

BROADBAND SEMIPARAMETRIC ESTIMATION OF THE MEMORY
PARAMETER OF A LONG-MEMORY TIME SERIES USING FRACTIONAL
EXPONENTIAL MODELS

BY CLIFFORD M. HURVICH AND JULIA BRODSKY

New York University and AT&T Labs

First Version received October 1998

Abstract. We consider a fractional exponential, or FEXP estimator of the memory parameter of a stationary Gaussian long-memory time series. The estimator is constructed by fitting a FEXP model of slowly increasing dimension to the log periodogram at all Fourier frequencies by ordinary least squares, and retaining the corresponding estimated memory parameter. We do not assume that the data were necessarily generated by a FEXP model, or by any other finite-parameter model. We do, however, impose a global differentiability assumption on the spectral density except at the origin. Because of this, and its use of all Fourier frequencies, we refer to the FEXP estimator as a broadband semiparametric estimator. We demonstrate the consistency of the FEXP estimator, and obtain expressions for its asymptotic bias and variance. If the true spectral density is sufficiently smooth, the FEXP estimator can strongly outperform existing semiparametric estimators, such as the Geweke–Porter–Hudak (GPH) and Gaussian semiparametric estimators (GSE), attaining an asymptotic mean squared error proportional to $(\log n)/n$, where n is the sample size. In a simulation study, we demonstrate the merits of using a finite-sample correction to the asymptotic variance, and we also explore the possibility of automatically selecting the dimension of the exponential model using Mallows' C_L criterion.

1. INTRODUCTION

We consider a Gaussian long-memory time series with spectral density

$$f(\lambda) = |1 - e^{-i\lambda}|^{-2d} f^*(\lambda) \quad \lambda \in [-\pi, \pi] \quad (1)$$

where the memory parameter $d \in (-\frac{1}{2}, \frac{1}{2})$ and the function $f^*(\lambda)$ govern the long- and short-term correlation structures of the series respectively. We assume that $f^*(\lambda)$ is positive, even, continuous, bounded above, and bounded away from zero on $[-\pi, \pi]$.

There are two main approaches to the estimation of d in current use. In the *parametric* approach, a finite-dimensional parametric model is assumed to hold for $f^*(\lambda)$. A key example is the ARFIMA(p, d, q) model (Granger and Joyeux, 1980; Hosking, 1981) in which $f^*(\lambda)$ is the spectral density of a stationary and invertible ARMA(p, q) process. The parameters of $f(\lambda)$, including d , may then be estimated using Gaussian maximum likelihood (Dahlhaus, 1989) or the Whittle likelihood (Fox and Taqqu, 1986; Giraitis and Surgailis, 1990). The

mean squared errors (MSEs) of the parametric estimators of d based on a sample of size n are typically $O(1/n)$ if the parametric model is correctly specified, but the estimator may be inconsistent if the model is misspecified. This drawback provides motivation for *semiparametric* estimators of d , which rely only on the behaviour of $f(\lambda)$ at low frequencies implied by (1), i.e.,

$$f(\lambda) \sim C\lambda^{-2d} \quad \lambda \rightarrow 0^+$$

where C is a positive constant. Examples of semiparametric estimators of d include the log-periodogram regression estimator of Geweke and Porter-Hudak (GPH) (Geweke and Porter-Hudak, 1983; Robinson, 1995a; Hurvich *et al.*, 1998), and the Gaussian semiparametric estimator (GSE) (Robinson, 1995b). These estimators have the advantage that, under certain conditions, they are consistent for d without the need to correctly specify a fully parametric model for $f(\lambda)$, but they typically have asymptotic MSEs proportional to $n^{-\alpha}$, with $\alpha \leq \frac{4}{5}$, no matter how smooth $f^*(\lambda)$ is. See Hurvich *et al.* (1998), Giraitis *et al.* (1997) and Henry and Robinson (1996).

Recently, Bhansali and Kokoszka (1999) have proposed a semiparametric estimator of d based on fitting a model which is fully parametric but not assumed to be correctly specified. They assume that the true spectral density obeys an ARFIMA($\infty, d, 0$) model, and then show that the estimate of d based on fitting potentially incorrect ARFIMA($p, d, 0$) (i.e., fractional autoregressive, or FAR) models using the Whittle estimator is consistent and asymptotically normal, assuming that $p \rightarrow \infty$ as $n \rightarrow \infty$. The FAR estimator may thus be viewed as providing a parametric approach to semiparametric estimation. Such an approach has proven to be effective in the somewhat analogous context of nonparametric spectral estimation. There, it has been shown, (Berk, 1974) that, if the true process obeys an AR(∞) model and an AR(p) model is fitted to the data by least squares, with $p \rightarrow \infty$ at a suitable rate, then the resulting spectral estimate is consistent for the true spectral density in the sense that the mean integrated squared error of the estimate tends to zero. Furthermore, if the true spectral density is infinitely differentiable, so that the autoregressive coefficients decline exponentially fast to zero, then taking p proportional to $\log n$ yields a mean integrated squared error of rate $(\log n)/n$, while the classical smoothed periodogram estimator can achieve a mean integrated squared error of rate no better than $n^{-4/5}$ no matter how smooth the true spectral density is. For a more detailed discussion of this phenomenon, see Newton (1988, p. 251). In the context of semiparametric estimation of d , we note that f^* is potentially quite smooth. For example, if one were prepared to assume (as Bhansali and Kokoszka as well as many other practitioners do in their simulations) that the process was ARFIMA(p, d, q) with p, d, q unknown, then f^* would be infinitely differentiable, and so it should not be surprising that estimating d using a potentially misspecified parametric model such as FAR could outperform the GPH estimator.

In this paper, we consider an explicitly defined semiparametric estimator of d based on the fitting of potentially misspecified parametric models. This

estimator is denoted as the FEXP estimator, as the fitted models are in the FEXP class described in Beran (1993, 1994). The FEXP estimator was originally proposed by Janacek (1982), and was also discussed by Robinson (1994), although neither of these papers provided asymptotic theory for the estimator. The spectral density of a FEXP model is of form

$$g(\lambda) = |1 - e^{-i\lambda}|^{-2d} g^*(\lambda)$$

where $g^*(\lambda)$ is given by the exponential model of Bloomfield (1973), i.e.,

$$\log g^*(\lambda) = \sum_{k=0}^h g_k \cos(\lambda k) \tag{2}$$

where g_0, \dots, g_h are real constants, and h is a positive integer, possibly with $h \rightarrow \infty$ as $n \rightarrow \infty$. For numerical and mathematical simplicity, we estimate the parameters (d, g_0, \dots, g_h) by least-squares regression of the log periodogram on the regressors implied by the model (details given below) over all nonzero Fourier frequencies between 0 and π . Since neither the FEXP nor the FAR estimator is explicitly restricted to a narrow band of frequencies around zero, we refer to them as *broadband semiparametric estimators*.

We will derive expressions for the bias and variance of the FEXP estimator of d in a semiparametric context. It will follow from these expressions that if $f^*(\lambda)$ is sufficiently smooth then the estimator may have MSE which decreases at a rate as fast as $(\log n)/n$, even though $f(\lambda)$ may not obey a FEXP model.

2. THE FEXP ESTIMATOR OF D

Suppose we have n observations y_0, y_1, \dots, y_{n-1} from a stationary Gaussian time series having spectral density $f(\lambda)$ given by (1). For the sake of definiteness and simplicity, we assume throughout this paper that n is even. The periodogram is given by

$$I_j = I(\omega_j) = \frac{1}{2\pi n} \left| \sum_{t=0}^{n-1} y_t \exp(-i\omega_j t) \right|^2 \quad j = 0, 1, \dots$$

where $\omega_j = 2\pi j/n$ is the j th Fourier frequency. In the special case of Gaussian white noise ($d = 0, f^* \equiv \text{Const}$), it is well known that at all nonzero Fourier frequencies between 0 and π , the periodogram ordinates $I_1, \dots, I_{\tilde{n}}$ (with $\tilde{n} = n/2 - 1$) are independently and identically distributed (i.i.d.) as $f(\omega_j) \frac{1}{2} \chi_2^2$. These properties no longer hold when $d \neq 0$ (Kunsch, 1986; Hurvich and Beltrao, 1993; Robinson, 1995a; Deo, 1997), although the fiction that they remain valid when $d \neq 0$ has helped to motivate both the GPH and GSE estimators.

Defining $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_{\tilde{n}})'$ and letting $\gamma = 0.577216 \dots$ denote Euler's constant, we have from (1) that

$$\begin{aligned} \log I(\boldsymbol{\omega}) &= \log f(\boldsymbol{\omega}) + \log \left[\frac{I(\boldsymbol{\omega})}{f(\boldsymbol{\omega})} \right] \\ &= -2d \log \left| 2 \sin \left(\frac{\boldsymbol{\omega}}{2} \right) \right| + \log f^*(\boldsymbol{\omega}) - \gamma + \boldsymbol{\varepsilon} \end{aligned} \tag{3}$$

where the entries of

$$\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{\tilde{n}})' = \left[\log \frac{I(\omega_1)}{f(\omega_1)} + \gamma, \dots, \log \frac{I(\omega_{\tilde{n}})}{f(\omega_{\tilde{n}})} + \gamma \right]'$$

are, in the case of Gaussian white noise, i.i.d. with $E[\varepsilon_j] = 0$, and $\text{Var}[\varepsilon_j] = \pi^2/6$.

We will fit a FEXP model (see (2) and the surrounding discussion) to $\log I(\boldsymbol{\omega})$. If such a model held exactly, then $\log f^*(\boldsymbol{\omega})$ in (3) would reduce to a linear combination of $\mathbf{V}_0, \mathbf{V}_1, \dots, \mathbf{V}_h$, where $\mathbf{V}_j = \cos(j\boldsymbol{\omega})$ for $j = 0, 1, 2, \dots$. This motivates the use of the FEXP estimator \hat{d}_{FEXP} of d defined as $-\frac{1}{2}$ times the coefficient of $\log |2 \sin(\boldsymbol{\omega}/2)|$ in a least-squares regression of $\log I(\boldsymbol{\omega})$ on $[\mathbf{V}_0, \mathbf{V}_1, \dots, \mathbf{V}_h, \log |2 \sin(\boldsymbol{\omega}/2)|]$. Equivalently, $-2\hat{d}_{\text{FEXP}}$ is the slope coefficient in a simple linear regression (without intercept) of $\log I(\boldsymbol{\omega})$ on \mathbf{r} , where

$$\mathbf{r} = \log \left| 2 \sin \left(\frac{\boldsymbol{\omega}}{2} \right) \right| - \sum_{k=0}^h b_k \mathbf{V}_k \tag{4}$$

and $\sum_{k=0}^h b_k \mathbf{V}_k$ is the orthogonal projection of $\log |2 \sin(\boldsymbol{\omega}/2)|$ on the linear space spanned by $[\mathbf{V}_0, \mathbf{V}_1, \dots, \mathbf{V}_h]$. Thus,

$$\hat{d}_{\text{FEXP}} = \frac{-\mathbf{r}' \log I(\boldsymbol{\omega})}{2\|\mathbf{r}\|^2} \tag{5}$$

where $\|\cdot\|^2$ denotes the sum of squares of an \tilde{n} -dimensional vector. An explicit formula for $\mathbf{b} = (b_0, \dots, b_h)'$ is given by

$$\mathbf{b} = (\mathbf{X}'_h \mathbf{X}_h)^{-1} \mathbf{X}'_h \log \left| 2 \sin \left(\frac{\boldsymbol{\omega}}{2} \right) \right| \tag{6}$$

where \mathbf{X}_h is the $[\tilde{n} \times (h + 1)]$ matrix with columns $[\mathbf{V}_0, \mathbf{V}_1, \dots, \mathbf{V}_h]$.

3. THEORETICAL RESULTS

We assume that $\log f^*(\cdot)$ has a convergent Fourier expansion,

$$\log f^*(\omega) = \sum_{k=0}^{\infty} \gamma_k \cos(k\omega) \quad \text{for } \omega \in [-\pi, \pi]$$

with $\sum_{k=0}^{\infty} |\gamma_k| < \infty$. We also assume that $f^*(\cdot)$ is differentiable on $[-\pi, \pi] - \{0\}$ with

$$|f^{*\prime}(\lambda)| \leq C/|\lambda| \quad \text{for } \lambda \in [-\pi, \pi] - \{0\} \tag{7}$$

for some finite constant C . Equation (7), which constitutes a global smoothness assumption on $f^*(\lambda)$, holds for many models in current use, including the ARFIMA, fractional Gaussian noise and FEXP models. A condition similar to (7) was assumed by Robinson (1995a), but only in a vanishingly small neighbourhood around zero frequency. Equation (7) together with the boundedness of f^* imply that $f(\cdot)$ is differentiable on $[-\pi, \pi] - \{0\}$ with

$$|f'(\lambda)| \leq \frac{Cf(\lambda)}{|\lambda|} \quad \text{for } \lambda \in [-\pi, \pi] - \{0\} \tag{8}$$

for some finite constant C .

From (3) and (5), we have

$$\hat{d}_{\text{FEXP}} = d - \frac{\mathbf{r}' \log f^*(\boldsymbol{\omega})}{2\|\mathbf{r}\|^2} - \frac{\mathbf{r}' \boldsymbol{\varepsilon}}{2\|\mathbf{r}\|^2} \tag{9}$$

where we have used the relation

$$-\frac{1}{2\|\mathbf{r}\|^2}(-2d)\mathbf{r}' \log \left| 2 \sin \left(\frac{\boldsymbol{\omega}}{2} \right) \right| = -\frac{1}{2\|\mathbf{r}\|^2}(-2d)\mathbf{r}' \left[\mathbf{r} + \sum_{k=0}^h b_k \mathbf{V}_k \right] = d$$

since \mathbf{r} is orthogonal to $\mathbf{V}_0, \dots, \mathbf{V}_h$.

From Bloomfield (1976, p. 43), it can be shown that, for all nonnegative integers j, k ,

$$\begin{aligned} \mathbf{V}'_k \mathbf{V}_j &= -\frac{1}{2} + \frac{1}{2}(-1)^{k+j-1} \\ &+ \frac{n}{4} \chi_{\{j \bmod n = k \bmod n \neq 0 \text{ or } n/2\}} + \frac{n}{4} \chi_{\{j \bmod n = -k \bmod n \neq 0 \text{ or } n/2\}} \\ &+ \frac{n}{2} \chi_{\{j \bmod n = k \bmod n = 0\}} + \frac{n}{2} \chi_{\{j \bmod n = k \bmod n = n/2\}} \end{aligned} \tag{10}$$

where χ_S denotes the indicator of the set S . It follows from Gradshteyn and Ryzhik (1994, p. 614 #7) that, for all nonnegative integers k ,

$$\int_{-\pi}^{\pi} \log \left| 2 \sin \left(\frac{\lambda}{2} \right) \right| \cos(k\lambda) d\lambda = \begin{cases} 0 & \text{if } k = 0 \\ -\frac{\pi}{k} & \text{if } k > 0 \end{cases} \tag{11}$$

This, together with the fact that $\log|2 \sin(\lambda/2)|$ is square integrable on $[-\pi, \pi]$, implies that for any fixed n ,

$$\log \left| 2 \sin \left(\frac{\boldsymbol{\omega}}{2} \right) \right| = -\sum_{j=1}^{\infty} \frac{1}{j} \mathbf{V}_j \tag{12}$$

To obtain the asymptotic bias and variance of \hat{d}_{FEXP} , we will need the following Lemmas, which are proved in the Appendix.

LEMMA 1

$$\mathbf{V}'_k \log \left| 2 \sin \left(\frac{\omega}{2} \right) \right| = \begin{cases} R_0 & \text{if } k = 0 \\ -\frac{n}{4k} + R_k & \text{if } 1 \leq k \leq \tilde{n} \\ R_{n/2} & \text{if } k = \frac{n}{2} \end{cases}$$

where the remainder terms R_k satisfy

$$\max_{k=0,1,\dots,n/2} |R_k| = O(\log n)$$

LEMMA 2. Let k be an integer, $k \geq 3$. Define the $(k \times k)$ matrices \mathbf{A} and \mathbf{B} by

$$\mathbf{A} = \text{diag} \left(\frac{n}{2}, \frac{n}{4}, \dots, \frac{n}{4} \right) \quad \text{and} \quad \mathbf{B}_{ij} = -1\chi_{\{i+j \text{ even}\}}$$

for $i, j = 0, \dots, k - 1$. Define

$$k_1 = \left\lfloor \frac{k}{2} \right\rfloor$$

$$k_2 = \left\lfloor \frac{(k-1)}{2} \right\rfloor$$

$$x_1 = \frac{1}{(n/4)(n/4 - k_1)}$$

$$x_2 = \frac{-1}{n/2} + \frac{1}{n/2 - 1 - k_2/(n/4 - k_2)}$$

$$x_3 = \frac{1}{(n/2 - 1)(n/4 - k_2) - k_2}$$

$$x_4 = \frac{1}{n/4 - k_2} \left[\frac{1}{n/4} + \frac{1}{(n/2 - 1)(n/4 - k_2) - k_2} \right]$$

Then $(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1} + \mathbf{M}$, where

$\mathbf{M} =$

$$\left\{ \begin{array}{l} \left[\begin{array}{cc} \text{diag}(x_2, x_1) & \mathbf{1}'_{k/2-1} \otimes \text{diag}(x_3, x_1) \\ \mathbf{1}_{k/2-1} \otimes \text{diag}(x_3, x_1) & \mathbf{1}_{k/2-1} \mathbf{1}'_{k/2-1} \otimes \text{diag}(x_4, x_1) \end{array} \right] \quad \text{if } k \text{ is even} \\ \left[\begin{array}{ccc} \text{diag}(x_2, x_1) & \mathbf{1}'_{(k-3)/2} \otimes \text{diag}(x_3, x_1) & (x_3, 0)' \\ \mathbf{1}_{(k-3)/2} \otimes \text{diag}(x_3, x_1) & \mathbf{1}_{(k-3)/2} \mathbf{1}'_{(k-3)/2} \otimes \text{diag}(x_4, x_1) & \mathbf{1}_{(k-3)/2} \otimes (x_4, 0)' \\ (x_3, 0) & \mathbf{1}'_{(k-3)/2} \otimes (x_4, 0) & x_4 \end{array} \right] \quad \text{if } k \text{ is odd} \end{array} \right.$$

LEMMA 3. *If $h/n \rightarrow 0$, then*

$$b_k = \begin{cases} R_0^* & \text{if } k = 0 \\ -\frac{1}{k} + R_k^* & \text{if } 1 \leq k \leq h \end{cases}$$

where the remainder terms R_k^* satisfy

$$\max_{k=1, \dots, h} |R_k^*| = O\left[\frac{(\log n)}{n}\right]$$

LEMMA 4

$$\left\| \log \left| 2 \sin \left(\frac{\omega}{2} \right) \right| \right\|^2 = \frac{n\pi^2}{24} + O(\log^2 n)$$

LEMMA 5. *If*

$$\frac{1}{h} + \frac{h}{n} \log^2 n \rightarrow 0$$

then

$$\|\mathbf{r}\|^2 = \frac{n}{4h} + o\left(\frac{n}{h}\right)$$

LEMMA 6. *If*

$$\frac{1}{h} + \frac{h}{n} \log^2 n \rightarrow 0$$

then

$$\begin{aligned} \mathbf{r}' \log f^*(\omega) &= \sum_{j=h+1}^{\infty} \frac{-n/4}{j \bmod n} \gamma_j \chi_{\{h+1 \leq j \bmod n \leq \tilde{n}\}} \\ &+ \sum_{j=n/2+1}^{\infty} \frac{-n/4}{n - j \bmod n} \gamma_j \chi_{\{n/2+1 \leq j \bmod n \leq n-h-1\}} + O(\log n) \end{aligned}$$

LEMMA 7.

$$\frac{E[J_j \bar{J}_k - f_j \chi_{\{j=k\}}]}{\sqrt{f_j f_k}} = O\left[j^{-1} \left(\frac{j}{k}\right)^{|d|} + k^{-1} \log n \left(\frac{j}{k}\right)^d \right] \quad \text{for } 1 \leq k \leq j \leq \tilde{n}$$

where

$$J_j = (2\pi n)^{-1/2} \sum_{t=0}^{n-1} y_t \exp(-i\omega_j t)$$

LEMMA 8

$$\frac{E[J_j J_k]}{\sqrt{f_j f_k}} = O \left[j^{-1} \left(\frac{j}{k} \right)^{|d|} + k^{-1} \log n \left(\frac{j}{k} \right)^d \right] \quad \text{for } 1 \leq k \leq j \leq \tilde{n}$$

LEMMA 9

$$E(\mathbf{r}' \boldsymbol{\varepsilon}) = O(\log^3 n)$$

THEOREM 1. *If*

$$\frac{1}{h} + \frac{h}{n} \log^2 n \rightarrow 0$$

then

$$\begin{aligned} \text{Bias}(\hat{d}_{\text{FEXP}}) &= \frac{h}{2} \sum_{j=h+1}^{\infty} \frac{\gamma_j}{j \bmod n} \chi_{\{h+1 \leq j \bmod n \leq \tilde{n}\}} \\ &+ \frac{h}{2} \sum_{j=n/2+1}^{\infty} \frac{\gamma_j}{n - j \bmod n} \chi_{\{n/2+1 \leq j \bmod n \leq n-h-1\}} + O\left(\frac{h \log^3 n}{n}\right) \end{aligned}$$

THEOREM 2. *Define Conditions 1–4:*

1. $(h/n) \log^6 n \rightarrow 0$.
2. $hn^{\delta-1} \log^4 n \rightarrow 0$.
3. $hn^{-\delta} \log^3 n \rightarrow 0$.
4. $hn^{2d(1-\delta)-\delta} \log^4 n \rightarrow 0$.

If Condition 1 holds, and if there exists a fixed $\delta \in (0, 1)$ such that Conditions 2 and 3 hold simultaneously, then

$$\text{Var}(\hat{d}_{\text{FEXP}}) = O\left(\frac{h}{n}\right)$$

If Condition 1 holds and if there exists a fixed $\delta \in (0, 1)$ such that Conditions 2, 3 and 4 hold simultaneously, then

$$\text{Var}(\hat{d}_{\text{FEXP}}) = \frac{\pi^2 h}{6 n} + o\left(\frac{h}{n}\right) \tag{13}$$

REMARK 1. If $h = O(\log n)$ then Condition 1 holds, and Conditions 2, 3 and 4 hold simultaneously for any d and any fixed $\delta \in (0.5, 1)$. Thus, (13) holds if $h = O(\log n)$. Suppose next that $h = O(n^\alpha)$ for some fixed $\alpha \in (0, \frac{1}{2})$. Then Condition 1 holds, and Conditions 2, 3 and 4 are equivalent to $\delta < 1 - \alpha$, $\delta > \alpha$ and $\delta > (\alpha + 2d)/(1 + 2d)$, respectively. Theorem 1, together with a consideration of simultaneous solutions of the inequalities above, yields the following conclusions:

$$\text{Var}(\hat{d}_{\text{FEXP}}) = O(h/n) \quad \forall \alpha \in (0, \frac{1}{2})$$

The more precise result of (13) holds for all $\alpha \in (0, \frac{1}{2})$ if $d \leq 0$. Equation (13) also holds for $0 < \alpha < (2d + 2)^{-1}$ if $d > 0$.

REMARK 2. We have not been able to establish the asymptotic normality of \hat{d}_{FEXP} . In view of (5), Lemmas 3 and 5, as well as (5.15) of Robinson (1995a), a sufficient condition for the asymptotic normality of \hat{d}_{FEXP} is that

$$\sum_{j=1}^{\tilde{n}} |r_j|^p = O\left(\frac{n}{h}\right)$$

for any fixed integer $p > 2$.

4. MEAN SQUARED ERROR OF \hat{d}_{FEXP} IN SPECIFIC SITUATIONS

It is instructive to compare the minimum MSE of \hat{d}_{FEXP} with that of the GPH estimator \hat{d}_{GPH} under various assumptions on the smoothness of $f^*(\cdot)$. The MSEs given below assume an optimal selection of the tuning constant, i.e., h for \hat{d}_{FEXP} and the number of frequencies for \hat{d}_{GPH} .

The case perhaps most widely considered in practice is where the γ_j decay exponentially fast to zero, so that $\log f^*(\cdot)$ has an infinite number of continuous derivatives. In this case, it follows from Theorems 1 and 2 that, if we take h proportional to $\log n$, with a suitably chosen proportionality constant, then

$$\text{MSE}(\hat{d}_{\text{FEXP}}) = O\left[\frac{\log n}{n}\right]$$

Furthermore, if we take h proportional to $(\log n)^{1+\epsilon}$ for any fixed $\epsilon > 0$ and using any positive proportionality constant, then

$$\text{MSE}(\hat{d}_{\text{FEXP}}) = O\left[\frac{(\log n)^{1+\epsilon}}{n}\right]$$

On the other hand, the fastest rate of convergence which can be attained in general for \hat{d}_{GPH} in this case is

$$\text{MSE}(\hat{d}_{\text{GPH}}) = O(n^{-4/5})$$

See Hurvich *et al.* (1998). The reason for this relatively slow rate of convergence for $\text{MSE}(\hat{d}_{\text{GPH}})$ is that, although $f^*(\cdot)$ is extremely smooth here, it does not in general satisfy the condition $f^{*''}(0) = 0$. Indeed, if the process is ARFIMA(p, d, q), then the γ_j decay exponentially fast, but $f^{*''}(0)$ need not be zero so that, even with an optimal selection of tuning constants, \hat{d}_{FEXP} strongly outperforms \hat{d}_{GPH} .

Another case of interest is where

$$\gamma_j \sim j^{2d^*-1} \quad \text{for some } d^* \in (-\infty, 0) \quad (14)$$

This situation could arise in econometric applications, as described below. If $d^* < -\frac{1}{4}$, Theorems 1 and 2 imply that the optimal MSE for \hat{d}_{FEXP} is

$$\text{MSE}(\hat{d}_{\text{FEXP}}) = O[n^{4d^*/(1-4d^*)}]$$

attained by taking $h = O[n^{1/(1-4d^*)}]$. The rate of convergence of $\text{MSE}(\hat{d}_{\text{FEXP}})$ can be made arbitrarily close to $O(n^{-1})$ if d^* is sufficiently negative. This seems sensible, since the smaller d^* is, the smoother $f^*(\cdot)$ is. Unfortunately, for the case $d^* \in [-\frac{1}{4}, 0)$, the results of Theorem 2 are not sufficiently strong to allow us to determine the optimal rate for $\text{MSE}(\hat{d}_{\text{FEXP}})$.

Next, we consider the best possible performance of \hat{d}_{GPH} under the assumptions of (14). Theorem 2 of Yong (1971), together with the results of Giraitis *et al.* (1997), imply that for $d^* \in [-1, 0)$ the GPH estimator with an optimally chosen tuning parameter attains the rate

$$\text{MSE}(\hat{d}_{\text{GPH}}) = O[n^{4d^*/(1-4d^*)}]$$

This is of the same order as the optimal value for $\text{MSE}(\hat{d}_{\text{FEXP}})$ on the subinterval where $\text{MSE}(\hat{d}_{\text{FEXP}})$ can be determined, i.e., for $d^* \in [-1, -\frac{1}{4})$. On the other hand, for $d^* < -1$, $\text{MSE}(\hat{d}_{\text{GPH}})$ remains at $O(n^{-4/5})$ in general, since (14) does not guarantee that $f^{*''}(0) = 0$, while $\text{MSE}(\hat{d}_{\text{FEXP}})$ retains the rate of $O[n^{4d^*/(1-4d^*)}]$, so that \hat{d}_{FEXP} outperforms \hat{d}_{GPH} in this case.

REMARK 3. The situation described in (14) could arise in an econometric context, from differencing the sum of two independent and potentially nonstationary long-memory time series. Specifically, suppose that $d_1 \in (0.5, 1.5)$ and $d_2 < d_1$. Let $x_{1,t}$ and $x_{2,t}$ be two independent processes whose first differences are stationary, having spectral densities $|1 - e^{-i\lambda}|^{-2(d_1-1)} f_1^*(\lambda)$ and $|1 - e^{-i\lambda}|^{-2(d_2-1)} f_2^*(\lambda)$ respectively, where $f_1^*(\lambda)$ and $f_2^*(\lambda)$ are the spectral densities of stationary and invertible ARMA processes. Let $x_t = x_{1,t} + x_{2,t}$. Define $y_t = x_t - x_{t-1}$ and let $d^* = d_2 - d_1 < 0$. The process $\{y_t\}$ is a stationary and invertible long-memory series with memory parameter $d_1 - 1$ and spectral density

$$f_y(\lambda) = |1 - e^{-i\lambda}|^{-2(d_1-1)} f_y^*(\lambda)$$

where

$$f_y^*(\lambda) = f_1^*(\lambda) + f_2^*(\lambda)|1 - e^{-i\lambda}|^{-2d^*}$$

It follows from Yong (1971, Theorem 2) that the Fourier coefficients γ_j of $\log f^*(\cdot)$ satisfy (14).

5. MONTE CARLO RESULTS

We conducted a Monte Carlo study to compare the asymptotic theory for \hat{d}_{FEXP} with its finite-sample performance, and also to compare \hat{d}_{FEXP} with the widely used log-periodogram regression estimator \hat{d}_{GPH} of Geweke and Porter-Hudak (1983). For clarity, we now introduce the notation \hat{d}_{FEXP}^h , where the superscript denotes the value of h used in computing \hat{d}_{FEXP} . Using the method of Davies and Harte (1987), we generated 500 realizations of a variety of zero-mean stationary ARFIMA(1, d , 0) models, with $d = 0.4$. The models are expressed as

$$(1 - \rho B)(1 - B)^d x_t = \varepsilon_t$$

where the ε_t are independent standard normal, and B is the backshift operator. We considered the three sample sizes $n = 512, 1024$ and 4096 , and the four values $\rho = 0.2, 0.4, 0.6$ and 0.8 .

For \hat{d}_{GPH} , we used the estimator originally recommended by Geweke and Porter-Hudak (1983). Specifically, \hat{d}_{GPH} is given by $-\frac{1}{2}$ times the slope estimator in a least-squares linear regression of $\{\log I_j\}_{j=1}^{\sqrt{n}}$ on $\{\log|2 \sin(\omega_j/2)|\}_{j=1}^{\sqrt{n}}$. The number of frequencies used in the regression is \sqrt{n} . This choice was found to work well in the simulation studies of Geweke and Porter-Hudak (1983), and is in widespread use today. In general, however, it is not an optimal choice, as shown by Hurvich *et al.* (1998), and by Giraitis *et al.* (1997). Note that we do not omit a block of the lowest frequencies from the regression. Such trimming was suggested by Robinson (1995a), but it was shown by Hurvich *et al.* (1998) that consistency and asymptotic normality of the estimator hold without trimming, and that finite-sample performance is better without trimming. For each realization, we evaluated \hat{d}_{GPH} , as well as \hat{d}_{FEXP}^h for $h = 1, \dots, 20$. We focus first on the performance of \hat{d}_{FEXP}^h .

As predicted by the theory, we found that for all values of n and ρ studied, $\text{Var}(\hat{d}_{\text{FEXP}}^h)$ increases with h . Moreover, for a given n and h , $\text{Var}(\hat{d}_{\text{FEXP}}^h)$ was virtually constant for all ρ , again in accordance with the theory. Unfortunately, for all n, ρ and h , $\text{Var}(\hat{d}_{\text{FEXP}}^h)$ was clearly larger than the asymptotic value $\pi^2 h/6n$ given in Theorem 2, although the disparity between the two values did decrease for larger values of n . For a given n and ρ , $\text{Bias}(\hat{d}_{\text{FEXP}}^h)$ was typically decreasing in h , especially for small values of h , as would be expected from Theorem 1. For larger values of h , $\text{Bias}(\hat{d}_{\text{FEXP}}^h)$ sometimes increased slightly with h , perhaps due to the gradual deterioration of some of the approximations used in Theorem 1. Not surprisingly, for a given n and h , $\text{Bias}(\hat{d}_{\text{FEXP}}^h)$ increased with ρ .

To illustrate some of these findings, Table 1 gives $\text{Bias}(\hat{d}_{\text{FEXP}}^h)$, $\text{Var}(\hat{d}_{\text{FEXP}}^h)$ and $\text{MSE}(\hat{d}_{\text{FEXP}}^h)$ for $n = 512, \rho = 0.6$ and $h = 1, \dots, 20$, based on the 500 simulated realizations of the process. Here, MSE, is the average of $(\hat{d}_{\text{FEXP}}^h - d)^2$ over the 500 realizations. In addition to the patterns described above, it should be noted that $\text{MSE}(\hat{d}_{\text{FEXP}}^h)$ was a convex function of h , minimized at $h = 3$. To see the inadequacies of the variance formula from

TABLE I
BIAS, VARIANCE AND MSE OF \hat{d}_{FEXP}^h FOR $n = 512$, $\rho = 0.6$,
 $h = 1, \dots, 20$

h	Bias	Variance	MSE
1	0.2394	0.0060	0.0633
2	0.1302	0.0103	0.0272
3	0.0797	0.0144	0.0208
4	0.0524	0.0196	0.0223
5	0.0361	0.0256	0.0268
6	0.0308	0.0340	0.0349
7	0.0253	0.0408	0.0413
8	0.0206	0.0479	0.0482
9	0.0151	0.0564	0.0565
10	0.0199	0.0670	0.0673
11	0.0200	0.0778	0.0781
12	0.0239	0.0889	0.0893
13	0.0316	0.1033	0.1041
14	0.0311	0.1140	0.1148
15	0.0218	0.1272	0.1274
16	0.0167	0.1426	0.1426
17	0.0177	0.1511	0.1511
18	0.0107	0.1635	0.1632
19	0.0125	0.1758	0.1756
20	0.0047	0.1874	0.1870

Theorem 2, note for example that, when $h = 5$, the formula gives 0.016, while the observed variance is 0.026.

A better approximation to $\text{Var}(\hat{d}_{\text{FEXP}}^h)$ than that provided by Theorem 2 can be obtained from the expression

$$\text{Var}(\hat{d}_{\text{FEXP}}^h) \approx \frac{\pi^2}{24\|\mathbf{r}\|^2} \quad (15)$$

This expression follows from (9) together with the simplifying approximation that $\varepsilon_1, \dots, \varepsilon_{\bar{n}}$ are i.i.d. with mean 0 and variance $\pi^2/6$. Note that $\|\mathbf{r}\|^2$ may be evaluated on a computer as the residual sum of squares in the least squares regression of $\log|2 \sin(\omega/2)|$ on $[\mathbf{V}_0, \dots, \mathbf{V}_h]$. Table 2 gives values of the approximate theoretical variances from (15) for $h = 1, \dots, 20$, assuming $n = 512$. Agreement of these values with the observed variances in Table 1 is good, and improves as n increases. In practice, we recommend that (15) be used instead of the formula from Theorem 2 to approximate $\text{Var}(\hat{d}_{\text{FEXP}}^h)$.

Next, we compare the performance of \hat{d}_{FEXP}^h with that of \hat{d}_{GPH} . Let h^{opt} be the value of h for which $\text{MSE}(\hat{d}_{\text{FEXP}}^h)$ is minimized. Clearly, h^{opt} would not be known in practice, but it is still of interest to see how \hat{d}_{FEXP} would behave under this optimal choice for h . Table 3 gives the bias and MSE for $\hat{d}_{\text{FEXP}}^{h^{\text{opt}}}$ and for \hat{d}_{GPH} . For a given value of ρ , h^{opt} tends to increase very slowly with n . This finding is consistent with the discussion in Section 4. In addition, for a given value of n , h^{opt} increases with ρ . This seems reasonable since, as ρ increases, it is necessary to use larger values of h to overcome bias in the

TABLE 2
 THEORETICAL VARIANCE APPROXIMATION FOR
 \hat{d}_{FEXP} , BASED ON EQUATION (15), $n = 512$,
 $h = 1, \dots, 20$

h	Variance approximation
1	0.0060
2	0.0106
3	0.0157
4	0.0214
5	0.0276
6	0.0344
7	0.0416
8	0.0495
9	0.0579
10	0.0668
11	0.0764
12	0.0866
13	0.0974
14	0.1089
15	0.1210
16	0.1338
17	0.1474
18	0.1617
19	0.1768
20	0.1927

TABLE 3
 PERFORMANCE OF $\hat{d}_{\text{FEXP}}^{h^{\text{opt}}}$ AND \hat{d}_{GPH}

n	ρ	h^{opt}	Bias($\hat{d}_{\text{FEXP}}^{h^{\text{opt}}}$)	MSE($\hat{d}_{\text{FEXP}}^{h^{\text{opt}}}$)	Bias(\hat{d}_{GPH})	MSE(\hat{d}_{GPH})
512	0.2	1	0.0345	0.0073	0.0273	0.0291
512	0.4	2	0.0444	0.0122	0.0359	0.0295
512	0.6	3	0.0797	0.0208	0.0617	0.0318
512	0.8	7	0.0995	0.0496	0.1725	0.0571
1024	0.2	1	0.0265	0.0035	0.0157	0.0182
1024	0.4	2	0.0384	0.0062	0.0201	0.0184
1024	0.6	4	0.0445	0.0108	0.0340	0.0191
1024	0.8	7	0.0929	0.0252	0.1026	0.0285
4096	0.2	1	0.0184	0.0010	0.0129	0.0073
4096	0.4	3	0.0140	0.0017	0.0139	0.0073
4096	0.6	4	0.0324	0.0031	0.0172	0.0074
4096	0.8	9	0.0409	0.0063	0.0364	0.0085

estimate of d . In all cases, $\text{MSE}(\hat{d}_{\text{FEXP}}^{h^{\text{opt}}})$ is less than $\text{MSE}(\hat{d}_{\text{GPH}})$, by a factor of as much as 7.3 when ρ is small. In these cases, \hat{d}_{GPH} is using far less than the optimal number of frequencies; see Hurvich *et al.* (1998). On the other hand, when ρ is large, \hat{d}_{GPH} becomes more competitive with $\hat{d}_{\text{FEXP}}^{h^{\text{opt}}}$ because for a given n , the optimal number of frequencies to use in \hat{d}_{GPH} tends to decrease as ρ is increased.

Unlike \hat{d}_{GPH} , the estimator $\hat{d}_{FEXP}^{h^{opt}}$ as summarized in Table 3 could not be used in practice, since h^{opt} would be unknown. Thus the practical effectiveness of \hat{d}_{FEXP}^h depends on the ability of the user to select an appropriate value for h solely on the basis of the observed data. One could try to suggest a choice for h which depends only on the sample size, in analogous fashion to the suggestion of Geweke and Porter-Hudak (1983) to use \sqrt{n} frequencies in their estimator. We prefer not to do this, however, since there will always be situations where any such data-independent choice will perform badly, even if n is large.

Instead, we propose a data-based choice for h , based on Mallows' C_L criterion. (Mallows, 1973; Li, 1987). Specifically, we propose to choose h so as to minimize

$$C_L(h) = RSS_h + 2h \frac{\pi^2}{6}$$

over the set of all candidate values for h , where RSS_h denotes the residual sum of squares in the regression used to calculate \hat{d}_{FEXP}^h . We defer any theoretical results on this choice to a later paper. The use of C_L in the context of long-memory processes was originally considered in Brodsky (1997).

Denote the value of h which minimizes $C_L(h)$ by h^{CL} . Then $\hat{d}_{FEXP}^{h^{CL}}$ provides a feasible estimator of d , based on a data-driven selection of h . Table 4 gives $Bias(\hat{d}_{FEXP}^{h^{CL}})$, $MSE(\hat{d}_{FEXP}^{h^{CL}})$, and the ratio $MSE(\hat{d}_{GPH})/MSE(\hat{d}_{FEXP}^{h^{CL}})$, based on the same simulated realizations described above. The ratio of MSEs provides a measure of the relative efficiency of $\hat{d}_{FEXP}^{h^{CL}}$ to \hat{d}_{GPH} . Table 4 indicates that, for a given value of ρ , as n is increased, the efficiency of $\hat{d}_{FEXP}^{h^{CL}}$ improves. This suggests that if n is made sufficiently large, then $\hat{d}_{FEXP}^{h^{CL}}$ will eventually outperform d_{GPH} . In the best case, $n = 4096$ and $\rho = 0.2$, the efficiency of $\hat{d}_{FEXP}^{h^{CL}}$ is 3.84. On the other hand, for a given n as ρ is increased the efficiency of $\hat{d}_{FEXP}^{h^{CL}}$

TABLE 4
BIAS AND MSE FOR $\hat{d}_{FEXP}^{h^{CL}}$

n	ρ	$Bias(\hat{d}_{FEXP}^{h^{CL}})$	$MSE(\hat{d}_{FEXP}^{h^{CL}})$	$\frac{MSE(\hat{d}_{GPH})}{MSE(\hat{d}_{FEXP}^{h^{CL}})}$
512	0.2	0.0224	0.0230	1.266
512	0.4	0.0501	0.0295	1.002
512	0.6	0.0862	0.0531	0.600
512	0.8	0.1982	0.1073	0.532
1024	0.2	0.0179	0.0093	1.949
1024	0.4	0.0412	0.0138	1.330
1024	0.6	0.0646	0.0208	0.919
1024	0.8	0.1391	0.0479	0.595
4096	0.2	0.0104	0.0019	3.844
4096	0.4	0.0185	0.0029	2.545
4096	0.6	0.0332	0.0047	1.590
4096	0.8	0.0658	0.0115	0.736

decreases, and can be substantially below unity. This may occur because for larger values of ρ the optimal number of frequencies to use in \hat{d}_{GPH} decreases, making the choice of \sqrt{n} frequencies commonly used in practice much more competitive.

6. POSTSCRIPT

After submitting the first version of this paper, we learned of the simultaneous and independent work of Eric Moulines and Philippe Soulier on the FEXP estimator. The authors kindly provided us with their manuscripts, which are now published as Moulines and Soulier (1999, 2000). In Moulines and Soulier (1999), under less restrictive conditions on the γ_j than assumed here, it is established that a generalized version of the FEXP estimator, which allows for pooling (averaging) of the periodogram prior to taking logarithms, is asymptotically normal. Our asymptotic variance expression (13), as well as the rates of convergence of the FEXP estimator discussed in Section 4, follow as special cases of their results, which include stronger bounds on the covariances between discrete Fourier transform values than we obtained in Lemmas 7 and 8. Their use of pooling reduces the asymptotic variance by a constant multiplicative factor, which can be arbitrarily close to $6/\pi^2$. Moulines and Soulier (2000) established the asymptotic efficiency of the C_L criterion for data-driven choice of h , although the loss function they considered was the global mean integrated squared error of the log of the estimated spectral density, rather than, say, $E(\hat{d}_{\text{FEXP}} - d)^2$, a quantity which would be more directly relevant from the point of view of estimation of d . For theoretical convenience, Moulines and Soulier (1999, 2000) used a set of regression frequencies which does not correspond to Fourier frequencies, even in the case of no pooling. Although their choice of regression frequencies makes no difference asymptotically, it remains to be seen whether it adds any appreciable bias in finite samples compared to the FEXP estimator as defined here.

APPENDIX

PROOF OF LEMMA 1. For any $k \in \{0, 1, \dots, n/2\}$ with n fixed, we have from (10) and (12) that

$$\begin{aligned} \mathbf{V}'_k \log \left| 2 \sin \left(\frac{\omega}{2} \right) \right| &= \mathbf{V}'_k \left[- \sum_{j=1}^{\infty} \frac{1}{j} \mathbf{V}_j \right] = - \lim_{M \rightarrow \infty} \sum_{j=1}^{nM} \frac{1}{j} \mathbf{V}'_k \mathbf{V}_j \\ &= - \lim_{M \rightarrow \infty} [B_{0M} + B_{1M} + B_{2M} + B_{3M} + B_{4M}] \end{aligned} \tag{16}$$

where M is a positive integer, and

$$\begin{aligned}
 B_{0M} &= \sum_{j=1}^{nM} \frac{1}{j} \left[-\frac{1}{2} + \frac{1}{2}(-1)^{k+j-1} \right] \\
 B_{1M} &= \frac{n}{4} \sum_{j=1}^{nM} \frac{1}{j} \chi_{\{j \bmod n = k \bmod n \neq 0 \text{ or } n/2\}} \\
 B_{2M} &= \frac{n}{4} \sum_{j=1}^{nM} \frac{1}{j} \chi_{\{j \bmod n = -k \bmod n \neq 0 \text{ or } n/2\}} \\
 B_{3M} &= \frac{n}{2} \sum_{j=1}^{nM} \frac{1}{j} \chi_{\{j \bmod n = k \bmod n = 0\}} \\
 B_{4M} &= \frac{n}{2} \sum_{j=1}^{nM} \frac{1}{j} \chi_{\{j \bmod n = k \bmod n = n/2\}}
 \end{aligned}$$

For any fixed positive integer α , as $M \rightarrow \infty$, we have

$$\begin{aligned}
 \sum_{j=1}^{\alpha M} \frac{1}{2j} &= \frac{1}{2} \log(\alpha M) + \frac{\gamma}{2} + o(1) \\
 \sum_{j=1}^{\alpha M} \frac{1}{2j-1} &= \sum_{j=1}^{2\alpha M} \frac{1}{j} - \frac{1}{2} \sum_{j=1}^{\alpha M} \frac{1}{j} \\
 &= \log(2\alpha M) + \gamma - \frac{1}{2} [\log(\alpha M) + \gamma] + o(1) \\
 &= \frac{1}{2} \log(\alpha M) + \frac{\gamma}{2} + \log 2 + o(1)
 \end{aligned}$$

where γ is Euler’s constant. Thus,

$$\begin{aligned}
 B_{0M} &= - \sum_{j=1}^{nM/2} \frac{1}{2j} \chi_{\{k \text{ is even}\}} - \sum_{j=1}^{nM/2} \frac{1}{2j-1} \chi_{\{k \text{ is odd}\}} \\
 &= -\frac{1}{2} \log\left(\frac{nM}{2}\right) - \frac{\gamma}{2} - \log 2 \chi_{\{k \text{ is odd}\}} + o(1)
 \end{aligned} \tag{17}$$

First, we consider the case $k = 0$. Then $B_{1M} = B_{2M} = B_{4M} = 0$, and

$$B_{3M} = \frac{n}{2} \left[\frac{1}{n} + \frac{1}{2n} + \dots + \frac{1}{nM} \right] = \frac{1}{2} \log M + \frac{\gamma}{2} + o(1)$$

Combining this with (16) and (17) yields

$$\mathbf{V}_0 \log \left| 2 \sin\left(\frac{\omega}{2}\right) \right| = R_0 = \frac{1}{2} \log\left(\frac{n}{2}\right)$$

Next, we consider the case $k = n/2$. Then $B_{1M} = B_{2M} = B_{3M} = 0$ and

$$\begin{aligned}
 B_{4M} &= \frac{n}{2} \left[\frac{1}{n/2} + \frac{1}{3n/2} + \dots + \frac{1}{(2M-1)n/2} \right] \\
 &= 1 + \frac{1}{3} + \dots + \frac{1}{2M-1} = \frac{1}{2} \log M + \log 2 + \frac{\gamma}{2} + o(1)
 \end{aligned}$$

Combining this with (16) and (17) yields

$$\begin{aligned}
 \mathbf{V}'_{n/2} \log \left| 2 \sin \left(\frac{\omega}{2} \right) \right| &= R_{n/2} \\
 &= \frac{1}{2} \log \left(\frac{n}{2} \right) - \log 2 \chi_{\{n/2 \text{ is even}\}}
 \end{aligned}$$

Finally, we treat the case $k \in \{1, \dots, \tilde{n}\}$. Then $B_{3M} = B_{4M} = 0$. Also,

$$\begin{aligned}
 B_{1M} &= \frac{n}{4} \left[\frac{1}{k} + \frac{1}{n+k} + \dots + \frac{1}{n(M-1)+k} \right] \\
 &= \frac{n}{4k} + \frac{n}{4k} \sum_{l=1}^{M-1} \frac{1}{1+Ln/k} \\
 &= \frac{n}{4k} + \frac{n}{4k} \sum_{l=1}^{M-1} \left[\frac{1}{Ln/k} - \frac{1}{(Ln/k)(1+Ln/k)} \right] \\
 &= \frac{n}{4k} + \frac{n}{4k} \frac{k}{n} (\log M + \gamma + o(1)) - \frac{n}{4k} \sum_{l=1}^{M-1} \left[\frac{1}{(Ln/k)(1+Ln/k)} \right] \\
 &= \frac{n}{4k} + \frac{1}{4} \log M + \frac{\gamma}{4} + o(1) - \frac{n}{4k} \sum_{l=1}^{M-1} \left[\frac{1}{(Ln/k)(1+Ln/k)} \right] \tag{18}
 \end{aligned}$$

and

$$\begin{aligned}
 B_{2M} &= \frac{n}{4} \left[\frac{1}{n-k} + \frac{1}{2n-k} + \dots + \frac{1}{nM-k} \right] \\
 &= \frac{n}{4k} \sum_{l=1}^M \frac{1}{-1+Ln/k} \\
 &= \frac{n}{4k} \sum_{l=1}^M \left[\frac{1}{Ln/k} + \frac{1}{(Ln/k)(-1+Ln/k)} \right] \\
 &= \frac{1}{4} \log M + \frac{\gamma}{4} + o(1) + \frac{n}{4k} \sum_{l=1}^M \left[\frac{1}{(Ln/k)(-1+Ln/k)} \right] \tag{19}
 \end{aligned}$$

Note that

$$\begin{aligned} & \left| -\frac{n}{4k} \sum_{L=1}^{M-1} \left[\frac{1}{(Ln/k)(1 + Ln/k)} \right] + \frac{n}{4k} \sum_{L=1}^M \left[\frac{1}{(Ln/k)(-1 + Ln/k)} \right] \right| \\ & \leq \frac{n}{4k} \sum_{L=1}^{\infty} \frac{1}{(Ln/k)^2} + \frac{nk}{4} \sum_{L=1}^M \frac{1}{Ln(-k + Ln)} \\ & \leq \frac{n}{4k} \sum_{L=1}^{\infty} \frac{1}{(Ln/k)^2} + \frac{nk}{4} \sum_{L=1}^{\infty} \frac{1}{(Ln/2)^2} \\ & = \frac{5k\pi^2}{4n6} \\ & \leq \frac{5\pi^2}{46} \end{aligned}$$

Combining this with (16), (17), (18) and (19) yields

$$\mathbf{V}'_k \log \left| 2 \sin \left(\frac{\omega}{2} \right) \right| = -\frac{n}{4k} + R_k$$

where

$$\max_{k=1, \dots, \bar{n}} |R_k| \leq \frac{1}{2} \log \left(\frac{n}{2} \right) + \frac{5\pi^2}{46} + \log 2 = O(\log n)$$

PROOF OF LEMMA 2. Using the relationships between x_1, x_2, x_3 and x_4 , it may be verified directly that the matrix $(\mathbf{A} + \mathbf{B})\mathbf{A}^{-1} + (\mathbf{A} + \mathbf{B})\mathbf{M}$ is the $(k \times k)$ identity matrix.

PROOF OF LEMMA 3. Define $c_k = \mathbf{V}'_k \log |2 \sin(\omega/2)|$ for $k = 0, 1, 2, \dots$ and let $\mathbf{c} = (c_0, c_1, \dots, c_h)'$. Then by (6), $\mathbf{b} = (\mathbf{X}'_h \mathbf{X}_h)^{-1} \mathbf{c}$. By (10), $\mathbf{X}'_h \mathbf{X}_h$ is of the form $\mathbf{A} + \mathbf{B}$, where \mathbf{A} and \mathbf{B} are the matrices defined in Lemma 2 with $k = h + 1$. It follows from Lemma 2 that $\mathbf{b} = \mathbf{A}^{-1} \mathbf{c} + \mathbf{M} \mathbf{c}$ where \mathbf{M} is the $(h + 1) \times (h + 1)$ matrix defined in Lemma 2. Since $h/n \rightarrow 0$ the upper lefthand corner of \mathbf{M} is $M_{00} = x_2 = O(1/n)$, and the maximum absolute value of all other entries of \mathbf{M} is $\max(|x_1|, |x_3|, |x_4|) = O(1/n^2)$, where x_1, x_2, x_3, x_4 are defined in Lemma 2. Thus,

$$b_0 = R_0^* = \frac{2}{n} c_0 + x_2 c_0 + \sum_{j=1}^h M_{0j} c_j$$

so by Lemmas 1 and 2,

$$\begin{aligned} |R_0^*| & \leq \left(\frac{2}{n} + |x_2| \right) |c_0| + |x_3| \left| \sum_{j=1}^h \right| - \frac{n}{4j} + R_j \Big| \\ & \leq \left(\frac{2}{n} + |x_2| \right) |R_0| + \frac{n}{4} |x_3| \text{Const} \cdot \log n + |x_3| h \max_{j=1, \dots, h} |R_j| \\ & = O \left[\frac{\log n}{n} \right] + O \left[\left(\frac{h}{n^2} \right) \log n \right] \\ & = O \left[\frac{\log n}{n} \right] \end{aligned}$$

By Lemmas 1 and 2, for $k = 1, \dots, h$

$$b_k = \frac{4}{n} c_k + M_{k0} c_0 + \sum_{j=1}^h M_{kj} c_j = -\frac{1}{k} + R_k^*$$

where

$$R_k^* = \frac{4}{n} R_k + M_{k0} R_0^* + \sum_{j=1}^h M_{kj} \left[-\frac{n}{4j} + R_j \right]$$

Thus,

$$\begin{aligned} \max_{k=1, \dots, h} |R_k^*| &\leq \frac{4}{n} \max_{k=1, \dots, h} |R_k| + |x_3| |R_0| + \max(|x_1|, |x_3|, |x_4|) [\text{Const} \cdot n \log n + h \max_{j=1, \dots, h} |R_j|] \\ &= O\left[\frac{\log n}{n}\right] \end{aligned}$$

Since

$$|R_0^*| = O\left(\frac{\log n}{n}\right)$$

it follows that

$$\max_{k=0, \dots, h} |R_k^*| = O\left(\frac{\log n}{n}\right)$$

PROOF OF LEMMA 4. The proof is reasonably straightforward, and is omitted for brevity.

PROOF OF LEMMA 5. Since

$$\mathbf{r} = \log \left| 2 \sin \left(\frac{\omega}{2} \right) \right| - \sum_{k=0}^h b_k \mathbf{V}_k$$

and since \mathbf{r} is orthogonal to the linear space spanned by $[\mathbf{V}_0, \mathbf{V}_1, \dots, \mathbf{V}_h]$, we have by Lemmas 1, 3 and 4 that

$$\begin{aligned}
 \|\mathbf{r}\|^2 &= \mathbf{r}'\mathbf{r} = \mathbf{r}' \log \left| 2 \sin \left(\frac{\boldsymbol{\omega}}{2} \right) \right| \\
 &= \left\| \log \left| 2 \sin \left(\frac{\boldsymbol{\omega}}{2} \right) \right| \right\|^2 - \sum_{k=0}^h b_k \mathbf{V}'_k \log \left| 2 \sin \left(\frac{\boldsymbol{\omega}}{2} \right) \right| \\
 &= \left\| \log \left| 2 \sin \left(\frac{\boldsymbol{\omega}}{2} \right) \right| \right\|^2 - R_0^* R_0 - \sum_{k=1}^h \left(-\frac{1}{k} + R_k^* \right) \left(-\frac{n}{4k} + R_k \right) \\
 &= \left\| \log \left| 2 \sin \left(\frac{\boldsymbol{\omega}}{2} \right) \right| \right\|^2 - \sum_{k=1}^h \frac{n}{4k^2} + O(\log^2 n) \\
 &= \frac{n\pi^2}{24} - \frac{n}{4} \sum_{k=1}^h \frac{1}{k^2} + O(\log^2 n) \\
 &= \frac{n}{4} \left(\frac{\pi^2}{6} - \sum_{k=1}^h \frac{1}{k^2} \right) + O(\log^2 n) \\
 &= \frac{n}{4} \left[\frac{1}{h} + o\left(\frac{1}{h}\right) \right] + O(\log^2 n) \\
 &= \frac{n}{4h} + o\left(\frac{n}{h}\right) + O(\log^2 n)
 \end{aligned}$$

PROOF OF LEMMA 6. From Lemma 3,

$$\mathbf{r} = \log \left| 2 \sin \left(\frac{\boldsymbol{\omega}}{2} \right) \right| - \sum_{k=0}^h b_k \mathbf{V}_k = \log \left| 2 \sin \left(\frac{\boldsymbol{\omega}}{2} \right) \right| + \sum_{k=1}^h \frac{1}{k} \mathbf{V}_k - \sum_{k=0}^h R_k^* \mathbf{V}_k$$

Therefore,

$$\mathbf{r}' \log f^*(\boldsymbol{\omega}) = \log f^*(\boldsymbol{\omega})' \mathbf{r} = \sum_{j=0}^{\infty} \gamma_j \mathbf{V}'_j \left[\log \left| 2 \sin \left(\frac{\boldsymbol{\omega}}{2} \right) \right| + \sum_{k=1}^h \frac{1}{k} \mathbf{V}_k - \sum_{k=0}^h R_k^* \mathbf{V}_k \right]$$

From Lemma 1, (10) and the fact that $V_j = V_{j \bmod n} = V_{n-j \bmod n}$ for all integers j , we obtain

$$\begin{aligned}
 \mathbf{V}'_j \log \left| 2 \sin \left(\frac{\boldsymbol{\omega}}{2} \right) \right| &= \mathbf{V}'_{j \bmod n} \log \left| 2 \sin \left(\frac{\boldsymbol{\omega}}{2} \right) \right| \\
 &= R_0 \chi_{\{j \bmod n=0\}} + \left[\frac{-n/4}{j \bmod n} + R_{j \bmod n} \right] \chi_{\{1 \leq j \bmod n \leq \bar{n}\}} \\
 &\quad + \left[\frac{-n/4}{n-j \bmod n} + R_{n-j \bmod n} \right] \chi_{\{n/2+1 \leq j \bmod n \leq n-1\}} \\
 &\quad + \mathbf{V}'_{n/2} \log \left| 2 \sin \left(\frac{\boldsymbol{\omega}}{2} \right) \right| \chi_{\{j \bmod n=n/2\}}
 \end{aligned}$$

From (10),

$$\begin{aligned} \mathbf{V}'_j \sum_{k=1}^h \frac{1}{k} \mathbf{V}_k &= \sum_{k=1}^h \frac{1}{k} \mathbf{V}'_k \mathbf{V}_{j \bmod n} \\ &= O(\log n) + \frac{n/4}{j \bmod n} \chi_{\{1 \leq j \bmod n \leq h\}} + \frac{n/4}{n - j \bmod n} \chi_{\{n-h \leq j \bmod n \leq n-1\}}. \end{aligned}$$

From Lemma 3 and (10),

$$\begin{aligned} \left| -\mathbf{V}'_j \sum_{k=0}^h R_k^* \mathbf{V}_k \right| &= \left| \sum_{k=0}^h R_k^* \mathbf{V}'_k \mathbf{V}_{j \bmod n} \right| \leq \text{Const} \cdot \left[\frac{\log n}{n} \right] \sum_{k=0}^h |\mathbf{V}'_k \mathbf{V}_{j \bmod n}| \\ &= O\left[\frac{\log n}{n} \right] \cdot O(h + n) \\ &= O(\log n) \end{aligned}$$

uniformly in j . Thus,

$$\begin{aligned} \mathbf{r}' \log f^*(\omega) &= \sum_{j=h+1}^{\infty} \frac{-n/4}{j \bmod n} \gamma_j \chi_{\{h+1 \leq j \bmod n \leq \tilde{n}\}} \\ &+ \sum_{j=n/2+1}^{\infty} \frac{-n/4}{n - j \bmod n} \gamma_j \chi_{\{n/2+1 \leq j \bmod n \leq n-h-1\}} + O(\log n) \\ &+ \sum_{j=0}^{\infty} \gamma_j [R_{j \bmod n} \chi_{\{0 \leq j \bmod n \leq \tilde{n}\}} + R_{n-j \bmod n} \chi_{\{n/2+1 \leq j \bmod n \leq n-1\}} \\ &+ R_{n/2} \chi_{\{j \bmod n = n/2\}}] \end{aligned}$$

Since $\max_j |R_{j \bmod n}| = O(\log n)$ and $\sum_{j=0}^{\infty} |\gamma_j| < \infty$, the final term above is $O(\log n)$.

PROOF OF LEMMA 7. Our proof follows the structure of the proof of Lemma 3.5 of Giraitis *et al.* (2000). In the discussion below, $C_2 - C_{13}$ are finite positive constants not depending on j, k, n . Define

$$E_{j,k}(\omega) = \frac{1}{2\pi n} D_n(\omega_j - \omega) D_n(\omega - \omega_k)$$

where

$$D_n(\omega) = \sum_{t=0}^{n-1} \exp(-i\omega t) = \exp\left[-i(n-1)\frac{\omega}{2}\right] \frac{\sin(n\omega/2)}{\sin\omega/2}$$

Note that

$$\int_{-\pi}^{\pi} E_{j,j}(\omega) d\omega = 1$$

and that

$$\int_{-\pi}^{\pi} E_{j,k}(\omega) d\omega = 0 \quad \text{for } 1 \leq k < j \leq \tilde{n}$$

Thus,

$$\begin{aligned}
 E[J_j \bar{J}_k - f_j \chi_{\{j=k\}}] &= \int_{-\pi}^{\pi} [f(\omega) - f_j] E_{j,k}(\omega) d\omega \\
 &= \int_{|\omega| < \omega_k/2} [f(\omega) - f_j] E_{j,k}(\omega) d\omega + \int_{\omega_k/2 \leq |\omega| \leq \pi} [f(\omega) - f_j] E_{j,k}(\omega) d\omega
 \end{aligned}$$

From the definition of $D_n(\cdot)$, it can be shown that

$$|D_n(\lambda)| \leq \frac{Cn}{1 + n|\lambda|} \quad \text{for } -3\pi/2 \leq \lambda \leq 3\pi/2$$

Thus, for $\omega_j - \omega$ and $\omega - \omega_k$ in this range, $|E_{j,k}(\omega)| \leq C e_{j,k}(\omega)$, where

$$e_{j,k}(\omega) = \frac{n}{1 + n|\omega_j - \omega|} \cdot \frac{1}{1 + n|\omega_k - \omega|}$$

First, we assume that $|\omega| < \omega_k/2$. Then

$$|f(\omega) - f_j| \leq C_2(|\omega|^{-2d} + \omega_j^{-2d})$$

and

$$\begin{aligned}
 e_{j,k}(\omega) &\leq \frac{1}{n(\omega_j - \omega)} \cdot \frac{1}{\omega_k - \omega} \\
 &= \frac{1}{n\omega_j\omega_k} \cdot \frac{1}{1 - \omega/\omega_j} \cdot \frac{1}{1 - \omega/\omega_k} \\
 &\leq \frac{1}{n\omega_j\omega_k} \cdot \frac{1}{1 - 0.5k/j} \cdot \frac{1}{1 - 1/2} \\
 &\leq \frac{C_3}{n\omega_j\omega_k}
 \end{aligned}$$

Thus,

$$\begin{aligned}
 \left| \int_{|\omega| < \omega_k/2} [f(\omega) - f_j] E_{j,k}(\omega) d\omega \right| &\leq \frac{C_2 C_3}{n\omega_j\omega_k} \int_{\omega = -\omega_k/2}^{\omega_k/2} [|\omega|^{-2d} + \omega_j^{-2d}] d\omega \\
 &= \frac{C_2 C_3}{n\omega_j\omega_k} \left[2 \left(\frac{\omega_k}{2} \right)^{-2d+1} + \omega_k \omega_j^{-2d} \right] \\
 &\leq \frac{C_4}{j} (\omega_k^{-2d} + \omega_j^{-2d}) \tag{20}
 \end{aligned}$$

Next, we suppose that $\omega_k/2 \leq \omega \leq \pi$. Then, by the mean value theorem and (8),

$$\begin{aligned}
 |f(\omega) - f_j| &\leq |\omega - \omega_j| \sup_{\omega_k/2 \leq \lambda \leq \pi} |f'(\lambda)| \\
 &\leq C_5 |\omega - \omega_j| \sup_{\omega_k/2 \leq \lambda \leq \pi} |\lambda|^{-2d} |\lambda|^{-1} \\
 &\leq C_6 |\omega - \omega_j| \omega_k^{-2d-1} \\
 &= C_6 \omega_k^{-1} f(\omega_k) |\omega - \omega_j| \frac{\omega_k^{-2d}}{f(\omega_k)} \\
 &\leq C_7 \omega_k^{-1} f(\omega_k) |\omega - \omega_j|
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 &\left| \int_{\omega_k/2}^{\pi} [f(\omega) - f_j] E_{j,k}(\omega) d\omega \right| \\
 &\leq C \cdot C_7 \omega_k^{-1} f(\omega_k) \int_{\omega_k/2}^{\pi} |\omega - \omega_j| e_{j,k}(\omega) d\omega \\
 &\leq C \cdot C_7 \omega_k^{-1} f(\omega_k) \int_{\omega_k/2}^{\pi} \frac{1}{1 + n|\omega_k - \omega|} d\omega \\
 &\leq C \cdot C_7 \omega_k^{-1} f(\omega_k) \left[\int_{\omega_k/2}^{\omega_{k-1}} \frac{1}{n(\omega_k - \omega)} d\omega + \int_{\omega_{k-1}}^{\omega_{k+1}} d\omega + \int_{\omega_{k+1}}^{\pi} \frac{1}{n(\omega - \omega_k)} d\omega \right] \\
 &\leq C_8 \omega_k^{-1} f(\omega_k) n^{-1} \log n \\
 &\leq C_9 \omega_k^{-2d} k^{-1} \log n
 \end{aligned} \tag{21}$$

Finally, we suppose that $-\pi \leq \omega \leq -\omega_k/2$. Then

$$\begin{aligned}
 |f(\omega) - f_j| &= |f(-\omega) - f_j| \leq |-\omega - \omega_j| \sup_{-\pi \leq \lambda \leq -\omega_k/2} |f'(-\lambda)| \\
 &= |\omega + \omega_j| \sup_{\omega_k/2 \leq \lambda \leq \pi} |f'(\lambda)| \leq |\omega + \omega_j| C_7 \omega_k^{-1} f(\omega_k)
 \end{aligned}$$

Thus,

$$\begin{aligned}
 &\left| \int_{-\pi}^{-\omega_k/2} [f(\omega) - f_j] E_{j,k}(\omega) d\omega \right| \\
 &\leq C_7 \omega_k^{-1} f(\omega_k) \frac{1}{2\pi n} \int_{-\pi}^{-\omega_k/2} |\omega_j + \omega| \left| \frac{\sin n(\omega_j - \omega)/2}{\sin(\omega_j - \omega)/2} \right| \left| \frac{\sin n(\omega - \omega_k)/2}{\sin(\omega - \omega_k)/2} \right| d\omega \\
 &\leq C_{10} \omega_k^{-1} f(\omega_k) \frac{1}{n} \int_{-\pi}^{-\omega_k/2} \left| \frac{\sin n(\omega - \omega_k)/2}{\sin(\omega - \omega_k)/2} \right| d\omega \\
 &\leq C_{11} \omega_k^{-1} f(\omega_k) n^{-1} \log n \\
 &\leq C_{12} \omega_k^{-2d} k^{-1} \log n
 \end{aligned} \tag{22}$$

where we have used the fact that

$$\sup_{\omega \in [-\pi, -\omega_k/2]} \left| \frac{\omega_j + \omega}{\sin(\omega_j - \omega)/2} \right| < C_{13}$$

The Lemma now follows from (20), (21) and (22).

PROOF OF LEMMA 8. Since

$$E[J_j J_k] = E[J_j \bar{J}_{-k}] = \int_{-\pi}^{\pi} f(\omega) E_{j,-k}(\omega) d\omega$$

and since

$$\int_{-\pi}^{\pi} E_{j,-k}(\omega) d\omega = 0 \quad \forall j, k \text{ such that } 1 \leq j \leq k \leq \tilde{n}$$

we have

$$E[J_j J_k] = \int_{-\pi}^{\pi} [f(\omega) - f_j] E_{j,-k}(\omega) d\omega \tag{23}$$

Breaking this integral into the ranges $|\omega| \leq \omega_k/2$, $\omega \in [\omega_k/2, \pi]$ and $\omega \in [-\pi, -\omega_k/2]$, and using arguments very similar to those given in the proof of Lemma 7, we can obtain bounds for the contribution to (23) from these ranges which are of the same order as given by the right-hand sides of (20), (21) and (22).

PROOF OF LEMMA 9. From (4) and Lemma 3, we conclude that $\max_j |r_j| = O(\log n)$. We have

$$\begin{aligned} |E(\mathbf{r}'\boldsymbol{\varepsilon})| &= \left| \sum_{j=1}^{\tilde{n}} r_j E(\varepsilon_j) \right| \\ &\leq \text{Const} \cdot \log n \cdot \sum_{j=1}^{\tilde{n}} |E(\varepsilon_j)| \\ &\leq \text{Const} \cdot \log n \left[\sum_{j=1}^{\log^2 n} |E(\varepsilon_j)| + \sum_{j=\log^2 n}^{\tilde{n}/\log n} |E(\varepsilon_j)| + \sum_{j=\tilde{n}/\log n+1}^{\tilde{n}} |E(\varepsilon_j)| \right] \end{aligned}$$

From Lemma 5 of Hurvich *et al.* (1998)

$$\lim_n \sup_{1 \leq j \leq \log^2 n} |E(\varepsilon_j)| < \infty$$

For $j \in [\log^2 n + 1, \dots, \tilde{n}/\log n]$, Lemma 6 of Hurvich *et al.* (1998) implies that $E(\varepsilon_j) = O[(\log j)/j]$ uniformly in j . For $j \in [\tilde{n}/\log n + 1, \dots, \tilde{n}]$, Lemmas 7 and 8 of this paper together with the proof of Lemma 2 of Hurvich *et al.* (1998) imply that $E(\varepsilon_j) = O[(\log n)/j]$. Thus,

$$|E(\mathbf{r}'\boldsymbol{\varepsilon})| = O \left(\log^3 n + \log^2 n \sum_{j=\log^2 n+1}^{\tilde{n}} j^{-1} \right) = O(\log^3 n)$$

PROOF OF THEOREM 1. The theorem follows immediately from (9), together with Lemmas 5, 6 and 9.

PROOF OF THEOREM 2. From (9) and Lemma 5,

$$\text{Var}(\hat{d}_{\text{FEXP}}) = \frac{1}{4\|r\|^4} \text{Var}(\mathbf{r}'\boldsymbol{\varepsilon}) = \left[\frac{4h^2}{n^2} + o\left(\frac{h^2}{n^2}\right) \right] \text{Var}(\mathbf{r}'\boldsymbol{\varepsilon}) \tag{24}$$

We therefore focus on $\text{Var}(\mathbf{r}'\boldsymbol{\varepsilon})$. For any fixed $\delta \in (0, 1)$, we have

$$\mathbf{r}'\boldsymbol{\varepsilon} = T_1 + T_2 + T_3 + T_4$$

where

$$\begin{aligned} T_1 &= \sum_{j=1}^{\log^2 n} r_j \varepsilon_j \\ T_2 &= \sum_{j=\log^2 n+1}^{n^\delta} r_j \varepsilon_j \\ T_3 &= \sum_{j=n^\delta+1}^{\tilde{n}/\log n} r_j \varepsilon_j \\ T_4 &= \sum_{j=\tilde{n}/\log n+1}^{\tilde{n}} r_j \varepsilon_j \end{aligned}$$

Thus,

$$\text{Var}(\mathbf{r}'\boldsymbol{\varepsilon}) = \sum_{j=1}^{\tilde{n}} r_j^2 \text{Var}(\varepsilon_j) + \theta_1 + \theta_2 + \theta_3 + \theta_4 + 2 \sum_{j=1}^3 \sum_{k=j+1}^4 \text{Cov}(T_j, T_k) \tag{25}$$

where

$$\begin{aligned} \theta_1 &= 2 \sum_{k=1}^{\log^2 n} \sum_{j=k+1}^{\log^2 n} r_j r_k \text{Cov}(\varepsilon_j, \varepsilon_k) \\ \theta_2 &= 2 \sum_{k=\log^2 n+1}^{n^\delta} \sum_{j=k+1}^{n^\delta} r_j r_k \text{Cov}(\varepsilon_j, \varepsilon_k) \\ \theta_3 &= 2 \sum_{k=n^\delta+1}^{\tilde{n}/\log n} \sum_{j=k+1}^{\tilde{n}/\log n} r_j r_k \text{Cov}(\varepsilon_j, \varepsilon_k) \\ \theta_4 &= 2 \sum_{k=\tilde{n}/\log n+1}^{\tilde{n}} \sum_{j=k+1}^{\tilde{n}} r_j r_k \text{Cov}(\varepsilon_j, \varepsilon_k) \end{aligned}$$

Consider the first term on the right-hand side of (25). By Lemma 5 of Hurvich *et al.* (1998),

$$\lim_n \sup_{1 \leq j \leq \log^2 n} \text{Var}(\varepsilon_j) < \infty$$

Combining Lemma 7 of Hurvich *et al.* (1998), the proof of Lemma 2 of Hurvich *et al.* (1998) and Lemmas 7 and 8 of this paper, we conclude that

$$\text{Var}(\varepsilon_j) = \frac{\pi^2}{6} + O\left[\frac{\log n}{j}\right]$$

uniformly for $\log^2 n \leq j \leq \tilde{n}$. Thus,

$$\begin{aligned} \sum_{j=1}^{\tilde{n}} r_j^2 \text{Var}(\varepsilon_j) &= \sum_{j=1}^{\log^2 n} r_j^2 \text{Var}(\varepsilon_j) + \sum_{j=\log^2 n+1}^{\tilde{n}} r_j^2 \text{Var}(\varepsilon_j) \\ &= O(\log^4 n) + \sum_{j=\log^2 n+1}^{\tilde{n}} r_j^2 \left\{ \frac{\pi^2}{6} + O\left[\frac{\log n}{j}\right] \right\} \\ &= \frac{\pi^2}{6} \sum_{j=1}^{\tilde{n}} r_j^2 + O(\log^4 n) \\ &= \frac{\pi^2}{6} \frac{n}{4h} + o\left(\frac{n}{h}\right) \end{aligned}$$

by Lemma 5 and Condition 1.

Next, we present bounds for $\text{Cov}(\varepsilon_j, \varepsilon_k)$. From the discussion above, $|\text{Cov}(\varepsilon_j, \varepsilon_k)|$ is uniformly bounded for $1 \leq k \leq j \leq \log^2 n$. From Lemma 2 of Hurvich *et al.* (1998) and the proof of Lemma 3 of Hurvich *et al.* (1998), it follows that

$$\text{Cov}(\varepsilon_j, \varepsilon_k) = O\{(f_j f_k)^{-1} [|\text{E}(J_j \bar{J}_k)|^2 + |\text{E}(J_j J_k)|^2]\} \quad \forall j, k$$

Thus from Theorem 2 of Robinson (1995a),

$$\text{Cov}(\varepsilon_j, \varepsilon_k) = O\left(\frac{\log^2 j}{k^2}\right)$$

uniformly for $\log^2 n \leq k < j \leq \tilde{n}/\log n$, and from Lemmas 7 and 8 of this paper,

$$\text{Cov}(\varepsilon_j, \varepsilon_k) = O[(jk)^{-1} + \log^2 n k^{-2} \left(\frac{j}{k}\right)^{2d}]$$

uniformly for $\tilde{n}/\log n \leq k < j \leq \tilde{n}$. We now present bounds for θ_1 , θ_2 , θ_3 and θ_4 . From the discussion above, we have

$$\theta_1 = O(\log^6 n) = o\left(\frac{n}{h}\right) \text{ by Condition 1}$$

$$\theta_2 = O\left(\log^4 n \sum_{k=\log^2 n+1}^{n^\delta} k^{-2} n^\delta\right)$$

$$= O(n^\delta \log^2 n)$$

$$= o\left(\frac{n}{h}\right) \text{ by Condition 2}$$

$$\theta_3 = O\left(\log^4 n \sum_{k=n^\delta+1}^{\tilde{n}/\log n} k^{-2} \frac{n}{\log n}\right)$$

$$= O(n^{1-\delta} \log^3 n)$$

$$= o\left(\frac{n}{h}\right) \text{ by Condition 3}$$

$$\theta_4 = O\left\{\log^2 n \sum_{k=\tilde{n}/\log n+1}^{\tilde{n}} \sum_{j=k+1}^{\tilde{n}} [(jk)^{-1} + \log^2 n k^{-2}(j/k)]\right\}$$

$$= O(\log^4 n) + O\left(\log^4 n \sum_{k=\tilde{n}/\log n+1}^{\tilde{n}} k^{-3} n^2\right)$$

$$= O\left[\frac{n^2 \log^4 n}{(n/\log n)^2}\right]$$

$$= O(\log^6 n)$$

$$= o\left(\frac{n}{h}\right) \text{ by Condition 1}$$

Finally, we consider the last term on the right-hand side of (25). We have

$$\text{Var}(T_1) = \sum_{j=1}^{\log^2 n} r_j^2 \text{Var}(\varepsilon_j) + \theta_1 = O(\log^4 n) + \theta_1 = o\left(\frac{n}{h}\right)$$

$$\text{Var}(T_2) = \sum_{j=\log^2 n+1}^{n^\delta} r_j^2 \text{Var}(\varepsilon_j) + \theta_2 = o\left(\frac{n}{h}\right)$$

$$\text{Var}(T_3) = O\left[\sum_{j=1}^{\tilde{n}} r_j^2 \text{Var}(\varepsilon_j)\right] + \theta_3 = \frac{\pi^2 n}{24 h} + o\left(\frac{n}{h}\right) = O\left(\frac{n}{h}\right)$$

$$\text{Var}(T_4) = O\left[\sum_{j=1}^{\tilde{n}} r_j^2 \text{Var}(\varepsilon_j)\right] + \theta_4 = O\left(\frac{n}{h}\right)$$

Thus, by the Cauchy–Schwartz inequality,

$$\text{Cov}(T_j, T_k) = o\left(\frac{n}{h}\right) \quad \text{for } (j, k) = (1, 2), (1, 3), (1, 4), (2, 3), (2, 4)$$

A direct calculation yields

$$\begin{aligned} \text{Cov}(T_3, T_4) &= O(\log^4 n) + O\left(\log^4 n \sum_{k=n^\delta+1}^{\tilde{n}/\log n} k^{-2-2d} \sum_{j=\tilde{n}/\log n+1}^{\tilde{n}} j^{2d}\right) \\ &= O(\log^4 n) + O[\log^4 n \cdot n^{2d+1} (n^\delta)^{-2d-1}] \\ &= o\left(\frac{n}{h}\right) \end{aligned}$$

by Conditions 1 and 4.

If Conditions 1–4 simultaneously hold, then we obtain from (25) that

$$\text{Var}(\mathbf{r}'\boldsymbol{\varepsilon}) = \sum_{j=1}^{\tilde{n}} r_j^2 \text{Var}(\varepsilon_j) + o\left(\frac{n}{h}\right) = \frac{\pi^2 n}{24 h} + o\left(\frac{n}{h}\right)$$

If Conditions 1–3 simultaneously hold, then we can still conclude that $\text{Var}(\mathbf{r}'\boldsymbol{\varepsilon}) = O(n/h)$ since

$$\text{Cov}(T_3, T_4) \leq [\text{Var}(T_3)\text{Var}(T_4)]^{1/2} = O\left(\frac{n}{h}\right)$$

and all other aspects of the above argument are unchanged. The theorem now follows from the above expressions for $\text{Var}(\mathbf{r}'\boldsymbol{\varepsilon})$ together with Equation (24).

REFERENCES

- BERAN, J. (1993) Fitting long-memory models by generalized linear regression. *Biometrika* 80, 817–22.
 — (1994) *Statistics for Long-Memory Processes*. New York: Chapman and Hall.
 BERK, K. N. (1974) Consistent autoregressive spectral estimates. *Ann. Stat.* 2, 489–502.

- BHANSALI, R. J. and KOKOSZKA, P. S. (1999) Estimation of the long memory parameter by fitting fractional autoregressive models. (Preprint).
- BLOOMFIELD, P. (1973) An exponential model for the spectrum of a scalar time series. *Biometrika* 60, 217–26.
- (1976) *Fourier Analysis of Time Series: An Introduction*. New York: Wiley.
- BRODSKY, J. (1997) *Parameter estimation, model selection and multi-step forecasting for a long memory time series*, Doctoral dissertation. New York University Stern School of Business.
- DAHLHAUS, R. (1989) Efficient parameter estimation for self-similar processes. *Ann. Stat.* 17, 1749–66.
- DAVIES, R. B. and HARTE, D. S. (1987) Tests for Hurst effect. *Biometrika* 74, 95–101.
- DEO, R. S. (1997) Asymptotic theory for certain regression models with long memory errors. *J. Time Ser. Anal.* 18, 385–93.
- FOX, R. and TAQQU, M. S. (1986) Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Stat.* 14, 517–32.
- GEWEKE, J. and PORTER-HUDAK, S. (1983) The estimation and application of long memory time series models. *J. Time Ser. Anal.* 4, 221–37.
- GIRAITIS, L. and SURGAILIS, D. (1990) A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probab. Theory Related Fields* 86, 87–104.
- GIRAITIS, L., ROBINSON, P. M. and SAMAROV, A. (1997) Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long-range dependence. *J. Time Ser. Anal.* 18, 49–60.
- (2000) Adaptive semiparametric estimation of the memory parameter. *J. Multivariate Anal.* 72, 183–207.
- GRADSHTEYN, I. S. and RYZHIK, I. M. (1994) *Table of Integrals, Series, and Products* (5th edn). Boston: Academic Press.
- GRANGER, C. W. J. and JOYEUX, R. (1980) An introduction to long memory time series models and fractional differencing. *J. Time Ser. Anal.* 1, 15–29.
- HENRY, M. and ROBINSON, P. M. (1996) Bandwidth choice in Gaussian semiparametric estimation of long range dependence. In *Athens Conference on Applied Probability and Time Series*. (eds P. M. Robinson and M. Rosenblatt). Springer Verlag, pp. 220–32.
- HOSKING, J. R. M. (1981) Fractional differencing. *Biometrika* 68, 165–76.
- HURVICH, C. M. and BELTRAO, K. I. (1993) Asymptotics for the low-frequency ordinates of the periodogram of a long memory time series. *J. Time Ser. Anal.* 14, 455–72.
- HURVICH, C. M., DEO, R. S. and BRODSKY, J. (1998) The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long memory time series. *J. Time Ser. Anal.* 19, 19–46.
- JANACEK, G. J. (1982) Determining the degree of differencing for time series via the log spectrum. *J. Time Ser. Anal.* 3, 177–83.
- KUNSCH, H. R. (1986) Discrimination between monotonic trends and long-range dependence. *J. Appl. Probab.* 23, 1025–30.
- LI, K.-C. (1987) Asymptotic optimality for C_p , C_L , cross-validation and generalized cross-validation. *Ann. Stat.* 15, 958–75.
- MALLOWS, C. L. (1973) Some comments on C_p . *Technometrics* 15, 661–75.
- MOULINES, E. and SOULIER, P. (1999) Broad band log-periodogram regression of time series with long range dependence. *Ann. Stat.* 27, 1415–39.
- (2000) Data-driven order selection for long range dependent time series. *J. Time Ser. Anal.* 21, 193–218.
- NEWTON, H. J. (1988) *Timeslab: A Time Series Laboratory*. New York: Wadsworth.
- ROBINSON, P. M. (1994) Time series with strong dependence. In *Advances in Econometrics: Sixth World Congress*. Vol. 1. (ed. C. Sims). Cambridge University Press, 47–95.
- (1995a) Log-periodogram regression of time series with long range dependence. *Ann. Stat.* 23, 1048–72.
- (1995b) Gaussian semiparametric estimation of long range dependence. *Ann. Stat.* 23, 1630–61.
- YONG, C.-H. (1971) On the asymptotic behavior of trigonometric series, I. *J. Math. Anal. App.* 33, 23–34.