h \geqslant c_0 \log^\delta n$ *for some* $\delta > (\xi - 1)/(\xi - 2)$ *and* $\xi > 2$,

$$D_K T_{\mathrm{LK,bc}} \sim_a \chi^2_{D_K \mu_n}.$$

The condition imposed on $\hat{\theta}$ in Theorem 2 is very mild as usually the parametric estimator can reach the convergence rate of $n^{-\frac{1}{2}}$. Thus, the generalised likelihood ratio statistics with bias correction perform better than those without any bias correction. The proofs of Theorems 1 and 2 are given in the Appendix.

5. NUMERICAL RESULTS

5·1. *Simulations*

Here and in §5·2, the Epanechnikov kernel is used for constructing nonparametric estimators. The bandwidth is chosen by the method of Fan & Kreutzberger (1998) for estimating the spectral density.

We use a simulated example to verify how well the sizes of the generalised test statistics are approximated by the bootstrap method and to compare the power of various versions of the generalised likelihood ratio statistics. Under the null hypothesis, this example satisfies the technical conditions listed in §4.

As expected, the generalised likelihood ratio test with bias correction performs better than the uncorrected counterpart. We also include the residual-based statistic T_{R} in §3·2 for comparisons.

Example 1. We consider the AR(3) model with $\theta_1 = 0{\cdot}8$, $\theta_2 = -0{\cdot}56$ and $\theta_3 = 0{\cdot}6$, confounded with a nonlinear function. We simulate a series of length T from the model

$$X_t = \{\theta_1(1-\beta) + \beta v(X_{t-3})\}X_{t-1} + \theta_2 X_{t-2} + \theta_3 X_{t-3} + \varepsilon_t, \tag{5·1}$$

where $v(x) = 0{\cdot}95I\ (-5 \leqslant x < 0) - 0{\cdot}18xI\ (0 \leqslant x \leqslant 5)$, the ε_t are independently and identically distributed $N(0, 1)$ random variables and β is a given parameter.

For each fixed β, we simulate a time series of length $T = 500$, and then test if the time series shares the same spectral density as that of an AR(3) model. In other words, we test the hypothesis (2·2) with

$$g_\theta(x) = (2\pi)^{-1}\sigma^2 \left| 1 - \sum_{j=1}^{3} \theta_j \exp(-ijx) \right|^{-2}$$

and evaluate the power of the test at the model (5·1). When $\beta = 0$, the power becomes the size of the test.

All of the preceding three generalised likelihood ratio tests are used and compared. The related bandwidth is taken to be 0·23, a data-driven choice based on a few simulations. The distributions of T_{LK}, $T_{\mathrm{LK,bc}}$ and T_{R} under the null hypothesis are approximated by 1000 Monte Carlo simulations.

To verify whether or not the Wilks phenomenon holds for reasonably large sample sizes, we deliberately set the parameters in the AR(3) model to zero and simulated the null distributions of the test statistics.

If the Wilks phenomenon holds, the null distributions of the test statistics should not depend sensitively on the nuisance parameters under the null hypothesis, and hence the above approximation should be reasonably accurate. Table 1 confirms this, especially for the bias-reduction-based approaches.

Table 1: *Example* 1. *The sizes of the tests under* H_0

α	$T_{\mathrm{LK,bc}}$	T_{LK}	T_{R}	LB(M)	CM
0·01	0·012	0·088	0·011	0·012	0·014
0·05	0·054	0·038	0·058	0·051	0·053
0·10	0·114	0·080	0·101	0·098	0·102

The proposed tests were performed at three different significance levels, namely 0·01, 0·05 and 0·1. The percentages of rejections by each test statistic are computed, based on 1000 simulations. Figure 1 (a) depicts the empirical power functions for various choices of β. When $\beta = 0$, the null and alternative hypotheses coincide and the power function becomes the probability of type I errors; see also Table 1.

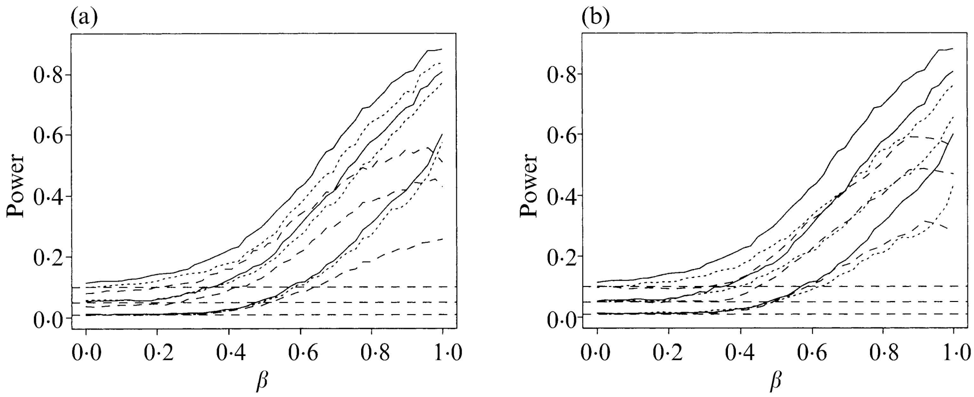

Fig. 1: Example 1. (a) Powers of the generalised likelihood ratio statistics at significance levels 0·01, 0·05 and 0·1, based on 1000 simulations for different choices of β; the solid lines are for the likelihood-based approach with bias correction, the dotted lines are for the residual-based approach, and the dashed lines are for the likelihood-based approach without bias correction. (b) Comparison between the likelihood-based test with bias reduction, the Cramér–von Mises test and the Ljung–Box test; the solid line is the power of the likelihood-based test with bias reduction at significance levels 0·01, 0·05 and 0·1, based on 1000 simulations, the dotted line is the power of the Ljung–Box test, and the dashed line is that of the Cramér–von Mises test.

Figure 1 (a) clearly indicates that the likelihood-based test with bias reduction is the most powerful one. The residual-based approach performs quite well too, the bias being automatically corrected.

Comparisons are also made with two other popular tests. The first is the Ljung & Box (1978) statistic, given by

$$\textsc{lb}(M) = T(T+2)\sum_{i=1}^{M} r_i^2/(T-i),$$

where r_i is the ith sample autocorrelation of the residuals, and $M=20$. The distribution of $\textsc{lb}(M)$ is obtained by standard bootstrap. The sizes of the Ljung–Box test at three different significance levels are given in Table 1, and the power at different β is plotted in Fig. 1 (b).

The second test is the Cramér–von Mises test which is based on

$$\textsc{cm} = T\{4\pi^4 G^2(\pi)\}^{-1}\sum_{i=1}^{\infty}\left\{\sum_{j=1}^{T-1}(r_j-\rho_j)(\rho_{i+j}-\rho_{i-j})/j\right\}^2,$$

where, with abuse of notation, r_j is the jth sample autocorrelation of X_t, ρ_j is the jth autocorrelation under the null hypothesis, and $G(\pi)=2\int_0^\pi g_\theta^2(x)dx$. Since $\textsc{cm}$ as written above involves unknown parameters, we use a version in which unknown parameters are replaced by their maximum likelihood estimators. Distribution of this $\textsc{cm}$ is approximated by bootstrap. Sizes and powers are presented in Table 1 and Fig. 1 (b). Figure 1 (b) shows that the likelihood-based test with bias reduction is the most powerful, and the Cramér–von Mises test performs worst.

We also did some simulations for the ARMA(1, 1) model; the conclusions were the same as that for Example 1.

5·2. *Real data examples*

Example 2. The annual record of the numbers of the Canadian lynx trapped in the Mackenzie River district of northwest Canada is a popular dataset in time series. The periodic fluctuation has a big impact on ecological theory. Moran (1953) used an AR(2) model

$$X_t = a_0 + a_1 X_{t-1} + a_2 X_{t-2} + \varepsilon_t$$

to fit the logarithms of the lynx data. However, he noted that the sum of squares of residuals corresponding to values of X_t greater than the mean is 1·781 while the sum of squares of residuals corresponding to values of X_t smaller than the mean is 4·007. Tong (1990, Ch. 7) suggested a TAR(2) model,

$$X_t = \begin{cases} a_0 + a_1 X_{t-1} + a_2 X_{t-2} + \varepsilon_t & (X_{t-2} \leqslant C), \\ b_0 + b_1 X_{t-1} + b_2 X_{t-2} + \varepsilon_t & (X_{t-2} > C). \end{cases}$$

To investigate whether the AR(2) model or the TAR(2) model is more suitable for this dataset, the residual-based test is employed. The null distribution of T_{R} is approximated by the bootstrap using 10 000 Monte Carlo simulations. The test gives p-value 0·0009 for the AR(2) model. This provides stark evidence against the AR(2) model. On the other hand, the test gives p-value 0·423 for the TAR(2) model. Thus the TAR(2) model fits the dataset, which is in line with the biological interpretation in Stenseth et al. (1998).

Example 3. Figure 2(a) displays the weekly egg prices at a German agricultural market in April 1967–May 1990. Since the data exhibit clear nonstationary features, we take first-order differences of the series. The differenced series are plotted in Fig. 2(b), which looks more stationary. Fan & Yao (2003, §3·6), concluded that ARMA(1, 2) and MA(7) models are suitable for this dataset.

We use the likelihood-based generalised likelihood ratio statistic with bias reduction to test the suitability of these models. The null distributions of the test statistic are approximated

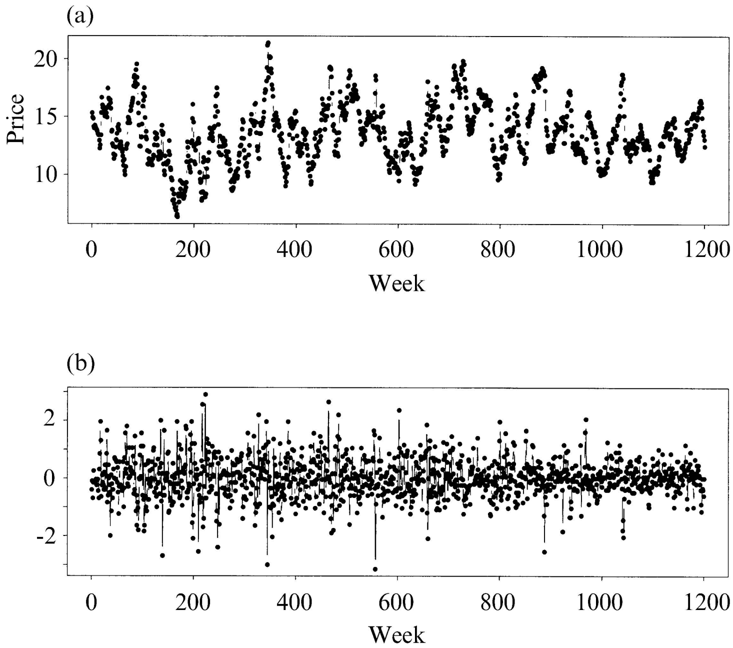

Fig. 2: Example 3. (a) Weekly egg prices (per egg in Deutsch Marks/100) at a German agricultural market from April 1967 to May 1990. (b) First differences of the egg prices.

by the bootstrap described in §2·5 using 10 000 Monte Carlo simulations. The test gives p-values of 0·370 for the ARMA(1, 2) model and 0·531 for the MA(7) model. These p-values lend further support to the conclusions made by Fan & Yao (2003).

ACKNOWLEDGEMENT

Fan's research was partially supported by grants from the United States National Science Foundation and from the Research Grants Council of the Hong Kong Special Administrative Region of the People's Republic of China.

APPENDIX

Proofs

LEMMA A1. *Under Condition* 2 *of* §4,

$$\max_{1\leqslant k\leqslant n} |R_{n,k}| = O_P(n^{-\frac{1}{2}} \log n).$$

See Kooperberg et al. (1995b) for the proof.

LEMMA A2. *Suppose that* $\varepsilon_1, \ldots, \varepsilon_n$ *are independently and identically distributed with* $E(\varepsilon_1)=0$ *and* $E(|\varepsilon_1|^s)<\infty$. *Suppose also that* $0<X_1<\ldots<X_n<1$ *are fixed designed points, that* $X_i - X_{i-1} = O(n^{-1})$ *and that* $K(t)$ *satisfies the Lipschitz condition and has a bounded support. Then, for any sequence* $\alpha_n \to \infty$,

$$\sup_{0\leqslant x\leqslant 1} n^{-1} \left| \sum_{i=1}^{n} K_h(X_i - x)\varepsilon_i \right| = o\{(nh)^{-\frac{1}{2}}(-\log h)^{\frac{1}{2}} \alpha_n\},$$

almost surely, provided that there exists $0<\eta<s$ *such that* $h^{-1} n^{(4-s+\eta)/(s-\eta)} \log n \to c$, *where* c *is a constant.*

Proof. This follows immediately from Theorem 11·2 in Müller (1988, p. 162). □

LEMMA A3. *Under Conditions* 1–3 *of* §4, *if* $nh^2 \to \infty$, *we have*

$$\sup_{0\leqslant x\leqslant \pi} |\hat{m}_{\mathrm{LK}}(x) - \hat{m}^{**}_{\mathrm{LK}}(x)| = O_P(n^{-\frac{1}{2}} \log n),$$

where $\hat{m}^{**}_{\mathrm{LK}}(w_k)$ *is the same as* $\hat{m}_{\mathrm{LK}}(w_k)$ *but replacing* Y_k *by* $Y_k^{**} = Y_k - r_k$.

Proof. Let

$$\beta = (nh)^{\frac{1}{2}}(a - m(x), h\{b - m'(x)\})^{\mathrm{T}}, \quad W_k = (1, (w_k - x)/h)^{\mathrm{T}},$$

and $\bar{m}_k = m(x) + m'(x)(w_k - x)$. Then the local likelihood (2·4) can be written as

$$L(\beta) = \sum_{k=1}^{n} [Y_k - \bar{m}_k - (nh)^{-\frac{1}{2}}\beta^{\mathrm{T}} W_k - \exp\{Y_k - \bar{m}_k - (nh)^{-\frac{1}{2}}\beta^{\mathrm{T}} W_k\}] K_h(w_k - x).$$

Using some simple algebra, we have

$$\begin{aligned} L(\beta) - L(0) &= \sum_{k=1}^{n} [-(nh)^{-\frac{1}{2}}\beta^{\mathrm{T}} W_k - \exp\{Y_k - \bar{m}_k - (nh)^{-\frac{1}{2}}\beta^{\mathrm{T}} W_k\} \\ &\qquad + \exp(Y_k - \bar{m}_k)] K_h(w_k - x) \\ &= \ell(\beta) - U_n(\beta), \end{aligned}$$

where $\ell(\beta)$ is the same as $L(\beta) - L(0)$ but with Y_k substituted by Y_k^{**}, and

$$U_n(\beta) = \sum_{k=1}^{n} R_{n,k}[\exp\{-\bar{m}_k - (nh)^{-\frac{1}{2}}\beta^{\mathrm{T}} W_k\} - \exp(-\bar{m}_k)] K_h(w_k - x).$$

For any fixed β, using Taylor's expansion and Lemma A1, we have

$$h \sup_{0\leqslant x\leqslant \pi} U_n(\beta) = O_P(h^{\frac{1}{2}} \log n). \tag{A·1}$$

Similarly, we can show that the following hold uniformly in x:

$$hU_n'(\beta) = O_P(h^{\frac{1}{2}} \log n), \quad hU_n''(\beta) = O_P(hU_n'(\beta)), \quad h\ell^{(3)}(\beta) = O_P(n^{-\frac{1}{2}} h^{-\frac{1}{2}}), \tag{A·2}$$

$$h\ell^{(4)}(\beta) = O_P(n^{-1}h^{-1}), \quad h\ell''(\tilde{0}) - A_1 = O_P(h^2 + n^{-\frac{1}{2}} h^{-\frac{1}{2}} \log h\alpha_n), \tag{A·3}$$

where α_n is any sequence such that $\alpha_n \to \infty$,

$$\tilde{0} = (0, 0)^{\mathrm{T}}, \quad A_1 = -\pi^{-1} \operatorname{diag}(1, \mu_2), \quad \mu_2 = \int \mu^2 K(u) du.$$

Let

$$\psi_n(x) = h(nh)^{-\frac{1}{2}} \sum_{k=1}^{n} \{-1 + \exp(Y_k^{**} - \bar{m}_k)\} W_k K_h(w_k - x).$$

By Taylor's expansion and (A·1), (A·2) and (A·3), the following hold uniformly in x:

$$h\ell(\beta) = \psi_n(x)^{\mathrm{T}} \beta + 2^{-1} \beta^{\mathrm{T}} A_1 \beta + \Delta_1(\beta), \quad h\{L(\beta) - L(0)\} = \psi_n(x)^{\mathrm{T}} \beta + 2^{-1} \beta^{\mathrm{T}} A_1 \beta + \Delta_2(\beta),$$

$$\Delta_1(\beta) = O_P(1), \quad \Delta_2(\beta) = O_P(1),$$

$$\Delta_1'(\beta) = O_P\{(nh)^{-\frac{1}{2}} \log h\alpha_n + h^2\}, \quad \Delta_2'(\beta) = \Delta_1'(\beta) + O_P(h^{\frac{1}{2}} \log n).$$

Using the same argument as that for the proof of Theorem 2 in Carroll et al. (1997) and the proof of the quadratic approximation lemma in Fan et al. (1995), we obtain that

$$(nh)^{\frac{1}{2}} \{\hat{m}_{\mathrm{LK}}^{**}(x) - m(x)\} = (\pi, 0)\psi_n(x) + O_P(h^{-\frac{1}{2}} n^{-\frac{1}{2}} \log h\alpha_n + h^2), \tag{A·4}$$

$$(nh)^{\frac{1}{2}} \{\hat{m}_{\mathrm{LK}}(x) - m(x)\} = (\pi, 0)\psi_n(x) + O_P(h^{\frac{1}{2}} \log n),$$

hold uniformly in x, which leads to Lemma A3. □

Proof of Theorem 1. Observe that

$$\begin{aligned} T_{\mathrm{LK}} &= \sum_{k=1}^{n} [m_\theta(w_k) - \hat{m}_{\mathrm{LK}}(w_k) + \exp\{Y_k - m_\theta(w_k)\} - \exp\{Y_k - \hat{m}_{\mathrm{LK}}(w_k)\}] \\ &= T_{\mathrm{LK}}^* + B_1 + B_2\{1 + O_P(1)\} - B_3, \end{aligned} \tag{A·5}$$

where T_{LK}^* is the same as T_{LK} but with Y_k and $\hat{m}_{\mathrm{LK}}(w_k)$ replaced respectively by Y_k^{**} and $\hat{m}_{\mathrm{LK}}^{**}(w_k)$,

$$B_1 = \sum_{k=1}^{n} [\exp\{Y_k - \hat{m}_{\mathrm{LK}}^{**}(w_k)\} - 1]\{\hat{m}_{\mathrm{LK}}^{**}(w_k) - \hat{m}_{\mathrm{LK}}(w_k)\},$$

$$B_2 = \sum_{k=1}^{n} \exp\{Y_k - \hat{m}_{\mathrm{LK}}^{**}(w_k)\}\{\hat{m}_{\mathrm{LK}}^{**}(w_k) - \hat{m}_{\mathrm{LK}}(w_k)\}^2,$$

$$B_3 = \sum_{k=1}^{n} R_{n,k}[\exp\{m_\theta(w_k) - \hat{m}_{\mathrm{LK}}^{**}(w_k)\} - 1]/g_\theta(w_k).$$

By Lemmas A1–A3, (A·4) and since $\exp(u) = 1 + u + O(u^2)$, we have $B_2 = O_P(\log^2 n)$, and

$$\begin{aligned} B_3 &\leqslant \max_{1\leqslant k\leqslant n} \{R_{n,k}/g_\theta(w_k)\} \sum_{k=1}^{n} |\exp\{m_\theta(w_k) - \hat{m}_{\mathrm{LK}}^{**}(w_k)\} - 1| \\ &\leqslant n^{\frac{1}{2}} h^{-\frac{1}{2}} \left\{ \sup_{0\leqslant x\leqslant \pi} |(-\pi, 0)\psi_n(x)| + O_P(1) \right\} \max_{1\leqslant k\leqslant n} R_{n,k}/g_\theta(w_k) \\ &= O_P[h^{-\frac{1}{2}} \log n\{\log^{\frac{1}{2}} h \log\log h + O(n^{\frac{1}{2}} h^{5/2})\}]. \end{aligned}$$

Similarly, we have

$$B_1 = B_{1,1} + \sum_{k=1}^{n} R_{n,k} \exp\{-\hat{m}^{**}_{\mathrm{LK}}(w_k)\}\{\hat{m}^{**}_{\mathrm{LK}}(w_k) - \hat{m}_{\mathrm{LK}}(w_k)\} = B_{1,1} + O_P(\log^2 n),$$

where

$$B_{1,1} = \sum_{k=1}^{n} [V_k \exp\{m_\theta(w_k) - \hat{m}^{**}_{\mathrm{LK}}(w_k)\} - 1]\{\hat{m}^{**}_{\mathrm{LK}}(w_k) - \hat{m}_{\mathrm{LK}}(w_k)\}.$$

By using Lemmas A1–A3 and (A·4) again, we have

$$B_{1,1} = \sum_{k=1}^{n} (V_k - 1)\{\hat{m}^{**}_{\mathrm{LK}}(w_k) - \hat{m}_{\mathrm{LK}}(w_k)\} + o_P(h^{-\frac{1}{2}} \log n \log h\alpha_n\} = o_P(h^{-\frac{1}{2}} \log n \log h\alpha_n).$$

As a special case of Theorem 10 in Fan et al. (2001), we have that

$$\sigma_n^{-1}(T^{*}_{\mathrm{LK}} - \mu_n + d_{1ng}) \to N(0, 1), \tag{A·6}$$

in distribution. Theorem 1 follows from (A·6) and the bounds for B_1, B_2 and B_3. □

Proof of Theorem 2. Obviously,

$$\tilde{Y}_k = z_k + r_k + m_\theta(w_k) - m_{\hat{\theta}}(w_k) = z_k + r^{*}_k$$

and $r^{*}_k = O_P(n^{-\frac{1}{2}} \log n)$ holds uniformly in k. Using exactly the same argument as in the proof of Theorem 1, and noting that for this case the d_{1ng} in (A·6) is zero, we prove Theorem 2. □

REFERENCES

ANDERSON, T. W. (1993). Goodness of fit tests for spectral distributions. *Ann. Statist.* **21**, 830–47.
BRILLINGER, D. R. (1981). *Time Series Analysis: Data Analysis and Theory*, 2nd ed. New York: Holt, Rinehart & Winston.
BROCKWELL, P. J. & DAVIS, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. New York: Springer-Verlag.
CARROLL, R. J., FAN, J., GIJBELS, I. & WAND, M. P. (1997). Generalised partially linear single-index models. *J. Am. Statist. Assoc.* **92**, 477–89.
CHEN, R. & TSAY, R. S. (1993). Functional-coefficient autoregressive models. *J. Am. Statist. Assoc.* **88**, 298–308.
DAVIS, H. T. & JONES, R. H. (1968). Estimation of the innovation variance of a stationary time series. *J. Am. Statist. Assoc.* **63**, 141–9.
DZHAPARIDZE, K. (1986). *Parameter Estimation and Hypothesis Testing on Spectral Analysis of Stationary Time Series.* New York: Springer-Verlag.
FAN, J. (1992). Design-adaptive nonparametric regression. *J. Am. Statist. Assoc.* **87**, 998–1004.
FAN, J. & GIJBELS, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J. R. Statist. Soc.* B **57**, 371–94.
FAN, J. & KREUTZBERGER, E. (1998). Automatic local smoothing for spectral density estimation. *Scand. J. Statist.* **25**, 359–69.
FAN, J. & YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods.* New York: Springer-Verlag.
FAN, J., HECKMAN, N. E. & WAND, M. P. (1995). Local polynomial kernel regression for generalised linear models and quasi-likelihood adaptation. *J. Am. Statist. Assoc.* **90**, 141–50.
FAN, J., ZHANG, C. & ZHANG, J. (2001). Generalised likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29**, 153–93.
GLAD, I. K. (1998). Parametrically guided non-parametric regression. *Scand. J. Statist.* **25**, 649–68.
HÄRDLE, W. & MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 882–904.
HASTIE, T. J. & TIBSHIRANI, R. (1990). *Generalised Additive Models.* London: Chapman and Hall.
HJORT, N. L. & GLAD, I. K. (1995). Nonparametric density estimation with a parametric start. *Ann. Statist.* **23**, 882–904.
INGSTER, Y. I. (1993a). Asymptotically minimax hypothesis testing for nonparametric alternatives I-III, *Math. Meth. Statist.* **2**, 85–114.
INGSTER, Y. I. (1993b). Asymptotically minimax hypothesis testing for nonparametric alternatives I-III. *Math. Meth. Statist.* **3**, 171–89.

INGSTER, Y. I. (1993c). Asymptotically minimax hypothesis testing for nonparametric alternatives I-III. *Math. Meth. Statist.* **4**, 249–68.

KOOPERBERG, C., STONE, C. J. & TRUONG, Y. K. (1995a). Logspline estimation of a possibly mixed spectral distribution. *J. Time Ser. Anal.* **16**, 359–88.

KOOPERBERG, C., STONE, C. J. & TRUONG, Y. K. (1995b). Rate of convergence for logspline spectral density estimation. *J. Time Ser. Anal.* **16**, 389–401.

LJUNG, G. M. & BOX, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika* **65**, 297–303.

MORAN, P. A. P. (1953). The statistical analysis of the Canadian lynx cycle, I: structure and prediction. *Aust. J. Zool.* **1**, 163–73.

MÜLLER, H.-G. (1988). *Nonparametric Regression Analysis of Longitudinal Data*, Lecture Notes in Statistics, **46**. Berlin: Springer-Verlag.

PAPARODITIS, E. (2000). Spectral density based goodness-of-fit tests in time series models. *Scand. J. Statist.* **27**, 143–76.

PAWITAN, Y. & O'SULLIVAN, F. (1994). Nonparametric spectral density estimation using penalised Whittle likelihood. *J. Am. Statist. Assoc.* **89**, 600–10.

PRESS, H. & TUKEY, J. W. (1956). *Power Spectral Methods of Analysis and their Application to Problems in Airplane Dynamics*, Bell Telephone System Monograph 2606.

STENSETH, N. C., FALCK, W., CHAN, K. S., BJORNSTAD, O. N., TONG, H., O'DONOGHUE, M., BOONSTRA, R., BOUTIN, S., KREBS, C. J. & YOCCOZ, N. G. (1998). From Patterns to processes: phase- and density-dependencies in Canadian lynx cycle. *Proc. Nat. Acad. Sci. USA* **95**, 15430–5.

TONG, H. (1990). *Non-Linear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.

WAHBA, G. (1980). Automatic smoothing of the log periodogram. *J. Am. Statist. Assoc.* **75**, 122–32.

WHITTLE, P. (1962). Gaussian estimation in stationary time series. *Bull. Int. Statist. Inst.* **39**, 105–29.

ZHANG, C. M. (2003). Calibrating the degrees of freedom for automatic data-smoothing and effective curve-checking. *J. Am. Statist. Assoc.* **98**, 609–28.

[*Received January* 2003. *Revised May* 2003]