NON-GAUSSIAN LOG-PERIODOGRAM REGRESSION

CARLOS VELASCO
Universidad Carlos III de Madrid

We show the consistency of the log-periodogram regression estimate of the long memory parameter for long range dependent linear, not necessarily Gaussian, time series when we make a pooling of periodogram ordinates. Then, we study the asymptotic behavior of the tapered periodogram of long range dependent time series for frequencies near the origin, and we obtain the asymptotic distribution of the log-periodogram estimate for possibly non-Gaussian observation when the tapered periodogram is used. For these results we rely on higher order asymptotic properties of a vector of periodogram ordinates of the linear innovations. Finally, we assess the validity of the asymptotic results for finite samples via Monte Carlo simulation.

1. INTRODUCTION

Long memory or long range dependent observations have been found in many fields of research (e.g., Robinson, 1994c; Beran, 1994). In this paper we consider semiparametric statistical inference for long range stationary dependent time series. In particular we concentrate on the estimate of the memory parameter based on the regression on the logarithm of the periodogram at Fourier frequencies close to the origin. This estimate, proposed initially by Geweke and Porter-Hudak (1983), has been very popular among practitioners because of its intuitive and computational appeal. At the same time, properties of maximum likelihood methods have been analyzed extensively for parametric models of long range dependence Gaussian and linear processes (see, e.g., Fox and Taqqu, 1986; Dahlhaus, 1989; Giraitis and Surgailis, 1990), obtaining equivalent efficiency results to the weak dependence situation. However, this approach involves a complete specification of the dynamics of the process, and if we are only interested in the estimation of long range dependence characteristics, semiparametric and nonparametric setups can be robust against any misspecification of the short run behavior of the time series.

Semiparametric models for long memory focus on some properties of the autocovariance sequence (hyperbolic decay) or of the spectral density (singularity at

the zero frequency). They are semiparametric because they do not make explicit assumptions on the behavior of the autocovariances at short lags or on the spectral density apart from the origin. We set our conditions in the frequency domain in terms of the spectral density since they are much neater and cover a broader range of possibilities. We assume that the spectral density satisfies

$$f(\lambda) \sim C\lambda^{-2d}, \qquad \text{as } \lambda \to 0^+, \tag{1}$$

where $d \in (0, \frac{1}{2})$ is the self-similarity parameter that governs the degree of strong dependence of the series. This is the interval of values of d for which the series exhibits long range dependence and is stationary. The basis for the log-periodogram regression estimate is the least squares estimation of the linear relationship implicit in expression (1) between the spectral density and the frequency in log-log coordinates with slope $-2d$ when the spectral density is estimated by the periodogram at Fourier frequencies close to the origin.

Robinson (1994a, 1995a, 1995b) and Lobato and Robinson (1996) have used similar assumptions to the ones we employ here to study the asymptotic behavior of several semiparametric estimates of d. Robinson (1995a) justified a modified version of the procedure proposed by Geweke and Porter-Hudak (1983), including multivariate and pooled periodogram versions. He proved the consistency and asymptotic normality of this estimate for Gaussian vector time series, which may seem very restrictive in view of the weak distribution assumptions under which the other estimates were investigated. In this paper we extend his consistency results for linear processes not necessarily Gaussian. To obtain an asymptotically normal estimate we need to taper the data to reduce the leakage in the periodogram ordinates from the zero frequency pole, and we need to pool the contribution for several adjacent frequencies to obtain better behaved regressors.

The paper is organized as follows. In the next section we present our main assumptions and definitions, discuss related references, and obtain the consistency of the log-periodogram estimate of d. The effects of tapering are discussed in Section 3. Section 4 is dedicated to the asymptotic normality of the estimate of d when we use the tapered periodogram. Finally, we report the results of a brief simulation exercise centered on the tapering and pooling techniques analyzed under different distributional settings. All the proofs and some technical Lemmas are given in the Appendix.

2. CONSISTENCY OF THE LOG-PERIODOGRAM REGRESSION ESTIMATE

Let $\{X_t, t = 1, 2, \dots\}$ be a covariance stationary process with spectral density satisfying (1). Given an observable sequence $X_t, t = 1, \dots, N$, we introduce the discrete Fourier transform at the frequency $\lambda_j = 2\pi j/N$, j integer,

$$w(\lambda_j) = (2\pi N)^{-1/2} \sum_{t=1}^{N} X_t e^{it\lambda_j},$$

and the periodogram is $I(\lambda_j) = |w(\lambda_j)|^2$. Define for $J = 1, 2, \ldots$, fixed, and some positive integers ℓ and m (assuming $(m - \ell)/J$ to be an integer),

$$Y_k^{(J)} = \log\left(\sum_{j=1}^{J} I(\lambda_{k+j-J})\right), \qquad k = \ell + J, \ell + 2J, \ldots, m.$$

The estimate of the memory parameter d considered in Robinson (1995a) is

$$\hat{d} = \left(\sum_k \Lambda_k^2\right)^{-1} \sum_k \Lambda_k Y_k^{(J)},$$

where $\Lambda_k = z_k - \bar{z}$, $z_k = -2\log\lambda_k$, and $\bar{z} = \{J/(m-\ell)\}\sum_k z_k$. Here m is an integer smaller than N and ℓ is a user-chosen trimming number. In the asymptotics both numbers tend to infinity with the sample size N but more slowly. We suppress in the notation reference to N or J.

We could substitute the (pooled) periodogram by nonparametric smoothed consistent estimates of the spectral density as was done in Velasco (1997) for long range dependent series or in, e.g., Hassler (1993) for antipersistent series ($d < 0$). However, when we consider in $\hat{d}$ fixed averages of the periodogram the analysis is much more complicated than in that situation. Here, we have to deal with the logarithm of a random variable that is not converging asymptotically to any constant and that can take values arbitrarily close to zero. Nonlinear functions (the logarithm in particular) of the periodogram of stationary sequences have been considered under different setups (see, e.g., Hannan and Nicholls, 1977; Taniguchi, 1979; Chen and Hannan, 1980; von Sachs, 1994; Janas and von Sachs, 1993; Comte and Hardouin, 1995; and the references given in these sources). These works assume Gaussianity to obtain the main results, except Chen and Hannan, and Janas and von Sachs, in which the researchers work with linear process conditions.

These last two references use higher order properties of the asymptotic distribution of the periodogram. Janas and von Sachs mainly applied the results for weakly dependent sequences of Götze and Hipp (1983), making it almost impossible to relax their assumptions for long range dependence situations. Instead, the approach of Chen and Hannan (1980) is based on the factorization of the periodogram of the observable sequence in the transfer function of the linear filter, times the periodogram of the independent and identically distributed (i.i.d.) innovations, plus a stochastic error term. The magnitude of this error depends on the smoothness of the spectral density and on the number of moments assumed for the innovations. Obviously the conditions they assumed ($\sum |j|^\delta |\alpha_j| < \infty$, $\delta > \frac{1}{2}$, see Assumption 3, which follows), rule out any long memory behavior or any singularity in the spectral density of X_t, but their results are based mainly on the properties of the periodogram of the i.i.d. innovations sequence, for which we assume the same set of conditions as in their Theorem 2 (see Assumption 4, which follows).

A related approach was used by Comte and Hardouin (1995) in a long memory environment but assuming Gaussianity. We use one of their ideas to avoid a modification of the estimate of d in the same spirit as the one proposed by Chen and Hannan (1980) for a different statistic to account for values of $I(\lambda_j)/f(\lambda_j)$ that are too small. Here, instead of redefining the periodogram with a truncation, we use an average of periodogram ordinates. Then we can use their higher order asymptotic approach and the long range dependence results of Robinson (1995b) to approximate the periodogram of X_t by that of the linear i.i.d. innovations times the long memory transfer function.

Tukey (1967) proposed tapering as an effective bias reduction technique for spectral inference to avoid leakage from remote frequencies. Under additional smoothness conditions on the behavior of the spectral density at the origin, we study the asymptotic effect of tapering the data prior to calculating the periodogram. We obtain the asymptotic normality of the estimate $\hat{d}$ based on the tapered periodogram. Von Sachs (1994) and Janas and von Sachs (1993) also used tapering for nonlinear functions of the periodogram, but their results do not apply to long memory time series. Robinson (1986), Dahlhaus (1988), and Hurvich and Ray (1995), among others, have proposed this technique to reduce the bias of several statistics when possibly nonstationary behaviors of the observed time series are suspected.

We now introduce some assumptions about the behavior of the spectral density around the origin, following Robinson (1995a, 1995b), but do not consider negative values of d. Later we strengthen these assumptions to obtain further results.

Assumption 1. X_t is covariance stationary and for some $d \in [0,\frac{1}{2})$, $\alpha \in (0,2]$ and $0 < G < \infty$,

$$f(\lambda) = G\lambda^{-2d} + O(\lambda^{\alpha-2d}), \qquad \text{as } \lambda \to 0^+.$$

Assumption 2. In a neighborhood $(0,\varepsilon)$ of the origin, $f(\lambda)$ is differentiable and

$$\left| \frac{d}{d\lambda} \log f(\lambda) \right| = O(\lambda^{-1}), \qquad \text{as } \lambda \to 0^+.$$

These conditions are standard in long memory research and are satisfied with $\alpha = 2$ by fractional ARIMA models, for which

$$f(\lambda) = \frac{\sigma^2}{2\pi} |1 - e^{i\lambda}|^{-2d} \left| \frac{a(e^{i\lambda})}{b(e^{i\lambda})} \right|^2, \qquad -\pi < \lambda \le \pi,$$

where $\sigma^2 > 0$ and a and b are polynomials of finite degree having no zeros in or on the unit circle and by the fractional Gaussian noise model with autocovariance sequence given by

$$\gamma(j) = \frac{\gamma(0)}{2} (|j+1|^{2d+1} - 2|j|^{2d+1} + |j-1|^{2d+1}), \qquad j = \pm 1, \pm 2, \ldots$$

Instead of Gaussianity we introduce a fourth order stationary linear process condition, with filter coefficients compatible with Assumptions 1 and 2:

Assumption 3. X_t satisfies

$$X_t = \sum_{j=0}^{\infty} \alpha_j \epsilon_{t-j}, \qquad \sum_{j=0}^{\infty} \alpha_j^2 < \infty, \qquad \left| \frac{d}{d\lambda} \alpha(\lambda) \right| = 0(\lambda^{-1} |\alpha(\lambda)|),$$

$$\alpha(\lambda) = \sum_{j=0}^{\infty} \alpha_j e^{ij\lambda},$$

as $\lambda \to 0^t$, where the ϵ_t are i.i.d. with $E[\epsilon_t] = 0$, $E[\epsilon_t^2] = 1$, and $E[\epsilon_t^4] < \infty$.

We assume zero mean for the series X_t without loss of generality, because we omit the periodogram at zero frequency in the definition of $\hat{d}$. Four bounded moments are enough for all our consistency results. We introduce the next assumption following Chen and Hannan (1980).

Assumption 4. ϵ_t has characteristic function $\hat{Q}(\theta) = E[e^{i\theta\epsilon_t}]$ satisfying

$$\sup_{|\theta| \geq \theta_0} |\hat{Q}(\theta)| = \delta(\theta_0) < 1, \qquad \forall \theta_0 > 0, \quad \text{and} \quad \int_{-\infty}^{\infty} |\hat{Q}(\theta)|^p \, d\theta < \infty,$$

for some integer $p > 1$.

The conditions of Assumption 4 are needed to prove the validity of an asymptotic approximation for the probability density of the discrete Fourier transform of the innovations ϵ_t (see Lemma 2 in the Appendix). The first line is a Cramér condition. The second condition is used to approximate the probability density, and it would not be necessary to approximate the probability distribution function. It implies that the probability distribution of ϵ_t has a bounded continuous density (see, e.g., Feller, 1971, Theorem 3, p. 509).

We now proceed to show the consistency of the estimate $\hat{d}$ when finite averages (for J fixed) of the periodogram of X_t are used under the linear process condition of Assumption 3. We approximate the logarithm of the periodogram of X_t by that of ϵ_t, times the transfer function, the error depending on the properties of the filter $\{\alpha_j\}$ and on the distribution of the linear innovations ϵ_t, but special care is needed because of the singularity of the logarithm function at the origin. However we are only able to deal with the case $J \geq 3$. The reason for this limitation is the following. To approximate the periodogram of X_t by that of ϵ_t we need to consider the inverse moments of the periodogram of ϵ_t at certain Fourier frequencies. The average of J periodogram ordinates of an i.i.d. sequence is asymptotically distributed as a χ^2_{2J} (up to constants). The key point is that if $Z \sim \chi^2_{2J}$, then $E[Z^{a-J}] < \infty$ for $0 < a < J$ (see Lemma 1, which follows). Of course, to approximate the moments of a random variable we need something more than its asymptotic distribution. That is why we approximate

the probability density of the Fourier transform of ϵ_t under the regularity conditions of Assumption 4. We conjecture that a related argument could be used to construct a proof for $J \leq 2$.

We introduce some more notation. Write $I_j = I(\lambda_j)$ and $f_j = f(\lambda_j)$ and, for the periodogram of the ϵ_t innovation sequence, $I_{\epsilon,j} = I_\epsilon(\lambda_j)$. Let J be a fixed positive integer. Define

$$\bar{I}_k = \sum_{j=1}^{J} I_{k+j-J}, \qquad \bar{I}_{\epsilon k} = \sum_{j=1}^{J} I_{\epsilon,k+j-J}, \qquad k = \ell + J, \ell + 2J, \ldots, m.$$

We suppress the dependence on J in the notation $\bar{I}_k$ and $\bar{I}_{\epsilon k}$.

We can write in the same spirit as in Comte and Hardouin (1995),

$$\log \bar{I}_k = \log f_k + \log 2\pi \bar{I}_{\epsilon,k} + \log \left(1 + \frac{\delta_k}{2\pi \bar{I}_{\epsilon,k}} \right), \tag{2}$$

where the error term $\delta_k = f_k^{-1} \bar{I}_k - 2\pi \bar{I}_{\epsilon,k}$ is analyzed in Lemma 3 making direct use of some results of Robinson (1995a, 1995b) based on the characteristics of the linear filter $\alpha(\lambda)$ under Assumptions 1 and 2. In fact, when approximating the observed pooled periodogram $\bar{I}_k$ by that of the innovations $\bar{I}_{\epsilon,k}$ we have two types of errors in δ_k; one is the bias due to the average across J frequencies, $f_k^{-1} I_j - f_j^{-1} I_j$, and the other is the stochastic error $f_j^{-1} I_j - 2\pi I_{\epsilon,j}$, for $k + 1 - J \leq j \leq k$. Then, for ℓ increasing with N, we show that both errors are negligible in (2) using a truncation argument similar to that in Chen and Hannan (1980), without need of modifying the definition of the estimate. Here the main problem is the small values of $\bar{I}_k$ and $\bar{I}_{\epsilon,k}$, which cause problems because of the logarithm. The following lemma is useful to show that these values will not occur too often if sufficient pooling is used, i.e., J is large enough.

LEMMA 1. *Under Assumption* 4, *for* $J \geq 1$, $k \neq 0 \pmod{N}$, *for all* $0 < a < J$,

$$E[\{\bar{I}_\epsilon(\lambda_k)\}^{a-J}] < \infty.$$

We now give sufficient conditions for the consistency of $\hat{d}$ for linear, possibly non-Gaussian, series under conditions 3 and 4. First, we introduce the following condition on the bandwidth numbers.

Assumption 5. As $N \to \infty$, for some $b > 0$, $J \geq 3$,

$$\frac{m^{1/2}}{\ell^{J/(J+2)-b}} + \frac{\ell (\log N)^2}{m} + \frac{m}{N} \to 0.$$

This assumption imposes quite a high trimming rate in $\hat{d}$, because for $J = 3$, ℓ has to grow faster than $m^{5/6}$. If J is large then ℓ only needs to grow slightly faster than $m^{1/2}$, which is the condition required by Robinson (1995b) for consistency and asymptotic normality of $\hat{d}$. Then our first result is the following theorem.

THEOREM 1. *Under Assumptions* 1–5, *with* $J \geq 3$, $\hat{d} \to_P d$.

The question of whether the asymptotic distribution of $\hat{d}$ remains the same once Gaussianity is no longer assumed is of evident interest, but the previous results are not enough to show that. First, it is necessary to improve the approximation results between the periodogram ordinates of the observable sequence and of the innovations. Then, a central limit theorem has to be proved for the random variable

$$\xi_N = \left(\sum_k \Lambda_k^2 \right)^{-1} \sum_k \Lambda_k \log 2\pi \bar{I}_{\epsilon,k}$$

that appears in the proof of Theorem 1. In the next section we propose tapering as a way of obtaining the previously mentioned approximation, and then we investigate the asymptotic distribution of $\hat{d}$.

3. TAPERED DISCRETE FOURIER TRANSFORM

In the previous section, we obtained the consistency of $\hat{d}$ with a pooling of periodogram ordinates. However, the bias of the periodogram makes it impossible to obtain the asymptotic distribution from the proof of Theorem 1, unless we translate the regression to frequencies $\lambda_j, \ell + 1 \leq j \leq \ell + m$, with $m/\ell \to 0$ as $N \to \infty$. A similar problem was observed for Gaussian series under stronger conditions by Comte and Hardouin (1995, Propositions 1 and 3).

Tapering the data is a well known method to reduce the leakage in the periodogram from other frequencies, and in this case it is a very effective way of reducing the bias of the periodogram. Tapering gives more relevance to observations in the central part of the observed sequence, downweighting those observations at both extremes by means of a smooth positive function h_t of time, $t = 1, \ldots, N, 0 \leq h_t \leq 1$. We need to strengthen Assumptions 1 and 2 on f to use these properties of tapering as in Assumption 3 of Robinson (1994b), with $1 < \alpha \leq 2$:

Assumption 6. To further Assumptions 1 and 2, denoting $g(\lambda) = G|\lambda|^{-2d}$, we assume, for some $0 < E_\alpha < \infty$, that as $\lambda \to 0^+$,

$$\frac{f(\lambda)}{g(\lambda)} = 1 + E_\alpha \times \lambda^\alpha + o(\lambda^\alpha), \qquad 1 < \alpha \leq 2.$$

Assumption 6 is satisfied with $\alpha = 2$ by the fractional ARIMA and fractional noise models. This condition is equivalent to assuming that $f(\lambda) = g(\lambda)h(\lambda)$, with $h(0) = 1$ and where $h(\lambda)$ is even and differentiable with derivative in Lip$(\alpha - 1)$ for $1 < \alpha \leq 2$ around the origin.

Therefore, this assumption is satisfied at frequencies $\lambda_j = 2\pi j/N, j = 1, 2, \ldots, m$, for N big enough. Then, for frequencies $|\lambda| \leq \lambda_j/2$, we can expand f in this way:

$$f(\lambda_j - \lambda) = f(\lambda_j) - \lambda \times f'(\lambda_j) + O(\lambda_j^{-2d-\alpha}|\lambda|^\alpha), \tag{3}$$

where the derivative of f satisfies $f'(\lambda_j) = O(\lambda_j^{-2d-1})$. This can be seen heuristically in the following way: making a Taylor expansion of f around λ_j, we are led to study the difference for $|\theta| \leq 1$ and $\lambda \in [\lambda_j/2, 3\lambda_j/2]$,

$$|\lambda f'(\lambda_j - \theta\lambda) - \lambda f'(\lambda_j)|. \tag{4}$$

Now, as we can write $f' = h'g + g'h$, this is not bigger than the sum of the differences (taking $\theta = 1$ to simplify notation, w.l.o.g.)

$$|h'(\lambda_j - \lambda)g(\lambda_j - \lambda) - h'(\lambda_j)g(\lambda_j)|$$
$$+ |h(\lambda_j - \lambda)g'(\lambda_j - \lambda) - h(\lambda_j)g'(\lambda_j)|, \tag{5}$$

times $|\lambda|$. The first term in (5) is bounded by

$$|h'(\lambda_j - \lambda) - h'(\lambda_j)||g(\lambda_j - \lambda)| + |g(\lambda_j - \lambda) - g(\lambda_j)||h'(\lambda_j)|,$$

and using the mean value theorem and that for these values of λ, $g = O(\lambda_j^{-2d}), g' = O(\lambda_j^{-2d-1})$, $g'' = O(\lambda_j^{-2d-2})$, $h = O(1)$, $h' = O(\lambda_j^{\alpha-1})$, and $|h'(\lambda_j - \lambda) - h'(\lambda_j)| = O(|\lambda|^{\alpha-1})$, this is bounded by

$$O(\lambda_j^{-2d}\lambda^{\alpha-1} + \lambda_j^{-2d-1}\lambda_j^{\alpha-1}\lambda) = O(\lambda_j^{-2d}\lambda^{\alpha-1}),$$

because $\alpha \in (1,2]$. Similarly, the second term in (5) is bounded by

$$|g'(\lambda_j - \lambda) - g'(\lambda_j)||h(\lambda_j - \lambda)| + |h(\lambda_j - \lambda) - h(\lambda_j)||g'(\lambda_j)|$$
$$= O(\lambda_j^{-2d-2}\lambda + \lambda_j^{\alpha-1}\lambda_j^{-2d-1}\lambda) = O(\lambda_j^{-2d-\alpha}\lambda^{\alpha-1}).$$

Then (4) is easily seen to be $O(\lambda_j^{-2d-\alpha}\lambda^{\alpha})$, multiplying the last two bounds by λ.

We consider the full cosine bell or hanning taper, as suggested by Hurvich and Ray (1995) for a related problem. A generalization of the results in this section can be possible for many smooth data tapers (see, e.g., Velasco, 1999, in a related context), but the hanning tapering has some desirable features that we use later to find the asymptotic distribution of the log-periodogram regression estimate with tapered observations. Tapering allows us to reduce the bias of the periodogram for frequencies close to the origin if we assume a spectral density smooth enough at these frequencies (i.e., $\alpha > 1$ in Assumption 6). Also, because the tapered Fourier transform can be written down as a linear combination of three (raw) Fourier transforms, we can still use the results of Chen and Hannan (1980) as before, with minor modifications.

The tapered discrete Fourier transform is defined as

$$w^T(\lambda_j) = \left(2\pi \sum h_t^2\right)^{-1/2} \sum_{t=1}^{N} h_t X_t \exp(i\lambda_j t),$$

where $h_t = \frac{1}{2}(1 - \cos[2\pi t/N])$ and the sum of the squared taper weights is $\sum h_t^2 = 3N/8$. This is called the *asymmetric* version of the cosine bell by Percival and Walden (1993, p. 325). The usual discrete Fourier transform $w(\lambda)$ is obtained setting $h_t \equiv 1, \forall t$.

Then, we can write (see Bloomfield, 1976, pp. 80–84; Percival and Walden, 1993, pp. 325–326) the tapered Fourier transform at $\lambda_j, 2 \leq j \leq N - 2$, as a linear combination of the usual Fourier transform at the frequencies λ_j, λ_{j-1}, and λ_{j+1},

$$w^T(\lambda_j) = \frac{1}{\sqrt{6}}[-w(\lambda_{j-1}) + 2w(\lambda_j) - w(\lambda_{j+1})].$$

As a consequence, tapering destroys the orthogonality relations between Fourier transforms at different frequencies if they are too close, even asymptotically, because the tapered Fourier transforms have a common component at Fourier frequencies that are less than two basic frequencies $\lambda_1 = 2\pi/N$ away. For that reason we only consider frequencies λ_j and λ_k such that $k < j - 2$, which we may expect to be uncorrelated, as in the general case. However, tapering allows us to obtain much neater results than previously, especially for the expectation of the periodogram, reducing notably its bias even for frequencies close to a singularity using expansion (3).

We now present the equivalent of Theorem 2 of Robinson (1995a) for the (univariate) tapered Fourier transform. Define $v^T(\lambda) = w^T(\lambda)/(G^{1/2}\lambda^{-d})$.

THEOREM 2.   *Under Assumption* $6[1 < \alpha \le 2]$, *for any sequence of positive integers* $j = j(N)$ *and* $k = k(N)$ *such that* $2 \le k < j + 2$ *and* $j/N \to 0$ *as* $N \to \infty$,

(a) $E[v^T(\lambda_j)\overline{v^T}(\lambda_j)] = 1 + O(j^{-\alpha} + [j/N]^\alpha)$,
(b) $E[v^T(\lambda_j)\underline{v^T}(\lambda_j)] = O(j^{-4})$,
(c) $E[v^T(\lambda_j)v^T(\lambda_k)] = O(k^{-1})$,
(d) $E[v^T(\lambda_j)v^T(\lambda_k)] = O(k^{-1})$.

Comparing this with Robinson's results for the expectation of the periodogram, the bound in (a) is improved from $O(\log j/j)$ to $O(j^{-\alpha})$ for $1 < \alpha \le 2$. This is the main bias reduction gain. The magnitude of this bound is determined by Assumption 6 and depends on the tapering, which makes all the other contributions of smaller order. This is also the reason why we have such an improved bound in part (b). This improvement is fundamental to approximate the (tapered and pooled) periodogram of the X_t sequence by the transfer function $|\alpha(\lambda)|$ times the periodogram of the innovations.

However, the bounds in Theorem 2 do not improve substantially for the correlations between Fourier transforms at different frequencies (just by a logarithm factor), because the frequencies can be arbitrarily close and tapering does not affect substantially the asymptotic behavior of the periodogram there. Improved bounds are possible if we consider explicitly the distance $|j - k|$ (see Giraitis, Robinson, and Samarov, 1997).

4. ASYMPTOTIC DISTRIBUTION

In this section we derive the asymptotic distribution of $\hat{d}$, for $J \ge 3$, when we use the tapered periodograms. As discussed in the previous section we modify the definition of $\hat{d}$ in this way: define for $J = 1, 2, \ldots$, fixed, (assuming $(m - \ell)/(3J)$ to be an integer),

$$Y_k^{(T,J)} = \log(\bar{I}^T(\lambda_k)), \qquad k = \ell + 3J, \ell + 6J, \ldots, m,$$

where

$$\bar{I}^T(\lambda_k) = \sum_{j=1}^{J} I^T(\lambda_{k+3(j-J)}), \qquad I^T(\lambda) = |w^T(\lambda)|^2$$

and

$$\hat{d}^T = \left(\sum_k \Lambda_k^2\right)^{-1} \sum_k \Lambda_k Y_k^{(T,J)}. \tag{6}$$

We consider a pooling of J tapered periodogram ordinates, and though each of the tapered Fourier transforms is a linear combination of the Fourier transform at three adjacent frequencies, this definition of $\hat{d}^T$ secures the asymptotic uncorrelatedness of $\bar{I}^T(\lambda_k)$ and the asymptotic independence of the regressors $Y_k^{(T,J)}$ at different frequencies.

Let us introduce the following condition concerning the bandwidth numbers.

Assumption 7. As $N \to \infty$, for some $a > 0$,

$$\frac{1}{\ell} + \frac{\ell(\log N)^2}{m} + \frac{m^{1+1/2\alpha}}{N} \to 0.$$

The basic difference with respect to Assumption 6 of Robinson (1995a) is the first condition: we only need the trimming number ℓ to increase with N but independently of m, as from Theorem 2 we can control the bias of the periodogram for closer frequencies to the origin thanks to tapering. Now we present our main result.

THEOREM 3. *Under Assumptions 3, 4, 6, and 7, if ϵ_t has moments of all orders, $J \geq 3$, $\alpha J/(J+2) > 1$,*

$$m^{1/2}(\hat{d}^T - d) \to_D N\left(0, \frac{3J}{4} \psi'(J)\right),$$

where $\psi'(x) = (d/dx)\psi(x) = (d/dx)\log \Gamma(x)$ is the digamma function.

Notice the trade-off between pooling and the smoothness condition on $f(\lambda)$, allowing for values $\alpha < 2$ when J is large enough. The proof of the theorem is based on the method of moments. Although for the estimation of the moments of the logarithm of the innovations (pooled and tapered) periodograms we only require four bounded moments of ϵ_t, this is not enough to approximate the moments of a normalized infinite average of such periodogram logarithms. Our moment assumption is then used to approximate with a sufficient degree of accuracy those moments by means of Edgeworth expansions for the probability density of the Fourier transform.

For the asymptotic normality proof we do not use any special properties of tapering or pooling the periodogram, apart from the bound in Lemma 1 (pooling) and Theorem 2 (tapering). These two devices are used to improve the approximations and behavior of the periodogram of the long range dependent time series.

Possibly, under stronger conditions on the dependence of the process X_t and/or its distribution, one or both of these techniques could be dispensed with.

5. SIMULATION WORK

In this section we present a simulation exercise to analyze the techniques of tapering and pooling in the log-periodogram estimate for non-Gaussian data. To that end we simulate 5,000 series following an ARFIMA$(0,d,0)$ model and innovations with different distributions, Uniform$[-\sqrt{3},\sqrt{3}]$, Exponential with parameter 1, recentered with zero mean, t_5 distribution, which only has four moments, so Theorem 1 holds, but not Theorem 3, and standard Gaussian for comparison

TABLE 1. Log-periodogram regression estimator $\hat{d}$ ($d = 0.45$, $N = 256$)

m	J	Bias	sd	th.sd	MSE	90%	95%	99%
				Gaussian				
30	1	0.0056	0.1416	0.1171	0.0201	83.46	89.50	96.48
30	2	−0.0295	0.1189	0.1037	0.0150	84.48	90.70	96.88
30	3	0.0287	0.1309	0.0994	0.0180	77.62	85.20	94.54
45	1	−0.0066	0.1104	0.0956	0.0122	85.30	91.06	97.04
45	2	−0.0342	0.0962	0.0847	0.0104	83.60	90.00	96.60
45	3	0.0097	0.1012	0.0811	0.0103	81.08	88.42	95.88
				Uniform				
30	1	0.0024	0.1437	0.1171	0.0207	81.80	88.90	96.64
30	2	−0.0325	0.1190	0.1037	0.0152	84.06	90.48	96.76
30	3	0.0241	0.1322	0.0994	0.0181	77.52	85.46	94.38
45	1	−0.0098	0.1120	0.0956	0.0126	84.24	90.98	96.78
45	2	−0.0374	0.0955	0.0847	0.0105	83.22	89.36	96.34
45	3	0.0059	0.1018	0.0811	0.0104	82.30	88.90	95.70
				Exponential				
30	1	0.0030	0.1426	0.1171	0.0204	82.40	89.18	96.36
30	2	−0.0318	0.1178	0.1037	0.0149	84.48	90.10	96.96
30	3	0.0252	0.1308	0.0994	0.0177	78.16	85.72	94.24
45	1	−0.0108	0.1124	0.0956	0.0127	84.34	91.02	96.62
45	2	−0.0373	0.0954	0.0847	0.0105	83.44	89.72	96.24
45	3	0.0063	0.1017	0.0811	0.0104	81.88	88.44	95.84
				t_5				
30	1	0.0046	0.1393	0.1171	0.0194	83.70	90.00	96.66
30	2	−0.0307	0.1162	0.1037	0.0144	84.84	91.02	97.04
30	3	0.0257	0.1277	0.0994	0.0170	79.04	86.40	95.00
45	1	−0.0052	0.1093	0.0956	0.0120	84.94	91.74	97.52
45	2	−0.0332	0.0932	0.0847	0.0098	84.60	90.94	96.48
45	3	0.0097	0.0987	0.0811	0.0098	82.74	89.24	96.40

TABLE 2. Log-periodogram regression estimator $\hat{d}$ ($d = 0.45$, $N = 512$)

m	J	Bias	sd	th.sd	MSE	90%	95%	99%
				Gaussian				
30	1	0.0075	0.1432	0.1171	0.0205	82.82	89.24	96.30
30	2	−0.0283	0.1198	0.1037	0.0152	84.54	90.40	96.50
30	3	0.0291	0.1330	0.0994	0.0185	77.22	84.90	94.12
60	1	0.0033	0.0942	0.0828	0.0089	85.10	91.50	97.50
60	2	−0.0211	0.0800	0.0733	0.0068	85.50	91.68	97.38
60	3	0.0168	0.0844	0.0703	0.0074	82.10	89.10	96.54
90	1	−0.0018	0.0753	0.0676	0.0057	86.32	92.12	97.58
90	2	−0.0210	0.0648	0.0599	0.0046	85.22	91.28	97.62
90	3	0.0082	0.0671	0.0574	0.0046	83.86	90.66	97.08
				Uniform				
30	1	0.0090	0.1411	0.1171	0.0200	83.06	89.70	96.50
30	2	−0.0265	0.1171	0.1037	0.0144	84.92	90.84	96.86
30	3	0.0322	0.1303	0.0994	0.0180	77.48	85.38	94.56
60	1	0.0047	0.0940	0.0828	0.0089	85.18	91.48	97.54
60	2	−0.0204	0.0798	0.0733	0.0068	86.10	92.14	97.34
60	3	0.0176	0.0843	0.0703	0.0074	82.22	89.04	96.28
90	1	−0.0019	0.0751	0.0676	0.0056	86.18	92.02	97.98
90	2	−0.0219	0.0638	0.0599	0.0046	85.74	91.72	97.46
90	3	0.0070	0.0663	0.0574	0.0044	84.38	90.62	97.32
				Exponential				
30	1	0.0088	0.1415	0.1171	0.0201	83.28	89.66	96.68
30	2	−0.0274	0.1175	0.1037	0.0145	84.70	90.74	96.90
30	3	0.0309	0.1302	0.0994	0.0179	77.56	85.86	94.40
60	1	0.0048	0.0927	0.0828	0.0086	85.56	92.10	97.62
60	2	−0.0202	0.0791	0.0733	0.0067	86.12	91.96	97.78
60	3	0.0183	0.0836	0.0703	0.0073	82.78	89.36	96.54
90	1	−0.0025	0.0736	0.0676	0.0054	86.92	92.88	98.14
90	2	−0.0220	0.0630	0.0599	0.0045	86.20	92.18	97.78
90	3	0.0070	0.0654	0.0574	0.0043	84.72	91.28	97.24
				t_5				
30	1	0.0069	0.1417	0.1171	0.0201	82.28	89.28	96.54
30	2	−0.0278	0.1166	0.1037	0.0144	84.68	90.66	96.82
30	3	0.0299	0.1289	0.0994	0.0175	77.64	85.54	94.54
60	1	0.0038	0.0945	0.0828	0.0089	84.94	91.14	97.50
60	2	−0.0202	0.0795	0.0733	0.0067	86.16	91.86	97.48
60	3	0.0173	0.0831	0.0703	0.0072	82.70	89.86	96.74
90	1	−0.0033	0.0747	0.0676	0.0056	86.42	92.56	97.68
90	2	−0.0224	0.0635	0.0599	0.0045	85.80	91.60	97.48
90	3	0.0064	0.0654	0.0574	0.0043	84.24	91.22	97.60

TABLE 3. Log-periodogram regression estimator (taper, $d = 0.45$, $N = 256$)

			Spacing, $\hat{d}^T$						No Spacing, $\hat{d}^{T,*}$					
m	J	Bias	sd	th.sd	MSE	90%	95%	99%	Bias	sd	MSE	90%	95%	99%
							Gaussian							
30	1	0.0133	0.2536	0.2028	0.0645	82.68	89.54	96.00	0.0067	0.2026	0.0411	90.32	94.52	98.64
30	2	−0.0043	0.2642	0.1796	0.0698	74.16	81.64	92.30	−0.0181	0.1728	0.0302	91.44	95.34	99.00
30	3	0.0657	0.3115	0.1721	0.1014	63.88	72.02	84.16	0.0191	0.1849	0.0346	87.80	92.76	97.96
45	1	−0.0030	0.1970	0.1656	0.0388	84.54	90.18	96.52	−0.009	0.1545	0.0240	92.46	96.48	99.20
45	2	−0.0200	0.2187	0.1466	0.0482	74.40	81.76	91.58	−0.0282	0.1356	0.0192	91.62	96.14	99.16
45	3	0.0209	0.2209	0.1405	0.0492	71.22	79.26	89.72	−0.0008	0.1405	0.0198	90.22	95.10	98.94
							Uniform							
30	1	−0.0015	0.2548	0.2028	0.0649	82.78	89.04	95.50	−0.0065	0.2033	0.0414	89.98	94.66	98.78
30	2	−0.0198	0.2762	0.1796	0.0767	73.02	80.60	90.92	−0.0296	0.1737	0.0311	90.84	95.12	99.04
30	3	0.0468	0.3175	0.1721	0.1030	63.10	71.66	83.86	0.0064	0.1863	0.0348	87.34	92.74	98.06
45	1	−0.0145	0.1955	0.1656	0.0384	85.00	90.96	96.70	−0.0211	0.1535	0.0240	92.62	96.12	99.40
45	2	−0.0325	0.2255	0.1466	0.0519	72.64	81.02	91.00	−0.0383	0.1347	0.0196	92.24	96.02	99.12
45	3	0.0111	0.2240	0.1405	0.0503	70.38	78.40	89.58	−0.0113	0.1397	0.0196	90.22	95.52	98.86

Exponential

30	1	−0.0017	0.2545	0.2028	0.0648	82.94	89.06	95.66	−0.0047	0.1980	0.0392	89.98	94.66	98.78
30	2	−0.0139	0.2686	0.1796	0.0724	73.86	82.00	91.02	−0.0275	0.1686	0.0292	90.84	95.12	99.04
30	3	0.0518	0.3093	0.1721	0.0984	63.78	72.24	84.64	0.0089	0.1806	0.0327	87.34	92.74	98.06
45	1	−0.0125	0.1977	0.1656	0.0393	84.52	90.62	96.62	−0.0193	0.1534	0.0239	92.62	96.12	99.40
45	2	−0.0282	0.2200	0.1466	0.0492	74.24	81.82	91.08	−0.0371	0.1341	0.0194	92.24	96.02	99.12
45	3	0.0128	0.2212	0.1405	0.0491	71.14	79.04	89.88	−0.0102	0.1385	0.0193	90.22	95.52	98.86

t_5

30	1	0.0033	0.2490	0.2028	0.0620	82.84	89.78	96.20	−0.0017	0.2000	0.0400	90.80	95.26	98.78
30	2	−0.0138	0.2678	0.1796	0.0719	74.38	81.74	91.66	−0.0254	0.1718	0.0302	91.16	95.52	98.86
30	3	0.0485	0.3125	0.1721	0.1000	63.80	71.68	84.14	0.0095	0.1828	0.0335	88.10	93.86	98.08
45	1	−0.0080	0.1932	0.1656	0.0374	84.76	91.04	97.02	−0.0140	0.1538	0.0239	92.30	96.50	99.40
45	2	−0.0250	0.2194	0.1466	0.0488	73.80	81.90	91.48	−0.0324	0.1350	0.0193	92.14	96.12	99.16
45	3	0.0126	0.2185	0.1405	0.0479	71.10	78.94	90.28	−0.0058	0.1390	0.0194	90.94	95.46	99.04

TABLE 4. Log-periodogram regression estimator (taper, $d = 0.45$, $N = 512$)

m	J	Spacing, $\hat{d}^T$							No Spacing, $\hat{d}^{T,*}$					
		Bias	sd	th.sd	MSE	90%	95%	99%	Bias	sd	MSE	90%	95%	99%
							Gaussian							
30	1	0.0068	0.2605	0.2028	0.0679	81.28	87.70	95.50	0.0019	0.2060	0.0424	89.88	94.48	98.44
30	2	−0.0113	0.2687	0.1796	0.0723	73.58	82.26	91.50	−0.0222	0.1777	0.0321	90.66	95.04	98.46
30	3	0.0534	0.3143	0.1721	0.1016	62.70	71.50	83.78	0.0138	0.1906	0.0365	86.66	92.58	97.60
60	1	0.0028	0.1670	0.1434	0.0279	84.88	91.04	97.14	0.0004	0.1316	0.0173	92.38	96.58	99.38
60	2	−0.0174	0.1817	0.1270	0.0333	76.22	83.66	92.72	−0.0167	0.1156	0.0136	92.38	96.32	99.38
60	3	0.0244	0.1989	0.1217	0.0402	69.04	77.56	88.58	0.0073	0.1194	0.0143	90.30	95.40	99.06
90	1	−0.0043	0.1301	0.1171	0.0169	86.74	92.80	97.48	−0.0062	0.1021	0.0105	93.80	97.40	99.64
90	2	−0.0219	0.1445	0.1037	0.0214	76.90	84.64	93.64	−0.0199	0.0901	0.0085	93.26	97.04	99.54
90	3	0.0029	0.1434	0.0994	0.0206	74.90	83.38	92.74	−0.0014	0.0919	0.0084	92.46	96.38	99.44
							Exponential							
30	1	0.0149	0.2525	0.2028	0.0640	82.38	88.74	95.88	0.0034	0.2043	0.0418	89.68	94.60	98.68
30	2	−0.0009	0.2710	0.1796	0.0734	74.16	81.32	91.38	−0.0182	0.1752	0.0310	90.72	94.98	98.82
30	3	0.0692	0.3129	0.1721	0.1027	62.62	71.08	83.66	0.0186	0.1871	0.0353	87.38	92.56	97.94
60	1	0.0068	0.1628	0.1434	0.0266	85.02	91.54	97.50	0.0009	0.1283	0.0165	93.66	97.24	99.52
60	2	−0.0137	0.1809	0.1270	0.0329	76.64	83.42	92.60	−0.0151	0.1117	0.0127	93.70	97.38	99.50
60	3	0.0302	0.1976	0.1217	0.0400	68.62	77.04	88.10	0.0093	0.1162	0.0136	91.04	95.90	99.38
90	1	−0.0023	0.1282	0.1171	0.0164	87.36	92.64	97.88	−0.0074	0.1011	0.0103	94.34	97.14	99.70
90	2	−0.0215	0.1436	0.1037	0.0211	78.14	84.68	93.44	−0.0201	0.0883	0.0082	94.20	97.20	99.52
90	3	0.0066	0.1430	0.0994	0.0205	74.42	82.56	92.94	−0.0012	0.0907	0.0082	93.22	96.48	99.34

Uniform

30	1	0.0141	0.2494	0.2028	0.0624	82.42	89.38	95.90	0.0075	0.2002	0.0402	89.90	95.36	99.26
30	2	−0.0059	0.2645	0.1796	0.0700	74.00	82.72	92.12	−0.0164	0.1725	0.0300	91.14	95.50	99.18
30	3	0.0617	0.3136	0.1721	0.1022	61.98	71.38	83.20	0.0204	0.1846	0.0345	86.80	93.22	98.18
60	1	0.0064	0.1635	0.1434	0.0268	85.30	91.62	97.50	0.0017	0.1276	0.0163	93.56	97.30	99.66
60	2	−0.0146	0.1807	0.1270	0.0329	75.42	83.44	92.90	−0.0157	0.1118	0.0127	93.84	97.12	99.66
60	3	0.0291	0.1989	0.1217	0.0404	68.44	76.40	88.30	0.0085	0.1161	0.0135	91.64	96.08	99.32
90	1	−0.0005	0.1305	0.1171	0.0170	85.62	92.32	97.72	−0.0062	0.1010	0.0102	94.32	97.42	99.62
90	2	−0.0207	0.1452	0.1037	0.0215	75.56	83.36	92.84	−0.0199	0.0890	0.0083	93.68	97.40	99.50
90	3	0.0073	0.1453	0.0994	0.0212	75.00	82.70	91.98	−0.0012	0.0911	0.0083	92.66	96.80	99.54

t_5

30	1	0.0099	0.2577	0.2028	0.0665	81.84	88.40	95.36	0.0103	0.2021	0.0410	90.56	95.36	99.00
30	2	−0.0051	0.2666	0.1796	0.0711	74.82	82.42	91.76	−0.0141	0.1730	0.0301	91.24	95.92	99.22
30	3	0.0656	0.3126	0.1721	0.1020	62.98	70.94	83.52	0.0231	0.1840	0.0344	87.20	93.16	98.50
60	1	0.0040	0.1660	0.1434	0.0276	86.06	91.62	97.40	0.0032	0.1278	0.0163	93.70	97.08	99.38
60	2	−0.0135	0.1824	0.1270	0.0335	75.60	83.66	92.80	−0.0140	0.1113	0.0126	93.90	97.12	99.54
60	3	0.0283	0.1970	0.1217	0.0396	68.08	77.34	88.38	0.0106	0.1152	0.0134	91.86	96.22	99.30
90	1	−0.0033	0.1326	0.1171	0.0176	86.40	92.24	97.56	−0.0047	0.1003	0.0101	94.48	97.68	99.68
90	2	−0.0190	0.1484	0.1037	0.0224	76.26	83.84	92.56	−0.0185	0.0881	0.0081	93.92	97.40	99.70
90	3	0.0058	0.1454	0.0994	0.0212	74.18	82.34	92.46	0.0005	0.0903	0.0082	92.84	96.86	99.48

purposes. We report only the results for $d = 0.45$, the conclusions being similar for other values in the interval $(0, \frac{1}{2})$. We use for the simulations a modification of the function `arima.fracdiff.sim` included in the SPLUS package.

The sample sizes tried are $N = 256$ and $N = 512$, and the bandwidth numbers considered are $m = 30,45$ and $m = 30,60,90$, respectively, with pooling numbers $J = 1,2,3$ for all cases. We do not perform any trimming, $\ell = 0$, this not being a decisive choice. For each of the time series simulated we calculate three different types of estimate for all the combinations of bandwidth and pooling choices. They are the nontapered log-periodogram regression estimate $\hat{d}$, the tapered log-periodogram estimate $\hat{d}^T$, as defined previously in (6) with asymptotically independent regressors (with "spacing"), and a modification of this last one, $\hat{d}^{T,*}$, considering all possible frequencies between the origin and λ_m ("no spacing"), with

$$\bar{I}^T(\lambda_k) = \sum_{j=1}^{J} I^T(\lambda_{k+j-J}), \qquad k = \ell + J, \ell + 2J, \ldots, m,$$

which we may expect to achieve some efficiency gains from the augmented number of elements in the regression.

We report in Tables 1–4 the results of the simulations for the three estimates, the first two tables for nontapered estimates, the other two for tapered data. We give for all the estimates calculated the bias and standard deviation (sd) across replications, the asymptotic standard deviation (th.sd) in the appropriate central limit theorem (CLT), the mean square error (MSE), and the true coverage for the 90%, 95%, and 99% confidence intervals calculated using the previous CLT's. For the third estimate we have not provided asymptotic theory, although its consistency could be shown using the same techniques. However the analysis of the asymptotic distribution of this estimate is more complicated, because without the additional spacing between regressors we cannot guarantee their asymptotic independence and our approach for $\hat{d}^T$ breaks down. Nevertheless we compute confidence intervals for this estimate using the asymptotic standard deviation of $\hat{d}^T$, pretending now that the increment in the variability of $\hat{d}^{T,*}$ caused by the regressors correlation induced by tapering when all periodograms are used is of about the same magnitude as the one resulting by the use of a reduced number of periodogram ordinates (though it also depends on J in a complicated way as a result of the averaging).

Following the discussion in Robinson (1995a), increasing J may produce asymptotic efficiency gains, because $J\psi'(J)$ is decreasing. This can be checked in the column for the theoretical standard deviation, th.sd. However, in practice and for these sample sizes, the gains are only apparent for the log-periodogram without tapering and $J = 2,3$, both giving similar reductions of the MSE but with larger standard deviations than the nonpooled estimates ($J = 1$), because of the reduced number of regressors. When we taper the observations, $\hat{d}^T$ is already using a reduced number of frequencies when $J = 1$, so setting $J > 1$ always increases the variance. The situation is much different when we do not space the regressors in $\hat{d}^{T,*}$, obtaining efficiency gains with larger values of J, with much reduced variances than $\hat{d}^T$ and only slightly smaller than the asymptotic values for $\hat{d}^T$.

In all cases considered, the variability decreases with m as expected, but the behavior of the bias is not uniform. For the smallest sample size, and for both tapered and nontapered periodograms, the minimum values are obtained for $J = 3$ and the maximum for $J = 2$. For $n = 512$ the overall result is less clear, but in most cases $J = 1$ seems to be the less biased option. In some situations the bias tends to decrease with m, because for these series the semiparametric model considered is a good approximation for the entire range of frequencies $(0, \pi)$. The bias tends to be negative for high m (all J) and for $J = 2$ (all m) for almost all choices of estimates and distributions.

The minimum MSE is then attained for the estimates with biggest m, and with $J = 2$ when tapering and spacing of frequencies are used $(\hat{d}^T)$, but with $J = 2, 3$ when no spacing is employed $(\hat{d}^{T,*})$, both values of J giving similar results. When no tapering is used, averaging with $J > 1$ seems to be the best option. All equal, tapering always increases the MSE, especially if spacing of frequencies is used.

The accuracy of the CLT deteriorates with tapering, because we are effectively reducing the number of observations in the log-periodogram regression, the true variability being much larger than predicted by the asymptotics. When no tapering is used, the asymptotic distribution approximates quite well the true distribution of the log-periodogram estimate for all m and J, especially for the largest sample size, even for the t_5 distribution, and it is not easy to distinguish differences across data distributions in the behavior of the estimates, Gaussianity not necessarily giving the best performance and the CLT being robust to bounded support, asymmetric or heavy-tailed innovation distributions. When tapering the same conclusions apply, here the CLT providing a very good approximation for the finite sample behavior of $\hat{d}^{d,*}$, though the true variability is slightly overestimated by our heuristic choice. We have not pursued other variance estimation methods using for example the (now autocorrelated) ordinary least squares (OLS) residuals, but this could also be tried.

REFERENCES

Beran, J. (1994) *Statistics for Long Memory Processes.* London: Chapman and Hall.

Bhattacharya, R.N. & R.R. Rao (1976) *Normal Approximation and Asymptotic Expansions.* New York: Wiley.

Bloomfield, P. (1976) *Fourier Analysis of Time Series: An Introduction.* New York: Wiley.

Chen, Z.-G. & E.J. Hannan (1980) The distribution of periodogram ordinates. *Journal of Time Series Analysis* 1, 73–82.

Comte, F. & C. Hardouin (1995) Regression on log-regularized periodogram for fractional models at low frequencies. Working paper, CREST.

Dahlhaus, R. (1988) Small sample effects in time series analysis: A new asymptotic theory and a new estimate. *Annals of Statistics* 16, 808–841.

Dahlhaus, R. (1989) Efficient parameter estimation for self-similar processes. *Annals of Statistics* 17, 1749–1766.

Feller, W. (1971) *An Introduction to Probability Theory and Its Applications*, vol. 2, 2nd ed. New York: Wiley.

Fox, R. & M.S. Taqqu (1986) Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals of Statistics* 14, 517–532.

Geweke, J. & S. Porter-Hudak (1983) The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4, 221–238.

Giraitis, L., P.M. Robinson, & A. Samarov (1997) Rate optimal semiparameteric estimation of the memory parameter of the Gaussian time series with long-range dependence. *Journal of Time Series Analysis* 18, 49–60.

Giraitis, L. & D. Surgailis (1990) A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate. *Probability Theory and Related Fields* 86, 87–104.

Götze, F. & C. Hipp (1983) Asymptotic expansions for sums of weakly dependent random vectors. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 64, 211–239.

Hannan, E.J. (1970) *Multiple Time Series.* New York: Wiley.

Hannan, E.J. & D.F. Nicholls (1977) The estimation of the prediction error variance. *Journal of the American Statistical Association* 72, 834–840.

Hassler, U. (1993) Regression on spectral estimators with fractionally integrated time series. *Journal of Time Series Analysis* 14, 369–380.

Hurvich, C.M. & B.K. Ray (1995) Estimation of the memory parameter for nonstationary or non-invertible fractionally integrated processes. *Journal of Time Series Analysis* 16, 17–42.

Janas, D. & R. von Sachs (1993) Consistency for non-linear functions of the periodogram of tapered data. *Journal of Time Series Analysis* 16, 585–606.

Percival, D.B. & A.T. Walden (1993) *Spectral Analysis for Physical Applications.* Cambridge: Cambridge University Press.

Lobato, I. & P.M. Robinson (1996) Averaged periodogram estimation of long memory. *Journal of Econometrics* 73, 303–324.

Robinson, P.M. (1986) On the errors-in-variables problem for time series. *Journal of Multivariate Analysis* 19, 240–250.

Robinson, P.M. (1994a) Semiparametric analysis of long-memory time series. *Annals of Statistics* 22, 515–539.

Robinson, P.M. (1994b) Rates of convergence and optimal spectral bandwidth for long range dependence. *Probability Theory and Related Fields* 99, 443–473.

Robinson, P.M. (1994c) Time series with strong dependence. In C.A. Sims (ed.), *Advances in Econometrics: Sixth World Congress*, vol. 1, pp. 47–95. Cambridge: Cambridge University Press.

Robinson, P.M. (1995a) Log-periodogram regression of time series with long range dependence. *Annals of Statistics* 23, 1048–1072.

Robinson, P.M. (1995b) Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* 23, 1630–1661.

von Sachs, R. (1994) Estimating non-linear functions of the spectral density, using a data-taper. *Annals of the Institute of Statistical Mathematics* 46, 453–474.

Taniguchi, M. (1979) On estimation of parameters of Gaussian stationary processes. *Journal of Applied Probability* 16, 575–591.

Tukey, J.W. (1967) An introduction to the calculation of numerical spectrum analysis. In B. Harris (ed.), *Advanced Seminar on Spectral Analysis of Time Series*, pp. 25–46. New York: Wiley.

Velasco, C. (1997) *Higher Order Asymptotic Theory for Nonparametric Time Series Analysis and Related Contributions.* Ph.D. Thesis, University of London.

Velasco, C. (1999) Non-stationary log-periodogram regression. *Journal of Econometrics* 91, 325–371.

APPENDIX

The following is a simplified version of Chen and Hannan's (1980) Lemma 2, where we only use the first two terms of an Edgeworth expansion for the probability density of the Fourier transform of ϵ_t, so only four bounded moments are required.

LEMMA 2. *Under Assumption 4, the probability distribution function Q_N of the vector*

$$W_N = N^{-1/2} \sum_{t=1}^{N} Y_t,$$

where

$$Y_t' = Y'(j(1), \ldots, j(k)) = \sqrt{2} \epsilon_t (\cos t\lambda_{j(1)}, \sin t\lambda_{j(1)}, \ldots, \cos t\lambda_{j(k)}, \sin t\lambda_{j(k)}),$$

has density q_N for all sufficiently large N and

$$\sup_{\mathbf{y} \in \mathbb{R}^{2k}} (1 + \|\mathbf{y}\|^4) \left| q_N(\mathbf{y}) - \sum_{r=0}^{1} N^{-r/2} P_r(-\phi : \bar{\chi}_{\nu,N})(\mathbf{y}) \right| = O(N^{-1}), \tag{A.1}$$

where P_r are polynomials in the average of the joint cumulants of Y_t $(1 \le t \le N)$ of order $\nu = (\nu_1, \ldots, \nu_{2k})$, $\bar{\chi}_{\nu,N}$, multiplied by the 2kth multivariate normal density ϕ and where $P_0(\mathbf{y}) = \phi(\mathbf{y})$.

Proof of Lemma 1. First, from Lemma 2, $2\pi \bar{I}_{\epsilon k}$ has the probability density of a $\frac{1}{2}\chi_{2J}^2$ distribution with error (using only P_0) of order $O((1 + \|\mathbf{y}\|^4)^{-1} N^{-1/2})$. Also the density of a χ_{2J}^2 is

$$\phi_{\chi_{2J}^2}(x) = \frac{x^{J-1} e^{-x/2}}{(J-1)! 2^J}, \qquad 0 \le x < \infty.$$

It is clear that if $X \sim \chi_{2J}^2$ then $E[X^{a-J}] < \infty$ if $0 < a < J$. Thus we only need to check that the error in the evaluation of the second inverse moment of $\bar{I}_{\epsilon k}$ using Lemma 2 is bounded. If we write

$$\bar{I}_{\epsilon k} = \sum_{j=1}^{J} I_\epsilon(\lambda_{k+j-J}) = \sum_{j=1}^{J} (y_{aj}^2 + y_{bj}^2)$$

we need

$$N^{-1/2} \int_{\mathbb{R}^{2J}} (1 + \|\mathbf{y}\|^4)^{-1} \left(\sum (y_{aj}^2 + y_{bj}^2) \right)^{a-J} d\mathbf{y} < \infty. \tag{A.2}$$

First, defining the sets $A = [-1, 1]^{2J}$ and A^c its complementary in $\mathbb{R}^{2J}$, for some $0 < C < \infty$,

$$\int_{\mathbb{R}^{2J}} (1 + \|\mathbf{y}\|^4)^{-1} \left(\sum (y_{aj}^2 + y_{bj}^2) \right)^{a-J} d\mathbf{y}$$

$$\le C \int_A \left(\sum (y_{aj}^2 + y_{bj}^2) \right)^{a-J} d\mathbf{y} + C \int_{A^c} (1 + \|\mathbf{y}\|^4)^{-1} d\mathbf{y}$$

$$\le \text{const.} \int_A \left(\sum (y_{aj}^2 + y_{bj}^2) \right)^{a-J} d\mathbf{y} + \text{const.},$$

because $(1 + \|\mathbf{y}\|^4)^{-1}$ and $(\sum(y_{aj}^2 + y_{bj}^2))^{a-J}$ are bounded from above in A and A^c, respectively. Next, to bound the remaining integral, if $\phi(\cdot)$ denotes the densities of the corresponding distributions, we have

$$\infty > \int_0^\infty x^{a-J} \phi_{\chi_{2J}^2}(x)\, dx = \int_{\mathbb{R}^{2J}} \left(\sum_j (y_{aj}^2 + y_{bj}^2) \right)^{a-J} \phi_{\mathcal{N}(0, I_{2J})}(\mathbf{y})\, d\mathbf{y}$$

$$> \int_A \left(\sum (y_{aj}^2 + y_{bj}^2) \right)^{a-J} \phi_{\mathcal{N}(0, I_{2J})}(\mathbf{y})\, d\mathbf{y}$$

$$\geq \text{const.} \int_A \left(\sum (y_{aj}^2 + y_{bj}^2) \right)^{a-J} d\mathbf{y},$$

as the normal density is bounded from below in A, and the lemma follows. ■

LEMMA 3. *For* $\delta_k = \bar{I}_k f_k^{-1} - 2\pi \bar{I}_{\epsilon, k}$, *under the conditions of Theorem* 1 (*no tapering*):

$$E|\delta_k| = O([k^{-1} \log k]^{1/2}), \qquad E[\delta_k^2] = O(k^{-1} \log k).$$

and under the conditions of Theorem 3 (*tapering*), $1 < \alpha \leq 2$:

$$E|\delta_k| = O(k^{-\alpha/2}), \qquad E[\delta_k^2] = O(k^{-\alpha}).$$

Proof of Lemma 3. Denote in either case

$$F_k = \bar{I}_k f_k^{-1} - H_k, \qquad H_k = \sum_{j=1}^J I_{k+j-J} f_{k+j-J}^{-1}.$$

Consider first the nontapered case. By the triangle inequality, $E|\delta_k| \leq E|\bar{I}_k f_k^{-1} - H_k| + E|H_k - 2\pi \bar{I}_{\epsilon, k}|$, where, for $j = 1 - J + k, \ldots, k$,

$$E|H_k - 2\pi \bar{I}_{\epsilon, k}| \leq \sum_j E|I_j f_j^{-1} - 2\pi \bar{I}_{\epsilon, j}| = O([k^{-1} \log k]^{1/2})$$

from Robinson (1995b, p. 1637) and using the mean value theorem,

$$|\bar{I}_k f_k^{-1} - H_k| \leq \max_j |f_j^{-1} - f_k^{-1}| \sum_j I_j \leq C f_k^{-1} k^{-1} \bar{I}_j,$$

so $E|\bar{I}_k f_k^{-1} - H_k| = O(k^{-1})$. Next, $E|\delta_k|^2 \leq 2E|\bar{I}_k f_k^{-1} - H_k|^2 + 2E|H_k - 2\pi \bar{I}_{\epsilon, k}|^2$, where $E|H_k - 2\pi \bar{I}_{\epsilon, k}|^2 = O(k^{-1} \log k)$ from Robinson (1995b, pp. 1648–1649) and $E|\bar{I}_k f_k^{-1} - H_k|^2 \leq C^2 f_k^{-2} k^{-2} E\bar{I}_j^2 = O(k^{-2})$.

When tapering is applied the results follow in the same way, using now part (a) of Theorem 2 to obtain that $E|H_k - 2\pi \bar{I}_{\epsilon, k}| = O(k^{-\alpha/2})$ and $E|H_k - 2\pi \bar{I}_{\epsilon, k}|^2 = O(k^{-\alpha})$. ■

Proof of Theorem 1. From Robinson (1995a) and the definition for the summation in k, we can obtain

$$\sum_k \Lambda_k^2 = \frac{4m}{J} \left(1 + O\left(\frac{\ell(\log N)^2}{m} \right) \right) = \frac{4m}{J} (1 + o(1)),$$

$$\sum_k |\Lambda_k|^p = O(m), \qquad p \geq 1, \tag{A.3}$$

and $\sup_k|\Lambda_k| = O(\log m)$. Hence, under Assumption 1, with the previous properties,

$$\left(\sum_k \Lambda_k^2\right)^{-1} \sum_k \Lambda_k \log f_k = d + O\left(\left[\frac{m}{N}\right]^\alpha\right). \tag{A.4}$$

Substituting in the definition of $\hat{d}$ and using (A.4),

$$\hat{d} = \left(\sum_k \Lambda_k^2\right)^{-1} \sum_k \Lambda_k[\log f_k + \log 2\pi\bar{I}_{\epsilon,k}] + R$$

$$= d + \left(\sum_k \Lambda_k^2\right)^{-1} \sum_k \Lambda_k \log 2\pi\bar{I}_{\epsilon,k} + O([mN^{-1}]^\alpha) + R$$

$$= d + \xi_N + R + o_P(1),$$

say. We first show that $R = o_p(1)$ splitting this remainder term into two terms as

$$R = \left(\sum_k \Lambda_k^2\right)^{-1} \sum_k \Lambda_k\{\log \bar{I}_k f_k^{-1} - \log 2\pi\bar{I}_{\epsilon,k}\} = \sum_{\bar{A}_T} + \sum_{A_T},$$

where $A_T = \{k:\bar{I}_k f_k^{-1} < m^{-c}\}$, for some $c > 0$, and let $n_T =$ the number of elements in A_T. Denote $\log \bar{I}_k^* = \log(\bar{I}_k f_k^{-1})\mathbb{I}\{\bar{I}_k f_k^{-1} \geq m^{-c}\}$, $\delta_k = \bar{I}_k f_k^{-1} - 2\pi\bar{I}_{\epsilon,k}$.

We first bound the contribution from $\bar{A}_T$. For $k \in \bar{A}_T$, with $J \geq 3$, from (A.5)–(A.7)

$$E|\log \bar{I}_k f_k^{-1} - \log 2\pi\bar{I}_{\epsilon,k}|$$

$$= E\left|\log\left(1 + \frac{\delta_k}{2\pi\bar{I}_{\epsilon,k}}\right)\right|$$

$$\leq E\left|\frac{\delta_k}{2\pi\bar{I}_{\epsilon,k}}\right| + E\left[(|\log \bar{I}_k^*| + |\log 2\pi\bar{I}_{\epsilon,k}|)\mathbb{I}\left\{|\delta_k/(2\pi\bar{I}_{\epsilon,k})| \geq \frac{1}{2}\right\}\right]$$

$$= O\left(\sqrt{\frac{\log k}{k}} + \log m \times P\left\{|\delta_k/(2\pi\bar{I}_{\epsilon,k})| \geq \frac{1}{2}\right\}^{1-\varepsilon}\right) = O(k^{\varepsilon-0.5}\log m),$$

for any $\varepsilon > 0$ because $E[\bar{I}_{\epsilon,k}] < \infty$ and $E[\bar{I}_{\epsilon,k}^{-1}] < \infty$, so

$$E\left|\left(\sum_k \Lambda_k^2\right)^{-1} \sum_{\bar{A}_T} \Lambda_k\{\log \bar{I}_k f_k^{-1} - \log 2\pi\bar{I}_{\epsilon,k}\}\right|$$

$$\leq \left(\sum_k \Lambda_k^2\right)^{-1} \sum_{\bar{A}_T} |\Lambda_k|E|\log \bar{I}_k f_k^{-1} - \log 2\pi\bar{I}_{\epsilon,k}|$$

$$= O\left(m^{-1}\log^2 m \sum_k k^{\varepsilon-0.5}\right) = O(m^{\varepsilon-0.5}\log^2 m),$$

which is $o(1)$ with Assumption 5. For any fixed integer $b \geq 1$

$$E[|\log \bar{I}_k^*|^b] \leq (c\log m)^b + E[|\log \bar{I}_k f_k^{-1}|^b\mathbb{I}\{\bar{I}_k f_k^{-1} > 1\}]$$

$$\leq (c\log m)^b + E[|\bar{I}_k f_k^{-1}|]$$

$$= (c\log m)^b + E[|2\pi\bar{I}_{\epsilon,k} + \delta_k|] = O(\log^b m), \tag{A.5}$$

using Lemma 3 and $E[\bar{I}_{\epsilon,k}] < \infty$. Then,

$$P\left\{|\delta_k/(2\pi\bar{I}_{\epsilon,k})| \geq \frac{1}{2}\right\} \leq 2E\left|\frac{\delta_k}{2\pi\bar{I}_{\epsilon,k}}\right| = O(\log k \times k^{-0.5}), \tag{A.6}$$

because from Lemma 3, $0 < a' < J, J \geq 3$,

$$E\left|\frac{\delta_k}{2\pi\bar{I}_{\epsilon,k}}\right| \leq E[\delta_k^2]^{1/2}E[(\bar{I}_{\epsilon,k})^{-2}]^{1/2} = O([k^{-1}\log k]^{1/2}), \tag{A.7}$$

because $E[\{\bar{I}_{\epsilon,k}\}^{a'-J}] < \infty$ from Lemma 1.

We now estimate n_T. Following Hannan and Chen (1980) we can show that for any sequence $\rho \to \infty$ as $N \to \infty$, for all $\eta > 0$,

$$p\lim_{N\to\infty} \rho^{-1}m^{-1/2}\sum_k\left\{\log 2\pi\bar{I}_{\epsilon,k}\mathbb{I}\{2\pi\bar{I}_{\epsilon,k} \leq \eta\} - \int_0^\eta e^{-x}\log x\,dx\right\} = 0.$$

Let n_T' be the number of values of k for which $2\pi\bar{I}_{\epsilon,k}$ is less than $2m^{-c}$, for fixed $c > 0$. Then we argue as in that reference that

$$p\lim_{N\to\infty}\sup \frac{n_T'\log m}{\rho m^{1/2}} = 0$$

so $n_T' = O_p(\rho m^{1/2}\log^{-1}m) = o_p(T)$, $T = m^{1/2}\log^{-1/2}m$. Then, $\bar{I}_k f_k^{-1} < m^{-c}$ implies that $2\pi\bar{I}_{\epsilon,k} < 2m^{-c}$ or $|\delta_k| \geq m^{-c}$. Then $n_T = O_p(T)$, with $c < \frac{1}{4}$.

Then denoting $L_k = (\log \bar{I}_k f_k^{-1} - \log 2\pi\bar{I}_{\epsilon,k})\log^{1/2}m$, for some $C > 0$ and any $\varepsilon > 0$ fixed as $N \to \infty$,

$$P\left\{\max_{k\in A_T}|L_k| > \varepsilon\right\} = P\left\{\max_{k\in A_T}|L_k| > \varepsilon \cap n_T \leq CT\right\} + P\left\{\max_{k\in A_T}|L_k| > \varepsilon \cap n_T > CT\right\}$$

$$\leq P\left\{\max_{k\in A_T}|L_k| > \varepsilon/n_T \leq CT\right\} + P\{n_T > CT\},$$

where $P\{n_T > CT\} = o(1)$ as $N \to \infty$. Then, with Bonferroni's inequality and $L_k^+ = L_k\mathbb{I}\{k \in A_T\}$,

$$P\left\{\max_{k\in A_T}|L_k| > \varepsilon/n_T \leq CT\right\} \leq P\left\{\max_k|L_k^+| > \varepsilon/n_T \leq CT\right\}$$

$$\leq CT\max_k P\{|L_k^+| > \varepsilon/n_T \leq CT\},$$

noting that $P\{|L_k^+| > \varepsilon/n_T \leq CT\} = 0$ for all but n_T values of k because $|L_k^+| = 0$ for all k but at most $n_T \leq CT$. Then, because $P\{n_T \leq CT\} \to 1$ as $N \to \infty$, we can write for some $s > 0$ that

$$P\{|L_k^+| > \varepsilon/n_T \leq CT\} \leq P\{|L_k^+| > \varepsilon \cap n_T \leq CT\}P\{n_T \leq CT\}^{-1}$$

$$\leq sP\{|L_k^+| > \varepsilon \cap n_T \leq CT\}$$

$$\leq sP\{|L_k| > \varepsilon\} = O(k^{-J/(J+2)+a}\log m),$$

for each $a > 0$ using (A.8), which follows, $J \geq 3$. Then $\max_k P\{|L_k^+| > \varepsilon/n_T \leq CT\} = O(\ell^{-J/(J+2)+a} \log m)$ and therefore $\max_{k \in A_T} |L_k| = o_p(1)$, $m^{1/2} \ell^{-J/(J+2)+a} \log^{1/2} m \to 0$ as $N \to \infty$ for some $a > 0$, with Assumption 5.

Then the contribution from A_T is

$$\left| \left(\sum_k \Lambda_k^2 \right)^{-1} \sum_{A_T} \Lambda_k \{ \log I_k^{(J)} f_k^{-1} - \log 2\pi I_{\epsilon,k}^{(J)} \} \right|$$

$$\leq \left(\sum_k \Lambda_k^2 \right)^{-1} \sum_{A_T} |\Lambda_k| |L_k| \log^{-1/2} m$$

$$= O_P \left(m^{-1} T \log^{1/2} m \max_{k \in A_T} |L_k| \right)$$

$$= o_P(m^{-1/2}) = o_P(1).$$

It remains to bound $P\{|L_k| > \varepsilon\}$. For $\varepsilon > 0$ fixed as $N \to \infty$ (any $a > 0$), on choosing $A_k \to 0$ optimally as $k \to \infty$, $J \geq 3$, $\varepsilon_N = \varepsilon \log^{-1/2} m$ and $a' > 0$ small enough,

$$P\{|L_k| > \varepsilon\} \leq P \left\{ \left| \frac{\delta_k}{2\pi \bar{I}_{\epsilon,k}} \right| > \varepsilon_N \right\} + P \left\{ \left| \frac{\delta_k}{2\pi \bar{I}_{\epsilon,k}} \right| > 1 - e^{-\varepsilon_N} \right\}$$

$$\leq P \left\{ \left| \frac{\delta_k}{2\pi \bar{I}_{\epsilon,k}} \right| > \varepsilon_N \right\} + P \left\{ \left| \frac{\delta_k}{2\pi \bar{I}_{\epsilon,k}} \right| > \varepsilon_N/2 \right\}$$

$$\leq 2P\{|2\pi \bar{I}_{\epsilon,k}| < 2A_k/\sqrt{\varepsilon_N}\} + 2P\{|\delta_k| > A_k \sqrt{\varepsilon_N}/2\}$$

$$= O(\varepsilon_N^{(a'-J)/2} A_k^{J-a'} + \varepsilon_N^{-1} A_k^{-2} k^{-1} \log k) = O(\varepsilon_N^{-2} k^{-J/(J+2)+a}). \tag{A.8}$$

Then to prove the consistency of the estimate $\hat{d}$ we only need to calculate the first two moments of the random variable ξ_N. To evaluate the moments of $\log \bar{I}_{\epsilon,k}$, we approximate the probability density $q_N(\mathbf{y})$ with Chen and Hannan's Lemma 2. This result uses some results in Bhattacharya and Rao (1976) to approximate the density of the Fourier transform $w_\epsilon(\lambda)$ of the sequence ϵ_t. These researchers employed a finite fifth moment of ϵ_t to get a stronger result. For our purposes Lemma 2 is enough.

Set $4\pi \bar{I}_{ek} = \sum_{j=1}^{J} (y_{aj}^2 + y_{bj}^2)$, where y_{aj} and y_{bj} correspond to the sine and cosine components, respectively, of $4\pi I_{\epsilon,j}$. Now, from Chen and Hannan (1980, Lemma 2),

$$P_1(-\phi : \bar{\chi}_{\nu,N})(\mathbf{y}) = \sum_{|\nu|=3} \frac{\bar{\chi}_{\nu,N}}{v!} \prod_{j=1}^{2k} \left(\frac{\partial}{\partial y_i} \right)^{\nu_j} \phi(\mathbf{y}), \qquad v! = \prod_j v_j!.$$

As $|\nu| = 3$ the terms in P_1 are of one of the following types when we are considering the joint distribution in $\mathbb{R}^{4J}$ of the sine and cosine components $4\pi \bar{I}_{\epsilon,k}$ and $4\pi \bar{I}_{\epsilon,k'}, k \neq k'$ (up to constants).

1. $H_3(y_s)\phi(\mathbf{y})$, where H_i are the Hermite polynomials of order i and $s \in \{1,\ldots,4J\}$,

   $$H_i(x)\phi(x) = (-1)^i \left(\frac{\partial}{\partial x} \right)^i \phi(x), \qquad x \in \mathbb{R}.$$

   Then this term is odd in the component y_s of $\mathbf{y}$ (because H_3 is odd and ϕ is even).
2. $H_2(y_s)H_1(y_r)\phi(\mathbf{y}), y_r \neq y_s$ and $r,s \in \{1,\ldots,4J\}$. Then this term is odd in y_r.
3. $H_1(y_s)H_1(y_r)H_1(y_u)\phi(\mathbf{y}), y_r, y_s, y_u$ all different. Then this term is odd in y_s, y_r, and y_u.

If $k = k'$, we consider only a distribution in $\mathbb{R}^{2J}$, and the typical terms of P_1 are as follows.

1. $H_3(y_s)\phi(\mathbf{y})$, where $s \in \{1,\ldots,2J\}$. Then this term is odd in the component y_s of $\mathbf{y}$.
2. $H_2(y_s)H_1(y_r)\phi(\mathbf{y}), r \neq s$ and $r, s \in \{1,\ldots,2\}$. Then this term is odd in the component y_r.

Then we have

$$E[\log 2\pi\bar{I}_{ek}] + \log 2 = \int_{\mathbb{R}^{2J}} \log\left(\sum_j (y_{aj}^2 + y_{bj}^2)\right) q_N(\mathbf{y})\, d\mathbf{y}$$

$$= \int_{\mathbb{R}^{2J}} \log\left(\sum_j (y_{aj}^2 + y_{bj}^2)\right) [\phi(\mathbf{y}) + N^{-1/2}P_1(\mathbf{y})]\, d\mathbf{y} + O(N^{-1})$$

$$= \psi(J) + \log 2 + O(N^{-1}),$$

because $\int_0^\infty (\log x)^h/(1 + x^5)\, dx < \infty$ and $\int_0^\infty (x \log x)^h e^{-x}\, dx < \infty$, for all $h \geq 0$. Here $\psi(z) = d/dz \log \Gamma[z]$ is the digamma function. The contribution from $P_1(\mathbf{y})$ is 0 because the interval of integration is $(-\infty,\infty)$ and P_1 is always odd in one component of $\mathbf{y}$ and the log term is even in all the components.

Consider now the covariance terms. Denote $E_k = E[\log 2\pi\bar{I}_{ek}]$. Then, $(k \neq k')$,

$\text{Cov}[\log 2\pi\bar{I}_{ek}, \log 2\pi\bar{I}_{ek'}]$

$$= \int_{\mathbb{R}^{4J}} \left[\log\left(\sum_j (y_{aj}^2 + y_{bj}^2)\right) - E_k\right]\left[\log\left(\sum_{j'} (y_{aj'}^2 + y_{bj'}^2)\right) - E_{k'}\right]$$

$$\times [\phi(\mathbf{y}) + N^{-1/2}P_1(\mathbf{y})]\, d\mathbf{y} + O(N^{-1})$$

$$= N^{-1/2} \int_{\mathbb{R}^{4J}} \left[\log\left(\sum_j (y_{aj}^2 + y_{bj}^2)\right) - E_k\right]$$

$$\times \left[\log\left(\sum_{j'} (y_{aj'}^2 + y_{bj'}^2)\right) - E_{k'}\right] P_1(\mathbf{y}) + O(N^{-1})$$

$$= O(N^{-1}),$$

as $\phi(\mathbf{y})$ is the density of the standard normal density in $\mathbb{R}^{4J}$ (with uncorrelated components) and because the contribution from P_1 cancels out by the same argument as before. The variance is

$$\text{Var}[\log 2\pi\bar{I}_{ek}] = \int_{\mathbb{R}^{2J}} \left[\log\left(\sum_j (y_{aj}^2 + y_{bj}^2)\right) - E_k\right]^2 q_N(\mathbf{y})\, d\mathbf{y}$$

$$= \int_{\mathbb{R}^{2J}} \left[\log\left(\sum_j (y_{aj}^2 + y_{bj}^2)\right) - E_k\right]^2 [\phi(\mathbf{y}) + N^{-1/2}P_1(\mathbf{y})]\, d\mathbf{y} + O(N^{-1})$$

$$= \psi'(J) + O(N^{-1}),$$

reasoning as before. Then, using (A.3), it is immediate that $E[\xi_N] = O(N^{-1})$ and that

$$\text{Var}[\xi_N] = \frac{J}{4m}\psi'(J) + O(N^{-1}) + o(m^{-1}) \sim \frac{J}{4m}\psi'(J).$$

Therefore $\xi_N = o_P(1)$ with Assumption 5, and the theorem is proved. ∎

Proof of Theorem 2. We use exactly the same method as in Robinson (1995a) or in Theorem 4 of Velasco (1999). These proofs are based on the properties of the Dirichlet kernel $D_N(\lambda) = \sum_{t=1}^{N} e^{it\lambda}$ and the Fejér kernel $K_N(\lambda) = (2\pi N)^{-1}|D_N(\lambda)|^2$. The spectral kernel for the tapered periodogram, corresponding to the Fejér kernel $K_N(\lambda)$ for the raw periodogram, is

$$K_N^T(\lambda) = \frac{1}{2\pi \sum h_t^2} \left| \sum_{t=1}^{N} h_t \exp\{it\lambda\} \right|^2 = \frac{1}{2\pi \sum h_t^2} |D_N^T(\lambda)|^2,$$

where $D_N^T(\lambda)$ is the equivalent of the Dirichlet kernel $D_N(\lambda)$ in the nontapered case. Obviously

$$D_N^T(\lambda_j) = \frac{1}{\sqrt{6}} \{2D_N(\lambda_j) - D_N(\lambda_{j-1}) - D_N(\lambda_{j+1})\}.$$

It can be seen that $K_N^T(\lambda)$ is even, positive, integrates to one and satisfies (see, e.g., Bloomfield, 1976; Hannan, 1970, p. 265) $\sup_{\lambda,N} |K_N^T(\lambda)| = O(\min\{N, N^{-5}|\lambda|^{-6}\})$, $\lambda \in [-\pi, \pi]$, where this property follows from $\sup_{\lambda,N} |D_N^T(\lambda)| = O(\min\{N, N^{-2}|\lambda|^{-3}\})$. This implies that the tapered periodogram has improved asymptotic properties with respect to the usual periodogram, because the tails of the kernel $K_N^T(\lambda)$ decrease much faster with the frequency and with the sample size than the tails of the Fejér kernel (with bound $O(N^{-1}|\lambda|^{-2})$), though they are no longer orthogonal. However, using the properties of the Dirichlet kernel $D_N(\lambda)$, we have that, for $3 \leq j + k \leq N - 3$,

$$\int_{-\pi}^{\pi} D_N^T(\lambda_j - \lambda) D_N^T(\lambda + \lambda_k) \, d\lambda = 0. \tag{A.9}$$

We consider the same intervals of integration as in Robinson(1995a) to analyze the bias of the tapered periodogram in

$$E[|w^T(\lambda_j)|^2] - f(\lambda_j) = \int_{-\pi}^{\pi} [f(\lambda) - f(\lambda_j)] K_N^T(\lambda_j - \lambda) \, d\lambda.$$

We only analyze here in detail the interval $[\lambda_j/2, 3\lambda_j/2]$ (for details see Velasco, 1999). Using (3),

$$\left| \int_{\lambda_j/2}^{3\lambda_j/2} [f(\lambda) - f(\lambda_j)] K_N^T(\lambda_j - \lambda) \, d\lambda \right|$$

$$= \left| \int_{-\lambda_j/2}^{\lambda_j/2} [f(\lambda_j - \lambda) - f(\lambda_j)] K_N^T(\lambda) \, d\lambda \right|$$

$$= \left| \int_{-\lambda_j/2}^{\lambda_j/2} [\lambda \times f'(\lambda_j) + O(\lambda_j^{-\alpha-2d}|\lambda|^\alpha)] K_N^T(\lambda) \, d\lambda \right|$$

$$= O\left(\lambda_j^{-\alpha-2d} \int_{-\lambda_j/2}^{\lambda_j/2} |\lambda|^\alpha K_N^T(\lambda) \, d\lambda \right),$$

because K_N^T is even and we integrate in a symmetric interval around 0. Now, with $\alpha \in (1,2]$,

$$\int_{-\lambda_j/2}^{\lambda_j/2} |\lambda|^\alpha K_N^T(\lambda) = 2\left\{ \int_0^{N^{-1}} + \int_{N^{-1}}^{\lambda_j/2} \right\} \lambda^\alpha K_N^T(\lambda)\, d\lambda$$

$$= O\left(N \int_0^{N^{-1}} \lambda^\alpha \, d\lambda + N^{-5} \int_{N^{-1}}^{\lambda_j/2} \lambda^{\alpha-6} \, d\lambda \right) = O(N^{-\alpha}),$$

and therefore this integral is $O(\lambda_j^{-\alpha-2d} \times N^{-\alpha}) = O(f(\lambda_j) \times j^{-\alpha})$. The analysis of the covariances between tapered Fourier transforms of X_t follows the same lines as in Robinson (1995a) and Velasco (1999). The additional term $O([j/N]^\alpha)$ that shows up in the theorem when we normalize with respect to $G\lambda_j^{-2d}$ (instead of with respect to $f(\lambda_j)$) follows as in Robinson (1995a). ∎

Proof of Theorem 3. We do this in three steps.

First. We argue that all the previous results concerning the asymptotic distribution of $\bar{I}_{\epsilon j}$, $j \in \{j_{(1)}, j_{(2)}, \ldots, j_{(k)}\}$, still go through for the tapered version $\bar{I}_{\epsilon,j}^T$ if $j_{(1)} + 2 < j_{(2)}, \ldots, j_{(k-1)} + 2 < j_{(k)}$.

The reason is the following: the results of Chen and Hannan (1980) are based on the exact uncorrelatedness of the discrete Fourier transform of the i.i.d. sequence of ϵ_t at different Fourier frequencies, so the periodogram ordinates are approximately independent. Therefore, the real and imaginary components of the tapered Fourier transforms of ϵ_t are still exactly uncorrelated if we consider only one periodogram ordinate of every three, as we did in the definition of $\hat{d}^T$. Then an equivalent Edgeworth expansion for the density of the vector of real and imaginary components of $\bar{I}_{\epsilon,k}^T$ is valid as before, because each of the tapered Fourier transforms in a (fixed) linear combination of three Fourier transforms with valid Edgeworth expansions for their densities.

Second. Following the proof of Theorem 1 we obtain, $\alpha \le 2$,

$$\hat{d}^T = d + \left(\sum_k \Lambda_k^2 \right)^{-1} \sum_k \Lambda_k \log 2\pi \bar{I}_{\epsilon,k}^T + O([mN^{-1}]^\alpha) + R^T$$

$$= d + \xi_N^T + R^T + o_P(m^{-1/2}),$$

say, using Assumption 7. Hence, the asymptotic distribution of $m^{1/2}(\hat{d}^T - d)$ can be approximated by that of $m^{1/2}\xi_N^T$ if $R^T = o_p(m^{-1/2})$, where

$$R^T = \left(\sum_k \Lambda_k^2 \right)^{-1} \sum_k \Lambda_k \{ \log \bar{I}_k^T f_k^{-1} - \log 2\pi \bar{I}_{\epsilon,k}^T \} = \sum_{\bar{A}_T} + \sum_{A_T}$$

with the same definitions as in the proof of Theorem 1 but with tapered observations.

Then for $k \in \bar{A}_T$, we have that for $\alpha > 1$ and $\varepsilon > 0$ sufficiently small,

$$E\left| \left(\sum_k \Lambda_k^2 \right)^{-1} \sum_{\bar{A}_T} \Lambda_k \{ \log \bar{I}_k^T f_k^{-1} - \log 2\pi \bar{I}_{\epsilon,k}^T \} \right|$$

$$\le \left(\sum_k \Lambda_k^2 \right)^{-1} \sum_{\bar{A}_T} |\Lambda_k| E |\log \bar{I}_k f_k^{-1} - \log 2\pi \bar{I}_{\epsilon,k}^T|$$

$$= O\left(m^{-1} \log^2 m \sum_k k^{\varepsilon-\alpha/2} \right) = O(m^{\varepsilon-\alpha/2} \log^2 m) = o(m^{-1/2}),$$

with Assumption 7, because from (A.10)–(A.12) we obtain, $J \geq 3$, for any $\varepsilon > 0$,

$$E|\log \bar{I}_k f_k^{-1} - \log 2\pi \bar{I}_{\epsilon,k}^T|$$

$$\leq E\left|\frac{\delta_k}{2\pi I_{\epsilon,k}^T}\right| + E\left[(|\log \bar{I}_k^*| + |\log 2\pi \bar{I}_{\epsilon,k}^T|)\mathbb{I}\left\{|\delta_k/(2\pi I_{\epsilon,k}^T)| \geq \frac{1}{2}\right\}\right]$$

$$= O\left(k^{-\alpha/2} + \log m \times P\left\{|\delta_k/(2\pi I_{\epsilon,k}^T)| \geq \frac{1}{2}\right\}^{1-\varepsilon}\right) = O(k^{-\alpha/2} \log m).$$

As in the nontapered case, for any fixed integer $b \geq 1$,

$$E[|\log I_k^*|^b] = O(\log^b m). \tag{A.10}$$

Then, for $\alpha \in (1,2]$

$$P\left\{|\delta_k/(2\pi \bar{I}_{\epsilon,k}^T)| \geq \frac{1}{2}\right\} \leq 2E\left|\frac{\delta_k}{2\pi \bar{I}_{\epsilon,k}^T}\right| = O(k^{-\alpha/2}), \tag{A.11}$$

because, using Lemma 3,

$$E\left|\frac{\delta_k}{2\pi \bar{I}_{\epsilon,k}^T}\right| \leq E[\delta_k^2]^{1/2} E[(I_{\epsilon,k}^T)^{-2}]^{1/2} = O(k^{-\alpha/2}), \tag{A.12}$$

because $E[\{\bar{I}_{\epsilon,k}^T\}^{a'-J}] < \infty$, from Lemma 1, $0 < a' < J$.

We now estimate n_T. As before, let n_T' be the number of values of k for which $2\pi \bar{I}_{\epsilon,k}^T$ is less than $2m^{-c}$, for any fixed $c < \frac{1}{4}$, so $n_T' = O_p(\rho m^{1/2} \log^{-1} m)$ for any increasing sequence ρ and $n_T = O_p(\rho m^{1/2} \log^{-1} m)$.

Denote $\gamma = (\alpha J/(J+2) - 1)/2 > 0$. Now we can proceed in a simpler way than without tapering. Denoting $L_k = (\log \bar{I}_k^T f_k^{-1} - \log 2\pi \bar{I}_{\epsilon,k}^T)\gamma - a$, for any $\varepsilon > 0$ fixed and for any $a > 0$ such that $a < \gamma$,

$$P\left\{\max_k |L_k| > \varepsilon\right\} \leq \sum_k P\{|L_k| > \varepsilon\} = O\left(\ell^{2(\gamma-a)} \sum_k k^{a-1-2\gamma}\right) = O(\ell^{-a}) = o(1),$$

using (A.13), which follows. So $\max_k |L_k| = o_p(1)$. Then for $a > 0$ small enough

$$\left|\left(\sum_k \Lambda_k^2\right)^{-1} \sum_{A_T} \Lambda_k\{\log \bar{I}_k^T f_k^{-1} - \log 2\pi \bar{I}_{\epsilon,k}^T\}\right|$$

$$\leq \left(\sum_k \Lambda_k^2\right)^{-1} \sum_{A_T} |\Lambda_k||L_k|\ell^{a-\gamma}$$

$$= O_P\left(\rho m^{-1/2} \ell^{a-\gamma} \max_{k \in A_T} |L_k|\right)$$

$$= o_P(\rho m^{-1/2} \ell^{a-\gamma}) = o_P(m^{-1/2}).$$

The bound for $P\{|L_k| > \varepsilon\}$ follows as (A.8), because for any $\varepsilon > 0$ fixed with N (and any $a > 0$), on choosing $A_k \to 0$ as $k \to \infty$ optimally and $a' > 0$ small enough, $\varepsilon_N = \varepsilon \ell^{a-\gamma}$,

$$P\{|L_k| > \varepsilon\} \le 2P\{|2\pi\bar{I}^T_{\epsilon,k}| < A_k/\sqrt{\varepsilon_N}\} + 2P\{|\delta_k| > A_k\sqrt{\varepsilon_N}\}$$

$$= O(\varepsilon_N^{(a'-J)/2}A_k^{J-a'} + \varepsilon_N^{-1}A_k^{-2}k^{-\alpha}) = O(\varepsilon_N^{-2}k^{a-1-2\gamma}). \tag{A.13}$$

Third. Denoting by $\xi_N^{T,*}$ and $\log 2\pi\bar{I}^{T,*}_{\epsilon,k}$ the corresponding random variables when the ϵ_t are Gaussian with the same first two moments, we follow related arguments to those of Robinson (1995a); we show that the moments of all orders of $m^{1/2}\xi_N^T$ converge to those of $m^{1/2}\xi_N^{T,*}$ that are bounded, from the proof of Theorem 3 of Robinson (1995a). Next, the uncorrelatedness of the real and imaginary components of $\bar{I}^{T,*}_{\epsilon,k}$ for different frequencies implies the independence and equal distribution (due to the Gaussianity) of $\bar{I}^{T,*}_{\epsilon,k}$ and of $\log \bar{I}^{T,*}_{\epsilon,k}$ at different frequencies. Therefore, $m^{1/2}\xi_N^{T,*}$ is a sum of i.i.d. variables with bounded moments and by the Lindberg–Feller CLT it is asymptotically normal (the first two moments can be obtained from the proof of Theorem 1). Because each moment of the variate $m^{1/2}\xi_N^{T,*}$ is bounded uniformly in N, all these moments converge to those of the corresponding normal distribution. Hence, as all moments of $m^{1/2}\xi_N^T$ converge to those of $m^{1/2}\xi_N^{T,*}$, $m^{1/2}\xi_N^T$ is easily found asymptotically normal distributed by the method of moments.

Therefore, it only remains to prove that the moments of all orders of $m^{1/2}\xi_N^T$ converge to those of $m^{1/2}\xi_N^{T,*}$, so there is not influence from the higher order cumulants of ϵ_t, but this is Lemma 4. ∎

LEMMA 4. *Under the assumptions of Theorem 3, the moments of all orders of $m^{1/2}\xi_N^T$ converge to those of $m^{1/2}\xi_N^{T,*}$ as $N \to \infty$.*

Proof of Lemma 4. To make our arguments clearer we consider in an initial stage the nontapered and nonpooled ($J = 1$) case. After this, we will show that the same conclusions apply for the tapered case for any $J > 1$.

Following the arguments of the second part in the proof of Theorem 1, we can check that the moments and cross moments of all orders of $\log 2\pi\bar{I}^T_{\epsilon,k}, k = \ell + 3J, \ldots, m$ converge to those we would obtain if the ϵ_t's were actually Gaussian, with error $O(N^{-1})$, because $\int_0^\infty (\log x)^a (1 + x^4)^{-1}\, dx < \infty$ for all $a \ge 0$.

However, this result is not enough to approximate the moments of $m^{1/2}\xi_N^T$, which is an (increasing) weighted sum of the $\log 2\pi\bar{I}^T_{\epsilon,k}$. When ϵ_t has bounded moments of all orders, we can obtain an Edgeworth expansion for the density of the Fourier transform of ϵ_t of any order s fixed, under the same assumptions of Lemma 2. In Chen and Hannan (1980) the second term P_2 is presented, although it is not totally correct in their notation. The exact shape of these higher order terms in a general Edgeworth approximation is fundamental for our proof, and we dedicate some space to that.

Edgeworth Approximation. The validity of an Edgeworth approximation for the real and imaginary components of the discrete Fourier transform of ϵ_t of any order $s > 1$, when enough moments exist, follows from Lemma 2 of Chen and Hannan (1980), because their proof generalizes immediately for any order of approximation, not just 2. For any $s = 0, 1, \ldots$, fixed, we can obtain that the vector W_N (see Lemma 2) has density q_N for all sufficiently large N and

$$\sup_{\mathbf{y}\in\mathbb{R}^{2k}} (1 - \|\mathbf{y}\|^s) \left| q_N(\mathbf{y}) - \sum_{r=0}^{s} N^{-r/2} P_r(-\phi : \bar{\chi}_{\nu,N})(\mathbf{y}) \right| = O(N^{-(s+1)/2}), \tag{A.14}$$

where P_r are polynomials with coefficients depending on the joint cumulants of $Y_t, \bar{\chi}_{\nu,N}$, multiplied by the $2k$th standard multivariate normal density ϕ (given the covariance structure of the discrete Fourier transforms). Following Bhattacharya and Rao (1976), we find that $P_r(-\phi : \bar{\chi}_{\nu,N}) = \tilde{P}_r(-D : \bar{\chi}_{\nu,N})\phi$, where, for nonnegative integer vectors $\nu = (\nu(1), \nu(2), \ldots, \nu(2k))$ of $2k$ dimensions,

$$D^\nu = \left(\frac{\partial}{\partial y_1}\right)^{\nu(1)} \cdots \left(\frac{\partial}{\partial y_{2k}}\right)^{\nu(2k)},$$

with

$$\tilde{P}_r(z : \bar{\chi}_{\nu,N}) = \sum_{n=1}^{r} \frac{1}{n!} \left\{ \sum_{j_1,\ldots,j_n}{}^* \frac{\chi_{j_1}(z)}{j_1!} \frac{\chi_{j_2}(z)}{j_2!} \cdots \frac{\chi_{j_n}(z)}{j_n!} \right\}$$

$$= \sum_{n=1}^{r} \frac{1}{n!} \left\{ \sum_{j_1,\ldots,j_n}{}^* \left(\sum{}^{**} \frac{\chi_{\nu_1} \cdots \chi_{\nu_n}}{\nu_1! \cdots \nu_n!} z^{\nu_1 + \cdots + \nu_n} \right) \right\}. \tag{A.15}$$

The summation $\sum^*$ is over all n-tuples of positive integers $(j_1, \ldots, j_n)$ satisfying

$$\sum_{i=1}^{n} j_i = r, \qquad j_i = 1, 2, \ldots, r \qquad (1 \le i \le n), \tag{A.16}$$

and the $\sum^{**}$ denotes summation over all n-tuples of nonnegative integral vectors $(\nu_1, \ldots, \nu_n)$ satisfying $|\nu_i| = j_i + 2$, $(1 \le i \le n)$, where we use the usual multivariate notation, $|\nu_i| = \sum_{j=1}^{2k} \nu_i(j)$ (for details see Bhattacharya and Rao, 1976). In particular, $\tilde{P}_0 \equiv 1$,

$$\tilde{P}_1(z : \bar{\chi}_{\nu,N}) = \frac{\chi_3(z)}{3!} = \sum_{|\nu|=3} \frac{\bar{\chi}_{\nu,N}}{\nu!} z^\nu,$$

$$\tilde{P}_2(z : \bar{\chi}_{\nu,N}) = \frac{\chi_4(z)}{4!} + \frac{1}{2!} \left(\frac{\chi_3(z)}{3!}\right)^2,$$

$$\tilde{P}_3(z : \bar{\chi}_{\nu,N}) = \frac{\chi_5(z)}{5!} + \frac{\chi_4(z)\chi_3(z)}{3!4!} + \frac{1}{3!} \left(\frac{\chi_3(z)}{3!}\right)^3,$$

and in general

$$\chi_r(z) = \sum_{|\nu|=r} \frac{\bar{\chi}_{\nu,N}}{\nu!} z^\nu,$$

where $\nu_i! = \nu_i(1)! \cdots \nu_i(2k)!$, $0! = 1$. Then $P_r(\mathbf{y})$ is a polynomial in the components of $\mathbf{y}$ (times ϕ), with coefficients that are functions of the joint cumulants of Y_t of order ν (i.e.,

of the components in the vector Y_t with exponent in ν different from zero), $\bar{\chi}_{\nu,N}$, and of the Hermite polynomials of order ν, $H_\nu(\mathbf{y})$, obtained from (the derivatives of) $\phi(\mathbf{y})$. Following the discussion in the proof of Theorem 1 and the preceding comments, we stress some properties that we will use later.

Using (A.14) the first term of the expectations of functions of the periodogram of ϵ_t is always exactly equal to the Gaussian expectation, so we only need to concentrate on the higher order terms of an approximation up to a finite order to be determined later.

When $|\nu|$ is odd, the polynomial function $H_\nu(\mathbf{y})$ is odd in at least one of the components of $\mathbf{y}$. Then, all the summands in P_r with r odd also will be odd in at least one of the components of $\mathbf{y}$. As we are going to consider the expectation of even functions of $\mathbf{y}$ (i.e., logarithm of the periodogram minus a constant), we need to consider only terms P_r with r even ($r = 0, 2, \ldots$).

The cumulants $\bar{\chi}_{\nu,N}$ are exactly zero in many situations (i.e., for many vectors ν), because of the special nature of the vector W_N, because $\sum_{t=1}^{N} e^{it\lambda_k} = 0$, $\forall k \neq 0 \pmod{N}$. In other cases these cumulants are different from zero only under linear restrictions on some of the frequencies $(\lambda_{j(1)}, \ldots, \lambda_{j(2k)})$ of the periodogram ordinates that we are considering in each moment.

Moments. Because $\sum_k \Lambda_k \equiv 0$, in contrast with the proof of Theorem 1, we substitute now in the definition of ξ_N the actual mean of $\log 2\pi I_{\epsilon,k}$ by the mean it would have in the Gaussian case, $\psi(J)$, obtaining (without need to make explicit the log 2 adjustment), with $J = 1$,

$$\xi_N = \left(\sum_k \Lambda_k^2\right)^{-1} \sum_k \Lambda_k(\log 2\pi I_{\epsilon,k} - \psi(1)).$$

Denote by E_s and E_s^* the sth moments of $m^{1/2}\xi_N^T$ and $m^{1/2}\xi_N^{T,*}$, respectively. Then, for $s = 3, 4, \ldots$,

$$E_s = m^{s/2} \left(\sum_k \Lambda_k^2\right)^{-s} \sum_{j=1}^{s} \sum_p c_p \sum_{k(1)} \sum_{k(2)\neq} \cdots \sum_{k(j)\neq} \Lambda_{k(1)}^{p(1)} \Lambda_{k(2)}^{p(2)} \ldots \Lambda_{k(j)}^{p(j)}$$

$$\times E\left[\prod_{q=1}^{j} (\log I_{\epsilon,k(q)}^T - \psi(1))^{p(q)}\right], \tag{A.17}$$

where the index $k(i) \neq$ means that the summation is for all the values of $k(i) \neq k(1), \ldots, k(i-1)$ (so we only make explicit frequencies that are always different) and the sum in p is for all vectors of positive integers $(p(1), \ldots, p(j))$ such that $\sum_{i=1}^{j} p(i) = s$, and c_p is a combinatorial number that depends only on p. Obviously, when $j = s$, all $p(i) = 1$ and when $j = 1$, $p(1) = s$. Recall also that $(\sum \Lambda_k^2)^{-1} = O(m^{-1})$.

The key idea is to substitute each of the expectations in (A.17) by an integral over $\mathbb{R}^{2j}$, approximating the true probability density of the vector of periodogram ordinates by a $2j$-dimensional Edgeworth expansion of the form (A.14). The first term (in N^0) of the Edgeworth approximation always gives the corresponding Gaussian expectation E_s^* whereas the odd terms cancel out, so

$$E_s - E_s^* = m^{s/2}\left(\sum \Lambda_k^2\right)^{-s} \sum_{j=1}^{s} \sum_p c_p \sum_{k(1)} \sum_{k(2)\neq} \cdots \sum_{k(j)\neq} \Lambda_{k(1)}^{p(1)} \Lambda_{k(2)}^{p(2)} \cdots \Lambda_{k(j)}^{p(j)}$$

$$\times \int_{\mathbb{R}^{2j}} \left\{ \sum_{r=2}^{r^{\max}} P_r(\mathbf{y}) N^{-r/2} + O(N^{-(r^{\max}+1)/2})(1 + \|\mathbf{y}\|^4)^{-1} \right\}$$

$$\times \prod_{q=1}^{j} (\log(y_{a,k(q)}^2 + y_{b,k(q)}^2) - \psi(1))^{p(q)} \, d\mathbf{y} \qquad \textbf{(A.18)}$$

(with $r_{\max} \le s$ to be determined later). Then we need to check that the contribution from all higher order terms with $r \ge 2$ in (A.18) is negligible. First, for s fixed, we study all the terms in (A.18) with different values of j.

Consider first the terms in (A.18) for which $j \le 1 + s/2$. Using (A.3) and that $\int_0^\infty (\log x)^b (1 + x^4)^{-1} \, dx < \infty$, $b \ge 0$, the contribution to $E_s - E_s^*$ of each of the higher order terms P_r, $r > 0$, is $O(m^{-s/2+j}N^{-1}) = O(mN^{-1}) = o(1)$, just using the order of magnitude of the error term of an Edgeworth approximation with only P_0 and P_2, because the term P_1 cancels out.

Therefore, we only need to consider terms where $j > 1 + s/2$. The main idea to deal with these terms is the following. Because we have $j > 1 + s/2$ summands, there should be some of them, h, say, with exponent $p(i) = 1$. In fact $h \ge 2j - s \ge 3$. Then, whenever $h > 0$ the leading term in the approximation for the corresponding expectation (the Gaussian part) is exactly zero (i.e., $E_s^* = 0$), given the uncorrelatedness of the discrete Fourier transform at Fourier frequencies (even in the non-Gaussian case). We will argue that this orthogonality property of the first (Gaussian) term is transferred to some extent to the higher order terms. The reason is that for each periodogram ordinate (i.e., for each pair of variables in $\mathbf{y}$), some of the contributions from the higher order terms in (A.14) are still the Gaussian ones given by $\phi(y_a)\phi(y_b)$ (i.e., we have not taken derivatives w.r.t. those variables), given null contribution for the whole expectation when this periodogram ordinate has exponent $p(i) = 1$. (The same argument is valid for any exponent $p(i)$ odd, but we do not need it.)

We illustrate this idea with an example. Consider P_2, with

$$\tilde{P}_2(z : \bar{\chi}_{\nu,N}) = \sum_{|\nu|=4} \frac{\bar{\chi}_{\nu,N}}{\nu!} z^\nu + \frac{1}{2}\left(\sum_{|\nu|=3} \frac{\bar{\chi}_{\nu,N}}{\nu!} z^\nu\right)^2,$$

$$= \sum_{|\nu|=4} \frac{\bar{\chi}_{\nu,N}}{\nu!} z^\nu + \frac{1}{2} \sum_{|\nu|=3} \left(\frac{\bar{\chi}_{\nu,N}}{\nu!}\right)^2 z^{2\nu}$$

$$+ \frac{1}{2} \sum_{|\nu|=3} \sum_{|\nu'|=3, \nu\neq\nu'} \frac{\bar{\chi}_{\nu,N}}{\nu!} \frac{\bar{\chi}_{\nu',N}}{\nu'!} z^{\nu+\nu'}.$$

When we substitute z by $-D$, to obtain P_2, we observe that in each of the terms of the last expression we take at most 4, 3, or 6 derivatives, respectively, with respect to the vector of $2j$ components $\mathbf{y}$. Therefore, all but at most 4, 3, or 6 functions $\phi(y_i)$ in $\phi \equiv \phi(y_1) \cdots \phi(y_{2j})$ are not affected by the differential operator.

Then, for each periodogram ordinate, with frequency $\lambda_{j(i)}$, and each of the terms in P_2 (with ν, 2ν or ν and ν'), we obtain the following.

If a periodogram has exponent $p(i) = 1$ and neither of its two components in **y** is included in ν (or in ν'), then we have in (A.18) an integral of the form

$$\int_{\mathbb{R}} \int_{\mathbb{R}} [\log(y_a^2 + y_b^2) - \psi(1)]\phi(y_a)\phi(y_b)\, dy_a\, dy_b \equiv 0, \tag{A.19}$$

and this term does not contribute, because the whole integral is zero because the variables y_a and y_b do not appear in any other factor of this particular summation term in (A.18).

For any exponent $p(i)$, if any component of the vector **y** is included in ν (or in ν') with odd coordinate, then this term in P_2 will be odd in that variable y_a, say, and again the contribution of these terms is null, because the periodogram is even in its real and imaginary components.

In conclusion, we need at least two derivatives with respect to one of the two variables that have $p(i) = 1$, that is, only the terms in P_r that consider Hermite polynomials with an even number in the order vector corresponding to one of the two variables with $p(i) = 1$ have contribution different from zero.

In the particular case of P_2 we only need to consider the following generic vectors ν. When $|\nu| = 4$, only those vectors ν with coordinates 0, 2, or 4. For the $|\nu| = 3$ terms, any combination is valid from this point of view, because all the terms are squared, and for the $|\nu|, |\nu'| = 3$ terms, coordinates 1 or 3 are allowed in ν only if they coincide with another coordinate 1 or 3 in ν', so that we always take an even number of derivatives w.r.t. to any of the variables y_i. However the number of such terms is limited by r.

Then with $|\nu| = 4$, the maximum number of frequencies affected by the derivative, MNFA, say (in the sense that we are taking an even number of derivatives w.r.t. any of the components of the periodogram at this particular frequency), is 2, and with $|\nu| = 3$ and/or $|\nu'| = 3$ this number is 3. Consider the different possible situations for the same s fixed.

For $j = s$, $p(i) = 1 \,\forall i$. Then for P_2, the contribution when $s > 3$ is zero, because there always will be at least one integral equal to zero by (A.19), as none of its components is included in the differentiation (i.e., there are at least four possible orthogonal conditions like (A.19), and only three can be destroyed by the differentiation of ϕ). When $s = 3$ and $j = 3$ we obtain that any term will contribute $O(m^{3/2}N^{-1})$.

When $j = s - 1$, so there are at least $s - 2$ exponents $p(i) = 1$, the contribution of P_2 for $s > 5$ is zero, for $s = 3$, is $O(m^{1/2}N^{-1})$, for $s = 4$, $O(mN^{-1})$, and for $s = 5$, $O(m^{3/2}N^{-1})$.

In general, for any $j > 1 + s/2$, because $h \geq 2j - s$ there are only terms in P_2 that contribute to E_s if

$$\min h = 2j - s \leq \text{MNFA} = \{2_{[\text{if}|\nu|=4]}, 3_{[\text{if}|\nu|=3]}\},$$

and when $|\nu| = 4$ their contribution is of order $O(m^{j-s/2}N^{-1}) = O(mN^{-1}) = o(1)$, and when $|\nu| = 3$ is of order $O(m^{j-s/2}N^{-1}) = O(m^{3/2}N^{-1})$. Although we could assume that $m^{3/2}N^{-1} \to 0$ as N increases, to make this last bound $o(1)$, the consideration of the form of the cumulants when $|\nu| = 3$ will allow us to obtain the same results with just $mN^{-1}(\log m)^c \to 0$, any finite $c > 0$, implied by the assumptions of the lemma.

Cumulants. The preceding bounds have been constructed for P_2 considering that three frequencies were affected in the term corresponding to cumulants with $|\nu| = 3$. The question is, when are these cumulants different from zero? For any three frequencies $\lambda_{j_1}, \lambda_{j_2}, \lambda_{j_3}$ (possibly repeated), the cumulant $\chi_{\nu,N}, |\nu| = 3$ is of any of the following four types, with κ_3 being the third cumulant of ϵ_t.

(i) $\dfrac{\kappa_3}{N} \displaystyle\sum_{t=1}^{N} \cos t\lambda_{j_1} \cos t\lambda_{j_2} \cos t\lambda_{j_3},$

(ii) $\dfrac{\kappa_3}{N} \displaystyle\sum_{t=1}^{N} \cos t\lambda_{j_1} \cos t\lambda_{j_2} \sin t\lambda_{j_3},$

(iii) $\dfrac{\kappa_3}{N} \displaystyle\sum_{t=1}^{N} \cos t\lambda_{j_1} \sin t\lambda_{j_2} \sin t\lambda_{j_3},$

(iv) $\dfrac{\kappa_3}{N} \displaystyle\sum_{t=1}^{N} \sin t\lambda_{j_1} \sin t\lambda_{j_2} \sin t\lambda_{j_3}.$

Now using the orthogonality of the Dirichlet kernel D_N at Fourier frequencies, the cumulants $\chi_{\nu,N}$ with $|\nu| = 3$ will only be different from zero if there is a linear restriction between the frequencies $\lambda_{j_1}, \lambda_{j_2}, \lambda_{j_3}$. (This same holds for any odd-order $|\nu|$ cumulant but not for even-order cumulants as a result of symmetries.) Then, all the bounds have to be multiplied by $m^{-1} \log m$, because one of the summations $\sum_k$ in (A.18) cancels out because of the linear restriction with the other (two) summation(s), and $\sup|\Lambda_k| = O(\log m)$. Finally, we obtain a contribution of $O(m^{1/2}N^{-1} \log m) = o(1)$, for any term with $|\nu| = 3$.

Let us now study the contribution from a generic polynomial $P_r, r \geq 4$. We only need to consider expansions up to $r \leq r_{\max} = 2[(s-1)/2]$ (where $[\cdot]$ means integer part), because the bound in (A.18) due to the error term in the Edgeworth expansion with $P_{2[(s-1)/2]}$ is immediately $o(1)$ from the exponent $N^{-1-[(s-1)/2]}$ in it and the boundedness of the corresponding integral.

Now from (A.15), the different terms in P_r will include terms with combinations of cumulants

$$\kappa_{r+2}, \kappa_{r+1}\kappa_3, \kappa_r\kappa_4, \ldots, (\kappa_4)^{r/2}, \ldots, (\kappa_3)^r,$$

corresponding to all possible combinations of frequencies in the vector **y**.

We will only need to consider combinations of cumulants of the form $(\kappa_4)^{(r-a)/2}(\kappa_3)^a$, for even a, $0 \leq a \leq r$ when the MNFA is now $r + a/2$, which requires at least $a/2$ restrictions for $\kappa_3(1) \neq 0, \ldots, \kappa_3(a) \neq 0$. The reason is that with $(\kappa_4)^{r/2}$ we maximize the number of frequencies affected without any restrictions, and, on the other hand, with $(\kappa_3)^r$ we maximize the number of frequencies affected, in general, with and without restrictions. We show subsequently that any other combination of higher order cumulants will always provide a smaller MNFA or more restrictions than this combination.

Denoting by NRES the (minimum) number of linear restrictions necessary to make the cumulants considered different from zero, the contribution to (A.18) of these terms is of order, for fixed $s, j > 1 + s/2$,

$$m^{-s/2} \sum_j^s m^j \sum_r N^{-r/2} \mathbb{I}\{2j - s \le \text{MNFA}\}(m^{-1} \log m)^{\text{NRES}}$$

$$= O\left(\max_a m^{-s/2} \sum_j^s m^j \sum_r N^{-r/2} \mathbb{I}\{2j - s \le r + a/2\}(m^{-1} \log m)^{a/2}\right)$$

$$= O\left(\max_{a,j} m^{j-s/2} \sum_r N^{-r/2} \mathbb{I}\{2j - s \le r + a/2\}(m^{-1} \log m)^{a/2}\right)$$

$$= O\left(\max_r m^{a/4+r/2} N^{-r/2}(m^{-1} \log m)^{a/2}\right) = O\left(\max_r m^{r/2} N^{-r/2}\right) = o(1),$$

with $a = 0$, so the contribution is always negligible.

The last three points that need justification are the choice of cumulants, the tapering, and the pooling:

Cumulants. Let us check that we do not need to consider other sets of cumulants. Consider the case with maximum number of frequencies affected, without restrictions: $(\kappa_4)^{r/2}$, so we have the typical term with contribution of the largest order of magnitude without restrictions. We check that the introduction of $b \ge 1$ restrictions among the frequencies $\lambda_{k(i)}$, $i = 1, \ldots, j$ cannot generate terms of larger order of magnitude in P_r than the one corresponding to $(\kappa_4)^{r/2}$ (for any j and s fixed).

Seeking the least favorable situation, the new b restrictions will be used to maximize the number of frequencies affected by the differentiation, substituting a certain number of powers of κ_4 in $(\kappa_4)^{r/2}$ with a generic term in the odd-order cumulants (to take advantage of the restrictions) such as

$$(\kappa_{c_1})^2 (\kappa_{c_2})^2 \cdots (\kappa_{c_b})^2,$$

where the $c_i \ge 3$ are odd, possibly equal. This will increase MNFA by $\sum_i c_i$, and the reduction in the exponent of κ_4, to satisfy (A.16), is of magnitude $\sum_i (c_i - 2)$. This reduction will lower MNFA (by the contribution of κ_4) in $2\sum_i (c_i - 2)$ units. The global effect on MNFA is finally

$$\Delta \text{MNFA} = \sum_{i=1}^b c_i - 2 \sum_{i=1}^b (c_i - 2) = 4b - \sum_{i=1}^b c_i.$$

Therefore in a generic bound (for any j) for the contribution of these terms, $O(m^{j-s/2} \times m^{-\text{NRES}} N^{-r/2}) = O(m^{\text{MNFA}/2} m^{-\text{NRES}} N^{-r/2})$ (because $2j - s \le \text{MNFA}$), the net effect of introducing the new b restrictions is $O((\log m)^b m^\xi)$ where

$$\xi = \Delta \text{MNFA}/2 - b = \frac{4b - \sum_i c_i}{2} - b = b - \frac{1}{2} \sum_i c_i.$$

Because $c_i \ge 3$, the final effect is (ignoring the logarithm term) at most of order $O(m^{b-3b/2}) = O(m^{-b/2}) = O(1)$, so the term with biggest contribution is that with $(\kappa_4)^{r/2}$, i.e., $a = 0$.

Tapering. An equivalent Edgeworth expansion for the real and imaginary components of tapered Fourier transform of ϵ_t is valid, because they are fixed linear combinations of the components of the usual Fourier transform. Also, because we consider frequencies that are $3J\lambda_1$ apart, at least, we guarantee the uncorrelatedness fo the different variables in **y**. In this way the Edgeworth expansion is based again on the standard normal density, so the differentiation process is performed separately for each variable in **y**. Furthermore, the comments about the restrictions to obtain odd-order cumulants different from zero apply equally in the tapered case, because for the frequencies considered, D_N^T has the same orthogonality properties as D_N.

Pooling. The difference is that each pooled periodogram (tapered or not) depends on $2J$ components of the basic vector **y** instead on just 2 (single periodogram) as before. This does not affect any of the results, because we have only used the fact that in each summand of (A.18) there are j different $\log \bar{I}_{\epsilon, k}$ functions, but not that the vector of variables **y** (in the Edgeworth expansion required to approximate each expectation) was of dimension $2j$ ($2jJ$ now). The same comments about the differentiation to obtain the Hermite polynomials and the cancellation of integrals go through here again, as we have considered the cases where just differentiation (an even number of times) w.r.t. to one single component of the periodogram destroys the orthogonality condition (A.19).   ∎