Whittle Pseudo-Maximum Likelihood Estimation for Nonstationary Time Series

Carlos VELASCO and Peter M. ROBINSON

Whittle pseudo-maximum likelihood estimates of parameters for stationary time series have been found to be consistent and asymptotically normal in the presence of long-range dependence. Generalizing the definition of the memory parameter d, we extend these results to include possibly nonstationary ($.5 \le d < 1$) or antipersistent ($-.5 < d < 0$) observations. Using adequate data tapers, we can apply this estimation technique to any degree of nonstationarity $d \geq .5$ without a priori knowledge of the memory of the series. We analyze the performance of the estimates on simulated and real data.

KEY WORDS: Frequency domain estimation; Long-range dependence; Nonstationary fractional models; Nonstationary long memory time series; Tapering.

1. INTRODUCTION

Exact and approximate Gaussian maximum likelihood ("Whittle") estimates of parametric stationary time series models have been shown to have the same first-order asymptotic properties under long memory as was earlier shown under short memory (e.g., Dahlhaus, 1989; Fox and Taqqu 1986; Giraitis and Surgailis 1990; Hosoya 1996; Solo 1989). A covariance stationary series with spectral density (SD) $f(\lambda)$ satisfying

$$f(\lambda) \sim G|\lambda|^{-2d} \quad \text{as} \quad \lambda \to 0, \tag{1}$$

where $G > 0, |d| < 1/2$, and "$\sim$" means the ratio of left and right sides tends to 1, is said to have long memory if $0 < d < 1/2$, short memory if $d = 0$, and negative memory if $-1/2 < d < 0$.

Nonstationary time series have commonly been assumed to belong to the autoregressive integrated moving average (ARIMA) class, such that a finite number of integer differences produces an autoregressive moving average (ARMA) short memory process, with the degree of differencing determined by diagnostics such as unit root tests (see Box and Jenkins 1976). More generally, fractional ARIMA (ARFIMA) models can be considered such that integer differencing is assumed to produce a series with spectrum satisfying (1), with $d = 0$ not assumed. Equivalently, a nonstationary ARFIMA series X_t is such that $(1-L)^d X_t$ is a stationary and invertible ARMA, where $d > 1/2$ is a real number and L is the lag operator.

Beran (1995) considered a time domain version of Whittle estimation to estimate d along with other parameters in nonstationary ARFIMA models. Ling and Li (1997) extended his approach to allow for conditional heteroscedasticity, and Beran, Bhansali, and Ocker (1998) discussed model selection in the autoregressive case. We discuss Beran's asymptotic justification herein, in view of which we analyze an alternative, discrete-frequency domain version of Whittle. As originally designed for stationary environments (see Hannan 1973), this of course involves the parameterized spectral density (SD). However, for nonstationary series, no SD exists. Nevertheless if $U_t^{(s)} = (1-L)^s X_t, s = \lfloor d + 1/2 \rfloor, t > 0$, is covariance stationary with mean μ and SD $f_{U^{(s)}}(\lambda)$ behaving as $\lambda^{-2(d-s)}$ around $\lambda = 0$, we define the "pseudo SD" (PSD) of X_t as

$$f(\lambda) := |1 - e^{i\lambda}|^{-2s} f_{U^{(s)}}(\lambda) \sim G|\lambda|^{-2d} \quad \text{as} \quad \lambda \to 0. \tag{2}$$

Note that if $2d \geq 1$, then $f(\lambda)$ is not integrable in $[-\pi, \pi]$, is not an SD, and cannot represent a decomposition of the (infinite) variance of the nonstationary time series. However, as suggested by Solo (1992) and Hurvich and Ray (1995), the PSD $f(\lambda)$ can be interpreted as the limit of the expected sample periodogram, as in the stationary framework. This property was used by Velasco (1999a,b) for $-1/2 < d < 1/2$, with tapering needed for large enough d or to eliminate polynomial trends.

We illustrate the analysis of possibly nonstationary long memory series with the first 500 observations of a series of annual tree-ring widths in Arizona from 548 A.D. onward obtained by D. A. Graybill in 1984 and maintained by R. Hyndman at www-personal.buseco.monash.edu.au/˜hyndman/TSDL. The time plot of Figure 1 shows the prototypical behavior of long-range dependent data, with several local trends raising doubts about stationarity. We first analyze this question from a semiparametric standpoint and compute Robinson's (1995b) Gaussian semiparametric estimate for bandwidths $m = 25$ and 50 with the original and differenced series, adjusting the value of d in the latter case; results are given in Table 1.

Although all estimates give values $\hat{d} > .5$, confidence intervals based on the asymptotic normal distribution and the standard errors in parentheses include stationary values. These are valid for both stationary and nonstationary series as far as $-.5 < d < .75$ (see Velasco 1999b). We also tried

Carlos Velasco is Associate Professor, Departamento de Estadistica y Econometria, Universidad Carlos III de Madrid, 28903 Getafe (Madrid), Spain. (E-mail: *cavelas@est-econ.uc3m.es*). Peter M. Robinson is Tooke Professor of Economic Science and Statistics and Leverhulme Research Professor, Department of Economics, London School of Economics & Political Science, London WC2A 2AE, U.K. (E-mail: *p.m.robinson@lse.ac.uk*). The first author's research was supported by Spanish Direccion General de Ensenanza Superior, Ref. No. PB98-0025. The second author's re-

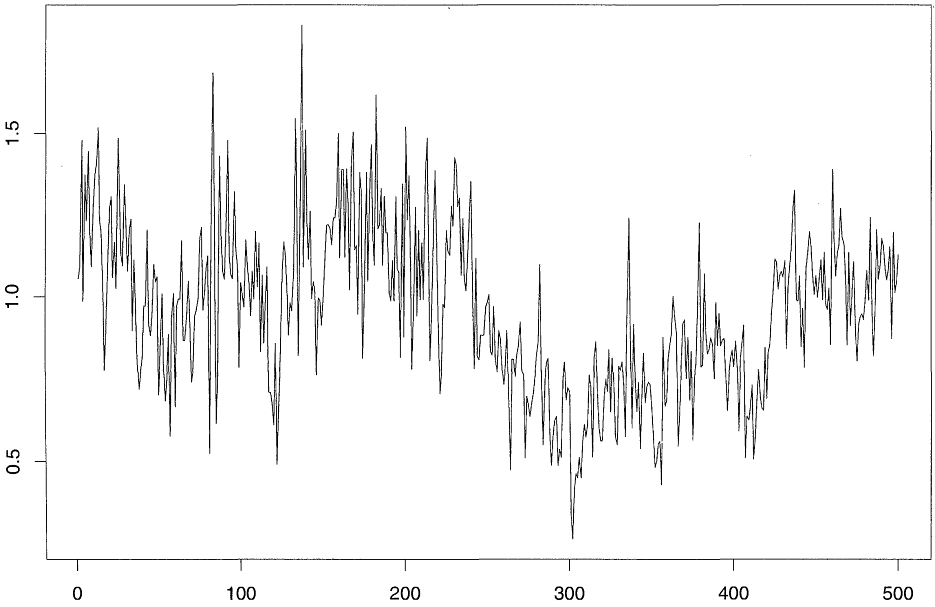

Figure 1. Arizona Tree-Ring Widths (548—1,047).

Robinson's (1994b) score tests of stationary and nonstationary hypotheses for fractional exponential [FEXP(q)] models with Bloomfield exponential modeling of short-range behavior (see Sec. 4.). They do not reject the null hypothesis $d = d_0 \leq .5$ comparing one-sided statistics to a standard normal for any low-order FEXP(q) model. However, for $q \geq 2$, the t ratios in Table 2 indicate the presence of some higher level of memory:

Turning to ARFIMA modeling, Beran's (1995) estimation procedure selects an AR(3) model with $\hat{d} = .611(.051)$, whereas a simpler AR(2) model gives $\hat{d} = .551(.057)$. These estimates use fractionally differenced data together with a time domain approximation to an ARMA Gaussian likelihood. In fact, if we assume that the series is nonstationary and take differences, then the results are very close: $\hat{d} = .627(.051)$ and $\hat{d} = .564(.057)$, with similar fits for the AR(3) and AR(2) parameters.

In this article we justify the $\sqrt{n}$ consistency and the use of normal approximations and asymptotic standard errors for the frequency domain (possibly tapered) Whittle estimates of nonstationary series with no prior assumptions on d. For the stationary increments of the tree-ring widths, the best fit of Whittle estimates is given by an AR(2) model with $\hat{d} = .563$. However, working with the original, nonstationary series and without constraining $\hat{d}$ to the stationary interval, the same criterion finds that $\hat{d} = .556$ with almost the same autoregressive parameters. This indicates the claimed robustness to nonstationarity of frequency domain estimation, which could be reinforced by tapering the data if the trending behavior were very strong. Using the cosine bell taper [(4) in Sec. 2], we confirm the small degree of nonstationarity. We also tried FEXP(2) models, obtaining similar values of $\hat{d}$, in agreement with the score tests. Results for both models are given in Table 3.

The rest of the article is organized as follows. Section 2 presents the parametric model and discusses the properties of the tapered Fourier transform of nonstationary time series. Section 3 defines the Whittle estimates and establishes their asymptotic properties. Section 4 examines the finite-sample properties of the estimates in a Monte Carlo experiment, and Section 5 applies the methods discussed to two empirical series. Appendix A gives technical assumptions and results, and Appendix B provides proofs.

Table 1. Memory of Tree-Ring Widths, Semiparametric Estimates

$\hat{d}_{Semip}$	m = 25	m = 50
X_t	.586 (.100)	.584 (.071)
ΔX_t	.599 (.100)	.594 (.071)

2. THE MODEL AND THE DISCRETE FOURIER TRANSFORM

We assume that the PSD of X_t satisfies (2) and belongs to the parametric class defined by

$$f(\lambda; \sigma^2, \boldsymbol{\theta}) = \frac{\sigma^2}{2\pi}\, k(\lambda; \boldsymbol{\theta}),$$

where $\boldsymbol{\theta} = (\theta^{(1)}, \ldots, \theta^{(a)})'$ (with $d = \theta^{(1)}$), and σ^2 are any admissible values of the unknown parameter vector $\boldsymbol{\theta}_0$ and

Table 2. Tests of Fractional Hypothesis for Tree-Ring Widths

d_0	.4	.5	.6	.7
q = 0	1.08	−2.16	−4.18	−5.50
q = 1	1.42	−1.03	−2.76	−3.97
q = 2	4.04	1.83	.14	−1.13
q = 3	3.87	1.99	.50	−.66

NOTE: q is the order of the FEXP(a) model maintained under the null.

Table 3. Memory of Tree-Ring Widths, Parametric Estimates

	AR(2)		FEXP(2)	
$\hat{d}_{Whittle}$	No taper	Cosine taper	No taper	Cosine taper
X_t	.556 (.057)	.536 (.082)	.617 (.071)	.613 (.099)
ΔX_t	.563 (.057)	.501 (.084)	.609 (.071)	.574 (.099)

scalar σ_0^2. Thus $f(\lambda) = f(\lambda, \sigma_0^2, \boldsymbol{\theta}_0)$. We assume that

$$\int_{-\pi}^{\pi} \log k(\lambda; \boldsymbol{\theta})\, d\lambda = 0, \quad \text{all} \quad \boldsymbol{\theta}. \tag{3}$$

In stationary series, with $d < 1/2$ (3) indicates that σ^2, functionally independent of $\boldsymbol{\theta}$, is the variance of the best linear predictor for a process with SD $f(\lambda, \sigma^2, \boldsymbol{\theta})$. This was used by Hannan (1973) in his treatment of short memory series and could be relaxed at cost of some extra complexity (Hosoya and Taniguchi 1982; Robinson 1978). However, (3) covers standard parameterizations of ARFIMA and FEXP models.

Define the tapered discrete Fourier transform (DFT) of X_t for $t = 1, \ldots, n$, and $\lambda_j = 2\pi j/n, j$ integer; a taper sequence $\{h_t\}_{t=1}^n$ as

$$w(\lambda_j) = w(X_t, h_t, \lambda_j) := \left(2\pi \sum_{t=1}^n h_t^2\right)^{-1/2} \sum_{t=1}^n h_t X_t e^{i\lambda_j t};$$

and the tapered periodogram as $I(\lambda_j) = |w(\lambda_j)|^2$. The usual DFT has $h_t \equiv 1$. Typically, h_t downweights the observations at both extremes of the sequence, leaving the central part of the data largely unchanged.

For short memory processes, the untapered periodogram is an inconsistent but asymptotically unbiased estimate at continuity points of the SD and approximately independent across frequencies λ_j. Robinson (1995a) extended such results for long-range dependent series, and Velasco (1999a) further extended them to certain nonstationary processes when the memory is not too high, $d < 1$, now with respect to the PSD (see App. A). However, the bias and dependence of the periodogram ordinates are affected by sharp peaks in the PSD. Tapering was suggested by Tukey (1967) to control leakage problems in spectral estimation when nonstationarity is suspected, as was checked in different frameworks by Zhurbenko (1979), Robinson (1986), and Dahlhaus (1988), among others.

Zhurbenko (1979) used a class of data weights $\{h_t^{(p)}\}$ suggested by Kolmogorov, with $p = 1, 2, \ldots$, and $N = n/p$ assumed to be integer, proportional to the coefficients $c_{p,N}(t)$, given by

$$\sum_{t=0}^{p(N-1)} z^t c_{p,N}(t+1) = (1 + z + \cdots + z^{N-1})^p = \left(\frac{1 - z^N}{1 - z}\right)^p.$$

These tapers can be obtained by increasingly smooth convolutions of the uniform density (see Alekseev 1996). When $p = 1$, they give the nontapered DFT weights, $h_t \equiv 1$; when $p = 3$, they are similar to the full cosine bell,

$$h_t = \frac{1}{2}\left(1 - \cos\frac{2\pi t}{n}\right); \tag{4}$$

and when $p = 4$, they are very close to Parzen's weights,

$$h_t = \begin{cases} 1 - 6\left[\left|\frac{2t-n}{n}\right|^2 - \left|\frac{2t-n}{n}\right|^3\right], & N < t < 3N \\ 2\left\{1 - \left|\frac{2t-n}{n}\right|\right\}^3, & 1 \le t \le N \\ & 3N \le t \le 4N. \end{cases}$$

The asymptotic properties of the taper sequences depend crucially on the kernel

$$D_h(\lambda) := \sum_{t=1}^n h_t e^{i\lambda t},$$

which is the Dirichlet kernel when $h_t = 1$, and we use them to characterize an extended class of data tapers. We say that a nonnegative, symmetric (around $\lfloor n/2 \rfloor$) and normalized ($\sup h_t = 1$) sequence of data tapers $\{h_t\}_1^n$ is of order p if the following two conditions are satisfied:

1. For n/p integer,

$$D_h(\lambda) = \frac{a(\lambda)}{n^{p-1}} \left(\frac{\sin[n\lambda/2p]}{\sin[\lambda/2]}\right)^p, \tag{5}$$

where $a(\lambda)$ is a complex function, whose modulus is bounded and bounded away from 0, with $p-1$ derivatives, all bounded for $\lambda \in [-\pi, \pi]$.

2. For some $\gamma_h, 0 < \gamma_h \le 1, \lim_{n\to\infty} n^{-1} \sum_{t=1}^n h_t^2 = \gamma_h$.

Then it can be shown that Parzen weights are of order $p = 4$, but cosine bell weights are of order $p = 1$, while sharing some properties with tapers of order $p = 3$, as discussed in Appendix A. Henceforth when $p = 1$, we will imply the usual DFT without tapering, and denote the tapered periodogram with a taper of order p as $I^p(\lambda_j)$.

As suggested by a referee, summation by parts yields, for a differentiable taper that vanishes at the boundaries, with derivative h_t',

$$w(X_t, h_t, \lambda) \approx \frac{e^{i\lambda}}{1 - e^{i\lambda}} \left[w(\Delta X_t, h_t, \lambda) + \frac{w(X_t, h_t', \lambda)}{n}\right],$$

for $\lambda \ne 0 \pmod{2\pi}$, explaining why a sufficiently smooth taper (i.e., a taper of sufficiently high-order p) can deal with arbitrary levels of memory d, justifying also definition (2). In fact, from work of Hurvich and Ray (1995) and Velasco (1999a), we can obtain Solo's (1992) inversion calculation for $f(\lambda)$ in the nonstationary case,

$$E[I^p(\lambda_{jp})] = \left(2\pi \sum_{t=1}^n h_t^2\right)^{-1} \int_{-\pi}^{\pi} |D_h(\lambda - \lambda_{jp})|^2 f(\lambda)\, d\lambda,$$
$$\sim f(\lambda_{jp}), \tag{6}$$

as $n \to \infty$. Then the tapered periodogram can be asymptotically unbiased for the PSD f of nonstationary series at Fourier frequencies $\lambda_{jp}, j \ne 0 \pmod N$, not too close to the origin, although with increased correlation between adjacent ordinates (see Ths. A.1–A.3 in Appendix A). Furthermore, using (5) for a data taper of order p,

$$w(t^l, h_t, \lambda_{jp}) = 0, \qquad l = 0, 1, \ldots, p-1, \tag{7}$$

so tapers also remove polynomial trends in the observed sequence as when, for example, the mean $\mu \neq 0$, if we concentrate on the same set of frequencies $\lambda_{jp}, j \neq 0$ (mod N).

3. WHITTLE ESTIMATES

To estimate $\boldsymbol{\theta}_0$, we use a possibly tapered version of Hannan's (1973) discrete frequency-domain Whittle objective function

$$Q_n(\boldsymbol{\theta}) = \frac{2\pi p}{n} \sum_{j(p)} \frac{I^p(\lambda_j)}{k(\lambda_j; \boldsymbol{\theta})}.$$

Here $\sum_{j(p)}$ is a sum over $j = p, 2p, \ldots, n-p$, assuming for simplicity that n/p is an integer. Thus we omit zero frequency for mean correction purposes in the stationary case, whereas the exclusion of frequencies λ_j between $\lambda_p, \lambda_{2p}, \ldots, \lambda_{n-p}$ is for (polynomial) trend correction and to guarantee the boundedness of the periodogram expectation under nonstationarity. This $Q_n(\boldsymbol{\theta})$ cannot be replaced by an integral, corresponding to the continuous Whittle objective function, but in any case the discrete form is computationally more convenient and makes more direct use of the fast Fourier transform and functional form for $k(\lambda; \boldsymbol{\theta})$. The omission of frequencies when $p > 1$ could be avoided to achieve greater efficiency; for example, if it is known that $d_0 < 3/2$ and $\mu = 0$. Following Hannan (1973), we do not require Gaussianity.

We estimate $\boldsymbol{\theta}_0$ by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \Theta} Q_n(\boldsymbol{\theta})$$

and estimate σ_0^2 by

$$\hat{\sigma}^2 = Q_n(\hat{\boldsymbol{\theta}}).$$

Here $\boldsymbol{\theta}_0 \in \Theta$, a compact set, and $d_0 \in \Theta^{(1)} = [\nabla_1, \nabla_2]$, a closed interval, $-1/2 < \nabla_1 < \nabla_2 < \infty$. Thus, following Beran (1995), we exclude noninvertible series. Unlike for the implicitly defined semiparametric estimators of d (Robinson 1995b; Velasco 1999b) there is no restriction on the upper limit ∇_2, but the maximum degree of memory d_0 that we can estimate consistently depends on the tapering applied.

In our statements of Theorems 1 and 2, we refer to assumptions listed in Appendix A.

Theorem 1. Under Assumptions A.1–A.4, with $p \geq \lfloor d_0 + 1/2 \rfloor + 1$ [only $p > d_0$ when $\mu = 0$ or $d_0 < 1/2$], $\hat{\boldsymbol{\theta}} \rightarrow_p \boldsymbol{\theta}_0$ and $\hat{\sigma}^2 \rightarrow_p \sigma_0^2$ as $n \rightarrow \infty$.

The estimates $\hat{\boldsymbol{\theta}}$ based on untapered observations ($p = 1$) are consistent for nonstationary observations and any $d_0 < 1$, but only if $\mu = 0$, thus covering nonstationary but "mean-reverting" data ($1/2 \leq d < 1$) without drift. When $\mu \neq 0$, we need increased tapering to eliminate the deterministic trend present in integrated series, $d \geq 1/2$, by (7). If μ is known to be 0 (and there are no other deterministic trends), then the tapering required for consistency is the minimum to obtain a periodogram with bounded expectation in (6) when the PSD diverges at the origin, $p > d_0$. In any case, more tapering is needed to obtain asymptotically normally distributed $\hat{\boldsymbol{\theta}}$.

Depending on the definition of Θ, in the proof we must consider separately the cases where it is possible that $d \leq d_0 - 1/2$ and those where $\nabla_1 > d_0 - 1/2$, because of the nonuniform behavior of $Q_n(\boldsymbol{\theta})$. A similar problem and solution appeared first in Robinson's (1995b) treatment of Gaussian semiparametric estimation for stationary and invertible long memory series with $d_0 \in (-1/2, 1/2)$. This question also affects Beran's (1995) treatment of nonstationary ARFIMA models.

Beran (1995) considered time domain approximate Gaussian maximum likelihood (ML) estimates based on untapered data, whatever the degree of nonstationarity. Of course, when stationarity is correctly assumed, his estimates are known to have the same asymptotic properties as ours with $p = 1$, because only different approximations to the Gaussian likelihood are being used. For the nonstationary case, Beran's definition of nonstationary processes in effect differs from ours; for the case of a simple fractionally differenced $(0, d, 0)$ model, he considered

$$(1-L)^{d_0} X_t = \varepsilon_t, \qquad t > 0; = 0, \qquad t \leq 0, \tag{8}$$

where ε_t is white noise, whereas we take

$$(1-L)^{s} X_t = U_t, \qquad t > 0; = 0, \qquad t \leq 0; \tag{9}$$

and

$$(1-L)^{d_0 - s} U_t = \varepsilon_t, \qquad t = 0, \pm 1, \ldots,$$

for $s = \lfloor d_0 + 1/2 \rfloor$. Beran considered the objective function $n^{-1} \sum_{t=2}^{n} [(1-L)^d X_t]^2$. Under (8), this is $n^{-1} \sum_{t=2}^{n} [(1-L)^{d-d_0} \varepsilon_t]^2$, and for consistency one must consider uniform convergence probability arguments with respect to the whole parameter space of admissible d, and existence of an asymptotic global minimum at $d = d_0$. This involves consideration of the processes $(1-L)^{d-d_0} \varepsilon_t$, which are stationary for $d > d_0 - 1/2$ and nonstationary otherwise. In fact, it is the neighborhood of $d - d_0 = -1/2$ that causes the most difficulty, because $(1-L)^{-1/2} \varepsilon_t$ is at the stationary/nonstationary border. Our alternative definition (9) of nonstationary processes, when combined with tapers, avoids this difficulty. The Taylor expansion used by Beran (1995, p. 670) to prove consistency seems to be circular, because the $o_p(1)$ error in the expansion for $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is justified only when $\hat{\boldsymbol{\theta}}$ is in a suitably small neighborhood of $\boldsymbol{\theta}_0$, which presupposes the consistency to be established, whereas for asymptotic normality of implicitly defined extremum estimates such as his, a rigorous previous proof of consistency is essential. At the same time tapering involves an efficiency loss (see Thm. 2), and Beran's simulations support his insight that ML estimates of fractional models have the classical $\sqrt{n}$ consistency, asymptotic normality, and efficiency properties. Indeed, it is consistent with Robinson's (1994b) findings that score tests for a unit root and many other stationary and nonstationary null hypotheses, when directed against fractional alternatives such as (8), have standard asymptotics, because the test statistic depends on only the null differenced data. In contrast, unit

Table 4. Bias of $\hat{d}$ for ARFIMA(2, d, 0) Models

		n = 512						n = 200					
d		−.4	.4	.6	.9	1.1	1.4	−.4	.4	.6	.9	1.1	1.4
No taper, p = 1													
$\hat{d}$	G-SEM	−.037	−.041	−.036	−.023	−.078	−.350	−.206	−.105	−.267	−.246	−.225	−.362
	W-p	−.006	.002	.002	−.025	−.137	−.395	−.023	.070	−.095	−.081	−.067	−.456
	W-2S	−.017	.002	−.003	.000	−.027	−.014	−.033	.116	−.034	−.037	−.037	−.073
	ML-2S	−.002	.004	.004	−.004	−.028	−.030	−.012	.128	−.022	−.020	−.016	−.109
Taper, p= 2													
$\hat{d}$	G-SEM	−.068	−.064	−.059	−.048	−.059	.007	−.270	−.261	−.253	−.237	−.222	−.188
	W-p	−.037	−.026	−.023	−.018	.006	.017	−.082	−.076	−.071	−.061	−.050	−.029
	W-2S	−.018	−.009	−.013	−.016	−.007	.009	−.035	−.031	−.033	−.036	−.037	−.031
	ML-2S	−.006	.007	.005	.007	.020	.030	−.014	−.010	−.011	−.009	−.004	.013

Table 5. Standard Deviation of $\hat{d}$ for ARFIMA(2, d, 0) Models

		n = 512							n = 200						
d			−.4	.4	.6	.9	1.1	1.4		−.4	.4	.6	.9	1.1	1.4
No taper, p = 1															
$\hat{d}$	G-SEM	(.079)	.088	.097	.102	.101	.086	.087	(.091)	.124	.178	.178	.174	.172	.081
	W-p	(.046)	.049	.053	.056	.056	.119	.065	(.073)	.093	.141	.137	.135	.132	.191
	W-2S	(.046)	.048	.056	.051	.050	.217	.049	(.073)	.090	.126	.090	.089	.089	.237
	ML-2S	(.046)	.049	.054	.055	.049	.120	.045	(.073)	.088	.129	.096	.097	.099	.196
Taper, p = 2															
$\hat{d}$	G-SEM	(.112)	.165	.151	.152	.154	.143	.165	(.129)	.201	.203	.202	.200	.199	.199
	W-p	(.069)	.078	.081	.080	.079	.074	.065	(.110)	.146	.147	.147	.146	.145	.146
	W-2S	(.046)	.049	.052	.051	.049	.049	.045	(.073)	.089	.091	.090	.088	.088	.094
	ML-2S	(.046)	.058	.058	.059	.061	.059	.055	(.073)	.099	.100	.101	.103	.106	.110

Table 6. Bias of $\hat{\phi}_1$ and $\hat{\phi}_2$ for ARFIMA(2, d, 0) Models

		n = 512						n = 200					
d		−.4	.4	.6	.9	1.1	1.4	−.4	.4	.6	.9	1.1	1.4
No taper, p = 1													
$\hat{\phi}_1$	W-p	−.004	−.004	−.005	−.081	−.316	−.601	.011	.038	.047	.041	.035	−.503
	W-2S	.001	−.003	.001	−.004	.003	.003	.015	.014	.014	.014	.014	.019
	ML-2S	−.007	−.003	−.001	.000	.006	.015	.004	.009	.009	.008	.005	.064
$\hat{\phi}_2$	W-p	.003	.006	.001	.112	.390	.585	.008	.004	.003	.004	.005	.575
	W-2S	.002	.005	−.004	.005	.002	.002	.007	.007	.007	.006	.006	.026
	ML-2S	.000	.002	−.008	.002	.000	−.002	.002	.002	.002	.001	.001	.009
Taper, p = 2													
$\hat{\phi}_1$	W-p	.010	.014	.013	.010	.002	.048	.042	.040	.038	.033	.029	.019
	W-2S	.001	.003	.005	.006	.002	.015	.017	.013	.013	.014	.014	.010
	ML-2S	−.005	−.005	−.004	−.005	−.008	.037	.005	.003	.003	.002	−.001	−.010
$\hat{\phi}_2$	W-p	−.001	−.006	−.005	−.005	.000	.009	.000	.001	.001	.002	.004	.008
	W-2S	.002	.000	−.001	−.001	.002	.027	.007	.008	.007	.006	.006	.010
	ML-2S	.000	−.003	−.004	−.004	−.001	.011	.002	.002	.001	.001	.001	.003

root tests against autoregressive alternatives have nonstandard asymptotics (see, e.g., Solo 1984).

We now discuss the asymptotic normality of $\hat{\boldsymbol{\theta}}$.

Theorem 2. Under Assumptions A.1–A.4, A.7–A.9, and (a) assumption A.5 if $p = 1$ ($\mu = 0$ or $d_0 < 1/2$), with $d_0 < 3/4$, and (b) assumption A.6 if $p > 1$ (any μ), such that $p \geq \lfloor d_0 + \frac{1}{2} \rfloor + 1, \beta > 1$, as $n \to \infty$,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \underset{d}{\to} \mathrm{N}(0, 4\pi p \Phi_p \boldsymbol{\Sigma}_0^{-1}).$$

The asymptotic variance formula is the same as that for stationary series, with

$$\boldsymbol{\Sigma}_0 = \int_{-\pi}^{\pi} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \log k(\lambda; \boldsymbol{\theta}_0) \right\} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}'} \log k(\lambda; \boldsymbol{\theta}_0) \right\} d\lambda,$$

and it may be shown that

$$\frac{2\pi}{n} \sum_{j=1}^{n-1} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \log k(\lambda_j; \hat{\boldsymbol{\theta}}) \right\} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}'} \log k(\lambda_j; \hat{\boldsymbol{\theta}}) \right\}$$

Table 7. Standard Deviation of $\hat{\phi}_1$ and $\hat{\phi}_2$ for Fractional ARIMA(2, d, 0) Models

			n = 512							n = 200					
d			−.4	.4	.6	.9	1.1	1.4		−.4	.4	.6	.9	1.1	1.4
No taper, p = 1															
$\hat{\phi}_1$	W-p	(.044)	.043	.047	.043	.098	.187	.088	(.071)	.078	.101	.107	.106	.106	.204
	W-2S	(.044)	.043	.048	.041	.049	.043	.046	(.071)	.078	.077	.077	.077	.077	.102
	ML-2S	(.044)	.043	.048	.043	.048	.042	.044	(.071)	.078	.079	.080	.081	.081	.147
$\hat{\phi}_2$	W-p	(.036)	.036	.039	.037	.123	.188	.065	(.057)	.058	.075	.081	.081	.081	.115
	W-2S	(.036)	.036	.038	.036	.038	.036	.031	(.057)	.058	.059	.059	.059	.059	.103
	ML-2S	(.036)	.035	.038	.036	.037	.035	.030	(.057)	.058	.058	.058	.058	.058	.061
Taper, p = 2															
$\hat{\phi}_1$	W-p	(.066)	.063	.071	.070	.070	.063	.215	(.106)	.115	.115	.115	.115	.114	.115
	W-2S	(.044)	.043	.046	.046	.045	.043	.197	(.071)	.078	.077	.077	.076	.077	.078
	ML-2S	(.044)	.045	.049	.049	.050	.046	.190	(.071)	.082	.083	.083	.084	.085	.086
$\hat{\phi}_2$	W-p	(.054)	.054	.050	.050	.050	.054	.061	(.085)	.082	.082	.082	.082	.082	.083
	W-2S	(.036)	.036	.031	.031	.031	.036	.128	(.057)	.058	.059	.059	.059	.059	.060
	ML-2S	(.036)	.035	.031	.030	.030	.035	.058	(.057)	.058	.058	.058	.058	.058	.058

Table 8. Bias of $\hat{d}$ for Bloomfield FEXP(2) Models

		n = 512						n = 200					
d		−.4	.4	.6	.9	1.1	1.4	−.4	.4	.6	.9	1.1	1.4
No taper, p = 1													
$\hat{d}$	G-SEM	−.072	.169	.258	.086	−.099	−.396	−.150	−.191	−.175	−.144	−.480	−.373
	W-p	−.149	−.009	.093	.026	−.137	−.426	−.259	−.359	−.379	−.397	−.623	−.411
	W-2S	−.175	−.121	−.156	−.208	−.215	−.020	−.313	−.361	−.386	−.380	−.639	−.362
Taper, p = 2													
$\hat{d}$	G-SEM	−.130	−.117	−.108	−.099	−.093	−.079	−.242	−.253	−.215	−.205	−.179	−.149
	W-p	−.074	−.064	−.056	−.053	−.085	−.121	−.216	−.225	−.203	−.208	−.239	−.335
	W-2S	.546	−.054	−.041	−.101	−.135	−.221	.365	−.202	−.214	−.144	−.176	−.374

Table 9. Standard Deviation of $\hat{d}$ for Bloomfield FEXP(2) Models

			n = 512							n = 200					
d			−.4	.4	.6	.9	1.1	1.4		−.4	.4	.6	.9	1.1	1.4
No taper, p = 1															
$\hat{d}$	G-SEM	(.079)	.102	.192	.149	.044	.020	.024	(.112)	.171	.161	.162	.158	.437	.198
	W-p	(.070)	.114	.167	.177	.089	.027	.018	(.113)	.193	.188	.190	.176	.349	.216
	W-2S	(.070)	.126	.133	.125	.124	.106	.187	(.113)	.200	.178	.184	.204	.392	.368
Taper, p = 2															
$\hat{d}$	G-SEM	(.119)	.144	.141	.141	.142	.150	.158	(.168)	.292	.279	.290	.278	.275	.285
	W-p	(.102)	.118	.113	.110	.112	.131	.149	(.163)	.267	.257	.266	.251	.256	.277
	W-2S	(.070)	.225	.147	.158	.144	.177	.105	(.113)	.263	.203	.203	.216	.247	.193

is a consistent estimate of $\mathbf{\Sigma}_0$. Here Φ_p is the taper variance inflation factor, $\Phi_p = \lim_{n\to\infty} \sum_{k=p,2p,\ldots}^{n} h^2(\lambda_k)$, with $h(\lambda) = (\sum_1^n h_t^2)^{-1} \sum_1^n h_t^2 \cos t\lambda$, which takes the values 1.05000, 1.00354, and 1.00086 for the Zhurbenko data tapers with $p = 2, 3, 4$, implying modest increments of the variance of 5%, .35%, and .09% for each of the data tapers (apart from the extra factor p due to the sampling of frequencies). Note that if we summed for $k = 1, 2, \ldots, n$ in Φ_p by considering all Fourier frequencies in Q_n, then by Parseval's identity, $\Phi_p = \lim_{n\to\infty} n(\sum_1^n h_t^2)^{-2} \sum_1^n h_t^4$ would be the usual tapering variance correction (see, e.g., Dahlhaus 1985) and $\Phi_1 = 1$ by orthogonality of the sine and cosine functions.

The same result holds for the cosine bell taper (4) when $d_0 < 3/2$ and $\mu = 0$ (or $d_0 < 1/2$ for any μ) are known, where in this case it is possible to include all frequencies $\lambda_j, 2 \leq j \leq n-2$, in Q_n as if actually $p = 1$, obtaining $\Phi_{\cos} = 35/18$. This accounts for a more than 33% reduction in the variance from setting p as 3, involving then only asymptotically uncorrelated periodogram ordinates in Q_n; see Theorem A.2 in App. A.

In fractional models, because of the separate modeling of short-run and long-run behavior, $\boldsymbol{\theta} = (d, \boldsymbol{\theta}^{(-1)})'$, $f(\lambda; \sigma^2, \boldsymbol{\theta}) = (\sigma^2/(2\pi))|1 - e^{i\lambda}|^{-2d} g(\lambda; \boldsymbol{\theta}^{(-1)})$, where g is a short memory SD, corresponding, for example, to an ARMA or Bloomfield (1973) exponential model [see (10)], the asymptotic variance of the parameter estimates is free of d_0, and

Table 10. Bias of $\hat{\theta}_1$ and $\hat{\theta}_2$ for Bloomfield FEXP(2) Models

		n = 512						n = 200					
d		−.4	.4	.6	.9	1.1	1.4	−.4	.4	.6	.9	1.1	1.4
No taper, *p* = 1													
$\hat{\theta}_1$	W-p	−.725	−2.119	−3.605	−4.703	−4.858	−4.944	−.752	−1.074	−1.365	−2.364	−1.813	−3.001
	W-2S	−.753	−.823	−.717	−.864	−1.580	−3.648	−.909	−1.127	−1.356	−1.589	−1.235	−2.813
$\hat{\theta}_2$	W-p	.905	1.653	2.468	2.949	2.989	3.018	1.150	1.527	1.757	2.345	1.708	1.793
	W-2S	.961	.924	.909	1.107	1.591	2.387	1.350	1.572	1.776	1.868	1.606	2.167
Taper, *p* = 2													
$\hat{\theta}_1$	W-p	.132	.113	.099	.077	−.007	−.256	.373	.393	.327	.278	.160	−.304
	W-2S	−3.580	−1.031	−1.155	−1.302	−1.691	.618	−3.711	−1.479	−1.435	−2.024	−2.457	2.008
$\hat{\theta}_2$	W-p	.040	.032	.029	.044	.209	−.875	.142	.154	.161	.240	.426	−1.142
	W-2S	2.088	.957	1.025	1.231	1.544	1.178	2.388	1.551	1.550	1.844	2.159	1.583

Table 11. Standard Deviation of $\hat{\theta}_1$ and $\hat{\theta}_2$ for Bloomfield FEXP(2) Models

			n = 512							n = 200					
d			−.4	.4	.6	.9	1.1	1.4		−.4	.4	.6	.9	1.1	1.4
No taper, *p* = 1															
$\hat{\theta}_1$	W-p	(.166)	.462	.994	1.889	.583	.444	.025	(.372)	.613	.722	.771	.973	2.064	2.297
	W-2S	(.166)	.551	.472	.477	.520	.763	1.260	(.266)	.664	.748	.726	.990	.941	1.685
$\hat{\theta}_2$	W-p	(.113)	.456	.549	.501	.233	.217	.016	(.253)	.555	.598	.597	.552	1.306	1.547
	W-2S	(.113)	.556	.526	.528	.533	.514	.490	(.181)	.627	.635	.605	.791	.714	.812
Taper, *p* = 2															
$\hat{\theta}_1$	W-p	(.249)	.261	.255	.251	.258	.290	.425	(.372)	.534	.504	.524	.514	.569	.731
	W-2S	(.166)	1.186	.604	.681	.776	.984	.408	(.266)	1.258	1.006	1.038	1.249	1.266	.568
$\hat{\theta}_2$	W-p	(.170)	.166	.169	.169	.170	.214	.527	(.253)	.311	.306	.304	.316	.412	.687
	W-2S	(.113)	.718	.521	.549	.565	.611	.471	(.181)	.731	.692	.691	.712	.646	.588

Table 12. Chemical Series-C. ARFIMA(1, d, 0)

		p = 1			p = 2			cos	
	G-SEM	W-p	W-2S	G-SEM	W-p	W-2S	G-SEM	W-p	W-2S
$\hat{d}$	.9928	1.0400	.9788	1.4410	.8676	1.0130	1.6370	.9686	.9930
	(.100)	(.091)		(.150)	(.137)		(.140)	(.128)	
$\hat{\phi}_1$		.1157	.8237		.8389	.7973		.8263	.8128
		(.116)			(.141)			(.143)	
$\hat{\sigma}^2$		.3171	.0186		.0162	.0189		.0150	.0187

thus of the degree of nonstationarity of the observed time series (apart from the effects of tapering, if used), and again is consistent with the nature of the score tests of Robinson (1994b). Initial differencing improves asymptotic efficiency only if a lower-order taper is used (with smaller p, because the contribution of Φ_p is of less significance), but this makes all estimates more sensitive to peaks or nonstationarity at other frequencies. (See the conclusions of Hauser 1999 for stationary ARFIMA models and various methods of approximate ML estimation, including tapered-Whittle estimates.) In any case, the steeper the $f(\lambda)$ at $\lambda = 0$ (i.e. the larger the d), the worse the asymptotic approximations that can be expected for finite samples.

4. SIMULATION RESULTS

In this section we investigate the performance of Whittle estimates for simulated stationary and nonstationary data. We generated independent samples of two Gaussian time series models with several values of the memory parameter d and two sample sizes ($n = 200$ and 512, with 1,000 and 100 replications). The short memory components are ARMA(2, 0) with autoregressive parameters $\phi_1 = .65, \phi_2 = -.6$, and $\sigma = 4$ and Bloomfield's (1973) exponential model as proposed by Robinson (1994a, p. 73), with parameters $\sigma^2 = 2\pi e^{-1}$ (which corresponds to $\theta_0 = -1$ in the usual parameterization), $\theta_1 = 5$, and $\theta_2 = -3$, leading to the FEXP(2) model

$$f(\lambda; \sigma^2, \boldsymbol{\theta}) = \frac{\sigma^2}{2\pi}|1 - e^{i\lambda}|^{-2d} e^{\theta_1 \cos\lambda + \theta_2 \cos 2\lambda}. \quad (10)$$

These models have variances of similar order of magnitude and PSD of similar shape, with a peak around $\pi/3$ and comparable behavior at the origin. The memory parameters used were $d = -.4$ to analyze series close to noninvert-

Table 13. Chemical Series-C. Memory Estimates $\hat{d}$, FEXP(q)

	p = 1		p = 2		cos	
	W-p	W-2S	W-p	W-2S	W-p	W-2S
q = 0	1.1300	1.7522	1.7032	1.7509	1.7966	1.7510
	(.052)	(.052)	(.075)	(.052)	(.073)	(.052)
q = 1	1.0688	1.5231	1.5110	1.6136	1.5685	1.6183
	(.083)	(.083)	(.120)	(.083)	(.116)	(.083)
q = 2	.9299	1.3703	1.2225	1.4082	1.2653	1.4125
	(.106)	(.106)	(.154)	(.106)	(.147)	(.106)
q = 3	.7529	1.2126	1.1313	1.3162	1.1964	1.3286
	(.125)	(.125)	(.181)	(.125)	(.174)	(.125)

ibility; $d = .4$ to compare with the stationary case; $d = .6$ and .9, nonstationary but still mean-reverting series; and $d = 1.1, 1.4$, difference stationary ($\mu = 0$).

Stationary ARFIMA series were simulated by the S-PLUS function arima.fracdiff.sim and then integrated an integer number of times if $d \geq .5$. For the exponential models (10), we first simulated the short memory model with $d = 0$ and then integrated fractionally to reach the appropriate value of d. For short memory simulation of the exponential models, we used the Davies and Harte (1987) algorithm, as programmed by Beran (1994), using the first 50 autocovariances obtained by numerical integration of the SD.

Nontapered ($p = 1$) and tapered (with Zhurbenko taper of order $p = 2$) data were considered. The cosine bell taper (4) was also used, but the results were similar to the taper with $p = 2$ and thus are not reported here. The following estimates were compared:

- G-SEM: Gaussian semiparametric narrow-band estimate of d (Robinson 1995b; Velasco 1999b) with bandwidth numbers $m = 30$ (20 for the FEXP model) and 40 for each sample size
- W-p: Whittle estimates $\hat{\boldsymbol{\theta}}$.

These estimates are consistent and asymptotically normal for all d that we tried when tapering is applied, but consistent only for $d < 1$, and asymptotically normal for $d < 3/4$ if the raw series is used. Using the Whittle memory estimates $\hat{\boldsymbol{\theta}}^{(1)} = \hat{d}(p)$ from W-p, it is possible to fractionally difference X_t to achieve approximate short memory stationarity and then use standard untapered stationary long memory methods to evaluate the first Whittle step, which uses possibly tapered nonstationary inputs. We propose two alternative two-step (asymptotically equivalent) procedures, where in both cases the second step's input is the untapered $\Delta^{\hat{d}(p)} X_t$:

- W-2S: two-step Whittle estimates, where the second step is Whittle (stationary) estimation, $p = 1$.
- ML-2S: two-step time domain (stationary) ML estimates for ARIMA stationary series, where the second step is implemented by the S-PLUS function arima.fracdiff.mle (see Haslett and Raftery 1989).

All parametric estimates (W-p, W-2S, and ML-2S) use the same (known) true model, because otherwise the estimates of d are not guaranteed to be consistent for the second step (even if enough tapering were applied). Moreover, tapering is used only for the first-step estimates, because it is hoped that differencing achieves stationarity of the second-step inputs.

For Whittle estimates (and the Gaussian semiparametric), the minimum of Q_n was found using the S-PLUS function nlmin. We report bias and standard error across replications. The asymptotic standard deviations for each particular sample size are in parentheses, taking into account the tapering applied and assuming that the two-step estimates have the ML asymptotic variance.

4.1 Autoregressive Integrated Moving Average Models

Summaries of the simulations are given in Tables 4 and 5 for the estimates of d and in Tables 6 and 7 for the estimates of ϕ_1 and ϕ_2. When no tapering is applied ($p = 1$), the estimation of d breaks down if $d > 1$, but even for these values of the memory parameter, the two-step procedures give consistent estimates, because the Whittle procedure tends to report $\hat{d} \approx 1$, so the differenced series with these initial estimates of d are stationary. The bias in Table 4 decreases for all estimates in sample size, and the large bias for semiparametric estimates can be explained in part by suboptimal bandwidth choices. The asymptotic standard deviation gives a good indication of the variability of the Whittle estimates but tends to increase slightly with the memory d (see Table 5). The simulations confirm the consistency of Whittle tapered estimates for all d. Nevertheless, the bias is larger than for two-step estimates, and the standard deviations are also slightly larger than expected. This increment in variability of tapered Whittle leads to an increase in the variance of the two-step ML estimates but not of the two-step Whittle estimates, so time domain estimation seems more sensitive to previous fractional differencing.

In Table 6, tapered Whittle estimation provides better results for the short memory ARMA parameters than for the memory parameter d, with behavior very close to that of the two-step procedures in terms of bias. In Table 7, the standard deviation, though larger, is very well approximated by the asymptotic outcome. However, for the first autoregressive parameter ϕ_1, the tapered estimates produce larger biases than the other methods in some particular cases. Here the invariance of the results across d is even more evident

Table 14. Chemical Series-C. Robinson's (1994) Tests of Nonstationarity

d_0	0	.25	.5	.75	1	1.25	1.5	1.75	2
q = 0	28.3448	3.0751	29.4260	26.2129	20.0105	11.7572	4.5624	−.0429	−2.7152
q = 1	13.1330	14.4261	12.6705	1.0989	6.5726	2.8689	1.3251	−1.3023	−2.8529
q = 2	7.2360	8.0680	7.4209	5.7350	3.5236	1.2513	−.6660	−2.0282	−2.9065
q = 3	4.1008	4.8137	4.4391	3.3034	1.8890	.3854	−.9920	−2.0581	−2.7742

Table 15. Chemical Series-A. ARFIMA(0, d, 1)

	p = 1			p = 2			cos		
	G-SEM	W-p	W-2S	G-SEM	W-p	W-2S	G-SEM	W-p	W-2S
$\hat{d}$	.4237	.4408	.4572	.4674	.5502	.4592	.5178	.4515	.4578
	(.100)	(.067)		(.150)	(.104)		(.140)	(.096)	
$\hat{\psi}_1$		.0183	.0570		.1839	.0500		.1116	.0577
		(.061)			(.093)			(.086)	
$\hat{\sigma}^2$		.0994	.0972		.0819	.0974		.0868	.0971

(except for $d > 1$ and $p = 1$ when untapered procedures yield inconsistent estimates). In conclusion, for the largest sample size, the asymptotic theory gives a good approximation to the finite-sample behavior of Whittle estimates, confirming the uniform behavior of the estimates across d, even for nonstationary series.

4.2 Exponential Models

Tables 8–11 report the results for exponential models with the same values of d as used before. The conclusions for $n = 200$ and for all untapered estimates of d are rather negative, with large biases (Table 8) and variability (Table 9) relating to the asymptotic value, probably due to a difficult distinction between the short memory and long memory components of this particular model. Nevertheless, tapered Whittle estimation for $n = 512$ produces for all d reasonable biases and standard deviations, the smallest across all methods and quite close to the asymptotic ones, whereas both two-step estimations break down in many cases.

The superiority of tapered Whittle (W-p) estimates for the memory parameter of fractional exponential Bloomfield models carries over also for the short memory parameters θ_1 and θ_2, for which the untapered two-step procedures completely fail in capturing the true model for many parameter value combinations (Tables 10 and 11). Here Zhurbenko weights with $p = 2$ for tapered estimates appear superior than the cosine bell in terms of bias for most values of d.

Our simulations agree with the finding of Dahlhaus (1988) that tapering is desirable in estimating short memory parameters when the SD has peaks because autoregressive roots are close to the unit circle, which are similar to the zero-frequency singularity of the PSD of fractionally integrated processes.

Table 16. Chemical Series-A. ARFIMA(0, d, 0)

	p = 1		p = 2		cos	
	W-p	W-2S	W-p	W-2S	W-p	W-2S
$\hat{d}$	.4286	.4217	.4171	.4179	.3692	.4207
	(.056)		(.083)		(.078)	
$\hat{\sigma}^2$	.0994	.0973	.0829	.0983	.0872	.0973

5. ILLUSTRATIVE EXAMPLES

In this section we analyze the two empirical series studied by Beran (1995)—chemical process temperature readings (series C) and chemical process concentration readings (series A) from Box and Jenkins (1976)—which are also among the series to which Robinson (1994b) applied his score tests against fractional alternatives. We use the same estimates as in the simulations ($m = 25$) and Zhurbenko's ($p = 2$) and cosine tapers. Both data tapers can deal with nonstationary series with $\mu = 0$, but only tapering of order $p = 2$ allows series with linear drift.

The conclusions are in line with findings of Robinson (1994b) and Beran (1995), and they contradict the finding of Box and Jenkins' (1976) of $\hat{d} = 1$ in series A and cast serious doubts about their $\hat{d} = 1$ in series C, values obtained by considering only integer degrees of differencing.

For series C, all procedures in Table 12 for an ARFIMA$(1, d, 0)$ model found values of d indistinguishable from 1 (except both tapered Gaussian semiparametric estimates) and a highly significant first-order autoregressive parameter of about .82 (in close agreement with Beran 1995), which may explain why the tapered semiparametric estimates gave larger estimates of d (clearly more than 1) than the corresponding parametric methods. However, Whittle and semiparametric estimates without tapering ($p = 1$) may be inconsistent for this level of memory, as is confirmed by the value of $\hat{\sigma}^2$ (though the two-step estimate of d is very close to the estimate with original data).

We also estimated FEXP(q) models of orders $q = 0, 1, 2,$ and 3 (Table 13), the best fit produced by $q = 2$. Estimates of d decrease with the order q, from about 1.75 ($q = 0$) to 1.2 ($q = 3$). As for ARIMA models, Whittle estimates with the raw series are then likely to be inconsistent. We also used Robinson's (1994b) score test in Table 14 (using the same Bloomfield exponential models to explain high-frequency behavior), completing his results for an extended set of null values of d. The values reported are one-sided test statistics, with standard normal asymptotic distribution.

Table 17. Chemical Series-A. Robinson's (1994) Tests of Nonstationarity

d_0	0	.25	.5	.75	1	1.25	1.5	1.75	2
q = 0	16.0836	4.5696	−1.5296	−4.0191	−5.1917	−5.8752	−6.3249	−6.6391	−6.8677
q = 1	8.0214	2.9196	−.5299	−2.5009	−3.4693	−4.0131	−3.4103	−4.0675	−4.3663
q = 2	4.8408	1.7333	−.9173	−2.6523	−3.6170	−4.1686	−4.5341	−4.7653	−4.7765
q = 3	2.8842	1.1621	−.6467	−2.0962	−3.0060	−3.5145	−3.8202	−3.9798	−3.7932

Table 18. Chemical Series-A. Memory estimates $\hat{d}$, FEXP(q)

	p = 1		p = 2		cos	
	W-p	W-2S	W-p	W-2S	W-p	W-2S
q = 0	.4286	.4217	.4171	.4179	.3692	.4207
	(.056)	(.056)	(.081)	(.056)	(.078)	(.056)
q = 1	.4421	.4571	.5645	.4554	.4623	.4582
	(.089)	(.089)	(.129)	(.089)	(.124)	(.089)
q = 2	.3771	.4323	.4198	.4208	.3206	.4255
	(.113)	(.113)	(.165)	(.114)	(.158)	(.113)
q = 3	.3483	.4470	.4397	.4469	.2704	.4301
	(.134)	(.134)	(.194)	(.134)	(.187)	(.134)

The score tests always reject the hypothesis $d = 2$ against $d < 2$ and the hypothesis $d = 1$ against $d > 1$, but often do not reject $d = 1.75, 1.5$, and 1.25. The tests show a similar pattern to the FEXP Whittle estimates, which contrasts to the ARIMA modeling in Table 12 but agrees with the semiparametric tapered estimates.

For series A, the results were much more uniform. In this case the memory is noticeably smaller, about .45 as estimated for an ARFIMA$(0, d, 1)$ model (Table 15), and now all estimates are expected to be consistent; Beran (1995) reported $\hat{d} = .445$). The tapered Whittle ($p = 2$) and the semiparametric (cosine bell) estimates reported slightly larger values than other procedures. Here the MA(1) ψ_1 parameter seems insignificant (except perhaps for the Whittle estimate with $p = 2$, which is the method with a highest estimate of d and largest trade-off between d and the short memory part of the model). If we eliminate the parameter ψ_1 in a reduced ARIMA$(0, d, 0)$ model (see Table 16), then the estimates of d now drop to about .41, also with reduced standard deviations.

Robinson's (1994b) tests always reject $d = 1$ and $d = .75$ and find some evidence in support of $d = .5$ and .25 for Series A (see Table 17), confirming the results for Whittle estimates of the FEXP model (Table 18), which, except in two cases, always give values between .34 and .47.

APPENDIX A: TECHNICAL ASSUMPTIONS AND RESULTS

In the following regularity conditions, statements concerning vector or matrix derivatives of $k(\lambda; \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ should be understood elementwise. They are similar to those in conditions A of Fox and Taqqu (1986) or in work of Dahlhaus (1989) for parametric estimates or that of Robinson (1995a,b) for semiparametric estimation of d, all holding for standard models such as ARFIMAs, fractional Gaussian noise, or fractional exponential models (see, e.g., Beran 1994; Robinson 1994a). Denote $a_\lambda(\lambda; \boldsymbol{\theta}) = (\partial/\partial\lambda)a(\lambda; \boldsymbol{\theta}), \mathbf{a}_{\lambda\boldsymbol{\theta}}(\lambda; \boldsymbol{\theta}) = (\partial/\partial\lambda\partial\boldsymbol{\theta})a(\lambda; \boldsymbol{\theta})$, and so on for any function a.

Assumption A.1

a. $\boldsymbol{\theta}_0$ is an interior point of Θ.

b. $k(\lambda; \boldsymbol{\theta}) \sim G_{\boldsymbol{\theta}}|\lambda|^{-2d}$ as $\lambda \to 0, 0 < G_{\boldsymbol{\theta}} < \infty$, and is continuous and positive at all $\lambda \neq 0$ and $\boldsymbol{\theta} \in \Theta$.

c. $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ implies that $k(\lambda; \boldsymbol{\theta}_1) \neq k(\lambda; \boldsymbol{\theta}_2)$ on a set of positive Lebesgue measure.

Assumption A.2. $k(\lambda; \boldsymbol{\theta})$ is differentiable in λ, with $k_\lambda(\lambda; \boldsymbol{\theta})$ continuous at all $(\lambda, \boldsymbol{\theta}), \lambda \neq 0$, and $k_\lambda(\lambda; \boldsymbol{\theta}) = O(|\lambda|^{-2d-1})$ as $\lambda \to 0$.

Assumption A.3. For each $\delta > 0, k(\lambda; \boldsymbol{\theta})$ is continuously differentiable in $\boldsymbol{\theta}$ at all $(\lambda, \boldsymbol{\theta}), \lambda \neq 0$, with $\mathbf{k}_{\boldsymbol{\theta}}^{-1}(\lambda; \boldsymbol{\theta}) = O(|\lambda|^{2d-\delta})$ as $\lambda \to 0$, and these derivatives are continuously differentiable in λ at all $(\lambda, \boldsymbol{\theta}), \lambda \neq 0$, with $\mathbf{k}_{\lambda\boldsymbol{\theta}}^{-1}(\lambda; \boldsymbol{\theta}) = O(|\lambda|^{2d-1-\delta})$ as $\lambda \to 0$.

The differentiability with respect to λ is required to approximate discrete sums by integrals, even when $f(\lambda)$ has a singularity at the origin. To describe the stationary differenced series, we introduce the following linear process assumption from Robinson (1995b), which is restrictive in the linearity that it imposes but not otherwise.

Assumption A.4. We assume that

$$U_t^{(s)} = \mu + \sum_{l=0}^{\infty} \alpha_l \varepsilon_{t-l}, \qquad \sum_{l=0}^{\infty} \alpha_l^2 < \infty,$$

with $\alpha_l = \alpha_l(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta$, where $E[\varepsilon_t|\mathcal{F}_{t-1}] = 0, E[\varepsilon_t^2|\mathcal{F}_{t-1}] = 1$, a.s., $t = 0, \pm 1, \ldots$, and $\mathcal{F}_t$ is the σ field of events generated by $\varepsilon_t, s \leq t$, and there exists a random variable ε, such that $E\varepsilon^2 < \infty$ and for all $\eta > 0$ and some $C > 0, P(|\varepsilon_t| > \eta) \leq CP(|\varepsilon| > \eta)$.

We use the following technical assumptions to derive the asymptotic distribution of periodogram averages and parameter estimates. First, we consider two smoothness conditions on f. When no taper is applied, we impose the following assumption.

Assumption A.5. $\alpha(\lambda) = \sum_{l=0}^{\infty} \alpha_l e^{il\lambda}$ is differentiable in λ at all $(\lambda, \boldsymbol{\theta}), \lambda \neq 0$ and $\alpha_\lambda(\lambda) = O(|\alpha(\lambda)||\lambda|^{-1})$ as $\lambda \to 0$.

Assumption A.5 implies the differentiability of $f(\lambda)$ as stated in Assumption A.2. This condition was imposed by Robinson (1995b) in a semiparametric context, with the observation that it applies to such parametric models as ARFIMAs. To use the tapering bias-reduction properties, we assume stronger smoothness conditions.

Assumption A.6. $k_\lambda(\lambda; \boldsymbol{\theta})$ is Lipschitz$(\beta - 1)$ in λ, for some $1 < \beta \leq 2$ and for all $(\lambda, \boldsymbol{\theta}), \lambda \neq 0$ and for some $0 < G_{\boldsymbol{\theta}}, E_{\boldsymbol{\theta}} < \infty, k(\lambda; \boldsymbol{\theta}) = G_{\boldsymbol{\theta}}|\lambda|^{-2d} + E_{\boldsymbol{\theta}}|\lambda|^{-2d+\beta} + o(|\lambda|^{-2d+\beta})$ as $\lambda \to 0$. In particular, with $\beta > 1$, Assumption A.6 implies that for $|\lambda| < \lambda_j/2, 0 < j < n/2$,

$$f(\lambda_j - \lambda) = f(\lambda_j) - \lambda f_\lambda(\lambda_j) + O(\lambda_j^{-\beta-2d}|\lambda|^\beta), \qquad (A.1)$$

as $\lambda \to 0$, which is the basis for the tapering bias reduction. For the asymptotic distribution of the estimates and related quadratic forms, we need also an extra condition about the fourth moments of the linear innovations, again taken from Robinson (1995b), and two additional conditions to approximate the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$.

Assumption A.7. Assumption A.4 holds and for finite constants μ_3 and $\mu_4, E[\varepsilon_t^3|\mathcal{F}_{t-1}] = \mu_3, E[\varepsilon_t^4|\mathcal{F}_{t-1}] = \mu_4$, a.s., $t = 0, \pm 1, \ldots$

Assumption A.8. $k(\lambda; \boldsymbol{\theta})$ has two continuous derivatives in $\boldsymbol{\theta}$ at all $(\lambda, \boldsymbol{\theta}), \lambda \neq 0$, with $\mathbf{k}_{\boldsymbol{\theta}\boldsymbol{\theta}'}^{-1}(\lambda; \boldsymbol{\theta}) = O(|\lambda|^{2d-\delta})$ as $\lambda \to 0$ for each $\delta > 0$, and these derivatives are continuously differentiable in λ at all $(\lambda, \boldsymbol{\theta}), \lambda \neq 0$, with $\mathbf{k}_{\lambda\boldsymbol{\theta}\boldsymbol{\theta}'}^{-1}(\lambda; \boldsymbol{\theta}) = O(|\lambda|^{2d-1-\delta})$ as $\lambda \to 0$.

Assumption A.9. $\int_{-\pi}^{\pi}\{k^{-1}(\lambda; \boldsymbol{\theta}) + \log k(\lambda; \boldsymbol{\theta})\}\, d\lambda$ can be continuously differentiated twice (with respect to $\boldsymbol{\theta}$) under the integral sign and $\boldsymbol{\Sigma}_0^{-1}$ exists.

We review now some results obtained by Robinson (1995a) and Velasco (1999a) for the (tapered) DFT of possibly nonstationary time series. The following conditions on $f_{U^{(s)}}$, which hold under Assumptions A.1 and A.2, were assumed by these authors.

Assumption A.10. For some $0 < G < \infty, d > -\frac{1}{2}, s = \lfloor d + \frac{1}{2} \rfloor, f_{U^{(s)}}(\lambda) = G|\lambda|^{-2(d-s)} + o(|\lambda|^{-2(d-s)})$ as $\lambda \to 0$.

Assumption A.11. $f_{U^{(s)}}(\lambda)$ has bounded derivative at all $\lambda \neq 0$, and $d > -1/2, s = \lfloor d + 1/2 \rfloor, (d/d\lambda) f_{U^{(s)}}(\lambda) = O(|\lambda|^{-2(d-s)-1})$ as $\lambda \to 0$.

First, we analyze the covariance matrix of the raw DFT $w(\lambda_j)$, for frequencies $\lambda_j \to 0$ and $\lambda_j \to \nu \in (0, \pi]$ as $n \to \infty$. Define $v(\lambda) = w(\lambda)/f(\lambda)^{1/2}$.

Theorem A.1 ($p = 1$). Under Assumptions A.10 and A.11, $d \in (-1/2, 1)$ ($\mu = 0$ if $d \geq 1/2$), for any sequences of positive integers $j = j(n)$ and $k = k(n)$ such that $1 \leq k < j \leq n/2$, defining $\gamma_{j,k} \equiv (jk)^{d-1} \log(k+1)$, as $n \to \infty$, (a) $E[v(\lambda_j)\bar{v}(\lambda_j)] = 1 + O(j^{-1}\log(j+1) + \gamma_{j,j})$, (b) $E[v(\lambda_j)v(\lambda_j)] = O(j^{-1}\log(j+1) + \gamma_{j,j})$, and (c) $E[v(\lambda_j)\bar{v}(\lambda_k)], E[v(\lambda_j)v(\lambda_k)] = O(k^{-1}\log j + \gamma_{j,k})$.

For values $d \geq 1$, the periodogram is not asymptotically unbiased for f as j increases. Tapering allows a reduction in the order of magnitude of the bounds in Theorem A.1, making possible approximation of the PSD with larger d. Thus with the cosine bell taper, similar results occur for any $d < 3/2$. Other tapers reduce the bias even more and allow consideration of values $d \geq 3/2$. However, the full advantage of the tapers shows up only when we assume further smoothness conditions on f.

Assumption A.12. $f(\lambda)$ satisfies a Lipschitz condition of degree $\beta \leq 1$ for all $\lambda \neq 0$, or $f(\lambda)$ is differentiable and $f_\lambda(\lambda)$ satisfies a Lipschitz condition of degree $\beta \in (1, 2]$ for all $\lambda \neq 0$, and for some $0 < G, E_\beta < \infty, d > -1/2, s = \lfloor d + 1/2 \rfloor$, as $\lambda \to 0, f_{U^{(s)}}(\lambda) = G|\lambda|^{-2(d-s)} + E_\beta|\lambda|^{-2(d-s)+\beta} + o(|\lambda|^{-2(d-s)+\beta})$.

This condition holds under Assumption A.6 for $\beta > 1$; see also (A.1). We consider now the full cosine bell taper (4) and define the normalized cosine-tapered DFT $v_{\cos}(\lambda) = w(\lambda)/f(\lambda)^{1/2}$.

Theorem A.2 (Cosine Bell). Under Assumptions A.11 and A.12, $d \in (-1/2, 3/2)$ (and $\mu = 0$ if $d \geq 1/2$), for any sequences of positive integers $j = j(n)$ and $k = k(n), 3 < k + 2 < j \leq n/2$, defining $\gamma_{j,k} \equiv (jk)^{d-3}\log(k+1)$, as $n \to \infty$, (a) $E[v_{\cos}\lambda_j)\overline{v_{\cos}}(\lambda_j)] = 1 + O(\min\{j^{-\beta}, j^{-1}\} + \gamma_{j,j})$, (b) $E[v_{\cos}(\lambda_j)v_{\cos}(\lambda_j)] = O(j^{-4} + \gamma_{j,j})$, and (c) $E[v_{\cos}(\lambda_j)\overline{v_{\cos}}(\lambda_k)], E[v_{\cos}(\lambda_j)v_{\cos}(\lambda_k)] = O(k^{-1}|j-k|^{-2} + \gamma_{j,k})$, and when $k = j+1$ and $k = j+2$, all of the previous statements are true with (c′) $E[v_{\cos}(\lambda_j)\overline{v_{\cos}}(\lambda_{j+1})] = -2/3 + O(j^{-1} + \gamma_{j,j})$ and (c″) $E[v_{\cos}(\lambda_j)\overline{v_{\cos}}(\lambda_{j+2})] = 1/6 + O(j^{-1} + \gamma_{j,j})$.

We now analyze the covariance matrix of the (normalized) tapered DFT with tapers of order $p > 1, v_p(\lambda)$. The periodogram is now asymptotically unbiased for any $p > d$ at frequencies λ_{jp}, j integer, but tapering destroys the orthogonality of the sine and cosine transforms at close frequencies.

Theorem A.3 [$p \geq 2$]. Under Assumptions A.11 and A.12, $d > -1/2$, for $f_{U^{(s)}}$, a data taper of order $p = 2, 3, \ldots$, with $p > d$ ($p \geq s+1$ if $\mu \neq 0$), for any sequences of positive integers $k = k(n)$ and $j = j(n), 1 \leq k < j \leq n/(2p)$, defining $\gamma_{j,k} \equiv (jk)^{d-p}\log(k+1)$, as $n \to \infty$, (a) $E[v_p(\lambda_{jp})\overline{v_p}(\lambda_{jp})] = 1 + O(\min\{j^{-\beta}, j^{-1}\} + \gamma_{j,j})$, (b) $E[v_p(\lambda_{jp})v_p(\lambda_{jp})] = O(j^{-p} + j^{-1-p}\log(j+1) + \gamma_{j,j})$, and (c) $E[v_p(\lambda_{jp})\overline{v_p}(\lambda_{kp})], E[v_p(\lambda_{jp})v_p(\lambda_{kp})] = O(k^{-1}|j-k|^{1-p} + k^{-1}|j-k|^{-p}\log n + |j-k|^{-p} + \gamma_{j,k})$.

In (c), the term $\log n$ appears only if $p = 2$. Theorem A.2's bounds are similar to Theorem A.3's for $p = 3$, at all Fourier frequencies but only for $d < 3/2$, so the cosine bell taper shares some properties with tapers of order $p = 3$, although it cannot filter out polynomial trends.

We now present two lemmas for the consistency and uniform consistency in probability of discrete averages of periodogram ordinates of possibly nonstationary (and tapered) observations, which can be seen as specific quadratic forms of $X_t, t = 1, \ldots, n$. All functions are assumed to be periodic of period 2π. Proofs are collected in Appendix B.

Lemma A.1. For an even function $\psi(\lambda)$, differentiable at all $\lambda \neq 0$, let $\psi(\lambda) = O(f^{-1}(\lambda)|\lambda|^{-\delta})$ and $\psi_\lambda(\lambda) = O(f^{-1}(\lambda)|\lambda|^{-1-\delta})$ as $\lambda \to 0, \delta \in (0, 1)$, and let $H = \int_{-\pi}^{\pi} \psi(\lambda) f(\lambda)\, d\lambda < \infty$ and $H_n = (2\pi p/n)\sum_{j(p)} \psi(\lambda_j) I^p(\lambda_j)$, for $p = 1, 2, \ldots$. Then, under assumptions A.1, A.2, and A.4, $H_n \to_p H$ as $n \to \infty$ if $p \geq \lfloor d_0 + (1/2) \rfloor + 1$, (only $p > d_0$ if $\mu = 0$ or $d_0 < 1/2$).

Lemma A.2. For an even function $\psi(\lambda; \boldsymbol{\theta})$, let $\psi(\lambda; \boldsymbol{\theta}) = O(f^{-1}(\lambda)|\lambda|^{-\delta})$ as $\lambda \to 0$ be continuously differentiable in λ and $\boldsymbol{\theta}$ at all $(\lambda, \boldsymbol{\theta}), \lambda \neq 0, \boldsymbol{\theta} \in \Theta_1$ compact, with $\psi_\lambda(\lambda; \boldsymbol{\theta}) = O(f^{-1}(\lambda)|\lambda|^{-1-\delta})$ and $\psi_{\boldsymbol{\theta}}(\lambda; \boldsymbol{\theta}) = O(f^{-1}(\lambda)|\lambda|^{-\delta}), \delta = \delta(\boldsymbol{\theta}) \in (0, 1)$ for all $\boldsymbol{\theta} \in \Theta_1$, and let $H(\boldsymbol{\theta}) = \int_{-\pi}^{\pi} \psi(\lambda; \boldsymbol{\theta}) f(\lambda)\, d\lambda < \infty$, and, for $p = 1, 2, \ldots, H_n(\boldsymbol{\theta}) = (2\pi p/n)\sum_{j(p)} \psi(\lambda_j; \boldsymbol{\theta}) I^p(\lambda_j)$. Then, under assumptions A.1, A.2, and A.4, $\sup_{\boldsymbol{\theta} \in \Theta_1} |H_n(\boldsymbol{\theta}) - H(\boldsymbol{\theta})| \to_p 0$ as $n \to \infty$ if $p \geq \lfloor d_0 + 1/2 \rfloor + 1$ (only $p > d_0$ if $\mu = 0$ or $d_0 < 1/2$).

The condition on $\delta(\boldsymbol{\theta})$ in Lemma A.2 restricts the permitted values of $\boldsymbol{\theta}$ in the compact set Θ_1. The next lemma analyzes the asymptotic distribution of the periodogram averages.

Lemma A.3. In addition to the assumptions of Lemma A.1 on ψ, where now $\delta > 0$ is arbitrarily small, under assumptions A.1, A.2, A.4, and A.7, assumptions (a) and (b) of Theorem A.2 and $H = 0$, as $n \to \infty, \sqrt{n} H_n \to_d \mathrm{N}(0, 4\pi p \Phi_p \int_{-\pi}^{\pi} \psi(\lambda)\psi'(\lambda) f^2(\lambda)\, d\lambda)$, where $\psi'(\lambda)$ stands for the transpose of $\psi(\lambda)$.

Lemma A.4. Under the conditions of Theorem A.1, for $0 < \varepsilon < 1/2, (2\pi p/n)\sum_{j(p)}\{I^p(\lambda_j) - f(\lambda_j)\}|\lambda_j|^{2(d_0+\varepsilon-1/2)} \to_p 0$, as $n \to \infty$.

Lemma A.5. For a function g, even and periodic (of period 2π), satisfying $g(\lambda) = O(|\lambda|^{-\delta})$ as $\lambda \to 0, 0 < \delta < 1$, and a Lipschitz condition of degree $\alpha \in (0, 1]$ with constant $O(|\lambda|^{-1-\delta})$ [i.e., for $\omega > 0, |g(\lambda + \omega) - g(\lambda)| = O(|\lambda|^{-1-\delta}\omega^\alpha)$ as $\lambda \to 0, p = 1, 2, \ldots$], as $n \to \infty, \int_0^{2\pi} g(\lambda)\, d\lambda - (2\pi p/n)\sum_{j(p)} g(\lambda_j) = O(n^{\delta-\alpha})$.

Lemma A.6. Under the assumptions of Lemma A.3, for $g = \psi f, \lim_{n\to\infty} V_n = \Phi_p 4\pi \int_{-\pi}^{\pi} g^2(\lambda)\, d\lambda$, where

$$V_n = 4\sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 (4\pi)^2 \left(\sum_1^n h_t^2\right)^{-2}$$

$$\times \frac{p}{n} \sum_{j=1}^{n^*} \sum_{k=1}^{n^*} g(\lambda_{jp}) \cos s\lambda_{jp} g(\lambda_{kp}) \cos s\lambda_{kp}.$$

Lemma A.7. Under assumptions A.1–A.4, A.8, and A.9, with $p \geq \lfloor d_0 + 1/2 \rfloor + 1$ (only $p > d_0$ if $\mu = 0$ or $d_0 < 1/2$), as $n \to$

$\infty, \boldsymbol{\theta} = \boldsymbol{\theta}_0 + (2\pi p/n) \sum_{j(p)} \rho_0(\lambda_j)[I^p(\lambda_j) - f(\lambda_j)] + o_p(n^{-1/2})$, where $\rho_0(\lambda) = \boldsymbol{\Xi}_0^{-1}\mathbf{k}_\theta^{-1}(\lambda; \boldsymbol{\theta}_0)$ and $\boldsymbol{\Xi}_0 = \sigma_0^2\boldsymbol{\Sigma}_0/(2\pi)$.

Lemma A.8. If the sequence $\{h_t\}$ is a data taper of order p, for $0 < j < n/2$, as $n \to \infty, h(\lambda_j) = O(j^{-p})$.

Lemma A.9. If the sequence $\{h_t\}$ is a data taper of order p, for $0 < j < n$, as $n \to \infty, \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \cos s\lambda_j = (1/2)(\sum_{t=1}^{n} h_t^2 \cos t\lambda_j)^2 + O(n)$.

APPENDIX B: PROOFS

Proof of Theorem 1

We follow the proof in two steps of theorem 1 of Robinson (1995b).

First Step. Denote $\boldsymbol{\theta}^{(1)} = d$ and $\Theta_1 = \{d : \nabla_1 \leq d \leq \nabla_2\} \times \Theta^{(-1)}$, if $\nabla_1 > d_0 - 1/2$, or, otherwise $\Theta_1 = \{d : d_0 - 1/2 + \varepsilon \leq d \leq \nabla_2\} \times \Theta^{(-1)}$, for some $0 < \varepsilon < 1/2$. Define $\hat{\boldsymbol{\theta}}_1 = \arg\min_{\boldsymbol{\theta}\in\Theta_1} Q_n(\boldsymbol{\theta})$ and $Q(\boldsymbol{\theta}) = \int_{-\pi}^{\pi} f(\lambda)k^{-1}(\lambda; \boldsymbol{\theta})\, d\lambda$.

$\hat{\boldsymbol{\theta}}_1 \to_p \boldsymbol{\theta}_0$ follows by a standard argument for consistency of implicitly defined extremum estimates if we can write $Q_n(\boldsymbol{\theta}) - Q_n(\boldsymbol{\theta}_0) = S(\boldsymbol{\theta}) - U(\boldsymbol{\theta})$, where $S(\boldsymbol{\theta})$ is nonstochastic and constant over n, such that for all $\varepsilon > 0$ there exists $\eta > 0$ such that $\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\varepsilon} S(\boldsymbol{\theta}) \geq \eta$ and also $\sup_{\boldsymbol{\theta}\in\Theta_1} |U(\boldsymbol{\theta})| \to_p 0$. Because there is a unique minimum of $Q(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ from the identifiability conditions in assumption A.1, setting $S(\boldsymbol{\theta}) = Q(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0)$, the condition on S follows from the uniform continuity of $Q(\boldsymbol{\theta})$ on Θ_1. The condition on $U(\boldsymbol{\theta}) = Q_n(\boldsymbol{\theta}) - Q_n(\boldsymbol{\theta}_0) - Q(\boldsymbol{\theta}) + Q(\boldsymbol{\theta}_0)$ follows because $\sup_{\boldsymbol{\theta}\in\Theta_1} |Q_n(\boldsymbol{\theta}) - Q(\boldsymbol{\theta})| \to_p 0$ using Lemmas A.2, $[\psi(\lambda; \boldsymbol{\theta}) = k^{-1}(\lambda; \boldsymbol{\theta})]$, and A.5, $[g(\lambda) = \psi(\lambda; \boldsymbol{\theta})f(\lambda)]$, to approximate uniformly in Θ_1 integrals with sums, and using this last lemma, we get that $\sup_{\boldsymbol{\theta}\in\Theta_1} |Q_n(\boldsymbol{\theta}_0) - Q(\boldsymbol{\theta}_0)| \to 0$ as $n \to \infty$.

Second Step. Recall that $\Theta_1 = \{d : \nabla \leq d \leq \nabla_2\} \times \Theta^{(-1)}$, where $\nabla = \nabla_1$ when $d_0 < 1/2 + \nabla_1$ and $d_0 - 1/2 < \nabla \leq d_0$ otherwise. When $d_0 \geq 1/2 + \nabla_1$, define $\Theta_2 = \{d : \nabla_1 \leq d < \nabla\} \times \Theta^{(-1)}$, or setting $\varepsilon = \nabla - (d_0 - 1/2)$ with $\varepsilon = \varepsilon(\boldsymbol{\theta}_0) > 0$, to be chosen later, in $(0, 1/2), \Theta_2 = \{d : \nabla_1 \leq d < d_0 - 1/2 + \varepsilon\} \times \Theta^{(-1)}$.

Then $\hat{\boldsymbol{\theta}}_1 \to_p \theta_0$ if $\Theta_1 = \Theta$, and the theorem is proved. Consider now the situation where Θ_2 is not empty. We want to show that $\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_1 \to_p 0$. For any $\delta > 0$,

$$\begin{aligned} P(\|\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_1\| \geq \delta) &\leq P(\inf_{\boldsymbol{\theta}\in\Theta_2} Q_n(\boldsymbol{\theta}) \leq \min_{\boldsymbol{\theta}\in\Theta_1} Q_n(\boldsymbol{\theta})) \\ &\leq P(\inf_{\boldsymbol{\theta}\in\Theta_2} Q_n(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0) \leq \delta') \\ &\quad + P(|Q_n(\hat{\boldsymbol{\theta}}_1) - Q(\boldsymbol{\theta}_0)| \geq \delta'), \quad \text{(B.1)} \end{aligned}$$

for any $\delta' > 0$. Because $\hat{\boldsymbol{\theta}}_1 \to_p \boldsymbol{\theta}_0$ (for any $\varepsilon > 0$ in the definition of Θ_1), the second probability tends to 0 as $n \to \infty$, for any $\delta' > 0$. Write $k^*(\lambda; \boldsymbol{\theta}) = |\lambda|^{2\theta^{(1)}} k(\lambda; \boldsymbol{\theta})$, so $f(\lambda; \sigma^2, \boldsymbol{\theta}) = (\sigma^2/2\pi)|\lambda|^{-2\theta^{(1)}} k^*(\lambda; \boldsymbol{\theta}), 0 < c_1 < k^*(\lambda; \boldsymbol{\theta}) < c_2 < \infty$, say, for all λ and $\boldsymbol{\theta}$, under assumption A.1. To show that the first probability in (B.1) is negligible, note that

$$\begin{aligned} \inf_{\boldsymbol{\theta}\in\Theta_2} Q_n(\boldsymbol{\theta}) &= \inf_{\boldsymbol{\theta}\in\Theta_2} \frac{2\pi p}{n} \sum_{j(p)} I^p(\lambda_j)k^{-1}(\lambda_j; \boldsymbol{\theta}) \\ &\geq \frac{2\pi p}{nc_2} \sum_{j(p)} I^p(\lambda_j)|\lambda_j|^{2(d_0+\varepsilon-1/2)}. \end{aligned}$$

The last sum converges in probability (see Lemmas A.4 and A.5 again) to

$$\begin{aligned} &\frac{2\pi p}{nc_2} \sum_{j(p)} f(\lambda_j)|\lambda_j|^{2(d_0+\varepsilon-1/2)} \\ &= \frac{\sigma^2}{2\pi} \frac{2\pi p}{nc_2} \sum_{j(p)} k^*(\lambda_j)|\lambda_j|^{2\varepsilon-1} \\ &\sim \frac{\sigma^2}{2\pi c_2} \int_{-\pi}^{\pi} k^*(\lambda)|\lambda|^{2\varepsilon-1}\, d\lambda \\ &\geq \frac{\sigma^2}{2\pi} \frac{c_1}{c_2} \int_{-\pi}^{\pi} |\lambda|^{2\varepsilon-1}\, d\lambda = \frac{\sigma^2}{2\pi} \frac{c_1}{c_2} \frac{\pi^{2\varepsilon}}{\varepsilon} = C(\varepsilon) > 0, \quad \text{say,} \end{aligned}$$

and $C(\varepsilon)$ can be made as large as desired for any f and $\boldsymbol{\theta}_0$, by choice of ε. Fix $\delta' > 0$ and then pick $\varepsilon > 0$ such that $C(\varepsilon) > Q(\boldsymbol{\theta}_0) + 2\delta'$, and define Θ_1 and $\hat{\boldsymbol{\theta}}_1 \to_p \boldsymbol{\theta}_0$ so the first term in (B.1) tends to 0 as $n \to \infty$, and thus $\hat{\boldsymbol{\theta}} \to_p \hat{\boldsymbol{\theta}}_1$. The consistency of $\hat{\sigma}^2$ follows from that of $\hat{\boldsymbol{\theta}}$ and Lemma A.2.

Proof of Theorem 2

Define $\psi_0(\lambda) = \mathbf{k}_\theta^{-1}(\lambda; \boldsymbol{\theta}_0)$, and then use Theorem 1 and Lemmas A.3 and A.7 and that $\boldsymbol{\Sigma}_0^{-1} \int_{-\pi}^{\pi} \psi_0(\lambda)\psi_0'(\lambda)k^2(\lambda; \boldsymbol{\theta}_0)\, d\lambda \boldsymbol{\Sigma}_0^{-1} = \boldsymbol{\Sigma}_0^{-1}$.

Proof of Lemma A.1

We prove the lemma by approximating the periodogram of the (possibly tapered) observed series by that of the (possibly tapered) linear innovations, $I^{p,\varepsilon}(\lambda_j)$, times the transfer function, including the unit root filters. Define $p = 1, 2, \ldots, H_n^\varepsilon = (2\pi)^2(p/n) \sum_{j(p)} \psi(\lambda_j)f(\lambda_j)I^{p,\varepsilon}(\lambda_j)$. So, using Theorems A.1 and A.3 in Appendix A, and assuming evenness of all functions,

$$\begin{aligned} &H_n - H_n^\varepsilon \\ &= \frac{4\pi}{n} \sum_{j=1}^{n^*} \psi(\lambda_{jp})[I^p(\lambda_{jp}) - 2\pi f(\lambda_{jp})I^{p,\varepsilon}(\lambda_{jp})] + o_p(1), \quad \text{(B.2)} \end{aligned}$$

where $n^* = \lfloor[(n/p) - 1]/2\rfloor$. We now distinguish the cases with and without tapering.

No Tapering [p = 1]. Consider the case with $d_0 < 1, \mu = 0$, or $d_0 < 1/2$, for any μ, and $s = 1$ or 0. Using the same arguments as those used in the proof of theorem 2 of Velasco (1999b) (see also the proof of thm. 1 of Robinson 1995b), $0 < j \leq n/2, d_0 \in (-1/2, 1), f(\lambda_j) = (2\pi)^{-1}|1 - e^{-i\lambda_j}|^{-2s}|\alpha(\lambda_j)|^2$, to show that $E|I(\lambda_j) - 2\pi f(\lambda_j)I^\varepsilon(\lambda_j)| = O(f(\lambda_j)[j^{-1/2}(\log j)^{1/2} + j^{d_0-1}(\log j)^{1/2}])$, we find from (B.2) that, using $\psi(\lambda_j) = O(f^{-1}(\lambda_j)|\lambda_j|^{-\delta}), 0 < \delta < 1$,

$$\begin{aligned} H_n - H_n^\varepsilon &= O_p(n^{-1/2}(\log n)^{1/2} + n^{\delta-1}(\log n)^{3/2}) \\ &\quad + O_p(n^{d_0-1}(\log n)^{1/2}) + o_p(1), \end{aligned}$$

which is $o_p(1)$ if $d_0 < 1$. The expectation of H_n^ε is with Lemma A.5, using the continuity of $f(\lambda)$ and $\psi(\lambda)$, and the integrability of $f(\lambda)\psi(\lambda), E[H_n^\varepsilon] = (2\pi/n)\sum_{j=1}^{n-1} \psi(\lambda_j)f(\lambda_j) \sim \int_{-\pi}^{\pi} \psi(\lambda)f(\lambda)\, d\lambda < \infty$, as $n \to \infty$. Now, by summation by parts, for a positive constant C,

$$\begin{aligned} &\left| \frac{4\pi}{n} \sum_{1}^{n^*} f(\lambda_j)\psi(\lambda_j)\{2\pi I^\varepsilon(\lambda_j) - 1\} \right| \\ &\leq \frac{C}{n} \sum_{r=1}^{n^*} |f(\lambda_r)\psi(\lambda_r) - f(\lambda_{r+1})\psi(\lambda_{r+1})| \end{aligned}$$

$$\times \left| \sum_{j=1}^{r} \{2\pi I^{\varepsilon}(\lambda_j) - 1\} \right|$$

$$+ \frac{C}{n} \left| \sum_{1}^{n^*} \{2\pi I^{\varepsilon}(\lambda_j) - 1\} \right| |f(\lambda_{n/2})\psi(\lambda_{n/2})|. \quad \text{(B.3)}$$

Following the discussion of Robinson (1995b, pp. 1637–1638), we obtain that for $1 \le r \le n/2$,

$$\left| \sum_{1}^{r} \{2\pi I^{\varepsilon}(\lambda_j) - 1\} \right| = o_p(r) + O_p(r^{1/2}), \quad \text{(B.4)}$$

and using the properties of $f(\lambda)$ and $\psi(\lambda)$ and the mean value theorem, (B.3) is $o_p(n^{-1} \sum_{r=1}^{n} \lambda_r^{-1-\delta} n^{-1} r + 1) = o_p(1)$.

Tapering [p > 1]. We obtained in the proof of theorem 5 of Velasco (1999b) that under the conditions of this lemma, $E|I^p(\lambda_{jp}) - 2\pi I^{p,\varepsilon}(\lambda_{jp})| = O(f(\lambda_{jp})[j^{-1/2} + j^{d_0-p}(\log j)^{1/2}])$, $0 < j < n/(2p)$, so $H_n - H_n^{\varepsilon} = O_p(n^{\delta-1}\log n + n^{-1/2} + n^{d_0-p}(\log n)^{1/2})$, which is $o_p(1)$ if $p > d_0$. The expectation of H_n^{ε} for $p > 1$ is calculated as for $p = 1$. Now we can write

$$2\pi I^{p,\varepsilon}(\lambda_{jp}) - 1 = \frac{1}{\sum h_t^2} \sum_{t=1}^{n} h_t^2(\varepsilon_t^2 - 1) \quad \text{(B.5)}$$

$$+ \frac{1}{\sum h_t^2} \sum_t \sum_{s \ne t} h_t h_s \varepsilon_t \varepsilon_s \cos(s-t)\lambda_{jp}. \quad \text{(B.6)}$$

With $\gamma_h = \lim_{n\to\infty} \sum h_t^2/n, 0 < \gamma_h < \infty$, the right side of (B.5) is

$$\frac{1}{\sum h_t^2} \left\{ \frac{1}{n} \sum_{t=1}^{n} (h_t^2\varepsilon_t^2 - \gamma_h) + \gamma_h - \frac{\sum h_t^2}{n} \right\},$$

which is $o_p(1)$ because $1/n \sum_{t=1}^{n} h_t^2\varepsilon_t^2 - \gamma_h \to_p 0$ from theorem 1 of Heyde and Seneta (1972) (cf. the proof of thm. 1 in Robinson 1995b), because the triangular array $h_t\varepsilon_t$ satisfies the same regularity conditions as ε_t because $|h_t| \le 1$ and $\gamma_h = \lim_{n\to\infty} 1/n \sum_{t=1}^{n} E[h_t^2\varepsilon_t^2|\mathcal{F}_{t-1}] > 0$ a.s.

Next, we consider the contribution of (B.6). For $0 < r < n/(2p)$ and $h(j,k) = \sum_{t=1}^{n}\sum_{s=1}^{n} h_t^2 h_s^2 \cos(s-t)\lambda_{jp} \cos(s-t)\lambda_{kp}$,

$$E\left(\sum_t \sum_{s\ne t} h_t h_s \varepsilon_t \varepsilon_s \sum_{j=1}^{r} \cos(s-t)\lambda_{jp} \right)^2$$

$$= 2\sum_t \sum_{s \ne t} h_t^2 h_s^2 \left(\sum_{j=1}^{r} \cos(s-t)\lambda_{jp} \right)^2$$

$$= 2\sum_{j=1}^{r}\sum_{k=1}^{r} \left(h(j,k) - \sum_t h_t^4 \right). \quad \text{(B.7)}$$

Then, changing variables and using trigonometric identities (see also lemma 7 of Velasco 1999b),

$$h(j,k) = \sum_{t=1}^{n} \sum_{s=1-t}^{n-t} h_t^2 h_{s+t}^2 \cos s\lambda_{jp} \cos s\lambda_{kp}$$

$$= \frac{1}{2} \sum_{t=1}^{n} \sum_{s=1-t}^{n-t} h_t^2 h_{s+t}^2 (\cos s\lambda_{(j+k)p} + \cos s\lambda_{(j-k)p})$$

$$= \sum_{t=1}^{n} \sum_{s=1}^{n-t} h_t^2 h_{s+t}^2 (\cos s\lambda_{(j+k)p} + \cos s\lambda_{(j-k)p}).$$

Using Lemmas A.8 and A.9, this is

$$\frac{1}{2}\left(\left[\sum_{1}^{n} h_t^2 \cos t\lambda_{(j+k)p} \right]^2 + \left[\sum_{1}^{n} h_t^2 \cos t\lambda_{(j-k)p} \right]^2 + O(n) \right)$$

$$= O(n^2[|j+k|^{-2p} + |j-k|^{-2p}] + n),$$

so (B.7) is $O(n^2 r + r^2 n), 1 \le r \le n/(2p)$. Therefore, (B.4) holds for $p > 1$, and the lemma now follows as when $p = 1$ using (B.3).

Proof of Lemma A.2

This proof follows from the pointwise convergence in Lemma A.1 and an equicontinuity argument using the compactness of Θ_1, and the differentiability of $\psi(\lambda, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ (cf. Hannan 1973).

Proof of Lemma A.3

We consider only the scalar case, the argument for the vector case being identical but notationally more complex, because the stochastic argument, $I^p(\lambda_j)$, is scalar. We follow the same procedure as in the proof of Lemma A.1.

No Tapering (p = 1). Using the second moments of the periodogram and Robinson's (1995b, pp. 1648–1651) procedure, in lemma 1 of Velasco (1999b), $1 \le r \le n/2, d \in (-1/2, 1)$, we find that

$$\sum_{j=1}^{r} \left\{ \frac{I(\lambda_j)}{f(\lambda_j)} - 2\pi I^{\varepsilon}(\lambda_j) \right\}$$

$$= O_p(r^{1/3}(\log n)^{2/3} + r^{1/2}n^{-1/4} + r^{1/(5-4d_0)}(\log r)^{2/(5-4d_0)} + r^{2d_0-1}\log r + n^{-1/2}r^{(1+d_0)/2}(\log n)^{5/4} + n^{-1/4}r^{d_0}(\log r)^{1/2}).$$

Now, using the same arguments and $\psi(\lambda) = O(f^{-1}(\lambda)|\lambda|^{-\delta})$ as $\lambda \to 0, H_n - H_n^{\varepsilon}$ is

$$O_p(n^{\delta-1}[n^{1/3}(\log n)^{1/3} + n^{1/4} + n^{1/(5-4d_0)}(\log n)^{2/(5-4d_0)} + n^{2d_0-1}\log n + n^{d_0/2}(\log n)^{5/4} + n^{d_0-1/4}(\log n)^{1/2}]),$$

which is $o_p(n^{-1/2})$ if $d_0 < 3/4$.

Tapering (p > 1). Velasco (1999b) obtained that $\beta > 1, 1 \le r < n/(2p)$,

$$\sum_{j=1}^{r} \left(\frac{I^p(\lambda_{jp})}{f(\lambda_{jp})} - 2\pi I^{p,\varepsilon}(\lambda_{jp}) \right)$$

$$= O_p(r^{1-\beta/2} + \log r + r^{d_0-p+1}(\log n)^{1/2}),$$

and so adapting the proof, $H_n - H_n^{\varepsilon} = O_p(n^{\delta-1}[\log n + n^{1-\beta/2} + n^{d_0-p+1}(\log n)^{1/2}]) = o_p(n^{-1/2})$ if $\beta > 1, p > d_0 + 1/2$.

We now consider simultaneously the situations $p = 1$ and $p > 1$, but stressing the tapering situation, the untapered case being simpler because many bounds are exactly 0 due to the exact orthogonality of the sine and cosine functions. We have for $g = f\psi$ that $\int_{-\pi}^{\pi} g(\lambda)\, d\lambda = 0$, so $(2\pi p/n)\sum_{j(p)} g(\lambda_j) = O(n^{\delta-1}) = o(n^{-1/2})$ from Lemma A.5, and $H_n^{\varepsilon} = H_n^* + o_p(n^{-1/2})$, with

$$H_n^* = \frac{4\pi p}{n} \sum_{j=1}^{n^*} g(\lambda_{jp})\{2\pi I^{p,\varepsilon}(\lambda_{jp}) - 1\}.$$

Then $E[H_n^*] = 0$ and $\sqrt{n/p}H_n^* = \sum_{t=1}^{n} z_t$, where $z_t = h_t\varepsilon_t \sum_{s=1}^{t-1} h_s\varepsilon_s c_{t-s}$ is a martingale difference sequence and $c_s =$

$4\pi(\sum h_t^2)^{-1}\sqrt{p/n}\sum_{j=1}^{n^*} g(\lambda_{jp})\cos(s\lambda_{jp})$. Now we follow the same method of proof as that of Robinson (1995b, thm. 2), to show the asymptotic normality of H_n^*. First, we need to show that

$$\sum_1^n E[z_t^2|\mathcal{F}_{t-1}] \underset{p}{\to} \Phi_p 4\pi \int_{-\pi}^{\pi} g^2(\lambda)\, d\lambda. \tag{B.8}$$

The left side is

$$\sum_{t=2}^n h_t^2 \sum_{s=1}^{t-1} h_s^2 \varepsilon_s^2 c_{t-s}^2 + \sum_{t=1}^n h_t^2 \sum_{s=1}^{t-1}\sum_{r\neq s}^{t-1} h_s\varepsilon_s h_r\varepsilon_r c_{t-s}c_{t-r}. \tag{B.9}$$

The first term in (B.9) is

$$\sum_{t=1}^{n-1} h_t^2(\varepsilon_t^2-1)\sum_{s=1}^{n-t} h_{s+t}^2 c_s^2 + \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 c_s^2 = B_1 + B_2,$$

say. B_1 is $o_p(1)$, because it has mean 0 and variance $O(\sum_1^{n-1}(\sum_{s=1}^{n-t} c_s^2)^2)$, and from Robinson (1995b, p. 1646), we obtain, using summation by parts, for any $\delta > 0, c_s = c_{n-s}$ and

$$\begin{aligned}|c_s| &\le C\left| n^{-3/2}\sum_{j=1}^{n^*}(g(\lambda_{jp}) - g(\lambda_{(j+1)p}))\sum_{l=1}^{j}\cos s\lambda_{tp} \right.\\ &\quad \left. + n^{-3/2} g(\lambda_{pn^*})\sum_{j=1}^{n^*}\cos s\lambda_{jp}\right|\\ &\le Cn^{-1/2}s^{-1}\left\{n^\delta \sum_{j=1}^{n^*-1} j^{-1-\delta} + 1\right\} = O(n^{\delta-1/2}s^{-1}),\end{aligned}$$

for $1 \le s \le n/2$, so $\sum_{s=1}^n c_s^2 = O(n^{2\delta-1})$. By Lemma A.6, $B_2 = V_n \sim \Phi_p 4\pi\int_{-\pi}^{\pi} g^2(\lambda)\, d\lambda$.

The second term in (B.9) can be shown to be $o_p(1)$, using the same argument (see also Velasco, 1999b, lemma 6), because it has mean 0 and variance

$$\begin{aligned}&2\sum_{t=2}^n h_t^2 \sum_{u=2}^n h_u^2 \sum_{s}^{\min\{t-1,u-1\}}\sum_{r\neq s} h_s^2 h_r^2 c_{t-r}c_{t-s}c_{u-r}c_{u-s}\\ &= 2\sum_{t=2}^n h_t^4 \sum_s\sum_{r\neq s} h_s^2h_r^2 c_{t-r}^2 c_{t-s}^2\\ &\quad +4\sum_{t=3}^n h_t^2\sum_{u=2}^{t-1} h_u^2 \sum_s^{u-1}\sum_{r\neq s}^{u-1} h_s^2h_r^2 c_{t-r}c_{t-s}c_{u-r}c_{u-s},\end{aligned}$$

because the weights $\{h_t\}$ are symmetric around $\lfloor n/2\rfloor$. Using the bounds for c_s and $\sum_1^n c_s^2$, and because $\sup_t |h_t| \le 1$, the first term is $O(n^{2\delta-1}) = o(1)$, and the second term has absolute value bounded by

$$\begin{aligned}&4\sum_{t=3}^n\sum_{u=2}^{t-1}\left(\sum_s^{u-1} c_{t-r}^2 \sum_{r\neq s}^{u-1} c_{u-r}^2\right)\\ &\le 4\left(\sum_1^n c_t^2\right)\left(\sum_{t=3}^n\sum_{u=2}^{t-1}\sum_{r=t-u+1}^{t-1} c_r^2\right)\\ &\le 4\left(\sum_1^n c_t^2\right)\left(2n\sum_{j=1}^{n^*} jc_j^2\right),\end{aligned}$$

and this is $O(n^{4\delta-1}\log n) = o(1)$. Thus (B.8) is proved.

Finally, we need to show that $\sum_1^n E[z_t^2 I(|z_t| > \rho)] \to 0$ for all $\rho > 0$, for which we can check the sufficient condition $\sum_1^n E[z_t^4] \to 0$. Following Robinson (1995b), $\sum_1^n E[z_t^4] \le Cn(\sum_1^n c_s^2)^2 = O(n^{2\delta-1}) = o(1)$, and the central limit theorem follows.

Proof of Lemma A.4

For $\varepsilon > 0, \psi(\lambda) = |\lambda|^{2(d_0+\varepsilon-1/2)} = O(f^{-1}(\lambda)|\lambda|^{2\varepsilon-1})$ satisfies the conditions of Lemma A.1, with $\delta = 2\varepsilon - 1, 0 < \delta < 1$.

Proof of Lemma A.5

This follows from the discussion of Robinson (1994a, p. 75).

Proof of Lemma A.6

First, using trigonometric identities, we have that V_n is

$$\begin{aligned}&4(4\pi)^2\frac{p}{n}\sum_{j=1}^{n^*}\sum_{k=1}^{n^*} g(\lambda_{jp})g(\lambda_{kp})\left(\sum_1^n h_t^2\right)^{-2}\\ &\quad\times \frac{1}{2}\sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2\{\cos s\lambda_{(j+k)p} + \cos s\lambda_{(j-k)p}\},\end{aligned}$$

and, using Lemma A.9 and $h(\lambda_j) = (\sum_1^n h_t^2)^{-1}\sum_1^n h_t^2\cos t\lambda_j$, this is

$$\begin{aligned}&\frac{(4\pi)^2 p}{n}\sum_{j=1}^{n^*}\sum_{k=1}^{n^*} g(\lambda_{jp})\\ &\quad\times g(\lambda_{kp})\{h^2(\lambda_{(j+k)p}) + h^2(\lambda_{(j-k)p})\},\end{aligned} \tag{B.10}$$

plus a term of smaller order of magnitude that is $O(n^{-3}\sum_j\sum_k g(\lambda_{jp})g(\lambda_{kp})) = o(1)$. From Lemma A.8, $h(\lambda_j) = O(|j|^{-p})$, and with $g(\lambda) = O(|\lambda|^{-\delta})$ for any $\delta > 0$, the term in $h^2(\lambda_{(j+k)p})$ of (B.10) is

$$O\left(n^{2\delta-1}\sum_{j=1}^n\sum_{k=1}^n (j+k)^{-2p}\right) = O\left(n^{2\delta-1}\sum_{j=1}^n\sum_{k=1}^n j^{-p}k^{-p}\right),$$

which is $O(n^{2\delta-1}) = o(1)$ and can be ignored. The other term with $h^2(\lambda_{(j-k)p})$ in (B.10) is $0 \le \eta \le n/2$,

$$\begin{aligned}&(4\pi)^2\frac{p}{n}\sum_{j=1}^{n^*} g^2(\lambda_{jp})\sum_{k:|j-k|\le\eta} h^2(\lambda_{(j-k)p})\\ &\quad + O\left(n^{2\delta-1}\sum_{j=1}^{n^*} j^{-\delta}\sum_{k:|j-k|>\eta}|j-k|^{-2p}\right)\\ &\quad + n^{-1}\sum_{j=1}^{n^*}\sum_{k:1\le|j-k|\le\eta}|g(\lambda_{jp})|\\ &\quad \times O(|\sup_{|j-k|\le\eta}|g(\lambda_{jp}) - g(\lambda_{kp})||j-k|^{-2p})\end{aligned}$$

by Lemma A.8. This is, using that $\sup_{|j-k|\le\eta}|g(\lambda_{jp}) - g(\lambda_{kp})| = O(|g(\lambda_{kp})||\lambda_k|^{-1}|\lambda_k - \lambda_j| + |g(\lambda_{jp})||\lambda_j|^{-1}|\lambda_k - \lambda_j|)$,

$$\begin{aligned}&(4\pi)^2\frac{p}{n}\sum_{j(p)} g^2(\lambda_j)\sum_{k=0,p,2p}^{n-p} h^2(\lambda_{kp})\\ &\quad + O(n^\delta\eta^{1-2p}) + O\left(n^{2\delta-1}\sum_{j=1}^{n^*} g^2(\lambda_{jp})\sum_{k>\eta}^n k^{-2p}\right)\\ &\quad + O\left(n^{2\delta-1}\sum_{j=1}^n j^{-1-2\delta}\sum_{k=1+j}^{j+\eta}|j-k|^{1-2p}\right)\end{aligned}$$

$$+ O\left(n^{2\delta-1}\left\{\sum_{j=1}^{2\eta} + \sum_{j=1+2\eta}^{n}\right\} j^{-\delta} \sum_{k=1+j}^{j+\eta} k^{-1-\delta}|j-k|^{1-2p}\right)$$

$$= \Phi_p 4\pi \int_{-\pi}^{\pi} f^2(\lambda)\psi^2(\lambda)\,d\lambda + O(n^{2\delta}\eta^{1-2p})$$

$$+ O(n^{2\delta-1} + n^{2\delta-1}\eta^{1-\delta} + n^{\delta}\eta^{-1-\delta} + n^{\delta}\eta^{1-2p}),$$

with all the error terms being $o(1)$ on choosing, for example, $\eta \sim \sqrt{n}$, with $p > 1$.

Proof of Lemma A.7

By the definition of $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}, \int \mathbf{k}_{\boldsymbol{\theta}}^{-1}(\lambda;\boldsymbol{\theta}_0)f(\lambda)d\lambda = (2\pi p/n)\sum \mathbf{k}_{\boldsymbol{\theta}}^{-1}(\lambda_j;\hat{\boldsymbol{\theta}})I^p(\lambda_j) = 0$. It follows by the mean value theorem that $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \tilde{\Xi}_n^{-1}(\mathbf{b}_n - \zeta)$, where $\zeta = (2\pi p/n)\sum \mathbf{k}_{\boldsymbol{\theta}}^{-1}(\lambda_j;\boldsymbol{\theta}_0)[I^p(\lambda_j) - f(\lambda_j)] = O_p(n^{-1/2})$ from Lemma A.3, $\mathbf{b}_n = (2\pi p/n)\sum \mathbf{k}_{\boldsymbol{\theta}}^{-1}(\lambda_j;\boldsymbol{\theta}_0)I^p(\lambda_j) - \int \mathbf{k}_{\boldsymbol{\theta}}^{-1}(\lambda;\boldsymbol{\theta}_0)f(\lambda)\,d\lambda$, and the lth row of $\tilde{\Xi}_n$ is the lth row of the matrix $\Xi_n(\boldsymbol{\theta}) = (2\pi p/n)\sum \mathbf{k}_{\boldsymbol{\theta\theta}'}^{-1}(\lambda_j;\boldsymbol{\theta})I^p$ evaluated at $\tilde{\theta}_l$, which is in the line segment between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}$. Thus $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = -\Xi_0^{-1}\zeta - \mathbf{a}_n\zeta + \tilde{\Xi}_n^{-1}\mathbf{b}_n$, with $\mathbf{a}_n = \tilde{\Xi}_n^{-1} - \Xi_0^{-1}$. The lemma follows if $\mathbf{a}_n = o_p(1), \|\tilde{\Xi}_n^{-1}\| = O_p(1)$ and $\mathbf{b}_n = o_p(n^{-1/2})$.

First, we bound $\mathbf{a}_n : \|\mathbf{a}_n\| \leq \|\tilde{\Xi}_n^{-1}\|\|\Xi_0^{-1}\|\|\Xi_0 - \tilde{\Xi}_n\|$. From assumption A.9, $\|\Xi_0^{-1}\| < \infty$, and under the conditions of the lemma, $\tilde{\theta}_l \to_p \boldsymbol{\theta}_0$, so using the continuity of the elements of $\Xi_n^{-1}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ (to substitute $\tilde{\theta}_l$ by $\boldsymbol{\theta}$), the differentiability in λ of the second derivatives of k^{-1} for Lemma A.1 (to substitute f for I^p), and approximating sums by integrals with Lemma A.5, we can show the elementwise convergence in probability to 0 of $\Xi_0 - \tilde{\Xi}_n$, and we obtain for n sufficiently large that $\|\tilde{\Xi}_n\| < \infty$ with probability approaching 1, and thus $\mathbf{a}_n = o_p(1)$.

The bound for b_n follows by the previous argument, using Lemmas A.1 and A.5 to approximate I^p by f and sums by integrals, using the differentiability in λ of $k(\lambda;\boldsymbol{\theta})$ and $\mathbf{k}_{\boldsymbol{\theta}}^{-1}(\lambda;\boldsymbol{\theta})$, whose derivatives are $O(|\lambda|^{-2d-1})$ and $O(|\lambda|^{2d-1-\delta})$ as $\lambda \to 0$, for some $\delta > 0$.

Proof of Lemma A.8

This proof follows using the properties of a taper of order p, as in lemma 1 of Velasco (1999b), and that $\sum_{t=1}^{n} h_t^2 \cos t\lambda_j = \int_{-\pi}^{\pi} D_h(\lambda_j - \lambda)D_h(\lambda)\,d\lambda$.

Proof of Lemma A.9

This is part (B) of lemma 7 of Velasco (1999b).

[Received October 1998. Revised June 1999.]

REFERENCES

Alekseev, V. G. (1996), "Jackson and Jackson–Vallée Poussin-Type Kernels and Their Probability Applications," *Theory of Probability and Its Applications*, 41, 137–142.

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

——— (1995), "Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short- and Long-Memory ARIMA Models," *Journal of the Royal Statistical Society*, Ser. B, 57, 659–672.

Beran, J., Bhansali, R. J., and Ocker, D. (1998), "On Unified Model Selection for Stationary and Nonstationary Short- and Long-Memory Autoregressive Processes," *Biometrika*, 85, 921–934.

Bloomfield, P. (1973), "An Exponential Model for the Spectrum of a Scalar Time Series," *Biometrika*, 60, 217–226.

Box, G. E., and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.

Dahlhaus, R. (1985), "Asymptotic Normality of Spectral Estimates," *Journal of Multivariate Analysis*, 16, 412–431.

——— (1988), "Small-Sample Effects in Time Series Analysis: A New Asymptotic Theory and a New Estimate," *The Annals of Statistics*, 16, 808–841.

——— (1989), "Efficient Parameter Estimation for Self-Similar Processes," *The Annals of Statistics*, 14, 517–532.

Davies, R. B., and Harte, D. S. (1987), "Tests for the Hurst Effect," *Biometrika*, 74, 95–101.

Fox, R., and Taqqu, M. S. (1986), "Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series," *The Annals of Statistics*, 14, 517–532.

Geweke, J., and Porter-Hudak, S. (1983), "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis*, 4, 221–238.

Giraitis, L., and Surgailis, D. (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.

Hannan, E. J. (1973), "The Asymptotic Theory of Linear Time Series Models," *Journal of Applied Probability*, 10, 130–145.

Haslett, J., and Raftery, A. E. (1989), "Space-Time Modeling With Long Memory Dependence: Assessing Ireland's Wind Power Resource" (With discussion), *Applied Statistics*, 38, 1–50.

Hauser, M. A. (1999), "Maximum Likelihood Estimates for ARFIMA Models: A Monte Carlo Study," *Journal of Statistical Planning and Inference*, 80, 229–255.

Hosoya, Y. (1996), "The Quasi-Likelihood Approach to Statistical Inference on Multiple Time Series With Long-Range Dependence," *Journal of Econometrics*, 73, 217–236.

Hosoya, Y., and Taniguchi, M. (1982), "A Central Limit Theorem for Stationary Processes and the Parameter Estimation of Linear Processes," *The Annals of Statistics*, 10, 132–153.

Hurvich, C. M., and Ray, B. K. (1995), "Estimation of the Memory Parameter for Nonstationary or Noninvertible Fractionally Integrated Processes," *Journal of Time Series Analysis*, 16, 17–42.

Ling, S., and Li, W. K. (1997), "On Fractionally Integrated Autoregressive Moving-Average Time Series Models With Conditional Heteroscedasticity," *Journal of the American Statistical Association*, 92, 1184–1194.

Robinson, P. M. (1978), "Alternative Models for Stationary Stochastic Processes," *Stochastic Processes and Their Applications*, 8, 141–152.

——— (1986), "On the Errors-in-Variables Problem for Time Series," *Journal of Multivariate Analysis*, 19, 240–250.

——— (1994a), "Time Series With Strong Dependence," in *Advances in Econometrics. Sixth World Congress*, Vol. 1, ed. C. A. Sims, Cambridge, U.K.: Cambridge University Press, pp. 47–95.

——— (1994b), "Efficient Tests of Nonstationary Hypotheses," *Journal of the American Statistical Association*, 89, 1420–1437.

——— (1995a), "Log-Periodogram Regression of Time Series With Long-Range Dependence," *The Annals of Statistics*, 23, 1048–1072.

——— (1995b), "Gaussian Semiparametric Estimation of Long-Range Dependence," *The Annals of Statistics*, 23, 1630–1661.

Solo, V. (1984), "The Order of Differencing in ARMA Models," *Journal of the American Statistical Association*, 79, 916–921.

——— (1989), "Asymptotics of Quadratic Forms for Time Series," unpublished manuscript.

——— (1992), "Intrinsic Random Functions and the Paradox of $1/f$ Noise," *SIAM Journal of Applied Mathematics*, 52, 270–291.

Tukey, J. W. (1967), "An Introduction to the Calculation of Numerical Spectrum Analysis," in *Advanced Seminar on Spectral Analysis of Time Series*, ed. B. Harris, New York: Wiley, pp. 25–46.

Velasco, C. (1999a), "Non-Stationary Log-Periodogram Regression," *Journal of Econometrics*, 91, 325–371, 1999.

——— (1999b), "Gaussian Semiparametric Estimation for Non-Stationary Time Series," *Journal of Time Series Analysis*, 20, 87–127.

Zhurbenko, I. G. (1979), "On the Efficiency of Estimates of a Spectral Density," *Scandinavian Journal of Statistics*, 6, 49–56.