Econometric Theory, **20**, 2004, 382–416. Printed in the United States of America.
DOI: 10.1017/S0266466604202067

A GENERALIZED PORTMANTEAU GOODNESS-OF-FIT TEST FOR TIME SERIES MODELS

WILLA W. CHEN
Texas A&M University

ROHIT S. DEO
New York University

We present a goodness-of-fit test for time series models based on the discrete spectral average estimator. Unlike current tests of goodness of fit, the asymptotic distribution of our test statistic allows the null hypothesis to be either a short- or long-range dependence model. Our test is in the frequency domain, is easy to compute, and does not require the calculation of residuals from the fitted model. This is especially advantageous when the fitted model is not a finite-order autoregressive model. The test statistic is a frequency domain analogue of the test by Hong (1996, *Econometrica* 64, 837–864), which is a generalization of the Box and Pierce (1970, *Journal of the American Statistical Association* 65, 1509–1526) test statistic. A simulation study shows that our test has power comparable to that of Hong's test and superior to that of another frequency domain test by Milhoj (1981, *Biometrika* 68, 177–187).

1. INTRODUCTION

Most conventional goodness-of-fit tests for time series models are based on the autocorrelations of residuals from the fitted model. Examples of such tests include the portmanteau statistic of Box and Pierce (1970) and its generalization, based on arbitrary kernel functions, by Hong (1996). The Box–Pierce statistic is obtained as a particular case of the Hong statistic by using the truncated uniform kernel. Simulations by Hong show that his statistic computed using kernels other than the truncated uniform kernel gives better power than the Box–Pierce statistic against autoregressive (AR) processes and fractionally integrated processes.

Box and Pierce (1970) derive the null distribution of their test for autoregressive moving average (ARMA) models, and Hong derives the null distribution only for finite-order autoregressive models. Both these results require assumptions that rule out long memory processes that have hyperbolically decaying

Address correspondence to: Willa W. Chen, Department of Statistics, Texas A&M University, College Station, Texas 77843, USA; e-mail: wchen@stat.tamu.edu.

  0266-4666/04 $12.00

correlation functions and spectral densities unbounded at the origin. Furthermore, both tests requires the computation of residuals from the fitted model, which can be quite tedious when the model does not have a finite-order autoregressive representation. Also, in such cases, the residuals are not uniquely defined.

A test statistic that circumvents the computation of residuals from the fitted model is proposed by Milhoj (1981). To test the hypothesis that the observations x_t, $t = 1,\dots,n$, are from a process with spectral density $f(\lambda)$, he suggests the test statistic $M_n^d = \{\sum_{j=1}^{n-1} V_j\}^{-2} \sum_{j=1}^{n-1} V_j^2$ where $V_j = I(\lambda_j)/f(\lambda_j)$, $I(\lambda) = (2\pi n)^{-1}|\sum_{t=1}^{n} x_t e^{-i\lambda t}|^2$ is the periodogram of the observations, and $\lambda_j = 2\pi j/n$ is the jth Fourier frequency. Though Milhoj's test statistic is easily computed, his theoretical results are restricted to short memory time series models with bounded spectral densities. Assuming Gaussianity, Beran (1992) extends Milhoj's results to long memory time series models that have unbounded spectral densities at the origin. Examples of long memory processes are the autoregressive fractionally integrated moving average (ARFIMA) process (see Hosking, 1981). Beran states that the null distribution of M_n^d in the presence of long memory is the same as that derived by Milhoj (1981) in the case of short memory. Beran obtains his results by claiming that M_n^d is asymptotically equivalent to its integral version $M_n = \{\int_0^{2\pi} V(\lambda)\, d\lambda\}^{-2} \int_0^{2\pi} V^2(\lambda)\, d\lambda$ where $V(\lambda) = I(\lambda)/f(\lambda)$.

However, Deo and Chen (2000) show that even in the case of Gaussian white noise, M_n^d and M_n do not have the same asymptotic distribution and that the variance of the asymptotic distribution of M_n is two-thirds that of the variance of the asymptotic distribution of M_n^d. Thus, the asymptotic distribution of M_n^d in the long memory case is still an open question.

In this paper, we introduce a test statistic that is a frequency domain analogue of Hong's statistic. We derive the asymptotic null distribution for both short memory models and long memory models. Because our test does not require the calculation of residuals, it can be easily applied to long memory processes such as the ARFIMA models that do not possess finite-order AR representations. Our test delivers uniformly better power than the periodogram-based test M_n^d of Milhoj.

In the next section, we define our test statistic and provide the theoretical results on its asymptotic null distribution for short and long memory models. The power properties of our test are studied in Section 3 through simulations. The proofs are relegated to the Appendix.

2. THE TEST STATISTIC

To motivate our test statistic, it is instructive to consider Hong's statistic to test the null hypothesis that the observations, x_t, $t = 1,2,\dots,n$, are from an AR(p) process, $x_t = \alpha_0 + \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + \varepsilon_t$, where ε_t are zero

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press

mean white noise. Let e_t be the residuals from the fitted model, $e_t = x_t - \hat{\alpha}_0 - \hat{\alpha}_1 x_{t-1} - \cdots - \hat{\alpha}_p x_{t-p}$, where $\hat{\alpha}_0, \hat{\alpha}_1, .., \hat{\alpha}_p$ are the estimates of the parameters $\alpha_0, \alpha_1, \ldots, \alpha_p$. The test statistic of Hong (1996) is

$$H_n = \sum_{j=1}^{n-1} k^2(j/p_n)\hat{\rho}_{e,j}^2,$$

where $k(\cdot)$ is a suitable kernel function such that $k(0) = 1$, $\hat{\rho}_{e,j} = \hat{\gamma}_{e,j}/\hat{\gamma}_{e,0}$ are the sample autocorrelations of the residuals, and $\hat{\gamma}_{e,j}$ are their sample autocovariances,

$$\hat{\gamma}_{e,j} = \frac{1}{n}\sum_{t=|j|+1}^{n}(e_t - \bar{e})(e_{t-|j|} - \bar{e}), \qquad j = 0, \pm 1, \ldots, \pm(n-1).$$

By Parseval's identity, H_n can be written as

$$\begin{aligned} H_n &= \frac{1}{2}\left(\sum_{j=-(n-1)}^{n-1} k^2(j/p_n)\hat{\rho}_{ej}^2 - 1\right) \\ &= \frac{1}{2}\left\{\left(\int_0^{2\pi} \hat{f}_e(\lambda)\, d\lambda\right)^{-2}\left(2\pi\int_0^{2\pi} \hat{f}_e^2(\lambda)\, d\lambda\right) - 1\right\}, \end{aligned} \tag{1}$$

where

$$\hat{f}_e(\lambda) = \frac{1}{2\pi}\sum_{|j|<n} k(j/p_n)\hat{\gamma}_{e,j}e^{-i\lambda j}. \tag{2}$$

The kernel function k here is also called the lag window and $\hat{f}_e(\lambda)$ is the lag-weights spectral density estimator. Let $I_{n,e}$ be the mean corrected periodogram of the residuals given by

$$I_{n,e}(\lambda) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n}(e_t - \bar{e})e^{-i\lambda t}\right|^2.$$

Using the relation

$$\hat{\gamma}_{e,j} = \int_0^{2\pi} I_{n,e}(\omega)e^{i\omega j}\, d\omega,$$

we have an equivalent form of $\hat{f}_e(\lambda)$ in the frequency domain,

$$\hat{f}_e(\lambda) = \int_0^{2\pi} W(\lambda - \omega)I_{n,e}(\omega)\, d\omega, \tag{3}$$

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press

where W, the spectral window corresponding to the lag window k, is its Fourier transform

$$W(\lambda) = \frac{1}{2\pi} \sum_{|h|<n} k(h/p_n) e^{-ih\lambda}. \tag{4}$$

Expressions (1) and (3) provide the motivation for our test statistic. To test a general null hypothesis that the observations x_t are from a process with spectral density $f(\cdot)$, we propose the following test statistic:

$$T_n = \left\{ \frac{2\pi}{n} \sum_{\ell=0}^{n-1} \tilde{f}_e(\lambda_\ell) \right\}^{-2} \left\{ \frac{2\pi}{n} \sum_{\ell=0}^{n-1} \tilde{f}_e^2(\lambda_\ell) \right\}, \tag{5}$$

where

$$\tilde{f}_e(\lambda) = \frac{2\pi}{n} \sum_{j=1}^{n-1} \frac{W(\lambda - \lambda_j) I(\lambda_j)}{f(\lambda_j)}$$

and I is the periodogram of the observations $x_1, \ldots, x_n$. Note that $\tilde{f}_e$ is a discrete version of $\hat{f}_e$ in (3) with $I_{n,e}$ replaced by I/f. Thus, we whiten the process in the frequency domain instead of in the time domain. This not only avoids the computation of residuals but also allows one to easily test for arbitrary spectral densities. Furthermore, T_n is obtained by discretizing the integral in (1) with $\hat{f}_e$ replaced by $\tilde{f}_e$. Also note that T_n is mean invariant because $\tilde{f}_e$ is evaluated only at Fourier frequencies. This is especially favorable in the presence of long memory, because the sample mean is not fully efficient in that case (see Beran, 1994, p. 6).

Hong (1996) establishes the asymptotic normality of H_n for AR models. We show that T_n is asymptotically normal under a null hypothesis that can be either short memory or long memory if the process is Gaussian. The properties of a long memory process differ substantially from those of a short memory process, and hence the proof of the asymptotic results for long memory models requires a more delicate approach than that for short memory models. We now state the assumptions we make and our main results.

Throughout the rest of this paper, we assume that $\{x_t\}$ is a stationary linear process of the form

$$x_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \tag{6}$$

where the innovations ε_t satisfy the following assumption.

Assumption 1. The series $\{\varepsilon_t\}$is independently and identically distributed with mean zero, variance σ^2, and $E(\varepsilon_t^8) < \infty$.

We also make the following assumptions about the kernel $k(\cdot)$ and the bandwidth p_n.

Assumption 2a. The kernel function $k: R \to [-1,1]$ is a symmetric function that is continuous at zero and at all but a finite number of points, with $k(0) = 1$. Furthermore, assume that for some $\delta \geq 1$, $z^{\delta}|k(z)| < \infty$ as $z \to \infty$.

Assumption 3. The bandwidth p_n satisfies $\log^6 n/p_n \to 0$ and $p_n^{3/2}/n \to 0$.

As can be seen from the proof of Lemma 2 in the Appendix Assumption 3 on the maximum rate of increase of the bandwidth p_n is made merely to ensure that our test statistic has the same limiting distribution as Hong's test statistic. If we were to relax this assumption, we would get a slightly different mean and variance for the limiting distribution of our test statistic. It is also worth noting that all the kernels used in practice satisfy Assumption 2a. The next theorem states the asymptotic distribution of T_n when $\{x_t\}$ is a short memory process.

THEOREM 1. *Let $x_1, \ldots, x_n$ be n observations from a stationary linear process defined by (6) with coefficients ψ_j such that $\sum_{j=0}^{\infty}|\psi_j| j^{1/2} < \infty$ and innovations ε_t satisfying Assumption 1. Let $f(\cdot)$ be the spectral density of the process such that* $\inf_\lambda f(\lambda) > 0$. *Let T_n be as in (5) and W be defined by (4) with kernel function k satisfying Assumption 2a and bandwidth p_n satisfying Assumption 3. Then*

$$\frac{n(T_n - C_n(k))}{D_n(k)^{1/2}} \to N(0,1)$$

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

$$C_n(k) = \frac{1}{n\pi} \sum_{j=1}^{n-1} (1 - j/n) k^2(j/p_n) + \frac{1}{2\pi}$$

and

$$D_n(k) = \frac{2}{\pi^2} \sum_{j=1}^{n-2} \{(1 - j/n)(1 - (j+1)/n)\} k^4(j/p_n).$$

It can be shown that a process satisfying the assumptions in Theorem 1 has bounded spectral density and autocovariances that are absolutely summable (Brockwell and Davis, 1996, ex. 3.9). Such a process is a short memory process, an example of which is the ARMA model. The assumptions on the process $\{x_t\}$ of Theorem 1 are satisfied by a broad range of short memory models, whereas the asymptotic theory of H_n is established only for AR models.

To establish the asymptotic normality of T_n when the process is a long memory process, we restrict the process $\{x_t\}$ to be Gaussian. We also require additional assumptions on k, which we state next.

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press

Assumption 2b. In addition to Assumption 2a, the kernel function k is differentiable almost everywhere and satisfies $\int |k'(z)k(z)|\,dz < \infty$.

All the kernels used in practice satisfy Assumption 2b. We now state the asymptotic distribution of T_n when $\{x_t\}$ is a long memory process. For the long memory case, we make the extra assumption that the process x_t is Gaussian. We feel that this assumption can be relaxed just as in the short memory case in Theorem 1, though at the expense of much greater complexity in the proof.

THEOREM 2. *Let $x_1,\dots,x_n$ be n observations from a stationary Gaussian linear process defined by (6) that has a spectral density $f(\lambda) = \lambda^{-2d} g^*(\lambda)$, $d \in (0,0.5)$ and $g^*(\cdot)$ is an even differentiable function on $[-\pi,\pi]$. Also let the spectral density satisfy $\inf_\lambda f(\lambda) > 0$. Let T_n be defined as in Theorem 1 with kernel function k satisfying Assumption 2b and bandwidth p_n satisfying Assumption 3. Then*

$$\frac{n(T_n - C_n(k))}{D_n(k)^{1/2}} \to N(0,1)$$

in distribution as $n \to \infty$, where $C_n(k)$ and $D_n(k)$ are as in Theorem 1.

A stationary linear process that has a spectral density satisfying the assumption of Theorem 2 is a long memory process. It can be shown that the autocovariances decay to zero hyperbolically and are not summable for such a process (Zygmund, 1959, Theorem 2.24). Examples of long memory processes satisfying the assumptions of Theorem 2 are ARFIMA models (Granger and Joyeux, 1980; Hosking, 1981) and fractional Gaussian noise (Mandelbrot and Van Ness, 1968).

In applications, the null hypothesis of interest is the composite hypothesis that the process has spectral density $f(\boldsymbol{\theta},\cdot)$ for some unknown $\boldsymbol{\theta}$ in the parameter space Θ. Under this composite null, the test statistic becomes

$$T_n(\hat{\boldsymbol{\theta}}) = \left\{\frac{2\pi}{n}\sum_{\ell=0}^{n-1} \tilde{f}_e(\hat{\boldsymbol{\theta}},\lambda_\ell)\right\}^{-2} \left\{\frac{2\pi}{n}\sum_{\ell=0}^{n-1} \tilde{f}_e^2(\hat{\boldsymbol{\theta}},\lambda_\ell)\right\}, \tag{7}$$

where

$$\tilde{f}_e(\hat{\boldsymbol{\theta}},\lambda) = \frac{2\pi}{n}\sum_{j=1}^{n-1} \frac{W(\lambda-\lambda_j)I(\lambda_j)}{f(\hat{\boldsymbol{\theta}},\lambda_j)}$$

and $\hat{\boldsymbol{\theta}}$ is some estimator of $\boldsymbol{\theta}$ based on the sample $x_1,\dots,x_n$. Under certain additional assumptions, we show in the next two theorems that the asymptotic null distribution of $T_n(\hat{\boldsymbol{\theta}})$ remains the same as that of T_n in Theorem 1 and in Theorem 2. We first state the additional assumptions we need.

Assumption 4. Let Θ_0 be a compact subset of Θ, where Θ is a finite-dimensional parameter space. Let the spectral density of the process $\{x_t\}$ be

$f(\boldsymbol{\theta}_0,\cdot)$, where $\boldsymbol{\theta}_0$ is the true parameter vector that lies in the interior of Θ_0. Assume that the estimator $\hat{\boldsymbol{\theta}} \in \Theta$ satisfies $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_p(n^{-1/2})$.

The following is an assumption on the spectral density for short memory process.

Assumption 5. The spectral density $f(\boldsymbol{\theta},\lambda)$ satisfies the following conditions for $(\boldsymbol{\theta},\lambda) \in \Theta \times [0,2\pi]$:

(i) $f(\boldsymbol{\theta},\lambda)$ and $f^{-1}(\boldsymbol{\theta},\lambda)$ are continuous at all $(\boldsymbol{\theta},\lambda)$.
(ii) $\partial/\partial\theta_j f^{-1}(\boldsymbol{\theta},\lambda)$ and $\partial^2/\partial\theta_j\partial\theta_k f^{-1}(\boldsymbol{\theta},\lambda)$ are continuous and finite at all $(\boldsymbol{\theta},\lambda)$.

It is very easy to establish that Assumptions 4 and 5 are satisfied by all ARMA models. The next theorem states the asymptotic distribution of $T_n(\hat{\boldsymbol{\theta}})$ when $\{x_t\}$ is a short memory process.

THEOREM 3. *Let $x_1,\dots,x_n$ be n observations from a stationary linear process satisfying the same assumptions as those of Theorem 1. Let the estimated parameter vector $\hat{\boldsymbol{\theta}}$ satisfy Assumption 4 and the spectral density of the process $\{x_t\}$ satisfy Assumption 5. Also let $T_n(\hat{\boldsymbol{\theta}})$ be defined by (7) with kernel function k and bandwidth p_n satisfying the same assumptions as those of Theorem 1. Then*

$$\frac{n(T_n(\hat{\boldsymbol{\theta}}) - C_n(k))}{D_n(k)^{1/2}} \to N(0,1)$$

in distribution as $n \to \infty$, where $C_n(k)$ and $D_n(k)$ are defined as in Theorem 1.

To establish the asymptotic distribution of $T_n(\hat{\boldsymbol{\theta}})$ when $\{x_t\}$ is a long memory process, we need the following assumption on $\hat{\boldsymbol{\theta}}$ and the spectral density $f(\boldsymbol{\theta},\cdot)$.

Assumption 6. Let Θ_0 be a compact subset of Θ, where Θ is a finite-dimensional parameter space in $\Re^s$ for some positive integer s. Let the spectral density of the process $\{x_t\}$ be $f(\boldsymbol{\theta}_0,\lambda) = f^*(d_0,\lambda)g^*(\boldsymbol{\beta}_0,\lambda)$, where f^* and g^* are even functions on $[-\pi,\pi]$, $f^*(d,\lambda) \sim a_d\lambda^{-2d}$ as $\lambda \to 0$ for some $a_d > 0$, $g^*(\boldsymbol{\beta},\lambda)$ is differentiable on $[-\pi,\pi]$, and $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0,d_0)'$ is the true parameter vector that lies in the interior of Θ_0. Furthermore, assume that the sth component of Θ_0 is contained in the segment $[\delta_1,0.5-\delta_1]$ for some $0 < \delta_1 < 0.25$ and that there exists an estimator $\hat{\boldsymbol{\theta}} \in \Theta_0$ that satisfies $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_p(n^{-1/2})$.

Assumption 7. Let $\boldsymbol{\theta} = (\boldsymbol{\beta},d)'$, where $(\boldsymbol{\beta},d) \in \Theta_0$. For any $\delta > 0$, the spectral density $f(\boldsymbol{\theta},\lambda)$ satisfies the following conditions.

(i) $f(\boldsymbol{\theta},\lambda)$ is continuous at all $(\boldsymbol{\theta},\lambda)$ except $\lambda = 0$, $f^{-1}(\boldsymbol{\theta},\lambda)$ is continuous at all $(\boldsymbol{\theta},\lambda)$, and

$$\sup_{\lambda}\sup_{\theta\in\Theta_0}|\lambda|^{2d}f(\boldsymbol{\theta},\lambda) = A \quad \text{for some } 0 < A < \infty.$$

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press

(ii) $\partial/\partial\theta_j f^{-1}(\boldsymbol{\theta},\lambda)$ and $\partial^2/\partial\theta_j\partial\theta_k f^{-1}(\boldsymbol{\theta},\lambda)$ are continuous at all $(\boldsymbol{\theta},\lambda)$ and

$$\sup_{\lambda}\sup_{\theta\in\Theta_0}|\lambda|^{-2d+\delta}\left|\frac{\partial}{\partial\theta_j}f^{-1}(\boldsymbol{\theta},\lambda)\right| = A \quad \text{for some } 0 < A < \infty,$$

$$\sup_{\lambda}\sup_{\theta\in\Theta_0}|\lambda|^{-2d+\delta}\left|\frac{\partial^2}{\partial\theta_j\partial\theta_k}f^{-1}(\boldsymbol{\theta},\lambda)\right| = A \quad \text{for some } 0 < A < \infty.$$

(iii) There exists a constant C with

$$|f(\boldsymbol{\theta}_1,\lambda) - f(\boldsymbol{\theta}_2,\lambda)| \le C\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| f(\boldsymbol{\theta}_2,\lambda)$$

uniformly for all λ and all $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1, d_1)'$ and $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_2, d_2)'$ such that $d_1 < d_2$.

All the conditions of Assumptions 6 and 7 are satisfied by fractional Gaussian noise and ARFIMA processes (see Dahlhaus, 1989). We now state the asymptotic distribution of $T_n(\hat{\boldsymbol{\theta}})$ when $\{x_t\}$ is a long memory process.

THEOREM 4. *Let $x_1,\dots,x_n$ be n observations from a stationary Gaussian linear process satisfying the same assumptions as those of Theorem 2. Let the estimated parameter vector $\hat{\boldsymbol{\theta}}$ satisfy Assumption 6 and the spectral density of $\{x_t\}$ satisfy Assumption 7. Also let $T_n(\hat{\boldsymbol{\theta}})$ be defined by (7) with kernel function k and bandwidth p_n satisfying the same assumptions as those of Theorem 2. Then*

$$\frac{n(T_n(\hat{\boldsymbol{\theta}}) - C_n(k))}{D_n(k)^{1/2}} \to N(0,1)$$

in distribution as $n \to \infty$, where $C_n(k)$ and $D_n(k)$ are defined as in Theorem 1.

The theoretical results that we have presented all address the asymptotic behavior of the test statistic when the null hypothesis is correctly specified. An additional question of interest is the power property of the test statistic when the spectral density given by the null hypothesis is actually misspecified. If both the true model and also the misspecified model under the null hypothesis are short memory models, it can be shown quite easily that the statistic T_n is consistent. We do not include the proof for this statement because it is simply tedious but does not have any technical hurdles. However, in the long memory case establishing consistency is a more complicated problem. The complexity of the problem arises because of the fact that when a model is misspecified for a long memory series, the rate of convergence of the parameter estimates of the misspecified model need not be $\sqrt{n}$-consistent and need not even be asymptotically normal. For example, it is known (see Yajima, 1993) that when an AR(1) model is fit to a long memory process with memory parameter $d \in (0.25,0.5)$, the estimate of the AR(1) parameter converges to the population lag 1 autocorrelation at a rate $n^{0.5-d}$ and has an asymptotic distribution that is not Gaussian but is instead the Rosenblatt process. Thus, the "usual" behavior of

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press

estimators of parameters of a misspecified model is not obtained, and a careful analysis has to be carried out on the behavior of goodness-of-fit tests under such misspecifications. We leave this problem of consistency for future research. Another interesting problem for further research is the behavior of the test under local alternatives, where the spectral density under the alternative hypothesis approaches the spectral density under the null hypothesis at some rate a_n. As pointed out earlier, the rate of convergence of the estimators of the null hypothesis model when the alternative is true depends on d. Hence, we would expect that the rate a_n at which the test will have nontrivial local power will depend on d, unlike the result obtained in Theorem 4 of Hong (1996) for the short memory case. However, we are currently unable to conjecture exactly how a_n will depend on d, and we leave that question for future work.

An additional question of interest is the choice of p_n. Because $C_n(k) \sim 1/(2\pi)$ and $D_n(k) \sim Ap_n$ for some constant A, we would expect based on our preceeding results, that under a misspecified model, the rate at which T_n would diverge from $1/(2\pi)$ would be $n/p_n^{1/2}$. Thus, one would expect in general that the slower p_n grows the more powerful the test would be though no optimal choice of p_n can be stated.

In our next section we study the finite-sample performance of our test through Monte Carlo simulations.

3. SIMULATION STUDIES

We generated 5,000 replications of Gaussian series of length $n = 128$ and 512 from a variety of AR and ARFIMA processes. The algorithm of Davies and Harte (1987) was used in the data generation of ARFIMA models. For each series, we computed the three test statistics: (i) Our statistic T_n. (ii) Hong's statistic H_n. (iii) The Milhoj statistic M_n. The statistics were suitably normalized so that they would have an asymptotic standard normal distribution under the null. For T_n and H_n, we used the following three kernels.

(i) Bartlett $k(z) = 1 - |z|, \; |z| \le 1,$
$= 0$ otherwise,

(ii) Tukey $k(z) = \frac{1}{2}(\cos(z\pi) + 1), \; |z| \le 1,$
$= 0$ otherwise,

(iii) Quadratic spectral (QS), $k(z) = \dfrac{25}{12z^2}\left(\dfrac{\sin(6\pi z/5)}{6\pi z/5} - \cos(6\pi z/5)\right), \; z \in (-\infty,\infty).$

For computing T_n and H_n, we used three bandwidths, $p_n = [3n^{0.2}]$, $[3n^{0.3}]$, and $[3n^{0.4}]$. Note that there is no bandwidth involved in computing M_n.

In Tables 1 and 2, we report the sizes of the three tests under the composite null hypothesis of an AR(1) and an ARFIMA$(0,d,0)$, respectively. The true AR(1) parameter was set to 0.8, and the true long memory parameter d in the ARFIMA$(0,d,0)$ was set at 0.4. Because the null hypothesis was a composite one, we had to estimate the parameters of the AR(1) model and the

Table 1. Rejection rates in percentage under an AR(1) model

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	3.08	5.02	4.04	6.12	4.90	7.80	3.82	5.82	4.32	6.98	5.06	8.02
	TUK	3.04	4.96	4.04	6.12	6.30	9.68	3.98	5.82	4.56	7.10	5.16	8.40
	QS	3.64	5.64	4.52	6.90	5.04	7.74	4.06	6.52	4.74	7.64	5.58	9.26
H_n	BAR	3.30	5.08	3.82	5.82	4.26	6.76	3.62	5.72	4.20	6.54	4.76	7.34
	TUK	3.16	4.90	3.78	5.92	4.46	6.96	3.76	5.78	4.26	6.84	4.88	7.48
	QS	3.52	5.52	4.22	6.44	4.82	7.40	4.02	6.20	4.36	7.12	5.08	8.36
M_n		4.34 at 5%			7.12 at 10%			5.14 at 5%			8.88 at 10%		

Note: Model $x_t - 0.8x_{t-1}\varepsilon_t$, $\varepsilon_t \sim N(0,1)$.

TABLE 2. Rejection rates in percentage under an ARFIMA $(0,d,0)$ model

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	2.62	4.08	3.52	5.28	4.90	7.42	2.64	4.60	3.74	6.00	4.80	8.20
	TUK	2.52	4.00	3.46	5.58	4.96	7.50	2.92	4.78	3.86	6.14	5.10	8.42
	QS	3.22	4.98	4.34	6.78	6.62	9.60	3.30	5.74	4.40	7.06	5.58	9.08
H_n	BAR	2.28	3.76	3.02	4.86	3.54	5.88	2.56	4.42	3.42	5.86	4.22	7.00
	TUK	2.20	3.52	3.20	5.10	3.90	5.88	3.12	5.32	4.14	6.52	4.70	7.86
	QS	2.82	4.46	3.66	5.36	4.10	7.04	2.72	4.54	3.70	5.98	4.44	7.44
M_n		4.70 at 5%			7.58 at 10%			4.50 at 5%			8.18 at 10%		

Note: Model $x_t = (1 - B)^{-0.4}\varepsilon_t$, $\varepsilon_t \sim N(0,1)$.

TABLE 3. Rejection rates in percentage under an AR(1) model with innovations from t distribution

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	2.90	4.66	3.28	5.42	3.98	6.64	3.42	5.16	3.88	6.08	4.76	8.12
	TUK	2.92	4.50	3.36	5.40	3.98	6.84	3.52	5.28	4.00	6.14	4.90	8.02
	QS	3.20	5.14	3.44	6.08	5.36	8.52	3.60	5.76	4.32	6.90	5.74	8.96
H_n	BAR	3.10	4.76	3.32	5.20	3.34	5.88	3.22	4.92	3.66	5.60	4.48	7.08
	TUK	3.08	4.90	3.24	5.26	3.52	5.94	3.28	5.20	3.68	5.82	4.56	7.32
	QS	3.20	5.14	3.32	5.54	4.00	6.38	3.38	5.42	3.84	6.34	4.86	8.26
M_n		3.80 at 5%			6.26 at 10%			4.60 at 5%			8.48 at 10%		

Note: Model $x_t - 0.8x_{t-1} = \varepsilon_t$, $\varepsilon_t \sim t_9$.

TABLE 4. Rejection rates in percentage under an ARFIMA $(0,d,0)$ model with innovations from t distribution

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	2.16	3.50	3.02	5.00	4.28	6.54	2.66	4.10	3.44	5.64	4.32	7.24
	TUK	2.08	3.56	3.08	5.00	4.46	6.78	2.86	4.30	3.64	6.04	4.46	7.52
	QS	2.66	4.40	3.92	6.06	5.58	8.84	3.12	5.12	3.90	6.64	5.10	8.66
H_n	BAR	1.86	3.30	2.64	4.74	3.48	5.60	2.70	4.18	3.24	5.32	4.00	6.78
	TUK	1.96	3.28	2.66	4.90	3.66	5.88	2.86	4.42	3.46	5.86	4.10	7.18
	QS	2.30	4.24	3.52	5.44	4.08	6.58	3.20	5.16	3.72	6.36	4.66	7.60
M_n		3.94 at 5%		7.04 at 10%				4.92 at 5%		8.56 at 10%			

Note: Model $x_t = (1 - B)^{-0.4}\varepsilon_t$, $\varepsilon_t \sim t_9$.

TABLE 5. Rejection rates in percentage under AR(2) alternative fitting model AR(1)

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	22.48	28.60	22.92	29.04	22.88	29.62	80.18	84.96	76.02	81.64	68.54	75.74
	TUK	21.94	28.16	22.80	28.58	22.40	28.76	79.96	82.18	74.76	81.06	65.66	73.26
	QS	22.44	28.64	22.80	28.96	22.74	29.74	78.20	83.42	70.56	77.68	61.04	69.96
H_n	BAR	23.58	30.22	23.42	29.66	22.42	28.70	80.62	85.46	75.84	81.86	68.24	75.26
	TUK	23.18	29.36	23.22	29.42	21.98	28.22	80.32	85.24	74.90	80.90	65.02	72.66
	QS	23.28	29.90	22.76	28.54	21.12	27.40	78.34	83.96	70.46	77.14	59.60	68.88
M_n		8.84 at 5%			13.78 at 10%			17.78 at 5%			25.96 at 10%		

Note: Model $x_t - 0.8x_{t-1} + 0.15x_{t-2} = \varepsilon_t$, $\varepsilon_t \sim N(0,1)$.

Table 6. Rejection rates in percentage under ARMA(1,1) alternative fitting model ARIMA(1,d,0)

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	9.50	13.38	8.44	12.58	8.96	13.04	31.84	40.74	28.36	36.56	25.10	34.24
	TUK	7.24	11.28	8.04	12.06	8.80	12.80	31.34	40.54	26.94	35.12	23.36	32.48
	QS	8.74	12.26	8.04	12.02	10.04	14.74	29.20	37.80	25.00	33.62	22.78	31.50
H_n	BAR	12.68	17.04	11.28	15.52	8.92	13.20	33.02	42.28	28.94	37.36	24.82	33.70
	TUK	7.98	12.12	8.20	12.32	8.66	12.96	32.66	41.88	27.70	35.72	23.12	31.88
	QS	11.48	15.70	10.18	14.40	9.06	13.68	30.32	38.70	25.30	33.72	21.78	30.10
M_n		5.42 at 5%			8.76 at 10%			6.44 at 5%			10.38 at 10%		

Note: Model $x_t = 0.8x_{t-1} + \varepsilon_t + 0.2\varepsilon_{t-1}$, $\varepsilon_t \sim N(0,1)$.

TABLE 7. Rejection rates in percentage under ARFIMA$(0,d,0)$ alternative fitting model ARMA$(1,1)$

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	6.46	9.38	7.50	11.10	8.90	13.30	37.28	44.78	38.54	46.16	37.02	45.06
	TUK	6.54	9.10	7.50	11.32	8.68	13.14	38.06	45.20	39.14	46.74	36.20	44.42
	QS	7.20	10.22	8.24	12.56	10.44	15.34	39.70	46.54	37.90	45.72	34.36	43.00
H_n	BAR	5.26	7.54	6.22	8.90	6.84	10.34	36.14	43.38	37.28	44.80	35.00	42.88
	TUK	5.32	7.48	6.46	9.12	6.92	10.54	37.16	44.08	37.90	45.46	34.22	42.36
	QS	6.04	8.64	6.68	10.18	7.32	10.78	38.42	45.56	36.72	44.56	32.26	40.10
M_n		5.34 at 5%			8.92 at 10%			11.56 at 5%			17.96 at 10%		

Note: Model $x_t = (1-B)^{-0.4}\varepsilon_t$, $\varepsilon_t \sim N(0,1)$.

TABLE 8. Rejection rates in percentage under ARFIMA$(1,d,0)$ alternative fitting model ARFIMA$(0,d,0)$

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	8.52	12.48	8.76	12.68	9.68	14.16	16.92	22.42	14.94	21.14	13.32	19.42
	TUK	8.16	12.10	8.10	12.14	9.10	13.60	16.26	21.78	14.50	20.50	12.80	18.32
	QS	8.24	11.74	8.82	12.86	10.88	15.54	15.76	21.34	13.16	19.22	12.62	18.24
H_n	BAR	7.54	10.84	7.54	11.42	7.98	11.56	16.22	21.78	14.22	20.14	12.32	17.74
	TUK	7.36	10.68	7.26	11.06	7.60	11.36	15.28	20.38	12.52	18.14	10.88	16.32
	QS	7.32	10.70	7.32	11.34	8.12	11.53	15.84	20.98	13.88	19.68	11.60	16.82
M_n		6.14 at 5%			9.92 at 10%			6.82 at 5%			11.40 at 10%		

Note: Model $x_t - 0.1x_{t-1} = (1-B)^{-0.4}\,\varepsilon_t$, $\varepsilon_t \sim N(0,1)$.

ARFIMA$(0,d,0)$ model, which was done using the Whittle likelihood in the frequency domain. From Tables 1 and 2, it can be seen that for both models, all three statistics are undersized at both the 5% and 10% levels. The amount by which they are undersized decreases as the bandwidth p_n increases. The M_n-statistic is least undersized, whereas the sizes of T_n are comparable to those of H_n.

Though our theory on the asymptotic distribution of the test statistic T_n has been established only under the assumption of Gaussianity for the case of long memory series, we believe that our result would still hold for non-Gaussian innovations that have a finite eighth moment. Hence, we simulated both a non-Gaussian AR(1) process and a non-Gaussian ARFIMA$(0,d,0)$ process in which the innovations came from a t distribution with 9 degrees of freedom. The AR(1) parameter was set to 0.8, and the long memory parameter d was set to 0.4 as in the earlier simulation for Gaussian data. Tables 3 and 4 present the sizes of the three tests under the composite null hypothesis of an AR(1) and an ARFIMA$(0,d,0)$, respectively, for the case of t distributed innovations. On comparing Tables 3 and 4 with Tables 1 and 2, it is seen that the performance of the tests with respect to size in the case of t distributed innovations is very similar to that of the tests when the data are Gaussian.

To compare the power of the tests, we considered the following four cases: (a) fitting an AR(1) to data generated by an AR(2), $x_t = 0.8x_{t-1} - 0.1x_{t-2} + \varepsilon_t$. (b) fitting an ARFIMA$(1,d,0)$ to data generated by an ARMA$(1,1)$, $x_t = 0.8x_{t-1} + \varepsilon_t + 0.2\varepsilon_{t-1}$. (c) fitting an ARMA$(1,1)$ to data generated by an ARFIMA$(0,d,0)$, $(1-B)^{0.4}x_t = \varepsilon_t$ where B denotes the backshift operator. (d) fitting an ARFIMA$(0,d,0)$ to data generated by an ARFIMA$(1,d,0)$, $(1-B)^{0.4}(1-0.1B)x_t = \varepsilon_t$. The results are reported in Tables 5, 6, 7, and 8, respectively. In all cases, the null hypotheses were composite, and the parameters of the model under the null hypothesis were estimated using the Whittle likelihood.

It is seen that both the tests T_n and H_n have significantly higher power than M_n in all the alternatives considered. This is not surprising, because the tests T_n and H_n give decreasing weights to higher lag sample correlations, whereas M_n gives uniform weight at all lags. It might be tempting to believe that this property of M_n may be useful in detecting long memory alternatives. This belief is however belied by Table 7, where we fit a short memory model to a long memory series and yet M_n is outperformed by a wide margin by both of the other tests. On the other hand, it is seen that the power of T_n is very similar to the power of H_n, with neither test outperforming the other significantly in any situation considered.

REFERENCES

Beran, J. (1992) A goodness-of-fit test for time series with long range dependence. *Journal of the Royal Statistical Society, Series B* **54**, 749–760.

Beran, J. (1994) *Statistic for Long Memory Processes*. Chapman & Hall.
Box, G.E.P. & D.A. Pierce (1970) Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* 65, 1509–1526.
Brockwell, P. & R. Davis (1996) *Time Series: Theory and Methods*, 2nd ed. Springer.
Dahlhaus, R. (1989) Efficient parameter estimation for self-similar processes. *Annals of Statistics* 17, 1749–1766.
Davies, R.B. & D.S. Harte (1987) Tests for Hurst effect. *Biometrika* 74, 95–102.
Deo, R.S. & W.W. Chen (2000) On the integral of the squared periodogram. *Stochastic Processes and Their Applications* 85, 159–176.
Granger, C.W.J. & R. Joyeux (1980) An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15–29.
Grenander, U. & M. Rosenblatt (1957) *Statistical Analysis of Stationary Time Series*. Wiley.
Hong, Y. (1996) Consistent testing for serial correlation of unknown form. *Econometrica* 64, 837–864.
Hosking, J.R.M. (1981) Fractional differencing. *Biometrika* 68, 165–176.
Hurvich, C.M., R.S. Deo, & J. Brodsky (1998) The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. *Journal of Time Series Analysis* 19, 19–46.
Mandelbrot, B.B. & J.W. Van Ness (1968) Fractional Brownian motions, fractional noises, and applications. *SIAM Review* 10, 422–437.
Milhoj, A. (1981) A test of fit in time series models. *Biometrika* 68, 177–187.
Moulines, E. & P. Soulier (1999) Broadband log-periodogram regression of time series with long-range dependence. *Annals of Statistics* 27, 1415–1439.
Yajima, Y. (1993) Asymptotic properties of estimates in incorrect ARMA models for long-memory time series. In *New Directions in Time Series Analysis*, part 2, pp. 375–382.
Zygmund, A. (1977) *Trigonometric Series*. Cambridge University Press.

APPENDIX: PROOFS

We will only provide the proofs for long memory models. The proofs for short memory models are similar though much simpler and are available from the authors. In this Appendix, we will often use the following decomposition of $I(\lambda)$:

$$I(\lambda) = |\psi(\lambda)|^2 I_\varepsilon(\lambda) + I(\lambda) - |\psi(\lambda)|^2 I_\varepsilon(\lambda),$$

where $\psi(\lambda) = \sum_{k=0}^{\infty} \psi_k e^{-i\lambda k}$ and $I_\varepsilon(\lambda)$ is the periodogram of the innovations ε_t in (6). Then

$$\frac{I(\lambda)}{f(\lambda)} = \frac{2\pi}{\sigma^2} I_\varepsilon(\lambda) + R(\lambda), \tag{A.1}$$

where

$$R(\lambda) = \frac{I(\lambda)}{f(\lambda)} - \frac{2\pi}{\sigma^2} I_\varepsilon(\lambda). \tag{A.2}$$

Let $\hat{\gamma}_{\varepsilon,j}$ be the jth sample covariance of the ε_t given by $\hat{\gamma}_{\varepsilon,j} = n^{-1}\sum_{t=|j|+1}^{n}(\varepsilon_t - \bar{\varepsilon}) \times (\varepsilon_{t-|j|} - \bar{\varepsilon})$, for $|j| \le n-1$.

Proof of Theorem 2. Let $I_\varepsilon(\lambda) = (2\pi n)^{-1}|\sum_{t=1}^{n}\varepsilon_t e^{i\lambda t}|^2$ be the periodogram of the innovations ε_t without mean correction. For the Fourier frequencies, λ_k, $k = 1,\dots,(n-1)$, we have $I_\varepsilon(\lambda_k) = I_{n,\varepsilon}(\lambda_k)$, where $I_{n,\varepsilon}$ is the periodogram of the mean corrected innovations $\varepsilon_t - \bar{\varepsilon}$. Also define

$$\hat{f}_{\varepsilon,d}(\lambda) = \frac{2\pi}{n}\sum_{j=1}^{n-1} W(\lambda - \lambda_j) I_\varepsilon(\lambda_j).$$

In Lemmas 1–3, which follow, we show that

$$\frac{n}{p_n^{1/2}}\left\{\frac{2\pi}{n}\sum_{\ell=0}^{n-1}\left(\tilde{f}_e^2(\lambda_\ell) - \frac{4\pi^2}{\sigma^4}\hat{f}_{\varepsilon,d}^2(\lambda_\ell)\right)\right\} = o_p(1),$$

$$\frac{n}{p_n^{1/2}}\left\{\frac{2\pi}{n}\sum_{\ell=0}^{n-1}\frac{4\pi^2}{\sigma^4}\hat{f}_{\varepsilon,d}^2(\lambda_\ell) - \frac{2\pi}{\sigma^4}\sum_{j=-(n-1)}^{n-1}k^2(j/p_n)\hat{\gamma}_{\varepsilon,j}^2\right\} = o_p(1),$$

and

$$\frac{n}{p_n^{1/2}}\left[\left\{\frac{2\pi}{n}\sum_{\ell=0}^{n-1}\tilde{f}_e(\lambda_\ell)\right\}^2 - \left(\frac{2\pi}{\sigma^2}\hat{\gamma}_{\varepsilon,0}\right)^2\right] = o_p(1).$$

Also, by Lemma 3, $\{(2\pi/n)\sum_{\ell=0}^{n-1}(2\pi/\sigma^2)\hat{f}_{\varepsilon,d}(\lambda_\ell)\}^2 = (4\pi^2/\sigma^4)\hat{\gamma}_{\varepsilon,0}^2$ and $\sqrt{n}(\hat{\gamma}_{\varepsilon,0} - \sigma^2) = O_p(1)$. The theorem now follows by Theorem 1 of Hong (1996) and the fact that $p_n^{-1}D_n(k) \to D(k) \equiv \int_0^\infty k^4(z)\,dz < \infty$ as $n \to \infty$ by Assumption 2a. ■

Proof of Theorem 4. By Theorem 2 it suffices to show that

$$\frac{n}{p_n^{1/2}}(T_n(\boldsymbol{\theta}_0) - T_n(\hat{\boldsymbol{\theta}})) = o_p(1), \tag{A.3}$$

which we do by establishing that

$$\frac{n}{p_n^{1/2}}\left\{\frac{2\pi}{n}\sum_{\ell=0}^{n-1}(\tilde{f}_e^2(\boldsymbol{\theta}_0,\lambda_\ell) - \tilde{f}_e^2(\hat{\boldsymbol{\theta}},\lambda_\ell))\right\} = o_p(1) \tag{A.4}$$

and

$$\frac{n}{p_n^{1/2}}\left[\left\{\frac{2\pi}{n}\sum_{\ell=0}^{n-1}\tilde{f}_e(\boldsymbol{\theta}_0,\lambda_\ell)\right\}^2 - \left\{\frac{2\pi}{n}\sum_{\ell=0}^{n-1}\tilde{f}_e(\hat{\boldsymbol{\theta}},\lambda_\ell)\right\}^2\right] = o_p(1). \tag{A.5}$$

We will prove only (A.4) because the proof of (A.5) is similar. Let

$$\mathcal{G}_{\boldsymbol{\theta}}(\lambda_j,\lambda_h) = \frac{1}{f_{\boldsymbol{\theta}}(\lambda_j)}\frac{1}{f_{\boldsymbol{\theta}}(\lambda_h)}.$$

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press

Then the LHS of (A.4) is

$$\frac{n}{p_n^{1/2}}\frac{2\pi}{n}\sum_{\ell=0}^{n-1}\left(\frac{2\pi}{n}\sum_{j=1}^{n-1}W(\lambda_{\ell-j})\frac{I(\lambda_j)}{f_{\boldsymbol{\theta}_0}(\lambda_j)}\right)^2-\left(\frac{2\pi}{n}\sum_{j=1}^{n-1}W(\lambda_{\ell-j})\frac{I(\lambda_j)}{f_{\hat{\boldsymbol{\theta}}}(\lambda_j)}\right)^2$$

$$=\frac{(2\pi)^2}{n^2p_n^{1/2}}\sum_{\ell=0}^{n-1}\sum_{j,h=1}^{n-1}W(\lambda_{\ell-j})W(\lambda_{\ell-h})I(\lambda_j)I(\lambda_h)(\mathcal{G}_{\boldsymbol{\theta}_0}(\lambda_j,\lambda_h)-\mathcal{G}_{\hat{\boldsymbol{\theta}}}(\lambda_j,\lambda_h)).$$

By a similar argument of deriving (A.25) in the proof of Lemma 1, which follows, the RHS of the preceding equation is

$$\frac{2\pi}{np_n^{1/2}}\sum_{j,h=1}^{n-1}I(\lambda_j)I(\lambda_h)(\mathcal{G}_{\boldsymbol{\theta}_0}(\lambda_j,\lambda_h)-\mathcal{G}_{\hat{\boldsymbol{\theta}}}(\lambda_j,\lambda_h))\mathcal{K}_n(\lambda_{j-h}), \tag{A.6}$$

where

$$\mathcal{K}_n(\lambda_s)=\sum_{p=-(n-1)}^{n-1}k_p^2e^{i\lambda_s p}+2\sum_{p=1}^{n-1}k_pk_{n-p}e^{i\lambda_s p}. \tag{A.7}$$

For every λ_j and λ_h, we have by a Taylor series expansion,

$$\mathcal{G}(\lambda_j,\lambda_h,\boldsymbol{\theta}_0)-\mathcal{G}(\lambda_j,\lambda_h,\hat{\boldsymbol{\theta}})$$

$$=\sum_u\left(\frac{1}{f_{\boldsymbol{\theta}_0}(\lambda_h)}\frac{\partial f^{-1}(\lambda_j,\boldsymbol{\theta}_0)}{\partial\theta_u}+\frac{1}{f_{\boldsymbol{\theta}_0}(\lambda_j)}\frac{\partial f^{-1}(\lambda_h,\boldsymbol{\theta}_0)}{\partial\theta_u}\right)(\hat{\theta}_u-\theta_{0_u})$$

$$+\frac{1}{2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\frac{\partial^2\mathcal{G}(\lambda_j,\lambda_h,\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}^2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0),$$

where $\tilde{\boldsymbol{\theta}}_{jh}=\boldsymbol{\theta}_0+\alpha_{jh}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ for some $0<\alpha_{jh}<1$ and

$$\frac{\partial^2\mathcal{G}(\lambda_j,\lambda_h,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}=\frac{1}{f_{\boldsymbol{\theta}}(\lambda_h)}\frac{\partial^2 f^{-1}(\lambda_j,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}+\frac{\partial f^{-1}(\lambda_j,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial f^{-1}(\lambda_h,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}'$$

$$\times\frac{1}{f_{\boldsymbol{\theta}}(\lambda_j)}\frac{\partial^2 f^{-1}(\lambda_h,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}+\frac{\partial f^{-1}(\lambda_h,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial f^{-1}(\lambda_j,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}' \tag{A.8}$$

To prove (A.4), we will show that (A.6) is $o_p(1)$ by verifying, for each u,

$$\frac{2\pi}{np_n^{1/2}}\sum_{j,h=1}^{n-1}I(\lambda_j)I(\lambda_h)\left(\frac{1}{f_{\boldsymbol{\theta}_0}(\lambda_j)}\frac{\partial f^{-1}(\lambda_h,\boldsymbol{\theta}_0)}{\partial\theta_u}\right)(\hat{\theta}_u-\theta_{0_u})\mathcal{K}_n(\lambda_{j-h})=o_p(1) \tag{A.9}$$

and

$$\frac{2\pi}{np_n^{1/2}}\sum_{j,h=1}^{n-1}I(\lambda_j)I(\lambda_h)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\frac{\partial^2\mathcal{G}(\lambda_j,\lambda_h,\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}^2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\mathcal{K}_n(\lambda_{j-h})=o_p(1). \tag{A.10}$$

We first show (A.9). Let

$$g(\lambda)=\frac{\partial\ln f(\lambda,\boldsymbol{\theta}_0)}{\partial\theta_u}; \tag{A.11}$$

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press

then

$$\frac{1}{f_{\boldsymbol{\theta}_0}(\lambda_j)} \frac{\partial f^{-1}(\lambda_h, \boldsymbol{\theta}_0)}{\partial \theta_u} = -\frac{1}{f_{\boldsymbol{\theta}_0}(\lambda_j)} \frac{1}{f_{\boldsymbol{\theta}_0}(\lambda_h)} g(\lambda_h).$$

Because $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = O_p(n^{-1/2})$, (A.9) is true if

$$\sum_{j,h=1}^{n-1} \frac{I(\lambda_j)}{f_{\boldsymbol{\theta}_0}(\lambda_j)} \frac{I(\lambda_h)}{f_{\boldsymbol{\theta}_0}(\lambda_h)} g(\lambda_h)\mathcal{K}_n(\lambda_{j-h}) = o_p(n^{3/2}p_n^{1/2}).$$

By (A.1), it is thus enough to show that

$$\sum_{j,h=1}^{n-1} I_\varepsilon(\lambda_j)I_\varepsilon(\lambda_h)g(\lambda_h)\mathcal{K}_n(\lambda_{j-h}) = o_p(n^{3/2}p_n^{1/2}), \tag{A.12}$$

$$\sum_{j,h=1}^{n-1} I_\varepsilon(\lambda_j)R_{\boldsymbol{\theta}_0}(\lambda_h)g(\lambda_h)\mathcal{K}_n(\lambda_{j-h}) = o_p(n^{3/2}p_n^{1/2}), \tag{A.13}$$

and

$$\sum_{j,h=1}^{n-1} R_{\boldsymbol{\theta}_0}(\lambda_j)R_{\boldsymbol{\theta}_0}(\lambda_h)g(\lambda_h)\mathcal{K}_n(\lambda_{j-h}) = o_p(n^{3/2}p_n^{1/2}). \tag{A.14}$$

Because $g(\lambda) = O(\lambda^{-\delta})$ by Assumption 7, (A.13) and (A.14) can be shown by an argument similar to that used to establish (A.26) and (A.27) in the proof of Lemma 1. To show (A.12), we let

$$a_m = \sum_{h=1}^{n-1} g(\lambda_h)e^{-i\lambda_h m}.$$

Using the fact that $\sum_{j=1}^{n-1} e^{-i\lambda_j p} = (n-1)I(p=0) - 1I(p \neq 0)$, the LHS of (A.12) is

$$\begin{aligned}
&\sum_{j,h=1}^{n-1} I_\varepsilon(\lambda_j)I_\varepsilon(\lambda_h)g(\lambda_h)\left(\sum_{p=-(n-1)}^{n-1} k_p^2 e^{i\lambda_{j-h}p} + 2\sum_{p=1}^{n-1} k_p k_{n-p} e^{i\lambda_{j-h}p}\right)\\
&\quad = \frac{1}{(2\pi n)^2} \sum_{p=-(n-1)}^{n-1} k_p^2 \sum_{s,t,u,v} \varepsilon_s\varepsilon_t\varepsilon_u\varepsilon_v \sum_{h=1}^{n-1} g(\lambda_h)e^{-i\lambda_h(u-v-p)} \sum_{j=1}^{n-1} e^{-i\lambda_j(s-t+p)}\\
&\qquad + \frac{2}{(2\pi n)^2} \sum_{p=1}^{n-1} k_p k_{n-p} \sum_{s,t,u,v} \varepsilon_s\varepsilon_t\varepsilon_u\varepsilon_v \sum_{h=1}^{n-1} g(\lambda_h)e^{-i\lambda_h(u-v-p)} \sum_{j=1}^{n-1} e^{-i\lambda_j(s-t+p)}\\
&\quad = \frac{1}{4\pi^2 n}\left(\sum_{p=-(n-1)}^{n-1} k_p^2 \sum_{t,u,v} a_{u-v-p}\varepsilon_t\varepsilon_{t-p}\varepsilon_u\varepsilon_v + 2\sum_{p=1}^{n-1} k_p k_{n-p} \sum_{t,u,v} a_{u-v-p}\varepsilon_t\varepsilon_{t-p}\varepsilon_u\varepsilon_v\right)\\
&\qquad - \frac{1}{4\pi^2 n^2}\left(\sum_{p=-(n-1)}^{n-1} k_p^2 \sum_{s,t,u,v} a_{u-v-p}\varepsilon_s\varepsilon_t\varepsilon_u\varepsilon_v + 2\sum_{p=1}^{n-1} k_p k_{n-p} \sum_{s,t,u,v} a_{u-v-p}\varepsilon_s\varepsilon_t\varepsilon_u\varepsilon_v\right).
\end{aligned} \tag{A.15}$$

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press

We will show that both terms of the last expression in (A.15) have second moments of order $o(n^3 p_n)$. By the Cauchy–Schwarz inequality, we have

$$E\left(\frac{1}{4\pi^2 n}\sum_{p=-(n-1)}^{n-1} k_p^2 \sum_{t,u,v} a_{u-v-p}\varepsilon_t\varepsilon_{t-p}\varepsilon_u\varepsilon_v + 2\sum_{p=1}^{n-1} k_p k_{n-p}\sum_{t,u,v} a_{u-v-p}\varepsilon_t\varepsilon_{t-p}\varepsilon_u\varepsilon_v\right)^2$$

$$= O\left(\frac{1}{16\pi^4 n^2}\sum_{p_1,p_2=-(n-1)}^{n-1} k_{p_1}^2 k_{p_2}^2 \sum_{t_1,t_2,u_1,u_2,v_1,v_2} a_{u_1-v_1-p_1} a_{u_2-v_2-p_2} \times E(\varepsilon_{t_1}\varepsilon_{t_2}\varepsilon_{t_1-p_1}\varepsilon_{t_2-p_2}\varepsilon_{u_1}\varepsilon_{u_2}\varepsilon_{v_1}\varepsilon_{v_2})\right)$$

$$+ O\left(\frac{1}{16\pi^4 n^2}\sum_{p_1,p_2=-(n-1)}^{n-1} k_{n-p_1}^2 k_{n-p_2}^2 \sum_{t_1,t_2,u_1,u_2,v_1,v_2} a_{u_1-v_1-p_1} a_{u_2-v_2-p_2} \times E(\varepsilon_{t_1}\varepsilon_{t_2}\varepsilon_{t_1-p_1}\varepsilon_{t_2-p_2}\varepsilon_{u_1}\varepsilon_{u_2}\varepsilon_{v_1}\varepsilon_{v_2})\right).$$

Because ε_t are independent with zero mean, the preceding expectation is positive only when the random variables inside the parentheses consist of products of even powers of the ε_t. Thus, the preceding expression is dominated by two cases: one is when $p_1 = p_2 = 0$, $u_1 = u_2$, and $v_1 = v_2$, and the other is when $p_1 = p_2 = 0$, $u_1 = v_1$, and $u_2 = v_2$. Using Lemma 6, which follows, the order of these two cases is

$$O\left(n^{-2}\sum_{t_1,t_2,u_1,u_2} a_{u_1-u_2}^2 + n^{-2}\sum_{t_1,t_2,u_1,u_2} a_0^2\right)$$

$$= O(n^{-2}n^{5+2\delta} + n^{-2}n^{4+2\delta})$$

$$= o(n^3 p_n).$$

It can be shown that the second moment of the second term in (A.15) is also of order $o(n^3 p_n)$ by similar arguments. We have thus established (A.9).

Next, we establish (A.10). Let $\mathcal{A}_{uv}(\lambda_j, \lambda_h, \tilde{\boldsymbol{\theta}}_{jh})$ denote the (u,v)th element of the matrix

$$\frac{\partial^2 \mathcal{G}(\lambda_j, \lambda_h, \tilde{\boldsymbol{\theta}}_{jh})}{\partial \boldsymbol{\theta}^2}.$$

Then, by (A.8) and Assumption 7,

$$|\lambda_j^{-2\tilde{d}_{jh}}\lambda_h^{-2\tilde{d}_{jh}}\lambda_j^{\delta}\lambda_h^{\delta}\mathcal{A}_{uv}(\lambda_j, \lambda_h, \tilde{\boldsymbol{\theta}}_{jh})| \le A \quad \text{for some } 0 < A < \infty \text{ with probability 1,} \tag{A.16}$$

where $\tilde{\boldsymbol{\theta}}_{jh} = (\tilde{\boldsymbol{\beta}}_{jh}, \tilde{d}_{jh})'$. Because $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = O_p(n^{-1/2})$, (A.10) will follow if for every (u,v)

$$\sum_{j,h=1}^{n-1} I(\lambda_j) I(\lambda_h) \mathcal{A}_{uv}(\lambda_j, \lambda_h, \tilde{\boldsymbol{\theta}}_{jh}) \mathcal{K}_n(\lambda_{j-h}) = o_p(n^2 p_n^{1/2}).$$

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press

To show this, it suffices, by (A.1), to prove that

$$\sum_{j,h=1}^{n-1} f(\boldsymbol{\theta}_0,\lambda_j)f(\boldsymbol{\theta}_0,\lambda_h)I_\varepsilon(\lambda_j)I_\varepsilon(\lambda_h)\mathcal{A}_{uv}(\lambda_j,\lambda_h,\tilde{\boldsymbol{\theta}}_{jh})\mathcal{K}_n(\lambda_{j-h}) = o_p(n^2p_n^{1/2}), \quad \textbf{(A.17)}$$

$$\sum_{j,h=1}^{n-1} f(\boldsymbol{\theta}_0,\lambda_j)f(\boldsymbol{\theta}_0,\lambda_h)I_\varepsilon(\lambda_j)R_{\boldsymbol{\theta}_0}(\lambda_h)\mathcal{A}_{uv}(\lambda_j,\lambda_h,\tilde{\boldsymbol{\theta}}_{jh})\mathcal{K}_n(\lambda_{j-h}) = o_p(n^2p_n^{1/2}),$$

and

$$\sum_{j,h=1}^{n-1} f(\boldsymbol{\theta}_0,\lambda_j)f(\boldsymbol{\theta}_0,\lambda_h)R_{\boldsymbol{\theta}_0}(\lambda_j)R_{\boldsymbol{\theta}_0}(\lambda_h)\mathcal{A}_{uv}(\lambda_j,\lambda_h,\tilde{\boldsymbol{\theta}}_{jh})\mathcal{K}_n(\lambda_{j-h}) = o_p(n^2p_n^{1/2}).$$

We will prove only the first of these, because the proof for the other two is similar. Letting

$$Y_n = \sum_{j,h=1}^{n-1} f(\boldsymbol{\theta}_0,\lambda_j)f(\boldsymbol{\theta}_0,\lambda_h)I_\varepsilon(\lambda_j)I_\varepsilon(\lambda_h)\mathcal{A}_{uv}(\lambda_j,\lambda_h,\tilde{\boldsymbol{\theta}}_{jh})\mathcal{K}_n(\lambda_{j-h}),$$

we have

$$Y_n = I(\hat{d} \ge d_0)Y_n + I(\hat{d} < d_0)Y_n. \quad \textbf{(A.18)}$$

First consider $\hat{d} \ge d_0$. Then $\tilde{d}_{jh} \ge d_0$ for all j,h. Hence, by Assumption 7 and (A.16), we have

$$|f(\boldsymbol{\theta}_0,\lambda_j)f(\boldsymbol{\theta}_0,\lambda_h)\mathcal{A}_{uv}(\lambda_j,\lambda_h,\tilde{\boldsymbol{\theta}}_{jh})| \le A\lambda_j^{-\delta}\lambda_h^{-\delta} \quad \textbf{(A.19)}$$

with probability 1 for some $0 < A < \infty$ for all j,h . Also, by the Cauchy–Schwarz inequality, $\sup_{j,h} E|I_\varepsilon(\lambda_j)I_\varepsilon(\lambda_h)| < K < \infty$, and it follows from (A.19) and (A.28) in the proof of Lemma 1 that

$$\begin{aligned} E|I(\hat{d} \ge d_0)Y_n| &\le E\left(\sum_{j,h=1}^{n-1} A\lambda_j^{-\delta}\lambda_h^{-\delta}\{n(j-h)^{-1}I(j\ne h) + p_n I(j=h)\}|I_\varepsilon(\lambda_j)I_\varepsilon(\lambda_h)|\right) \\ &= O\left(\sum_{j,h=1}^{n-1} \lambda_j^{-\delta}\lambda_h^{-\delta}\{n(j-h)^{-1}I(j\ne h) + p_n I(j=h)\}\right) = O(n^2\log n) \\ &= o(n^2p_n^{1/2}). \end{aligned} \quad \textbf{(A.20)}$$

Now consider $\hat{d} < d_0$. Then $0 < \tilde{d}_{jh} < d_0$ for all j,h. By part (iii) of Assumption 7 we get that

$$f(\boldsymbol{\theta}_0,\lambda_j)f(\boldsymbol{\theta}_0,\lambda_h)f^{-1}(\tilde{\boldsymbol{\theta}}_{jh},\lambda_j)f^{-1}(\tilde{\boldsymbol{\theta}}_{jh},\lambda_h) = (1+\Delta_j)(1+\Delta_h),$$

where

$$|\Delta_j| \le K\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|\lambda_j^{-2d_0} \quad \textbf{(A.21)}$$

for all j. Furthermore,

$$|f(\tilde{\boldsymbol{\theta}}_{jh},\lambda_j)f(\tilde{\boldsymbol{\theta}}_{jh},\lambda_h)\mathcal{A}_{uv}(\lambda_j,\lambda_h,\tilde{\boldsymbol{\theta}}_{jh})| \le A\lambda_j^{-\delta}\lambda_h^{-\delta} \tag{A.22}$$

with probability 1 for some $0 < A < \infty$ by (A.16). Using these bounds and (A.28) we get

$$\begin{aligned}|I(\hat{d} < d_0)Y_n| &\le \sum_{j,h=1}^{n-1} A(1+\Delta_j)(1+\Delta_h)\lambda_j^{-\delta}\lambda_h^{-\delta}|\mathcal{K}_n(\lambda_{j-h})|I_\varepsilon(\lambda_j)I_\varepsilon(\lambda_h)\\
&= \sum_{j,h=1}^{n-1} A\lambda_j^{-\delta}\lambda_h^{-\delta}|\mathcal{K}_n(\lambda_{j-h})|I_\varepsilon(\lambda_j)I_\varepsilon(\lambda_h)\\
&\quad + 2\sum_{j,h=1}^{n-1} A\Delta_j\lambda_j^{-\delta}\lambda_h^{-\delta}|\mathcal{K}_n(\lambda_{j-h})|I_\varepsilon(\lambda_j)I_\varepsilon(\lambda_h)\\
&\quad + \sum_{j,h=1}^{n-1} A\Delta_j\Delta_h\lambda_j^{-\delta}\lambda_h^{-\delta}|\mathcal{K}_n(\lambda_{j-h})|I_\varepsilon(\lambda_j)I_\varepsilon(\lambda_h)\\
&\equiv T_1 + T_2 + T_3.\end{aligned}$$

From (A.20) we have that $T_1 = o_p(n^2 p_n^{1/2})$. Also, by (A.21),

$$T_2 \le A\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \sum_{j,h=1}^{n-1} \lambda_j^{-2d_0-\delta}\lambda_h^{-\delta}|\mathcal{K}_n(\lambda_{j-h})|I_\varepsilon(\lambda_j)I_\varepsilon(\lambda_h).$$

An argument similar to that in (A.20) shows that $\sum_{j,h=1}^{n-1}\lambda_j^{-2d_0-\delta}\lambda_h^{-\delta}\times |\mathcal{K}_n(\lambda_{j-h})|I_\varepsilon(\lambda_j)I_\varepsilon(\lambda_h) = O_p(n^2\log n)$ and because $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_p(n^{-1/2})$, we get $T_2 = O_p(n^{3/2}\log n) = o_p(n^2 p_n^{1/2})$. Arguing in the same vein, we establish that $T_3 = I(d_0 < \frac{1}{4})O_p(n\log n) + I(d_0 \ge \frac{1}{4})O_p(n^{4d_0+2\delta}\log n) = o_p(n^2 p_n^{1/2})$. These bounds on T_1, T_2, and T_3 yield

$$|I(\hat{d} < d_0)Y_n| = o_p(n^2 p_n^{1/2}). \tag{A.23}$$

Thus, (A.17) follows from (A.18), (A.20), and (A.23). ■

LEMMA 1. *Under the assumptions in Theorem 2,*

$$\frac{n}{p_n^{1/2}}\left\{\frac{2\pi}{n}\sum_{\ell=0}^{n-1}\left(\tilde{f}_e^2(\lambda_\ell) - \frac{4\pi^2}{\sigma^4}\hat{f}_{\varepsilon,d}^2(\lambda_\ell)\right)\right\} = o_p(1). \tag{A.24}$$

Proof. The LHS of (A.24) is

$$\begin{aligned}&\frac{n}{p_n^{1/2}}\frac{2\pi}{n}\sum_{\ell=0}^{n-1}\left[\left(\frac{2\pi}{n}\sum_{j=1}^{n-1}W(\lambda_{\ell-j})\frac{I(\lambda_j)}{f(\lambda_j)}\right)^2 - \left(\frac{2\pi}{\sigma^2}\frac{2\pi}{n}\sum_{j=1}^{n-1}W(\lambda_{\ell-j})I_\varepsilon(\lambda_\ell)\right)^2\right]\\
&\quad = \frac{(2\pi)^3}{n^2 p_n^{1/2}}\sum_{\ell=0}^{n-1}\sum_{j,h=1}^{n-1}W(\lambda_{\ell-j})W(\lambda_{\ell-h})\{I_\varepsilon(\lambda_j)R(\lambda_h) + I_\varepsilon(\lambda_h)R(\lambda_j) + R(\lambda_j)R(\lambda_h)\}.\end{aligned}$$

Letting $k_s = k(s/p_n)$ and $\Phi(\lambda_j,\lambda_h) = I_\varepsilon(\lambda_j)R(\lambda_h) + I_\varepsilon(\lambda_h)R(\lambda_j) + R(\lambda_j)R(\lambda_h)$, the last line of the preceding equation becomes

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press

$$\frac{2\pi}{n^2 p_n^{1/2}} \sum_{\ell=0}^{n-1} \sum_{j,h=1}^{n-1} \sum_{p,q=-(n-1)}^{n-1} k_p k_q e^{-i(\lambda_{\ell-j}p+\lambda_{\ell-h}q)} \Phi(\lambda_j, \lambda_h)$$

$$= \frac{2\pi}{n^2 p_n^{1/2}} \sum_{j,h=1}^{n-1} \sum_{p,q=-(n-1)}^{n-1} k_p k_q e^{i(\lambda_j p+\lambda_h q)} \Phi(\lambda_j, \lambda_h) \sum_{\ell=0}^{n-1} e^{-i\lambda_\ell (p+q)}$$

$$= \frac{2\pi}{n p_n^{1/2}} \left(\sum_{j,h=1}^{n-1} \Phi(\lambda_j, \lambda_h) \left(\sum_{p=-(n-1)}^{n-1} k_p^2 e^{i\lambda_{j-h}p} + 2 \sum_{p=1}^{n-1} k_p k_{n-p} e^{i\lambda_{j-h}p} \right)\right). \qquad \textbf{(A.25)}$$

We will show that (A.25) is $o_p(1)$ by verifying

$$E\left| \sum_{j,h=1}^{n-1} R(\lambda_j) I_\varepsilon(\lambda_h) \mathcal{K}_n(\lambda_{j-h}) \right| = o(n p_n^{1/2}) \qquad \textbf{(A.26)}$$

and

$$E\left| \sum_{j,h=1}^{n-1} R(\lambda_j) R(\lambda_h) \mathcal{K}_n(\lambda_{j-h}) \right| = o(n p_n^{1/2}), \qquad \textbf{(A.27)}$$

where $\mathcal{K}_n(\lambda_{j-h})$ is defined in (A.7). To prove the preceding two equations, we will need a bound for $\mathcal{K}_n(\lambda_s)$. We first note that from page 2 of Zygmund (1977)

$$\sum_{\ell=1}^{b} e^{i\lambda\ell} = \sum_{\ell=1}^{b} \cos\lambda\ell + i \sum_{\ell=1}^{b} \sin\lambda\ell$$

$$= \frac{\sin(b+\frac{1}{2})\lambda}{2\sin\lambda/2} - \frac{1}{2} + i\,\frac{\cos\lambda/2 - \cos(b+\frac{1}{2})\lambda}{2\sin\lambda/2} = O(\lambda^{-1})$$

uniformly in b for $0 < \lambda < \pi$. Using this bound, in conjunction with the fact that $p_n^{-1}\sum|k_p| = O(1)$ and by applying summation by parts and by Assumption 2b, for $s \neq 0$ we obtain

$$\sum_{p=-(n-1)}^{n-1} k_p^2 e^{i\lambda_s p} = 2\sum_{p=1}^{n-1} k_p^2 \cos(\lambda_s p) + 1$$

$$= 2\sum_{p=1}^{n-2} (k_p^2 - k_{p+1}^2) \sum_{u=1}^{p} \cos(\lambda_s u) + 2k_{n-1}^2 \sum_{u=1}^{n-1} \cos(\lambda_s u) + 1$$

$$= 2\sum_{p=1}^{n-2} (k_p - k_{p+1})(k_p + k_{p+1}) \sum_{u=1}^{p} \cos(\lambda_s u) + O(1)$$

$$= O\left(\sum_{m=1}^{n-2} \frac{1}{p_n} k'_{\tilde{p}} (k_p + k_{p+1}) \sum_{u=1}^{p} \cos(\lambda_s u) \right)$$

$$= O(\lambda_s^{-1}),$$

where $p < \tilde{p} < p+1$. Similarly,

$$\sum_{p=1}^{n-1} k_p k_{n-p} e^{i\lambda_s p} = O(\lambda_s^{-1}),$$

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press

and hence

$$\mathcal{K}_n(\lambda_s) = O(\lambda_s^{-1})I(s \neq 0) + O(p_n)I(s = 0). \tag{A.28}$$

We shall only derive (A.26) and (A.27) when $j \neq h$, because the proofs for $j = h$ are similar and simpler. To prove (A.26) we note that the LHS of (A.26) is bounded by

$$E\left|\sum_{j=1}^{\log^2 n}\sum_{h\neq j} R(\lambda_j)I_\varepsilon(\lambda_h)\mathcal{K}_n(\lambda_{j-h})\right| + 2E\left|\sum_{j=\log^2 n+1}^{n-1}\sum_{h=1}^{\log^2 n} R(\lambda_j)I_\varepsilon(\lambda_h)\mathcal{K}_n(\lambda_{j-h})\right|$$
$$+ E\left|\sum_{j=\log^2 n+1}^{n-1}\sum_{h=\log^2 n+1}^{n-1} R(\lambda_j)I_\varepsilon(\lambda_h)\mathcal{K}_n(\lambda_{j-h})\right|.$$

Using the Cauchy–Schwarz inequality, Lemma 5, equation (A.28), and the fact that $\max_j E(I_\varepsilon^2(\lambda_j)) < \infty$, the first term and second term of the preceding equation are of the order

$$O\left(\sum_{j=1}^{\log^2 n}\sum_{h\neq j}|\mathcal{K}_n(\lambda_{j-h})| + \sum_{j=\log^2 n+1}^{n-1}\sum_{h=1}^{\log^2 n}|\mathcal{K}_n(\lambda_{j-h})|\right)$$
$$= O\left(\sum_{j=1}^{\log^2 n}\sum_{h\neq j}\lambda_{j-h}^{-1} + \sum_{j=\log^2 n+1}^{n-1}\sum_{h=1}^{\log^2 n}\lambda_{j-h}^{-1}\right)$$
$$= O(n\log^3 n).$$

To verify the third term is $o(np_n^{1/2})$, we will show that

$$E\left|\sum_{j=\log^2 n+1}^{n-1}\sum_{j\neq h=\log^2 n+1}^{n-1} R(\lambda_j)I_\varepsilon(\lambda_h)\mathcal{K}_n(\lambda_{j-h})\right|^2 = o(n^2 p_n).$$

By Assumption 3, Lemma 4 and (A.28),

$$E\left|\sum_{j=\log^2 n+1}^{n-1}\sum_{j\neq h=\log^2 n+1}^{n-1} R(\lambda_j)I_\varepsilon(\lambda_h)\mathcal{K}_n(\lambda_{j-h})\right|^2$$
$$= 2\sum_{h_1=\log^2 n+1}^{n-1}\sum_{j_1=\log^2 n+1}^{h_1-1}\sum_{h_2=\log^2 n+1}^{n-1}\sum_{j_2=\log^2 n+1}^{h_2-1} E(R(\lambda_{j_1})I_\varepsilon(\lambda_{h_1})R(\lambda_{j_2})I_\varepsilon(\lambda_{h_2}))$$
$$\times \mathcal{K}_n(\lambda_{j_1-h_1})\mathcal{K}_n(-\lambda_{j_2-h_2})$$
$$+ 2\sum_{h_1=\log^2 n+1}^{n-1}\sum_{j_1=h_1+1}^{n-1}\sum_{h_2=\log^2 n+1}^{n-1}\sum_{j_2=h_2+1}^{n-1} E(R(\lambda_{j_1})I_\varepsilon(\lambda_{h_1})R(\lambda_{j_2})I_\varepsilon(\lambda_{h_2}))$$
$$\times \mathcal{K}_n(\lambda_{j_1-h_1})\mathcal{K}_n(-\lambda_{j_2-h_2})$$
$$= O\left(\sum_{h_1=\log^2 n+1}^{n-1}\sum_{j_1=\log^2 n+1}^{h_1}\sum_{h_2=\log^2 n+1}^{n-1}\sum_{j_2=\log^2 n+1}^{h_2} j_1^{-d}h_1^{d-1}j_2^{-d}h_2^{d-1}\log h_1\log h_2\,\lambda_{j_1-h_1}^{-1}\lambda_{j_2-h_2}^{-1}\right)$$
$$= O(n^2\log^6 n) = o(n^2 p_n).$$

Thus (A.26) is proved. The proof of (A.27) is similar to that of (A.26). ■

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press

LEMMA 2. *Under Assumptions 1, 2a, and 3,*

$$\frac{n}{p_n^{1/2}}\left\{\frac{2\pi}{n}\sum_{\ell=0}^{n-1}\frac{4\pi^2}{\sigma^4}\hat{f}^2_{\varepsilon,d}(\lambda_\ell)-\frac{2\pi}{\sigma^4}\sum_{j=-(n-1)}^{n-1}k^2(j/p_n)\hat{\gamma}^2_{\varepsilon,j}\right\}=o_p(1).$$

Proof. Because $I_\varepsilon(\lambda_j)=I_{n,\varepsilon}(\lambda_j)$ and $I_{n,\varepsilon}(0)=0$, we have

$$I_\varepsilon(\lambda_j)=\frac{1}{2\pi}\sum_{h=-(n-1)}^{n-1}\hat{\gamma}_{\varepsilon,h}e^{-i\lambda_j h},\quad \text{for } j=1,\dots,(n-1).$$

Now

$$\begin{aligned}&\frac{2\pi}{n}\sum_{\ell=0}^{n-1}\hat{f}^2_{\varepsilon,d}(\lambda_\ell)\\&=\left(\frac{2\pi}{n}\right)^3\frac{1}{4\pi^2\sigma^4}\sum_{\ell,j_1,j_2=0}^{n-1}\sum_{p_1,p_2,h_1,h_2=-(n-1)}^{n-1}k_{p_1}k_{p_2}\hat{\gamma}_{\varepsilon,h_1}\hat{\gamma}_{\varepsilon,h_2}e^{-i\lambda_{\ell-j_1}p_1}e^{i\lambda_{\ell-j_2}p_2}e^{-i\lambda_{j_1}h_1}e^{i\lambda_{j_2}h_2}\\&=\frac{2\pi}{n^3\sigma^4}\sum_{p_1,p_2,h_1,h_2=-(n-1)}^{n-1}k_{p_1}k_{p_2}\hat{\gamma}_{\varepsilon,h_1}\hat{\gamma}_{\varepsilon,h_2}\sum_{\ell=0}^{n-1}e^{-i\lambda_{p_1-p_2}\ell}\sum_{j_1=0}^{n-1}e^{-i\lambda_{h_1-p_1}j_1}\sum_{j_2=0}^{n-1}e^{-i\lambda_{h_2-p_2}j_2}\\&=\frac{2\pi}{\sigma^4}\sum_{p=-(n-1)}^{n-1}k_p^2\hat{\gamma}^2_{\varepsilon,p}+\frac{8\pi}{\sigma^4}\sum_{p=1}^{n-1}(k_p^2+k_pk_{n-p})\hat{\gamma}_{\varepsilon,p}\hat{\gamma}_{\varepsilon,n-|p|}+\frac{4\pi}{\sigma^4}\sum_{p=1}^{n-1}(k_{n-p}^2+2k_pk_{n-p})\hat{\gamma}^2_{\varepsilon,p}.\end{aligned}$$

Hence, to show Lemma 2, it is sufficient to prove that

$$\sum_{p=1}^{n-1}(k_{n-p}^2+k_pk_{n-p})\hat{\gamma}^2_{\varepsilon,p}=o_p(n^{-1}p_n^{1/2}) \quad \textbf{(A.29)}$$

and

$$\sum_{p=1}^{n-1}(k_p^2+k_pk_{n-|p|})\hat{\gamma}_{\varepsilon,p}\hat{\gamma}_{\varepsilon,n-|p|}=o_p(n^{-1}p_n^{1/2}). \quad \textbf{(A.30)}$$

In the steps that follow, we will assume that k has unbounded support. If k has bounded support, all terms involving $k_pk_{n-|p|}$ are zero in both (A.29) and (A.30) and the proof is extremely simple. By Assumptions 2a and 3,

$$\begin{aligned}E\left|\frac{n}{p_n^{1/2}}\sum_{p=1}^{n-1}(k_{n-p}^2+k_pk_{n-p})\hat{\gamma}^2_{\varepsilon,p}\right|&\le\frac{n}{p_n^{1/2}}\sum_{p=1}^{n-1}|k_pk_{n-p}|\frac{n-p}{n^2}+\frac{1}{np_n^{1/2}}\sum_{p=1}^{n-1}pk_p^2\\&\le\frac{n}{p_n^{1/2}}\sum_{p=1}^{n/2}|k_pk_{n-p}|\left(\frac{n-p}{n^2}+\frac{p}{n^2}\right)\\&\quad+\frac{1}{np_n^{1/2}}\left(\sum_{p=1}^{p_n}pk_p^2+\sum_{p=p_n+1}^{n-1}pk_p^2\right)\\&\le\frac{1}{p_n^{1/2}}\sum_{p=1}^{n/2}|k_p|\frac{p_n^\delta}{(n-p)^\delta}+O\left(\frac{p_n^{3/2}}{n}\right)=o(1)\end{aligned}$$

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press

because $p_n^{-1}\sum_{p=1}^{n/2}|k_p| = O(1)$. We now verify equation (A.30).

$$E\left(\sum_{p=1}^{n-1}(k_p^2 + k_p k_{n-|p|})\hat{\gamma}_{\varepsilon,p}\hat{\gamma}_{\varepsilon,n-p}\right)^2$$

$$= \sum_p (k_p^2 + k_p k_{n-|p|})^2 E(\hat{\gamma}_{\varepsilon,p}^2\hat{\gamma}_{\varepsilon,n-p}^2)$$

$$+ \sum_{p\neq q}(k_p^2 + k_p k_{n-|p|})(k_q^2 + k_q k_{n-|q|})E(\hat{\gamma}_{\varepsilon,p}\hat{\gamma}_{\varepsilon,n-p}\hat{\gamma}_{\varepsilon,q}\hat{\gamma}_{\varepsilon,n-q}). \quad \textbf{(A.31)}$$

By Lemma 1 on page 186 of Grenander and Rosenblatt (1957), $E(\hat{\gamma}_{\varepsilon,p}^2\hat{\gamma}_{\varepsilon,n-p}^2) = O(n^{-2})$ and $E(\hat{\gamma}_{\varepsilon,p}\hat{\gamma}_{\varepsilon,n-p}\hat{\gamma}_{\varepsilon,q}\hat{\gamma}_{\varepsilon,n-q}) = O(n^{-3})$. Hence, by Assumption 2a, the first term of (A.31) is

$$O\left(\sum_p (k_p^2 + k_p k_{n-|p|})^2 \frac{1}{n^2}\right) = O\left(\frac{p_n}{n^2}\right),$$

and the second term of (A.31) is

$$O_p\left(\frac{1}{n^3}\sum_{p\neq q}(k_p^2 + k_p k_{n-|p|})(k_q^2 + k_q k_{n-|q|})\right) = O_p\left(\frac{p_n^2}{n^3}\right),$$

and the lemma is established. ■

LEMMA 3. *Under the assumptions in Theorem 2,*

$$\left\{\frac{2\pi}{n}\sum_{\ell=0}^{n-1}\tilde{f}_e(\lambda_\ell)\right\}^2 - \left(\frac{2\pi}{\sigma^2}\right)^2\hat{\gamma}_{\varepsilon,0}^2 = O_p(n^{-1}\log^2 n) \quad \textbf{(A.32)}$$

and

$$\left\{\frac{2\pi}{n}\sum_{\ell=0}^{n-1}\hat{f}_{\varepsilon,d}(\lambda_\ell)\right\}^2 = \hat{\gamma}_{\varepsilon,0}^2.$$

Proof. The proof of the second claim of the lemma is contained in the proof of the first claim, which we show subsequently. By (A.1),

$$\frac{2\pi}{n}\sum_{\ell=0}^{n-1}\tilde{f}_e(\lambda_\ell) = \frac{2\pi}{n}\sum_{j=1}^{n-1}\frac{2\pi}{\sigma^2}I_\varepsilon(\lambda_j) + \frac{2\pi}{n}\sum_{j=1}^{n-1}R(\lambda_j).$$

Let $I_{n,\varepsilon}$ be the mean corrected periodogram of ε_t. Then $I_\varepsilon(\lambda_j) = I_{n,\varepsilon}(\lambda_j) = (1/2\pi)\sum\hat{\gamma}_{\varepsilon,h}e^{-i\lambda_j h}$ and $I_{n,\varepsilon}(0) = 0$. We have the first term of the last line,

$$\frac{2\pi}{n}\frac{2\pi}{\sigma^2}\sum_{j=0}^{n-1}I_{n,\varepsilon}(\lambda_j) = \frac{2\pi}{n}\frac{2\pi}{\sigma^2}\frac{1}{2\pi}\sum_{j=0}^{n-1}\sum_{h=-(n-1)}^{n-1}\hat{\gamma}_{\varepsilon,h}e^{-i\lambda_j h}$$

$$= \frac{1}{n}\frac{2\pi}{\sigma^2}\sum_{h=-(n-1)}^{n-1}\hat{\gamma}_{\varepsilon,h}\sum_{j=0}^{n-1}e^{-i\lambda_j h}$$

$$= \frac{2\pi}{\sigma^2}\hat{\gamma}_{\varepsilon,0}.$$

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press

Thus, the LHS of (A.32) is

$$2\cdot\frac{2\pi}{\sigma^2}\hat{\gamma}_{\varepsilon,0}\left(\frac{2\pi}{n}\sum_{j=1}^{n-1}R(\lambda_j)\right)+\left(\frac{2\pi}{n}\sum_{j=1}^{n-1}R(\lambda_j)\right)^2. \tag{A.33}$$

We will show that the second term is $O_p(n^{-2}\log^4 n)$. It follows by Chebyshev's inequality and the fact that $\hat{\gamma}_{\varepsilon,0}=O_p(1)$ that the first term is $O_p(n^{-1}\log^2 n)$. Now

$$E\left(\sum_{j=1}^{n-1}R(\lambda_j)\right)^2=E\left(\sum_{j=1}^{\log^2 n}R(\lambda_j)\right)^2+2E\left(\sum_{j=1}^{\log^2 n}\sum_{h=\log^2 n}^{n-1}R(\lambda_j)R(\lambda_h)\right)$$
$$+E\left(\sum_{j=\log^2 n}^{n-1}R(\lambda_j)\right)^2.$$

By Lemma 5, which follows, the first term is $O(\log^2 n)$, the second term is $O(\log^4 n)$, and the third term is $O(\log^4 n)$, and hence (A.33) is $O_p(n^{-1}\log^2 n)$. ■

LEMMA 4. *Under the assumptions in Theorem 2,*

$$E[R(\lambda_j)I_\varepsilon(\lambda_h)R(\lambda_k)I_\varepsilon(\lambda_\ell)]=O(j^{-d}h^{d-1}k^{-d}\ell^{d-1}\log h\log\ell) \tag{A.34}$$

and

$$E[R(\lambda_j)R(\lambda_h)R(\lambda_k)R(\lambda_\ell)]=O(j^{-d}h^{d-1}k^{-d}\ell^{d-1}\log h\log\ell) \tag{A.35}$$

uniformly for $\log^2 n\le j<h\le n$, $\log^2 n\le k<\ell\le n$.

Proof. The development of this proof closely matches that of Lemma 2 of Hurvich, Deo, and Brodsky (1998). We shall use the following notation:

$$I_j=I(\lambda_j),\qquad f_j=f(\lambda_j),\qquad \text{and } I_{\varepsilon j}=I_\varepsilon(\lambda_j).$$

The LHS of (A.34) is

$$E\left[\left(\frac{I_j}{f_j}-2\pi\sigma^{-2}I_{\varepsilon j}\right)I_{\varepsilon h}\left(\frac{I_k}{f_k}-2\pi\sigma^{-2}I_{\varepsilon k}\right)I_{\varepsilon\ell}\right]$$
$$=E\left[\left(\frac{I_j}{f_j}-1-2\pi\sigma^{-2}I_{\varepsilon j}+1\right)I_{\varepsilon h}\left(\frac{I_k}{f_k}-1-2\pi\sigma^{-2}I_{\varepsilon k}+1\right)I_{\varepsilon\ell}\right]$$
$$=E\left[\left(\frac{I_j}{f_j}-1\right)I_{\varepsilon h}\left(\frac{I_k}{f_k}-1\right)I_{\varepsilon\ell}\right]-E\left[(2\pi\sigma^{-2}I_{\varepsilon j}-1)I_{\varepsilon h}\left(\frac{I_k}{f_k}-1\right)I_{\varepsilon\ell}\right]$$
$$-E\left[\left(\frac{I_j}{f_j}-1\right)I_{\varepsilon h}(2\pi\sigma^{-2}I_{\varepsilon k}-1)I_{\varepsilon\ell}\right]$$
$$+E[(2\pi\sigma^{-2}I_{\varepsilon j}-1)I_{\varepsilon h}(2\pi\sigma^{-2}I_{\varepsilon k}-1)I_{\varepsilon\ell}]. \tag{A.36}$$

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press

Note that the last expectation of (A.36) is zero. Let

$$E\left[\left(\frac{I_j}{f_j}-1\right)I_{\varepsilon h}\left(\frac{I_k}{f_k}-1\right)I_{\varepsilon \ell}\right]=E(\zeta_j\xi_h\zeta_k\xi_\ell)$$

and

$$\boldsymbol{v}=\left(\frac{A_{xj}}{f_j^{1/2}},\frac{B_{xj}}{f_j^{1/2}},\frac{\sqrt{2\pi}}{\sigma}A_{\varepsilon h},\frac{\sqrt{2\pi}}{\sigma}B_{\varepsilon h},\frac{A_{xk}}{f_k^{1/2}},\frac{B_{xk}}{f_k^{1/2}},\frac{\sqrt{2\pi}}{\sigma}A_{\varepsilon\ell},\frac{\sqrt{2\pi}}{\sigma}B_{\varepsilon\ell}\right)'$$
$$=(\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3,\boldsymbol{v}_4,\boldsymbol{v}_5,\boldsymbol{v}_6,\boldsymbol{v}_7,\boldsymbol{v}_8)',$$

where

$$A_{aj}=\frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n}a_t\cos(\lambda_j t),\qquad B_{aj}=\frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n}a_t\sin(\lambda_j t).$$

The vector $\boldsymbol{v}$ has a eight-dimensional multivariate Gaussian distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}$. Define $\boldsymbol{\Psi}=\boldsymbol{\Sigma}^{-1}$. Partition $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ as

$$\boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\quad\text{and}\quad\boldsymbol{\Psi}=\begin{bmatrix}\boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12}\\ \boldsymbol{\Psi}_{21} & \boldsymbol{\Psi}_{22}\end{bmatrix},$$

where Σ_{ij} and Ψ_{ij} are 4×4 matrices. By the formulas for the inverse of a partitioned matrix,

$$\boldsymbol{\Psi}_{11}=\boldsymbol{\Sigma}_{11}^{-1}+\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1},$$
$$\boldsymbol{\Psi}_{12}=-\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1},$$

and

$$\boldsymbol{\Psi}_{22}=(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}.$$

Letting $\mathcal{V}_{aj}=A_{aj}$ or B_{aj}, we have from Lemma 4 of Moulines and Soulier (1999)

$$E\left(\frac{\mathcal{V}_{xj}}{f_j^{1/2}}\frac{\mathcal{V}_{xk}}{f_k^{1/2}}\right)=O(j^{-d}k^{d-1}\log k)\tag{A.37}$$

for $1\le j<k\le n/2$. Following arguments similar to those in this lemma, it can be shown that for $1\le j<k\le n/2$

$$E\left(\frac{\mathcal{V}_{xj}^2}{f_j}\right)=\frac{1}{2}+O(j^{-1}\log j),\tag{A.38}$$

$$E\left(\frac{\mathcal{V}_{\varepsilon j}\mathcal{V}_{xk}}{f_k^{1/2}}\right)=O(j^{-d}k^{d-1}\log k)\quad\text{and}\quad E\left(\frac{\mathcal{V}_{xj}\mathcal{V}_{\varepsilon k}}{f_j^{1/2}}\right)=O(j^{-d}k^{d-1}\log k).\tag{A.39}$$

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press

Letting

$$\mathbf{R} = \boldsymbol{\Sigma} - \frac{1}{2}\mathbf{I}_8 = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix},$$

where $\mathbf{I}_8$ is a 8×8 identity matrix, we see from (A.37)–(A.39) that $\mathbf{R} = o(1)$ for $\log^2 n < j < h \le n/2$, $\log^2 n < k < \ell \le n/2$. By the fact that $(\mathbf{I} + \mathbf{A})^{-1} = \mathbf{I} - (\mathbf{I} + \mathbf{A})^{-1}\mathbf{A}$, we get $\boldsymbol{\Psi} = 2\mathbf{I}_8 - 2\mathbf{R}(\mathbf{I}_8 + 2\mathbf{R})^{-1} = O(1)$. Let

$$\hat{\boldsymbol{\Psi}} = \begin{bmatrix} \boldsymbol{\Psi}_{11} & 0 \\ 0 & \boldsymbol{\Psi}_{22} \end{bmatrix}$$

and define $\bar{\Psi} = \Psi - \tilde{\Psi}$. We have

$$E(\zeta_j \xi_h \zeta_k \xi_\ell) = (2\pi)^4 |\boldsymbol{\Psi}|^{1/2} \int \ldots \int \zeta_j \xi_h \zeta_k \xi_\ell e^{(-1/2)\boldsymbol{v}'\boldsymbol{\Psi}\boldsymbol{v}}\, d\boldsymbol{v}$$

$$= (2\pi)^4 |\boldsymbol{\Psi}|^{1/2} \int \ldots \int \zeta_j \xi_h \zeta_k \xi_\ell e^{(-1/2)\boldsymbol{v}'\hat{\boldsymbol{\Psi}}\boldsymbol{v}}\, d\boldsymbol{v} \tag{A.40}$$

$$+ (2\pi)^4 |\boldsymbol{\Psi}|^{1/2} \int \ldots \int \zeta_j \xi_h \zeta_k \xi_\ell e^{(-1/2)\boldsymbol{v}'\hat{\boldsymbol{\Psi}}\boldsymbol{v}} \{e^{(-1/2)\boldsymbol{v}'\bar{\boldsymbol{\Psi}}\boldsymbol{v}} - 1\}\, d\boldsymbol{v}. \tag{A.41}$$

Let $\boldsymbol{v}_{(jh)} = (v_1, v_2, v_3, v_4)'$, $\boldsymbol{v}_{(k\ell)} = (v_5, v_6, v_7, v_8)'$; the first term of the preceding equation is

$$(2\pi)^4 |\boldsymbol{\Psi}|^{1/2} \iiiint \zeta_j \xi_h e^{(-1/2)\boldsymbol{v}'_{(jh)}\boldsymbol{\Psi}_{11}\boldsymbol{v}_{(jh)}}\, d\boldsymbol{v}_{(jh)} \iiiint \zeta_k \xi_\ell e^{(-1/2)\boldsymbol{v}'_{(k\ell)}\boldsymbol{\Psi}_{22}\boldsymbol{v}_{(k\ell)}}\, d\boldsymbol{v}_{(k\ell)}. \tag{A.42}$$

The first quadruple integral of (A.42) is

$$\iiiint \zeta_j \xi_h e^{(-1/2)\boldsymbol{v}'_{(jh)}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{v}_{(jh)}} e^{(-1/2)\boldsymbol{v}'_{(jh)}\mathbf{M}_{11}\boldsymbol{v}_{(jh)}}\, d\boldsymbol{v}_{(jh)}, \tag{A.43}$$

where

$$\mathbf{M}_{11} = \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}.$$

Let τ_{11} be the largest absolute entry of $\mathbf{M}_{11}$. Because $|e^u - 1| \le |u| e^{|u|}$ for all u,

$$e^{(-1/2)\boldsymbol{v}'_{(jh)}\mathbf{M}_{11}\boldsymbol{v}_{(jh)}} = 1 + O\{\tau_{11}\|\boldsymbol{v}_{(jh)}\|^2 e^{(3/2)\tau_{11}\|\boldsymbol{v}_{(jh)}\|^2}\}.$$

Thus (A.43) is equal to

$$\iiiint \zeta_j \xi_h e^{(-1/2)\boldsymbol{v}'_{(jh)}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{v}_{(jh)}}\, d\boldsymbol{v}_{(jh)}$$

$$+ O\left\{ \iiiint |\zeta_j \xi_h| \tau_{11} \|\boldsymbol{v}_{(jh)}\|^2 e^{(-1/2)\boldsymbol{v}'_{(jh)}(\boldsymbol{\Sigma}_{11}^{-1} - 3\tau_{11}\mathbf{I}_4)\boldsymbol{v}_{(jh)}}\, d\boldsymbol{v}_{(jh)} \right\}. \tag{A.44}$$

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press

The second term is $O(\tau_{11}) = O(j^{-2d}k^{2d-2}\log^2 k\mathbf{1}_{(j<k)} + k^{-2d}j^{2d-2}\log^2 k\mathbf{1}_{(j>k)})$ by (A.37)–(A.39). Note that

$$\boldsymbol{\Sigma}_{11}^{-1} = 2\mathbf{I}_4 - 2\mathbf{R}_{11}(\mathbf{I}_4 + 2\mathbf{R}_{11})^{-1} = 2\mathbf{I}_4 + o(1).$$

Let η_{11} be the largest absolute entry of $2\mathbf{R}_{11}(\mathbf{I}_4 + 2\mathbf{R}_{11})^{-1}$,

$$e^{(1/2)\boldsymbol{v}'_{(jh)}(2\mathbf{R}_{11}(\mathbf{I}_4+2\mathbf{R}_{11})^{-1})\boldsymbol{v}_{(jh)}} = 1 + O\{\eta_{11}\|\boldsymbol{v}_{(jh)}\|^2 e^{(3/2)\eta_{11}\|\boldsymbol{v}_{(jh)}\|^2}\}.$$

Thus the first term of (A.44) is

$$\begin{aligned}
&\iiiint \zeta_j \xi_h e^{(-1/2)\boldsymbol{v}'_{(jh)}2\mathbf{I}_4\boldsymbol{v}_{(jh)}}\, d\boldsymbol{v}_{(jh)} \\
&\qquad + O\left\{\iiiint |\zeta_j\xi_h|\eta_{11}\|\boldsymbol{v}_{(jh)}\|^2 e^{(1/2)\boldsymbol{v}'_{(jh)}((2-3\eta_{11})\mathbf{I}_4)\boldsymbol{v}_{(jh)}}\, d\boldsymbol{v}_{(jh)}\right\} \\
&\quad = \iint \zeta_j e^{(-1/2)\boldsymbol{v}'_{(j)}2\mathbf{I}_2\boldsymbol{v}_{(j)}}\, d\boldsymbol{v}_{(j)} \iint \zeta_j e^{(-1/2)\boldsymbol{v}'_{(h)}2\mathbf{I}_2\boldsymbol{v}_{(h)}}\, d\boldsymbol{v}_{(h)} \\
&\qquad + O\left\{\iiiint |\zeta_j\xi_h|\eta_{11}\|\boldsymbol{v}_{(jh)}\|^2 e^{(1/2)\boldsymbol{v}'_{(jh)}((2-3\eta_{11})\mathbf{I}_4)\boldsymbol{v}_{(jh)}}\, d\boldsymbol{v}_{(jh)}\right\}.
\end{aligned}$$

The first term of the RHS of the preceding equation is zero because the first double integral is the expectation of ζ_j assuming the covariance matrix is $0.5\mathbf{I}_4$. The second term is $O(\eta_{11}) = O(j^{-d}h^{d-1}\log h)$. We have shown that the first quadruple integral of (A.42) is $O(j^{-d}h^{d-1}\log h + j^{-2d}k^{2d-2}\log^2 k\mathbf{1}_{(j\le k)} + j^{2d-2}k^{-2d}\log^2 j\mathbf{1}_{(k\le j)})$. It can be shown in the same fashion that the second quadruple integral of (A.42) is $O(k^{-d}\ell^{d-1}\log \ell + j^{-2d}k^{2d-2}\log^2 k\mathbf{1}_{(j\le k)} + j^{2d-2}k^{-2d}\log^2 j\mathbf{1}_{(j>k)})$. Hence (A.40) is $O(j^{-d}h^{d-1}k^{-d}\ell^{d-1}\log h\log \ell)$.

Now we consider (A.41). By the mean value theorem, $|e^u - 1 - u| \le \frac{1}{2}u^2 e^{|u|}$ for all u. Thus

$$e^{(-1/2)\boldsymbol{v}'\bar{\boldsymbol{\Psi}}\boldsymbol{v}} - 1 = -\frac{1}{2}\boldsymbol{v}'\bar{\boldsymbol{\Psi}}\boldsymbol{v} + O(\tau^2\|\boldsymbol{v}\|^4 e^{2\tau\|\boldsymbol{v}\|^2}),$$

where τ is the largest absolute entry of $\bar{\boldsymbol{\Psi}}$. Note that $\tau^2 = O(j^{-2d}k^{2d-2}\log^2 k\mathbf{1}_{(j\le k)} + j^{2d-2}k^{-2d}\log^2 j\mathbf{1}_{(k\le j)})$. Hence (A.41) is

$$\begin{aligned}
&(2\pi)^4|\boldsymbol{\Psi}|^{1/2}\int\cdots\int \zeta_j\xi_h\zeta_k\xi_\ell - \frac{1}{2}\boldsymbol{v}'\bar{\boldsymbol{\Psi}}\boldsymbol{v}e^{(-1/2)\boldsymbol{v}'\tilde{\boldsymbol{\Psi}}\boldsymbol{v}}\, d\boldsymbol{v} \\
&\qquad + O\left\{\tau^2\int\cdots\int |\zeta_j\xi_h\zeta_k\xi_\ell|\,\|\boldsymbol{v}\|^4 e^{-(1/2)\boldsymbol{v}'(\tilde{\boldsymbol{\Psi}}-4\tau\mathbf{I}_8)\boldsymbol{v}}\, d\boldsymbol{v}\right\}.
\end{aligned}$$

The second term is $O(\tau^2)$. The first term is the linear combination of $E_{\tilde{\Psi}}[\zeta_j\xi_h\zeta_k\xi_\ell A_jA_k]$, $E_{\tilde{\Psi}}[\zeta_j\xi_h\zeta_k\xi_\ell A_jB_k]$, $E_{\tilde{\Psi}}[\zeta_j\xi_h\zeta_k\xi_\ell A_jA_\ell]$, $E_{\tilde{\Psi}}[\zeta_j\xi_h\zeta_k\xi_\ell A_jB_\ell],\ldots,$etc., where $E_{\tilde{\Psi}}$ denotes the expectation assuming that $\boldsymbol{v}$ is multivariate normal with mean zero and covariance matrix $\tilde{\boldsymbol{\Psi}}$. Note that $\mathrm{cov}(\boldsymbol{v}) = \tilde{\boldsymbol{\Psi}}$ implies that the vectors (A_j, B_j, A_h, B_h), $(A_k, B_k, A_\ell, B_\ell)$ are independent. Thus, for example, $E_{\tilde{\Psi}}[\zeta_j\xi_h\zeta_k\xi_\ell A_jA_k] = E_{\tilde{\Psi}}[\zeta_j\xi_hA_j]E_{\tilde{\Psi}}[\zeta_k\xi_\ell A_k]$, and both of these expectations are zero because the $\zeta_j\xi_h$ and $\zeta_k\xi_\ell$ are even functions of

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press

(A_j, B_j, A_h, B_h), respectively, and because the densities for (A_j, B_j, A_h, B_h) and $(A_k, B_k, A_\ell, B_\ell)$ are also even functions. We have shown that (A.41) is $O(\tau^2) = O(j^{-2d}k^{2d-2}\log^2 k\mathbf{1}_{(j\le k)} + j^{2d-2}k^{-2d}\log^2 j\mathbf{1}_{(k\le j)})$. Hence

$$E\left[\left(\frac{I_j}{f_j}-1\right)I_{\varepsilon h}\left(\frac{I_k}{f_k}-1\right)I_{\varepsilon \ell}\right] = O(j^{-d}h^{d-1}k^{-d}\ell^{d-1}\log h\log \ell).$$

It can be shown in a similar way that the rest of the second and the third expectations of (A.36) are both $O(j^{-d}h^{d-1}k^{-d}\ell^{d-1}\log h\log\ell)$ uniformly in $\log^2 n \le j < h \le n/2$, $\log^2 n \le k < \ell \le n/2$. The order in (A.35) can be derived following the same lines as previously. ∎

LEMMA 5. *Under the assumptions of Theorem 2,*

$$E[R(\lambda_j)R(\lambda_h)] = O(j^{-1}h^{-1}\log h\log j + j^{-2d}h^{2d-2}\log^2 h)$$

and

$$E[R^2(\lambda_j)] = O(j^{-1}\log j)$$

uniformly for $\log^2 n \le j < h \le n$. *Also* $\max_{1\le j\le n} E[R^2(\lambda_j)] < \infty$.

The proof of the first two bounds stated in this lemma is similar to that of Lemma 4. The last bound is obtained by using the bounds (A.37)–(A.39) and the Gaussianity of the observations.

LEMMA 6. *Let* $g(\lambda)$ *be defined as (A.11). Then, under Assumption 7,*

$$\sum_{h=1}^{n-1} g(\lambda_h)e^{-i\lambda_h m} = O(n^{\delta}) \quad \text{if } m = 0,$$
$$= O(n^{1+\delta}m^{-1}) \quad \text{if } m \ne 0.$$

Proof. We shall prove the lemma by showing that

$$\frac{2\pi}{n}\sum_{h=1}^{n-1} g(\lambda_h)e^{-i\lambda_h m} = O(n^{-1+\delta}) \quad \text{if } m = 0,$$
$$= O(n^{\delta}m^{-1}) \quad \text{if } m \ne 0 \quad \text{and} \quad |m| \le n. \tag{A.45}$$

We first derive the result for $m = 0$. Note that

$$\int_0^{2\pi} g(\lambda)\,d\lambda = 0.$$

Hence, the LHS of (A.45) is

$$\frac{2\pi}{n}\sum_{h=1}^{n-1} g(\lambda_h) = \left(\frac{2\pi}{n}\sum_{h=1}^{n-1} g(\lambda_h) - \int_0^{2\pi} g(\lambda)\,d\lambda\right)$$
$$= \sum_{h=1}^{n-1}\int_{\lambda_{h-1}}^{\lambda_h}(g(\lambda_h)-g(\lambda))\,d\lambda - \int_{\lambda_{n-1}}^{\lambda_n} g(\lambda)\,d\lambda$$
$$= \sum_{h=1}^{n-1} g'(\lambda_{\tilde h})\int_{\lambda_{h-1}}^{\lambda_h}(\lambda_h-\lambda)\,d\lambda - \int_{\lambda_0}^{\lambda_1} g(\lambda)\,d\lambda,$$

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press

where $\lambda_{h-1} < \lambda_{\tilde{h}} < \lambda_h$ and we use the fact that $g(\lambda)$ is symmetric around $\pi/2$. By Assumption 7, the last equation is

$$O\left(\sum_{h=1}^{n} \lambda_h^{-1-\delta} \cdot \frac{1}{2}\left(\frac{2\pi}{n}\right)^2 + \lambda_1^{1-\delta}\right) = O(n^{-1+\delta}).$$

For $m \neq 0$, we have by summation by parts

$$\begin{aligned}
&\frac{2\pi}{n} \sum_{h=1}^{n-1} g(\lambda_h) e^{i\lambda_h m} \\
&\quad = \frac{2\pi}{n} \sum_{h=1}^{n-2} (g(\lambda_h) - g(\lambda_{h+1})) \sum_{\ell=1}^{h} e^{i\lambda_\ell m} + \frac{2\pi}{n} g(\lambda_{n-1}) \sum_{\ell=1}^{n-1} e^{i\lambda_\ell m} \\
&\quad = \frac{2\pi}{n} \sum_{h=1}^{n-2} g'(\lambda_{\tilde{h}})(\lambda_h - \lambda_{h+1}) \sum_{\ell=1}^{h} e^{i\lambda_\ell m} + \frac{2\pi}{n} g(\lambda_{n-1})(-1).
\end{aligned}$$

Because $\sum_{\ell=a}^{b} e^{i\lambda\ell} = \sum_{\ell=1}^{b} e^{i\lambda\ell} - \sum_{\ell=1}^{a-1} e^{i\lambda\ell} = O(\lambda^{-1})$ uniformly in a and b for $0 < \lambda < \pi$ (see the proof of Lemma 1), this is

$$O\left(\frac{1}{n} \sum_{h=1}^{n-2} \lambda_h^{-1-\delta} \lambda_1 \lambda_m^{-1} + \frac{1}{n^{1-\delta}}\right) = O\left(\frac{n^\delta}{m}\right). \qquad \blacksquare$$

https://doi.org/10.1017/S0266466604202067 Published online by Cambridge University Press