Nonparametric frequency domain analysis of nonstationary multivariate time series

Carlos Velasco

Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Avenida de la Universidad 30, E-28911 Leganés, Madrid, Spain

Received 3 December 1999; received in revised form 15 February 2002; accepted 29 April 2002

Abstract

We analyse the properties of nonparametric spectral estimates when applied to long memory and trending nonstationary multiple time series. We show that they estimate consistently a generalized or pseudo-spectral density matrix at frequencies both close and away from the origin and we obtain the asymptotic distribution of the estimates. Using adequate data tapers this technique is consistent for observations with any degree of nonstationarity, including polynomial trends. We propose an estimate of the degree of fractional cointegration for possibly nonstationary series based on coherence estimates around zero frequency and analyse its finite sample properties in comparison with residual-based inference. We apply this new semiparametric estimate to an example vector time series.

MSC: primary 62M10; secondary 60G18; 62G05

Keywords: Long memory; Long-range dependence; Fractional cointegration; Coherence; Semiparametric estimation; Spectral density

1. Introduction

In many empirical studies, it is found that the most relevant property of spectral density estimates of observed time series is a marked peak at zero frequency. This feature is often associated with long-range or trending nonstationary behaviours. However these estimates, usually of nonparametric nature and designed for short-memory series, are constructed without detrending or explicit account of their long-run properties. This makes difficult the application of standard inference rules and the interpretation of such features, well documented otherwise, since the nonstationarity may affect the properties of spectral estimates.

E-mail address: cavelas@est-econ.uc3m.es (C. Velasco).

To describe low-frequency behaviour it is often assumed that the spectral density $f(\lambda)$ of an observed covariance stationary sequence satisfies for a positive constant G,

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

where $d < \frac{1}{2}$ is the parameter that governs the degree of memory of the series, but $f(\lambda)$ can be smooth outside a neighbourhood of the origin. For $d < \frac{1}{2}$ the process is stationary, and (1) allows spectral densities that either diverge, are positive or zero at $\lambda = 0$. If $d \in (0, \frac{1}{2})$ we say that the series exhibits *long-memory* or *long-range dependence*. If $d = 0$ the spectral density is bounded at $\lambda = 0$ and the process is called *short memory* or weakly dependent. When $d < 0$ the spectral density is zero at the origin and the series displays negative memory or *antipersistent* behaviour, due in most cases to overdifferentiation of observed time series. Related properties can be expressed alternatively in the time domain in terms of the autocovariance sequence and hold for fractional processes such as ARFIMA. See e.g. Robinson (1994a) or Beran (1994) for a review of the literature on long-memory or long-range dependent processes.

After integer differencing, many nonstationary series are transformed into (second-order) stationary ones with spectral density satisfying (1), as is the case of standard ARFIMA models with $d \geqslant \frac{1}{2}$. Then we consider the transfer function of the difference operator to define a generalized or pseudo-spectral density (PSD) $f(\lambda)$ with power law behaviour at the origin as in (1), but adding to d the number of integer differences taken to achieve stationarity. This PSD, though with similar shape to the spectral density of the stationary increments for frequencies away from the origin, is not integrable 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), $f(\lambda)$ has an interpretation as the limit of the expectation of the sample periodogram as occurs for stationary series. We show in this paper that this PSD is the quantity actually estimated in practice by smoothed spectral estimates, completing the analysis of estimates of full and semiparametric long-memory models without assumptions about the degree of the possible nonstationarity (see Velasco and Robinson, 2000; Velasco, 1999a, b).

We analyse in this paper the properties of standard nonparametric smoothed spectral estimates for both frequencies close and away from the zero-frequency singularity for possibly long-memory and nonstationary or trending series. In a multivariate context, for estimates based on discrete averages of periodogram ordinates we found similar asymptotic results as those for stationary set-ups and bounded spectral densities itemized in, e.g. Hannan (1970, Section V.5) or Brillinger (1975, Section 4.2). Assuming only local conditions around the frequency of interest we show the consistency and asymptotic normality of the nonparametric estimates for linear processes. Robinson (1994b) considered periodogram averages around the origin to estimate the spectral measure. Hidalgo (1996) and Marinucci (2000) analysed the properties of spectral estimators based on autocovariances for stationary and nonstationary long-memory processes, respectively, under different sets of assumptions, but periodogram-based estimates may be more natural in many contexts as they are often better designed to avoid leaking from remote frequencies. When the memory is very high, tapering the data (Tukey, 1967) might be necessary to reduce the bias in the nonparametric estimation

or to eliminate stochastic and deterministic trends, confirming the desirable resolution properties of such technique for stationary series found by, e.g. Zhurbenko (1979), Dahlhaus (1985) or Robinson (1986).

For the analysis of multivariate long-memory or (fractionally) integrated time series, possibly nonstationary, a key concept is that of cointegration. A vector of time series with equal memory components is cointegrated if a linear combination of them has smaller memory, the order of cointegration being this reduction in persistence measured by the memory parameter. The usual assumption is that the original series have a unit root, $d = 1$, and a linear combination is weakly dependent, $d = 0$, but other possibilities are also plausible as suggested originally by Granger (1981) and Engle and Granger (1987). When no assumption is made about the memory of the series, an additional inference problem is the determination of the cointegration order. This entails estimation of the memory of the original series (and testing for same memory) and of the cointegrating errors, mostly through residuals in an estimated regression model. Using nonparametric estimates of the coherence between two series at frequencies close to the origin we propose in this paper narrow band estimates of the order of cointegration in the spirit of Robinson and Marinucci's (2001) or Marinucci's (2000) slope estimates. We discuss inference based of such estimates and compare its finite sample performance with residual-based semiparametric alternatives.

The paper is organized as follows. We first present in the next section the main definitions and properties of nonstationary long-memory time series and data taper sequences. In Section 3 we define the nonparametric estimates of the PSD and find sufficient conditions for their consistency and asymptotic normality. Section 4 proposes the new estimate of the cointegration order and in the next two sections its finite sample properties are analysed, and it is applied to an example vector time series. Proofs are given in the appendix.

2. Nonstationary time series and data tapers

Following Hurvich and Ray (1995) in a univariate context, we propose a general model for possibly nonstationary integrated vector processes with components $\{X_{rt}\}$, $r = 1, \ldots, R$, each with memory parameter $d_r > -\frac{1}{2}$. We say that the observed sequence X_{rt} has memory $d_r > -\frac{1}{2}$ if $U_{rt} = \Delta^{D_r} X_{rt}$, $D_r = \lfloor d_r + \frac{1}{2} \rfloor$, is stationary with mean μ_r, possibly different from zero, and spectral density $f_{U_r}(\lambda) = g_{rr}(\lambda)$ behaving as $G_{rr}\lambda^{-2(d_r - D_r)}$ around the origin, $-\frac{1}{2} \leqslant d_r - D_r < \frac{1}{2}$. Here $\Delta = 1 - L$, where L is the lag operator. However, the definition of long-memory or fractional nonstationary models in terms of partial sums of stationary long-memory processes we adopt here is not the only possibility to obtain processes with similar long-run properties. Thus, for example, Robinson and Marinucci (2001) and Tanaka (1999) use truncated fractional difference filters that generate nonstationary series for any value of d.

Define the PSD of X_{rt} as

$$f_{rr}(\lambda) = |1 - \exp(\mathrm{i}\lambda)|^{-2D_r} g_{rr}(\lambda) \sim G_{rr}\lambda^{-2d_r} \quad \text{as } \lambda \to 0^+, \tag{2}$$

$0 < G_{rr} < \infty$. When $2d_r \geqslant 1$, $f_{rr}(\lambda)$ is not integrable in $[-\pi, \pi]$ and it is not a spectral density. We assume that $g_{rr}(\lambda)$ is the spectral density of a stationary process, but not necessarily ARMA, and it can be zero or unbounded at frequencies $\lambda \neq 0$, but integrable for second-order stationarity. Note that if $\mu_r \neq 0$ the observed time series has a deterministic component and if $D_r \geqslant 1$ this is a polynomial trend.

Similarly, we define the (pseudo) cross-spectral density of a pair of series (X_{rt}, X_{st}) as

$$f_{rs}(\lambda) = (1 - \exp(\mathrm{i}\lambda))^{-D_r}(1 - \exp(-\mathrm{i}\lambda))^{-D_s} g_{rs}(\lambda) \sim G_{rs}\lambda^{-d_r - d_s} \quad \text{as } \lambda \to 0^+,$$

$0 \leqslant |G_{rs}| < \infty$, where $g_{rs}(\lambda)$ is the cross-spectral density of (U_{rt}, U_{st}), and if $|G_{rs}| = 0$ we account for zero coherence between U_{rt} and U_{st} at zero frequency. See Lobato (1997) for a discussion on multivariate long-memory semiparametric models.

The basic statistic for our frequency domain analysis is the tapered discrete Fourier transform (DFT). The DFT of X_{rt} for a deterministic taper sequence h_t and n observations $t = 1, \ldots, n$ and $r = 1, \ldots, R$, $(\lambda_j = 2\pi j/n)$, is defined as

$$w_r(\lambda_j) = w(X_{rt}, h_t, \lambda_j) = \left(2\pi \sum_{t=1}^{n} h_t^2\right)^{-1/2} \sum_{t=1}^{n} h_t X_{rt} \exp(\mathrm{i}\lambda_j t),$$

and the (cross) periodogram of X_{rt} and X_{rt} is

$$I_{rs}(\lambda_j) = w_r(\lambda_j)\overline{w_s(\lambda_j)},$$

where the overline indicates complex conjugation. Tapering downweights the observations and both extremes of the observed data sequence to control leakage from frequencies where nonstationarity is suspected in the observed time series.

The usual DFT is obtained setting $h_t \equiv 1$, $t = 1, \ldots, n$, while the cosine or Hanning taper is given by $h_t = \frac{1}{2}(1 - \cos[2\pi t/n])$. For sample size $n = 4N$, where N is an integer, the weights h_t^{P} of the Parzen window are

$$h_t^{\mathrm{P}} = \begin{cases} 2\{1 - |(2t - n)/n|\}^3, & 1 \leqslant t \leqslant N \text{ or } 3N \leqslant t \leqslant 4N, \\ 1 - 6[\{(2t - n)/n\}^2 - |(2t - n)/n|^3], & N < t < 3N. \end{cases}$$

Zhurbenko (1979) used a class of data tapers $\{h_t^{\mathrm{Z}}\}$ suggested by Kolmogorov, indexed by the order $p = 1, 2, \ldots$, assuming $N = n/p$ integer. For $p = 3$, Zhurbenko's weights are similar to the cosine window, and when $p = 4$, h_t^{Z} are very close to h_t^{P}, sharing similar asymptotic properties. If $p = 2$, Zhurbenko taper is equal to Barlett's triangular window and when $p = 1$ they are constant.

We denote as $I_{rs}^{p}(\lambda_j) = w_r(\lambda_j)\overline{w_s(\lambda_j)}$ the (cross) periodogram with a taper of order p according with the following definition:

Definition. A sequence of positive data tapers $\{h_t\}_1^n$ symmetric around $\lfloor n/2 \rfloor$ is of *order* p if

(i) $\max_t h_t = 1$, and $\lim_{n\to\infty} (1/n) \sum_{t=1}^{n} h_t^2 = B$ for some $0 < B < \infty$.

(ii) For $N = n/p$ (which we assume integer),

$$\sum_{t=1}^{n} h_t \exp\{\mathrm{i}t\lambda\} = \frac{a(\lambda)}{n^{p-1}} \left(\frac{\sin[n\lambda/2p]}{\sin[\lambda/2]} \right)^p, \tag{3}$$

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

The higher the order, the more dramatic is the effect of tapering, being possible to deal with series with arbitrary high memory if enough tapering is applied, i.e. if p is sufficiently large (Velasco, 1999a). The raw DFT weights, the identity, are of order $p = 1$, and from now on when $p = 1$ we will imply the usual DFT, without tapering.

The effect on tapering can be illustrated by the following properties. Thus, summation by parts yields for a differentiable taper which vanishes at the boundaries, with derivative h_t',

$$w(X_{rt}, h_t, \lambda) \approx \frac{1}{1 - \exp(\mathrm{i}\lambda)} \left[w(\Delta X_{rt}, h_t, \lambda) + \frac{1}{n} w(X_{rt}, h_t', \lambda) \right],$$

explaining, if the term multiplied by $1/n$ is negligible, how a sufficiently smooth taper (i.e. of sufficiently high order p) can deal with arbitrarily high levels of memory d, justifying definition (2). In fact, from Hurvich and Ray (1995) and Velasco (1999a), we obtain in Theorems 1 and 2 (to follow) Solo's (1992) inversion calculation for nonstationary $f_{rr}(\lambda)$,

$$E[I_{rr}^p(\lambda_{jp})] = \left(2\pi \sum_{t=1}^{n} h_t^2 \right)^{-1} \int_{-\pi}^{\pi} |D_h(\lambda - \lambda_{jp})|^2 f_{rr}(\lambda)\, \mathrm{d}\lambda \to f_{rr}(\lambda_{jp}), \tag{4}$$

as $n \to \infty$, where $D_h(\lambda) = \sum_{t=1}^{n} h_t \exp\{\mathrm{i}\lambda t\}$. Then, the tapered periodogram is asymptotically unbiased for the PSD of nonstationary series at Fourier frequencies λ_{jp}, $j \neq 0 \pmod N$.

Furthermore, tapers of order p allow inference for time series with polynomial trends of orders up to $p-1$ without need of identification or estimation because for a data taper of order p, the DFT is invariant to these trends,

$$w(t^\ell, h_t, \lambda_{jp}) = 0, \quad \ell = 0, 1, \ldots, p-1. \tag{5}$$

See Lobato and Velasco (2000) for an application of this property.

We now review in a multivariate context some results obtained in Robinson (1995a) and Velasco (1999a) for the (tapered) DFT of possibly nonstationary time series. Here we are only concerned with positive Fourier frequencies $0 < \lambda_j \leqslant \pi$, since we can analyse negative ones by complex conjugation and symmetry. The regularity conditions on the behaviour of $g_{rr}(\lambda)$ around the frequency ν of interest, $0 \leqslant \nu < \pi$, are summarized in Section 3. The case $\nu = 0$ is of interest for the analysis of the persistence properties of the observed vector series, including cointegration properties.

We first analyse the covariance matrix of the raw $w_r(\lambda_j)$ with no tapering. Define $v_r(\lambda) = w_r(\lambda)/f_{rr}^{1/2}(\lambda)$.

Theorem 1 ($p = 1$). *Under Assumptions* 1 *and* 4, (4′ *for* $v = 0$), $d_r \in (-\frac{1}{2}, 1)$, $r = 1, \ldots, R$, $\mu = 0$, *for sequences of positive integers* $j = j(n)$ *and* $k = k(n)$ *such that* $0 < k < j < n/2$, $\lambda_j, \lambda_k \to v \in [0, \pi]$ *as* $n \to \infty$ *and defining* $\gamma_{j,k} = j^{d_r - 1} k^{d_s - 1} \log(1 + j)$,

(a) $E[v_r(\lambda_j)\bar{v}_s(\lambda_j)] = H_{rs}(\lambda_j) + \mathrm{O}(j^{-1}\log(1 + j) + \gamma_{j,j})$,
(b) $E[v_r(\lambda_j)v_s(\lambda_j)] = \mathrm{O}(j^{-1}\log(1 + j) + \gamma_{j,j})$,
(c) $E[v_r(\lambda_j)\bar{v}_s(\lambda_k)], E[v_r(\lambda_j)v_s(\lambda_k)] = \mathrm{O}(k^{-1}\log(1 + j) + \gamma_{j,k})$.

$H_{rs}(\lambda)$ is the coherence at frequency λ between U_{rt} and U_{st} and $\gamma_{j,k}$ bounds the nonstationarity bias due to the nonintegrability of $f_{rr}(\lambda)$ for $d_r \geqslant 0.5$. For values $d_r \geqslant 1$ the periodogram is not unbiased for the PSD $f_{rr}(\lambda)$, though its expectation is finite for $d < 1.5$ (see Hurvich and Ray, 1995). Tapering helps to control this bias using (3) to deal with nonstationary series and deterministic trends. However, the full advantage of the tapering only shows up when we assume further smoothness conditions on $f(\lambda)$. Denote by $\beta = 1, 2$ the number of derivatives of $f(\lambda)$ around v and set the normalized tapered Fourier transform $v_r^{\mathrm{T}}(\lambda) = w_r^{\mathrm{T}}(\lambda)/f_{rr}^{1/2}(\lambda)$.

Theorem 2 ($p \geqslant 2$). *Under Assumptions* 1, 4, 6 (4′ *and* 6′ *for* $v = 0$), $d_r > -\frac{1}{2}$, *a data taper of order* $p = 2, 3, \ldots$, *with either*

(i) $\mu_r = \mu_s = 0$ *and* $p > \max\{d_r, d_s\}$;
(ii) $p \geqslant \max\{D_r, D_s\} + 1$ *any* μ_r, μ_s;

for sequences of positive integers $k = k(n)$ *and* $j = j(n)$, *and* $\eta \equiv j - k$, $1 \leqslant k < j \leqslant n/(2p)$, $\lambda_{jp}, \lambda_{kp} \to v \in [0, \pi]$ *as* $n \to \infty$, $\gamma_{j,k} \equiv j^{d_r - p} k^{d_s - p} \log(1 + j)$,

(a) $E[v_r^{\mathrm{T}}(\lambda_{jp})\overline{v_s^{\mathrm{T}}}(\lambda_{jp})] = H_{rs}(\lambda_{jp}) + \mathrm{O}(j^{-\beta} + \gamma_{j,j})$,
(b) $E[v_r^{\mathrm{T}}(\lambda_{jp})\overline{v_s^{\mathrm{T}}}(\lambda_{jp})] = \mathrm{O}(j^{-p} + \gamma_{j,j})$,
(c) $E[v_r^{\mathrm{T}}(\lambda_{jp})\overline{v_s^{\mathrm{T}}}(\lambda_{kp})], E[v_r^{\mathrm{T}}(\lambda_{jp})v_s^{\mathrm{T}}(\lambda_{kp})] = \mathrm{O}(k^{-1}\eta^{1-p} + k^{-1}\eta^{-p}\{\log n\}_{[p=2]} + \eta^{-p} + \gamma_{j,k})$.

In part (c) the $\log n$ factor only appears when $p = 2$ but not otherwise. Thus, the tapered periodogram (with a taper of order p) is unbiased at Fourier frequencies λ_{jp} for any $d_r < p$ if $\mu_r = 0$, (i), or with some extra tapering if there are deterministic polynomial trends in time ($\mu_r \neq 0$ in (ii)).

3. Nonparametric estimates of the PSD matrix

We analyse in this section the properties of traditional nonparametric kernel spectral estimates at frequencies v, $0 < v < \pi$, fixed in the asymptotics, and at frequencies in a

degenerating band around the origin with $v \to 0$ as the sample size increases, as when e.g. $v = \lambda_j$ with j increasing slowly.

We define the following class of statistics, based on a discrete average of the (possibly tapered) periodogram ordinates $I_{rs}^p(\lambda_j)$ at the Fourier frequencies closest to the frequency of interest v:

$$\hat{f}_{rs}^M(v) = \frac{2\pi p}{n} \sum_j K_M(v - \lambda_{jp}) I_{rs}^p(\lambda_{jp}),$$

where $K_M(x) = MK(Mx)$ and $K(x)$ integrates to 1 and is of compact support inside $[-\pi, \pi]$. M is a bandwidth number which increases with the sample size in the asymptotics. The summation in j runs for all the λ_{jp} in the support of K_M, including $O(nM^{-1})$ Fourier frequencies (that is, $j = v, v \pm 1, \ldots, v \pm \lfloor n/2Mp \rfloor$, if the support of K is exactly $[-\pi, \pi]$, where λ_{pv} is the closest frequency to v for each n, v integer). When the series is not stationary, the frequency domain estimates $\hat{f}_{rs}^M$ are not necessarily asymptotically equivalent to estimates constructed in terms of the sample (cross) autocovariances, since our analysis depends crucially on the properties of the periodogram at Fourier frequencies (Theorems 1 and 2).

We need the following regularity conditions for the asymptotic analysis of the properties of $\hat{f}_{rs}^M(v)$.

Assumption 1. For $d_r > -\frac{1}{2}$, $r, s = 1, \ldots, R$,

$$g_{rs}(\lambda) = G_{rs} \lambda^{-d_r - d_s + D_r + D_s}(1 + o(1)) \quad \text{as } \lambda \to 0^+,$$

for some $0 < G_{rr} < \infty$ and $0 \leqslant |G_{rs}| < \infty$, $r \neq s$.

Assumption 2. $\mathbf{U}_t = \mu + \sum_{j=0}^{\infty} \mathbf{A}_j \epsilon_{t-j}$ with $\sum_{j=0}^{\infty} \|\mathbf{A}_j\|^2 < \infty$, where $\|\cdot\|$ denotes the supremum norm and ϵ_t satisfies a.s. $E(\epsilon_t | \mathfrak{I}_{t-1}) = 0$; $E(\epsilon_t \epsilon_t' | \mathfrak{I}_{t-1}) = \mathbf{\Sigma}$, $\Sigma_{rr} = 1$; $E(\epsilon_a(t)\epsilon_b(t)\epsilon_c(t) | \mathfrak{I}_{t-1}) = \mu_{abc}$ with $|\mu_{abc}| < \infty$ for $a, b, c = 1, \ldots, R$; $E(\epsilon_a(t)\epsilon_b(t)\epsilon_c(t)\epsilon_d(t) | \mathfrak{I}_{t-1}) = \mu_{abcd}$, where $|\mu_{abcd}| < \infty$ for $a, b, c, d = 1, \ldots, R$ and $\mathfrak{I}_{t-1}$ is the σ-field of events generated by $\{\epsilon_s, s \leqslant t-1\}$.

Assumption 3. The function K is even, has compact support inside $[-\pi, \pi]$, satisfies a Lipschitz condition and

$$\int_{-\pi}^{\pi} K(x)\,\mathrm{d}x = 1, \quad \|K\|_2^2 = \int_{-\pi}^{\pi} K^2(x)\,\mathrm{d}x < \infty.$$

Assumption 4. $|g_{rs}(\lambda)| > 0$ and $g_{rs}(\lambda)$ is boundedly differentiable for $\lambda \in (v - \varepsilon, v + \varepsilon)$, some $\varepsilon > 0$, $r, s = 1, \ldots, R$.

Assumption 5. $\mathbf{A}_{rs}(\lambda)$ is boundedly differentiable for $\lambda \in (v - \varepsilon, v + \varepsilon)$, some $\varepsilon > 0$, r, $s = 1, \ldots, R$.

Assumption 6. $g_{rs}(\lambda)$ is twice boundedly differentiable for $\lambda \in (v - \varepsilon, v + \varepsilon)$, some $\varepsilon > 0$, $r, s = 1, \ldots, R$.

Assumption 1 deals with the possible long memory or nonstationarity of the observed series, while Assumptions 4–6 impose some smoothness on the spectral density g_{rs} around the frequency of interest. Only Assumptions 1 and 4 are needed for the analysis of the covariance of the usual DFT in Theorem 1, but Assumption 6 is used to control smoothing bias and to fully use tapering properties when $p > 1$ in Theorem 2. Similar conditions are used for parametric and semiparametric inference on long-memory processes, imposing here only local conditions around the frequency of interest, allowing for PSDs with (integrable) poles or zeroes at remote frequencies.

Assumption 2 imposes linearity of the (differenced) stationary zero mean series $\mathbf{U}_t^{(\star)} = \mathbf{U}_t - \mu$, $\mathbf{U}_t = (U_{1t}, \ldots, U_{Rt})'$. It was introduced by Robinson (1995b) and Lobato (1997, 1999) to analyse semiparametric estimates of d for stationary long-memory processes and it does not restrict the form of $f_U(\lambda) = g(\lambda)$ in any way and is only restrictive in the linearity it imposes. Note that the variance of the components of ϵ_t is set to one for identifiability in Assumption 2. Define $\mathbf{A}(\lambda) = \sum_{j=0}^{\infty} \mathbf{A}_j e^{ij\lambda}$, and denote each of its rows by $\mathbf{A}_r(\lambda) = (A_{r1}(\lambda), \ldots, A_{rR}(\lambda))$. Then the spectral density matrix of ϵ_t is $\mathbf{f}_\epsilon(\lambda) = (2\pi)^{-1}\mathbf{\Sigma}$, so $\mathbf{g}(\lambda) = (2\pi)^{-1}\mathbf{A}(\lambda)\mathbf{\Sigma}\mathbf{A}^*(\lambda)$, with typical element

$$g_{rs}(\lambda) = \frac{1}{2\pi}\mathbf{A}_r(\lambda)\mathbf{\Sigma}\mathbf{A}_s^*(\lambda) = \frac{1}{2\pi}\sum_{a=1}^{R}\sum_{b=1}^{R} A_{ra}(\lambda)\Sigma_{ab}A_{sb}(-\lambda),$$

where $*$ stands for simultaneous transposition and complex conjugation. Denote $\mathbf{B}_r(\lambda) = (1 - e^{i\lambda})^{-D_r}\mathbf{A}_r(\lambda)$, $r = 1, \ldots, R$, so $f_{rs}(\lambda) = (2\pi)^{-1}\mathbf{B}_r(\lambda)\mathbf{\Sigma}\mathbf{B}_s^*(\lambda)$. Assumption 5 imposes smoothness on the components of $\mathbf{A}$ (equivalently $\mathbf{B}$) around the frequency of interest, implying Assumption 4.

Assumption 3 is standard in nonparametric kernel estimation and is satisfied by many kernels employed in spectral analysis with compact support, like the uniform and Barlett–Priestley kernels.

The first result of the paper is about the consistency and asymptotic distribution of the nonparametric estimate $\hat{f}_{rs}^M(\nu)$ for $|\nu| > 0$ fixed with n. To centre the asymptotic distribution in the actual value of $f_{rs}(\nu)$, we need to undersmooth the nonparametric estimates and use Assumption 6 for bias control. We recall that for complex quantities, the covariance is defined conjugating the second term in the expectation, so the variance is defined as the expectation of the squared modulus of the mean-corrected variates (see e.g. Brillinger (1975, p. 89) for the J-dimensional complex normal distribution, denoted as N_J^{C}, J fixed). Set $d_* = \max_{r=1,\ldots,R} d_r$ and for $r(i), s(i) \in \{1, \ldots, R\}$,

$$F_{\mathbf{rs}}(\nu) = (f_{r(1)s(1)}(\nu_1), \ldots, f_{r(J)s(J)}(\nu_J))',$$

where $\hat{F}_{\mathbf{rs}}^M(\nu)$ is defined accordingly, and

$$\Phi_p = \lim_{n\to\infty}\left(\sum_1^n h_t^2\right)^{-2}\sum_{k=0,p,2p,\ldots}^{n-p}\left[\sum_1^n h_t^2\cos t\lambda_k\right]^2.$$

All theorems are proved in the appendix.

Theorem 3. *For* $|\nu| > 0$ *as* $n \to \infty$ *under Assumptions* 1–5 *and*

$$Mn^{-1} + M^{-1} \to 0 \quad \text{as } n \to \infty, \tag{6}$$

and either

(i) $\mu_r = \mu_s = 0$ *and* $p > \max\{d_r, d_s\}$,
(ii) $p \geqslant \max\{D_r, D_s\} + 1$, *any* μ_r, μ_s,

then

$$\hat{f}^M_{rs}(v) - f_{rs}(v) \to_p 0.$$

Further, for $v_i \in (-\pi, \pi) - \{0\}$, $i = 1, \ldots, J$, *fixed with* n, *with Assumption* 6,

$$M^3 n^{-1} \log n + M^{-5} n \to 0 \quad \text{as } n \to \infty, \tag{7}$$

and either

(iii) $\mu = 0$ *and* $p = 1$ *and*

$$M^{-1} n^{4d_* - 3} \log^2 n \to 0 \quad \text{as } n \to \infty, \tag{8}$$

(iv) $\mu = 0$ *and* $p > d_*$ *and*

$$M^{-1} n^{2(d_* - p) + 1} \log n \to 0 \quad \text{as } n \to \infty. \tag{9}$$

(v) $p \geqslant \lfloor d_* + \frac{1}{2} \rfloor + 1$ *and* $p > 1$ (*any* μ),

we obtain that

$$\sqrt{\frac{n}{M}} \{\hat{F}^M_{\mathbf{rs}}(v) - F_{\mathbf{rs}}(v)\} \to_d N^{\mathrm{C}}_J(0, 2\pi p \Phi_p \|K\|^2 \Omega(v)) \quad \text{as } n \to \infty,$$

where $\Omega(v) = [\sigma_{ij}(v)]$, $\sigma_{ij}(v) = \delta(v_i - v_j) f_{r(i)r(j)}(v_i) f_{s(i)s(j)}(-v_i) + \delta(v_i + v_j) f_{r(i)s(j)}(v_i)$ $f_{r(j)s(i)}(v_i)$, $r(i), r(j), s(i), s(j) \in \{1, \ldots, J\}$.

We decided for simplicity not to include the cases $v_i = \pm\pi$ (but the standard results hold, see e.g. Brillinger, 1975, Theorem 7.4.3). A condition like (6) is also minimal for nonparametric estimation of smooth spectral densities. Tapering allows the consistent estimation of f with trending observations without need of initial detrending, (ii), and without any kind of tapering it is possible to estimate f consistently for nonstationary but *transitory* processes with $d < 1$, (i).

The taper variance inflation factor Φ_p is smaller than 1.05 for Zhurbenko kernels with $p > 1$ ($\Phi_1 \equiv 1$), implying moderate increments in the asymptotic variance of the estimates (apart from the p factor due to the reduced number of frequencies used in $\hat{f}^M_{rs}$). Note that $\Phi_p = \Upsilon_p$, where Υ_p is the usual tapering variance correction (see e.g. Dahlhaus, 1985) if the sum in k in the definition of Φ_p were running for all the possible values,

$$\Upsilon_p = \lim_{n \to \infty} n \left(\sum_1^n h_t^2 \right)^{-2} \sum_1^n h_t^4,$$

by Parseval's identity. The results also hold for the cosine bell taper with $\Phi_{\mathrm{c}} = \Upsilon_{\mathrm{c}} = \frac{35}{18}$ when $-\frac{1}{2} < d_* < \frac{3}{2}$, $\mu = 0$, considering all possible frequencies, like if $p = 1$ (see the

discussion in Velasco, 1999b). Notice that spectral estimates at different frequencies are asymptotically independent for $|\nu_h| > 0$ as in the weak dependence case with bounded spectral density.

When considering estimation in a degenerating band of zero frequency we need to adapt our assumptions:

Assumption 4′. $g_{rs}(\lambda)$ is differentiable for $\lambda \in (0, \varepsilon)$, some $\varepsilon > 0$, $r, s = 1, \ldots, R$, and

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} g_{rs}(\lambda) = \mathrm{O}(\lambda^{-d_r - d_s + D_r + D_s - 1}) \quad \text{as } \lambda \to 0^+.$$

Assumption 5′. $\mathbf{A}_r(\lambda)$ is differentiable for $\lambda \in (0, \varepsilon)$, some $\varepsilon > 0$, $r = 1, \ldots, R$, and

$$\frac{\mathrm{d}\mathbf{A}_r(\lambda)}{\mathrm{d}\lambda} = \mathrm{O}(\lambda^{-1} \|\mathbf{A}_r(\lambda)\|) \quad \text{as } \lambda \to 0^+.$$

Assumption 6′. $g_{rs}(\lambda)$ is twice differentiable for $\lambda \in (0, \varepsilon)$, some $\varepsilon > 0$, $r, s = 1, \ldots, R$, with

$$\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2} g_{rs}(\lambda) = \mathrm{O}(\lambda^{-d_r - d_s + D_r + D_s - 2}) \quad \text{as } \lambda \to 0^+.$$

The corresponding result for these frequencies is stated as

Theorem 4. *For $|\nu| \to 0$ as $n \to \infty$ under Assumptions* 1–3, 4′, 5′, (6),

$$(|\nu| M)^{-1} + (n|\nu|)^{-1} \log n + (n|\nu|)^{2(d-p)} \log n \to 0 \quad \textit{as } n \to \infty \tag{10}$$

and either (i) *or* (ii) *of Theorem* 3 *then,*

$$\frac{\hat{f}_{rs}^M(\nu)}{f_{rs}(\nu)} - 1 \to_p 0.$$

If we further take Assumption 6′,

$$M^3 n^{-1} \log n + M^{-5} |\nu|^{-4} n \to 0 \quad \textit{as } n \to \infty, \tag{11}$$

and either

(iii) $\mu = 0$, $d_* < \frac{3}{4}$ *and* $p = 1$;
(iv) *take* (iv) *or* (v) *of Theorem* 3 *and*

$$M^{-1} n (n|\nu|)^{2(d_* - p)} \log n \to 0 \quad \textit{as } n \to \infty, \tag{12}$$

we obtain that

$$\sqrt{\frac{n}{M}} \frac{\hat{f}_{rs}^M(\nu) - f_{rs}(\nu)}{(f_{rr}(\nu) f_{ss}(\nu))^{1/2}} \to_d N_1^C(0, 2\pi p \Phi_p \|K\|^2) \quad \textit{as } n \to \infty.$$

The condition $(|v|M)^{-1} + (|v|n)^{-1} \log n \to 0$ as $n \to \infty$ in (10) is needed now to avoid periodogram ordinates too close to the singularity of the PSD at the origin. The last condition in (10) controls the degree of nonstationarity when d is close to the degree of tapering applied, p. When $v = \lambda_j$, then j may grow as n^{ε}, any $0 < \varepsilon < 1$, for consistency. Condition (11) further restricts the range of low frequencies v for which we can obtain the asymptotic normality of $\hat{f}_{rs}^{M}(v)$, though if $v = \lambda_j$ it would be possible to consider $j \sim n^{5/6+\varepsilon}$ for any $\varepsilon > 0$.

The restrictions on the value of d_* (i.e. the degree of nonstationarity) depend on the tapering degree p in a parallel way as was found for semiparametric estimates a similar environment by Velasco (1999a, b) and by Velasco and Robinson (2000) for parametric estimates. Thus, for any allowed choice of M, $p > d_* + \frac{1}{2}$ is sufficient for (12). When $p = 1$ and all v_i are fixed, it is possible to find sequences M which lead to asymptotically normal estimates if $d_* < \frac{5}{6}$. Weaker conditions on the smoothing bandwidth M would be sufficient for a central limit theorem if we substitute $F_{\mathbf{rs}}(v)$ by $E[\hat{F}_{\mathbf{rs}}(v)]$ or if we employ higher order kernels or estimate higher order bias terms.

Finally, note that the asymptotic variance in Theorem 3 changes the sign of the frequency when $v_i = -v_j$ in $f_{r(j)s(i)}(v_i)$ with respect to Brillinger (1975, Theorem and Corollary 7.4.3), otherwise when $v_i = v_j$ and considering $\hat{f}_{r(j)s(j)}(-v_1) = \hat{f}_{s(j)r(j)}(v_1)$ we would obtain a contradiction. However, this is correctly stated in his equation (7.2.14), but not in the second line of (7.2.13).

All standard results for linear and nonlinear functions of the PSD (real and imaginary parts, modulus, coherency, phase, transfer function) can be deduced from Theorems 3 and 4. See e.g. Hannan (1970, Section V.5). These spectral estimates can be used for long memory estimation (Hassler, 1993; Chen et al., 1994; Reisen, 1994) and for efficient Hannan's (1963) regression for long-memory and nonstationary series (Robinson and Hidalgo, 1998; Marinucci, 2000), apart from nonparametric descriptive analysis. In the next section we take this further and analyse the behaviour of coherence measures for cointegrated series, and estimate a semiparametric model for them.

4. Spectral analysis of cointegrated time series

We will denote a time series whose stationary increments have spectral density satisfying Assumption 1 as integrated of order d_r, $I(d_r)$, generalizing the usual $I(0)$ and $I(1)$ terminology (see Engle and Granger, 1987). The parameter d_r determines the main long-run properties of $I(d_r)$ processes. Let the observable bivariate time series (Y_t, X_t) be $I(d)$ (i.e. $d_Y = d_X = d > 0$ in Assumption 1) and satisfy

$$Y_t = bX_t + Z_t \tag{13}$$

for some $b \neq 0$, where the cointegrating error Z_t is $I(d - \alpha)$, $0 < \alpha \leqslant d$, and may be correlated with X_t at some frequencies (all $\mu = 0$). If such Z_t exists with $\alpha > 0$ we say that the pair (Y_t, X_t) is cointegrated because a linear combination of them is less nonstationary and can be interpreted as a long-run relationship where the Z_t are departures from equilibrium. Often this set-up is only sensible if the errors Z_t are

transitory, i.e. if $d - \alpha < 1$ so shocks do not have a permanent effect on the long-run equilibrium. The main inference issues in cointegrated systems are the estimation of the long-run relationship b and of the memory of the series involved, d and $d - \alpha$.

We strengthen Assumption 1 and suppose that the PSD matrix $\mathbf{f}$ of X_t and Z satisfies

$$\mathbf{f}(\lambda) = \lambda^{-2d} \begin{pmatrix} G_{xx} & G_{xz}\lambda^{\alpha} \\ G_{zx}\lambda^{\alpha} & G_{zz}\lambda^{2\alpha} \end{pmatrix} (1 + \mathrm{O}(\lambda^2)) \quad \text{as } \lambda \to 0^+$$

for some constants $|G_{ab}| < \infty$, $a, b \in \{x, z\}$, where the matrix $\mathbf{G} = \{G_{ab}\}$ is Hermitian and nonsingular.

The memory d of the observables can be estimated directly from either X_t or Y_t, e.g. using the semiparametric estimates of Robinson (1995a, b). However, Z_t is not observed in order to estimate $d - \alpha$ or α. The regression model (13) can be estimated (for example using spectral estimates as in Marinucci, 2000) and residuals can be used instead, but asymptotic properties of estimates of b depend crucially on α (Robinson and Marinucci, 2001) and the same can be expected for residual-based estimates (see e.g. Hassler et al., 2001). Here we propose an estimation method for α related to our previous nonparametric analysis and avoiding intermediate steps.

Define the coherence $H_{ab}(\lambda)$ between two time series a_t and b_t at frequency λ as

$$H_{ab}(\lambda) = \frac{f_{ab}(\lambda)}{(f_{aa}(\lambda) f_{bb}(\lambda))^{1/2}},$$

so $|H_{zx}(\lambda) - H_{zx}(0)| = \mathrm{O}(|\lambda|^2)$ as $\lambda \to 0^+$, where $|H_{zx}(0)|^2 = |G_{zx}|^2/(G_{zz}G_{xx})$, which holds, e.g. for certain ARFIMA processes (cf. Assumption 3 of Robinson, 1995a).

Then, employing model (13), and pretending that the series are stationary to calculate the autocovariances (otherwise, integer difference (13) a sufficient number of times and then multiply by the unit root transfer functions) the PSD of Y_t is

$$f_{yy}(\lambda) = b^2 f_{xx}(\lambda) + f_{zz}(\lambda) + 2b \operatorname{Re} f_{zx}(\lambda) \sim b^2 G_{xx}\lambda^{-2d} \quad \text{as } \lambda \to 0^+,$$

and the cross-PSD of X_t and Y_t satisfies

$$f_{xy}(\lambda) = b f_{xx}(\lambda) + f_{xz}(\lambda) \sim b G_{xx}\lambda^{-2d} \quad \text{as } \lambda \to 0^+.$$

Therefore, X_t and Y_t have coherence equal to one at zero frequency, $H_{xy}(0) = 1$, and the PSD matrix of (X_t, Y_t) is singular at $\lambda = 0$. Note that the generalized coherence $H_{xy}(\lambda)$ defined in terms of the PSDs for nonstationary series belongs to the interval $[0, 1]$ for all λ as in the stationary case, independent of whether the PSDs are unbounded or zero at some frequencies.

After straightforward manipulations using (13) we can write the square coherence as

$$|H_{xy}(\lambda)|^2 = 1 - \frac{f_{zz}(\lambda)}{f_{yy}(\lambda)} + \frac{|f_{zx}(\lambda)|^2}{f_{yy}(\lambda) f_{xx}(\lambda)}. \tag{14}$$

Substituting in (14) the approximation of f_{yy} as $\lambda \to 0^+$,

$$|H_{xy}(\lambda)|^2 = 1 - \left(\frac{G_{zz}}{G_{xx}} - \frac{|G_{zx}|^2}{G_{xx}^2}\right)\left(1 - 2\lambda^\alpha \frac{\operatorname{Re} G_{zx}}{G_{xx}}\right)\lambda^{2\alpha} + \mathrm{O}(\lambda^{4\alpha} + \lambda^{2+2\alpha})$$
$$\sim 1 - G_{\mathrm{H}}\lambda^{2\alpha}$$

for a real constant $0 < G_{\mathrm{H}} < \infty$,

$$G_{\mathrm{H}} = \frac{G_{zz}}{G_{xx}}\left[1 - \frac{|G_{zx}|^2}{G_{xx}G_{zz}}\right],$$

depending on the (normalized) noise-to-signal ratio and on the coherence at zero between X_t and Z_t. Taking logs, we have

$$\log(1 - |H_{xy}(\lambda)|^2) \sim \log G_{\mathrm{H}} + 2\alpha \log \lambda \quad \text{as } \lambda \to 0^+, \tag{15}$$

and we may try to estimate α using consistent estimates of $|H_{xy}(\lambda)|^2$ at frequencies λ_j in a degenerating band around the origin. Notice that the smaller the α the worse is the above approximation for $|H_{xy}(\lambda)|$ based on the leading terms of the expansion, $1 - G_{\mathrm{H}}\lambda^{2\alpha}$, but in this case also estimates of b have slower rates of convergence (see e.g. Robinson and Marinucci, 2001). This approach is valid for both stationary and nonstationary series (tapering might be used to eliminate some intercept or polynomial trend in (13) or to cover very nonstationary situations, $d \geqslant 1$) and it is not affected asymptotically by the endogeneity of the residuals ($H_{zx}(\lambda) \neq 0$) because of its semiparametric nature. However, if X_t and Z_t are incoherent at zero frequency, $H_{zx}(0) = 0$, so $G_{zx} = 0$ and $G_{\mathrm{H}} = G_{zz}G_{xx}^{-1}$, then $|H_{xy}(\lambda)|^2 = 1 - G_{\mathrm{H}}\lambda^{2\alpha} + \mathrm{O}(\lambda^{4\alpha})$ reducing the bias of the semiparametric model (15). In any case we can consider terms of order $\lambda^{3\alpha}$, etc. for greater accuracy. For a general $R \times 1$ vector time series similar approximations should be possible in terms of multiple correlation coefficients based on the coherence matrix $\mathbf{H}(\lambda)$.

Denote by $\hat{\alpha}$ the least-squares estimate of α based on the regression (15) of $\log(1 - |\hat{H}_{xy}(\lambda_j)|^2)$ on $W_j = 2\log\lambda_j$, for frequencies λ_j, $j = \ell, \ldots, m$,

$$\hat{\alpha} = \left(\sum_{j=\ell}^{m} \tilde{W}_j^2\right)^{-1} \sum_{j=\ell}^{m} \tilde{W}_j \log(1 - |\hat{H}_{xy}(\lambda_j)|^2)$$

with $\tilde{W}_j = W_j - \bar{W}$, where $\bar{W}$ is the sample mean of the W_j, and

$$|\hat{H}_{xy}(\lambda_j)|^2 = \frac{|\hat{f}_{xy}^M(\lambda_j)|^2}{\hat{f}_{xx}^M(\lambda_j)\hat{f}_{yy}^M(\lambda_j)}.$$

We may call this estimate *log-coherence regression estimate* in parallel to Geweke and Porter-Hudak's (1983) log-periodogram regression estimate of the memory parameter d. As in Robinson (1995a) we introduce a trimming of the very first ℓ coherence estimates, which may not have very desirable asymptotic properties. The analysis of $\hat{\alpha}$

is complicated with respect to the log-periodogram regression estimate of d due to the nonlinear and nonparametric nature of sample coherences $|\hat{H}_{xy}(\lambda_j)|^2$. We show first the consistency of $\hat{\alpha}$ under conditions similar to those of Theorem 3 and then approximate its variability for large samples. We can write, $W_2 = \sum_{j=\ell}^{m} \tilde{W}_j^2$,

$$\hat{\alpha} - \alpha = W_2^{-1} \sum_{j=\ell}^{m} \tilde{W}_j \log \frac{1 - |\hat{H}_{xy}(\lambda_j)|^2}{1 - |H_{xy}(\lambda_j)|^2} + \Lambda_m,$$

where the bias term is $\Lambda_m = \mathrm{O}(m^{-1} \sum_j |\tilde{W}_j| \lambda_j^{\alpha}) = \mathrm{O}((m/n)^{\alpha})$ as in Robinson (1995a). We obtain then

$$\hat{\alpha} - \alpha = W_2^{-1} \sum_{j=\ell}^{m} \tilde{W}_j \hat{S}_j + W_2^{-1} \sum_{j=\ell}^{m} \tilde{W}_j \hat{C}_j + \mathrm{O}\left(\left(\frac{m}{n}\right)^{\alpha}\right), \tag{16}$$

where $\hat{S}_j = \hat{A}_j - \hat{B}_j$ is the linear part,

$$\hat{A}_j = \frac{\hat{f}_{xx}^M(\lambda_j)\hat{f}_{yy}^M(\lambda_j) - |\hat{f}_{xy}^M(\lambda_j)|^2}{f_{xx}(\lambda_j) f_{yy}(\lambda_j)(1 - |H_{xy}(\lambda_j)|^2)}, \quad \hat{B}_j = \frac{\hat{f}_{xx}^M(\lambda_j)\hat{f}_{yy}^M(\lambda_j)}{f_{xx}(\lambda_j) f_{yy}(\lambda_j)},$$

and

$$\hat{C}_j = \log \frac{1 - |\hat{H}_{xy}(\lambda_j)|^2}{1 - |H_{xy}(\lambda_j)|^2} - \hat{S}_j = (\log \hat{A}_j - \hat{A}_j + 1) - (\log \hat{B}_j - \hat{B}_j + 1).$$

We now analyse the properties of the spectral estimates included in $\hat{\alpha}$ under some stronger conditions.

Lemma 1. *Under Assumptions* 1–3, 4′–6′ (p=1, μ=0) *for Gaussian* X_t, Y_t, $0 < \alpha < \frac{1}{2}$, $0 < d < \frac{3}{4}$, $M^{-1} + Mn^{-1} + (m - \ell)^{-1} + mn^{-1} \to 0$ *as* $n \to \infty$, *and if for some* $\tau \geqslant 1$,

$$(n^{2\alpha-1}\ell^{-2\alpha}M + n^{2(1+\alpha)}\ell^{-2(1+\alpha)}M^{-2})\log m + M^{\tau/2} n^{\tau(2\alpha-0.5)} \ell^{1-2\alpha\tau} \log^{\tau} m \to 0, \tag{17}$$

as $n \to \infty$, *then* $a, b \in \{X, Y\}$,

$$\max_{\ell \leqslant j \leqslant m} (1 - |H_{xy}(\lambda_j)|^2)^{-1} (f_{aa}(\lambda_j) f_{bb}(\lambda_j))^{-1/2} |\hat{f}_{ab}^M(\lambda_j) - f_{ab}(\lambda_j)|$$
$$= \mathrm{o_p}(\log^{-1} m). \tag{18}$$

We impose $d < \frac{3}{4}$ and Gaussianity to simplify proofs and avoid conditions on the moments of the linear innovations. Note that under (17), condition (10) for the consistency of $\hat{f}_{ab}^M(\nu)$ holds for $\nu = \lambda_j$, $\ell \leqslant j \leqslant m$. The implications of the conditions of Lemma 1 are very strong on the trimming ℓ to obtain the uniform consistency of the spectral estimates in the frequency band of the log-coherence regression, though in practical applications this may not be needed as long as zero frequency periodograms

are avoided. Thus, if $\ell \sim An^a$, $m \sim Bn^b$, $M \sim Cn^c$, $0 < a \leqslant b < 1$, $0 < c < 1$, (17) holds if

$$\max\{(1+\alpha)(1-a)-c, 4\alpha(1-a)+c-1\} < 0,$$

which, when e.g. $\alpha = 0.2$ holds for $a = 0.6$ and $c = 0.5$.

Then under the assumptions of Lemma 1, $\hat{\alpha}-\alpha = \mathrm{o_p}(1)$ as $n \to \infty$, because using that $\max_j |\tilde{W}_j| = \mathrm{O}(\log m)$ and $\max_j |\hat{A}_j - 1| = \mathrm{o_p}(\log^{-1} m)$, $\max_j |\hat{B}_j - 1| = \mathrm{o_p}(\log^{-1} m)$ both implied by (18), we obtain that $\max_j |\hat{S}_j| = \mathrm{o_p}(\log^{-1} m)$ and $\max_j |\hat{C}_j| = \mathrm{o_p}(\log^{-1} m)$.

Now we proceed heuristically. Since the spectral estimates are approximately independent if they include periodogram ordinates at non-overlapping frequencies, so are the coherence estimates, the log-coherence regression estimate would be approximately normal in large samples. To estimate its variance we can approximate $\mathrm{Var}[\log(1 - |\hat{H}_{xy}(\lambda_j)|^2)]$ by $4\,\mathrm{Var}[\tanh^{-1}(|\hat{H}_{xy}(\lambda_j)|)]$ using that $\log(1-x^2) + 2\tanh^{-1}(x) = \log(1+x) \to 2\log 2$ as $x \to 1^-$, so

$$\mathrm{Var}[\hat{\alpha}] \approx \left(\sum \tilde{W}_j^2\right)^{-2} 4 \sum_j \sum_k \tilde{W}_j \tilde{W}_k$$
$$\times \mathrm{Cov}[\tanh^{-1}(|\hat{H}_{xy}(\lambda_j)|), \tanh^{-1}(|\hat{H}_{xy}(\lambda_k)|)]. \tag{19}$$

Here the transformation $\tanh^{-1}$ is variance stabilizing because $\hat{H}_{xy}$ is a sort of correlation coefficient in the frequency domain, and when $\hat{H}_{xy}$ is defined using spectral estimates with uniform weights over $2q+1$ Fourier frequencies we can write (see e.g. Brillinger, 1975, p. 312),

$$\mathrm{Var}[\tanh^{-1}(|\hat{H}_{xy}(\lambda_j)|)] \approx \frac{1}{2(2q+1)}. \tag{20}$$

We can also approximate for $|t| \leqslant 2q$:

$$\mathrm{Cov}[\tanh^{-1}(|\hat{H}_{xy}(\lambda_j)|), \tanh^{-1}(|\hat{H}_{xy}(\lambda_{j+t})|)] \approx \frac{2q+1-|t|}{2(2q+1)^2}, \tag{21}$$

and if the estimates $\hat{H}_{xy}$ are evaluated at frequencies sufficiently far apart we can suppose they are asymptotically uncorrelated. Plugging (20) and (21) in (19) we can estimate the sampling variance of $\hat{\alpha}$ for each m and q. For tapered series this approximation can be adjusted by Φ_p and p as for $\hat{f}$ (cf. Theorems 3 and 4).

We can also justify these variance estimates using the linear approximation to $\hat{\alpha}$ given by the first term on the right-hand side of (16). Thus,

$$\mathrm{Var}[\hat{S}_j]\{f_{xx}(\lambda_j) f_{yy}(\lambda_j)(1 - |H_{xy}(\lambda_j)|^2)\}^2$$
$$= |H_{xy}(\lambda_j)|^4\, \mathrm{Var}[\hat{f}_{xx}^M(\lambda_j)\hat{f}_{yy}^M(\lambda_j)]$$
$$- 2|H_{xy}(\lambda_j)|^2\, \mathrm{Cov}[\hat{f}_{xx}^M(\lambda_j)\hat{f}_{yy}^M(\lambda_j), |\hat{f}_{xy}^M(\lambda_j)|^2] + \mathrm{Var}[|\hat{f}_{xy}^M(\lambda_j)|^2],$$

and for uniform weights this is approximately equal to $1/(2q+1)$ times

$$|H_{xy}(\lambda_j)|^4\{2f_{xx}(\lambda_j)f_{yy}(\lambda_j)|f_{xy}(\lambda_j)|^2+8f_{xx}^2(\lambda_j)f_{yy}^2(\lambda_j)\}$$
$$-20|H_{xy}(\lambda_j)|^2f_{xx}(\lambda_j)f_{yy}(\lambda_j)|f_{xy}(\lambda_j)|^2$$
$$+2f_{xx}(\lambda_j)f_{yy}(\lambda_j)|f_{xy}(\lambda_j)|^2+8|f_{xy}(\lambda_j)|^4,$$

under appropriate conditions on the higher order cumulants of ϵ_t, e.g. assuming Gaussianity. Thus for small λ_j,

$$\mathrm{Var}[\hat{S}_j]\approx\frac{2}{2q+1}|H_{xy}(\lambda_j)|^2\approx\frac{2}{2q+1}. \tag{22}$$

Alternatively, we can write that

$$\hat{S}_j=\frac{\hat{f}_{xx}^M(\lambda_j)\hat{f}_{yy}^M(\lambda_j)}{f_{xx}(\lambda_j)f_{yy}(\lambda_j)}\frac{|H_{xy}(\lambda_j)|^2-|\hat{H}_{xy}(\lambda_j)|^2}{1-|H_{xy}(\lambda_j)|^2}\approx\frac{|H_{xy}(\lambda_j)|^2-|\hat{H}_{xy}(\lambda_j)|^2}{1-|H_{xy}(\lambda_j)|^2},$$

and then use the fact that

$$\mathrm{Var}[|\hat{H}_{xy}(\lambda_j)|^2]\approx\frac{2}{2q+1}|H_{xy}(\lambda_j)|^2(1-|H_{xy}(\lambda_j)|^2)^2 \tag{23}$$

(e.g. Brillinger, 1975, p. 309) to obtain again the approximation (22). Similar approximations can be used for the covariances between $\hat{S}_j$ and $\hat{S}_k$, $j\neq k$.

5. Simulation results

In this section we simulate the performance of the estimate $\hat{\alpha}$ of the cointegration degree in comparison with semiparametric procedures based on OLS residuals. In particular, we use the log-periodogram regression estimate (Geweke and Porter-Hudak, 1983; Robinson, 1995a) and an estimate based on a local Gaussian or Whittle likelihood (Künsch, 1987; Robinson, 1995b). These estimates are consistent for nonstationary series when $d<1$, or $d<p$ if tapering of order p is applied ($\mu=0$), see Velasco (1999a, b). We use the Zhurbenko taper of order $p=2$, which is valid for $d<1.5$ for memory estimation with any μ.

We have simulated cointegrated Gaussian series (X_t,Y_t) of lengths $n=128$ and 256 according to (13) with three pairs of cointegration values, CI(1,0) ($\alpha=1$), CI(1.3, 0.9) ($\alpha=0.4$), and CI(1.1,0.4) ($\alpha=0.7$). All the observed series are nonstationary while the residuals are weakly dependent, nonstationary but mean reverting and stationary long memory, respectively. The X_t series are all ARFIMA$(0,d,0)$, while Z_t are ARFIMA$(2,d-\alpha,0)$ with autoregressive coefficients $\phi_1=0.34$ and $\phi_2=-0.9$, guaranteeing that the PSDs of Z_t shows a peak at $\lambda=4\pi/9$ (Models 1–3), or ARFIMA$(1,d-\alpha,0)$, (Models 1′–3′) with $\phi_1=0.3$ and 0.6. The innovations are zero mean Gaussian independent sequences η_X, η_Z with standard deviations (sd's) $\sigma_X=1$, $\sigma_Z=2$, respectively, and

correlation 0.5. Nonstationary series are obtained by integration of series with memory parameter $d-1$. Similar data generation processes have been used previously in the CI$(0,1)$ case by Robinson and Marinucci (2001).

The bandwidths were $m=6,12,18$ for $n=128$ and $m=12,24,36$ for $n=256$ while for coherence estimation we used uniform weights $q=1,2$. The estimates $\hat{\alpha}$ are calculated from the original data and from tapered data with Zhurbenko taper of order $p=2$. Note that if no taper is used $\hat{f}^M$ is not consistent for our simulated series ($d_X \geqslant 1$). We also construct estimates $\hat{\alpha}_\Delta$ based on the increments $(\Delta X_t, \Delta Y_t)$. For AR(1) series we only report the estimates based on nontapered series with $N=256$.

For comparison purposes we consider alternative estimates of α based on OLS residuals. Notice that for these series the OLS estimate satisfies $\hat{b}-b=\mathrm{O_p}(n^{-\alpha})$ (cases II, IV and III, respectively, of Robinson and Marinucci, 2001). We consider two semiparametric estimation procedures with the same bandwidths as for $\hat{\alpha}$: the log-periodogram regression ($\hat{\alpha}_\mathrm{L}$) and the local Gaussian semiparametric estimate ($\hat{\alpha}_\mathrm{G}$). These estimates are implemented with three different input series. We first estimate d starting with X_t and with the OLS residuals $\hat{Z}_t$ we estimate the order d_Z of integration of Z_t and set the estimate of α as $\hat{d}-\hat{d}_\mathrm{Z}$. We also substitute X_t by ΔX_t and $\hat{Z}_t$ by $\Delta\hat{Z}_t$ and finally we only differentiate the observed series but work with the original residuals, adapting the estimates of α accordingly. Note that some of these estimates are not consistent for the models considered, but that some systematic biases may cancel out.

We report the mean, sd and mean square error (mse) of the estimates across 500 replications. We also give in parentheses the approximations of the sd's of $\hat{\alpha}$ based on (19) for both values of q and each m, taking into account the tapering applied.

The main conclusions for the ARFIMA$(2,d,0)$ cointegrating series are as follows (see Tables 1–3 for $N=256$ and 128 and Models 1–3, respectively). Coherence-based estimates with $q=1$ perform slightly better than those with $q=2$, except for Model 2 where the situation is reversed, though the improvement in the sd is smaller than that predicted by (19). The estimates $\hat{\alpha}_\Delta$ based on (stationary) increments work uniformly much worse than those with original data, except for Model 2, where the similar performance is explained in terms of the nonstationarity of the cointegrating residuals, so the differenced residuals are invertible, in contrast with the other two cases.

The variance approximation (19) gives a good indication of the sample variability of $\hat{\alpha}$ for both n and q and all m, though it underestimates the sample variance for the smallest values of m, especially for $n=128$. With tapering the variance increment is only slightly overestimated by (19) for large m, but the bias performance is more erratic than without tapering, leading to larger mse for all estimates considered. For the sample sizes considered the best results were attained for the largest values of m, both in terms of sample bias and sd.

Coherence estimates have similar properties than residual-based estimates for Models 1 and 3 but do not achieve results close to the best performances of log-periodogram and Gaussian estimates for Model 2. Among the alternatives to construct these residual estimates, the uniformly best is to use differenced data and original residuals, though, as expected, the second one using both differenced data and residuals works better for Model 2, while using both original data and residuals seems to have no

Table 1
Model 1, $\alpha = 1$, CI$(1,0)$, $\phi_1 = 0.34$, $\phi_2 = -0.9$

p	m			$q=2$		$q=1$		$X, \hat{Z}_t$		$\Delta X_t, \Delta\hat{Z}_t$		$\Delta X_t, \hat{Z}_t$	
				$\hat{\alpha}$	$\hat{\alpha}_\Delta$	$\hat{\alpha}$	$\hat{\alpha}_\Delta$	$\hat{\alpha}_L$	$\hat{\alpha}_G$	$\hat{\alpha}_L$	$\hat{\alpha}_G$	$\hat{\alpha}_L$	$\hat{\alpha}_G$
$n = 256$													
1	12	Mean		1.167	0.335	0.976	0.282	0.980	0.683	0.249	0.175	1.000	0.890
		sd (0.29)	(0.24)	0.321	0.369	0.297	0.354	0.341	0.266	0.405	0.245	0.337	0.267
		mse		0.131	0.578	0.089	0.637	0.117	0.171	0.729	0.741	0.114	0.083
	24	Mean		1.122	0.481	1.011	0.411	1.025	0.743	0.423	0.299	1.019	0.849
		sd (0.19)	(0.16)	0.205	0.313	0.196	0.306	0.213	0.169	0.334	0.226	0.199	0.169
		mse		0.057	0.368	0.038	0.441	0.046	0.094	0.444	0.543	0.040	0.051
	36	Mean		1.087	0.524	1.009	0.465	1.077	0.828	0.550	0.404	1.048	0.882
		sd (0.15)	(0.13)	0.159	0.266	0.156	0.270	0.172	0.138	0.303	0.220	0.153	0.133
		mse		0.033	0.297	0.025	0.359	0.035	0.049	0.294	0.404	0.026	0.032
2	12	Mean		1.074	0.870	0.965	0.856	1.027	0.708	0.937	0.630	0.955	0.653
		sd (0.47)	(0.42)	0.482	0.460	0.466	0.433	0.644	0.518	0.629	0.488	0.634	0.500
		mse		0.238	0.229	0.219	0.208	0.416	0.353	0.399	0.375	0.404	0.370
	24	Mean		1.048	0.919	0.979	0.907	1.036	0.742	0.977	0.686	0.991	0.700
		sd (0.29)	(0.26)	0.284	0.294	0.278	0.270	0.338	0.273	0.340	0.258	0.329	0.257
		mse		0.083	0.093	0.077	0.082	0.115	0.141	0.116	0.165	0.108	0.156
	36	Mean		0.984	0.890	0.942	0.888	1.070	0.822	1.029	0.780	1.037	0.785
		sd (0.22)	(0.20)	0.211	0.219	0.211	0.208	0.252	0.214	0.257	0.204	0.251	0.203
		mse		0.045	0.060	0.048	0.056	0.069	0.077	0.067	0.090	0.064	0.087
$n = 128$													
1	6	Mean		1.141	0.305	0.869	0.221	0.995	0.686	0.461	0.411	1.042	0.955
		sd (0.37)	(0.35)	0.537	0.467	0.468	0.452	0.466	0.336	0.546	0.296	0.471	0.324
		mse		0.309	0.701	0.236	0.812	0.217	0.212	0.588	0.434	0.224	0.107
	12	Mean		1.172	0.419	0.979	0.345	1.004	0.697	0.352	0.249	0.994	0.849
		sd (0.29)	(0.24)	0.331	0.372	0.295	0.365	0.361	0.268	0.434	0.279	0.333	0.262
		mse		0.139	0.476	0.088	0.562	0.130	0.164	0.608	0.642	0.111	0.092
	18	Mean		1.126	0.461	0.985	0.394	1.059	0.784	0.483	0.356	1.039	0.876
		sd (0.23)	(0.19)	0.246	0.314	0.230	0.320	0.268	0.211	0.370	0.255	0.245	0.202
		mse		0.076	0.389	0.053	0.469	0.075	0.091	0.404	0.480	0.062	0.056
2	6	Mean		1.056	0.817	0.807	0.658	1.024	0.740	0.966	0.675	0.977	0.691
		sd (0.60)	(0.68)	0.592	0.548	0.749	0.718	0.954	0.476	0.878	0.461	0.878	0.468
		mse		0.354	0.334	0.598	0.633	0.912	0.294	0.771	0.318	0.772	0.315
	12	Mean		1.045	0.846	0.942	0.822	1.020	0.726	0.927	0.642	0.974	0.681
		sd (0.47)	(0.42)	0.461	0.462	0.440	0.431	0.634	0.492	0.599	0.475	0.608	0.483
		mse		0.215	0.237	0.197	0.218	0.402	0.318	0.364	0.354	0.370	0.335
	18	Mean		0.975	0.817	0.910	0.813	1.083	0.784	0.998	0.715	1.029	0.740
		sd (0.36)	(0.32)	0.308	0.310	0.311	0.302	0.441	0.360	0.427	0.346	0.430	0.349
		mse		0.096	0.130	0.105	0.126	0.201	0.177	0.182	0.201	0.186	0.190

advantage in any case. Gaussian semiparametric estimates have less variability than log-periodogram ones, but are in general more biased. Here again tapering increases sd's and mse's.

Table 2
Model 2, $\alpha = 0.4$, CI$(1.3, 0.9)$, $\phi_1 = 0.34$, $\phi_2 = -0.9$

p	m			$q=2$		$q=1$		$X_t, \hat{Z}_t$		$\Delta X_t, \Delta\hat{Z}_t$		$\Delta X_t, \hat{Z}_t$	
				$\hat{\alpha}$	$\hat{\alpha}_\Delta$	$\hat{\alpha}$	$\hat{\alpha}_\Delta$	$\hat{\alpha}_L$	$\hat{\alpha}_G$	$\hat{\alpha}_L$	$\hat{\alpha}_G$	$\hat{\alpha}_L$	$\hat{\alpha}_G$
$n = 256$													
1	12	Mean		0.344	0.304	0.254	0.279	0.316	0.242	0.370	0.274	0.465	0.346
		sd (0.29)	(0.24)	0.351	0.304	0.308	0.292	0.354	0.220	0.333	0.220	0.359	0.251
		mse		0.126	0.101	0.116	0.100	0.132	0.073	0.112	0.064	0.133	0.066
	24	Mean		0.336	0.317	0.276	0.299	0.285	0.198	0.395	0.287	0.447	0.318
		sd (0.19)	(0.16)	0.206	0.189	0.194	0.188	0.220	0.134	0.201	0.137	0.220	0.147
		mse		0.046	0.043	0.053	0.045	0.062	0.059	0.041	0.032	0.050	0.028
	36	Mean		0.385	0.331	0.327	0.315	0.335	0.224	0.443	0.337	0.470	0.353
		sd (0.15)	(0.13)	0.154	0.141	0.151	0.143	0.170	0.113	0.156	0.116	0.172	0.120
		mse		0.024	0.024	0.028	0.028	0.033	0.044	0.026	0.017	0.035	0.017
2	12	Mean		0.443	0.281	0.362	0.267	0.427	0.331	0.369	0.302	0.321	0.242
		sd (0.47)	(0.42)	0.511	0.456	0.474	0.443	0.665	0.461	0.634	0.447	0.648	0.459
		mse		0.263	0.222	0.226	0.214	0.443	0.217	0.402	0.210	0.426	0.235
	24	Mean		0.426	0.322	0.366	0.302	0.429	0.307	0.384	0.280	0.358	0.248
		sd (0.29)	(0.26)	0.295	0.282	0.285	0.271	0.323	0.234	0.313	0.221	0.325	0.229
		mse		0.087	0.086	0.082	0.083	0.105	0.064	0.098	0.063	0.107	0.075
	36	Mean		0.421	0.345	0.376	0.328	0.476	0.352	0.445	0.332	0.424	0.303
		sd (0.22)	(0.20)	0.218	0.212	0.216	0.209	0.248	0.183	0.238	0.172	0.243	0.177
		mse		0.048	0.048	0.047	0.049	0.067	0.036	0.059	0.034	0.060	0.041
$n = 128$													
1	6	Mean		0.403	0.241	0.269	0.208	0.359	0.308	0.404	0.318	0.473	0.376
		sd (0.37)	(0.35)	0.616	0.422	0.486	0.432	0.407	0.259	0.474	0.307	0.483	0.324
		mse		0.380	0.204	0.254	0.224	0.167	0.075	0.224	0.101	0.239	0.106
	12	Mean		0.418	0.297	0.312	0.265	0.337	0.26	0.374	0.277	0.469	0.354
		sd (0.29)	(0.24)	0.337	0.295	0.295	0.283	0.346	0.227	0.318	0.220	0.355	0.264
		mse		0.114	0.098	0.095	0.098	0.123	0.071	0.102	0.064	0.131	0.072
	18	Mean		0.457	0.329	0.358	0.292	0.372	0.263	0.433	0.325	0.490	0.373
		sd (0.23)	(0.19)	0.245	0.215	0.227	0.213	0.258	0.169	0.247	0.175	0.266	0.192
		mse		0.064	0.051	0.053	0.057	0.068	0.047	0.062	0.036	0.079	0.038
2	6	Mean		0.434	0.252	0.271	0.145	0.444	0.369	0.354	0.318	0.313	0.274
		sd (0.60)	(0.68)	0.650	0.602	0.734	0.724	1.335	0.461	1.121	0.455	1.140	0.479
		mse		0.424	0.385	0.556	0.589	1.785	0.213	1.260	0.214	1.306	0.245
	12	Mean		0.476	0.308	0.390	0.278	0.440	0.332	0.377	0.304	0.334	0.253
		sd (0.47)	(0.42)	0.467	0.451	0.438	0.450	0.627	0.431	0.585	0.42	0.619	0.445
		mse		0.224	0.212	0.192	0.217	0.395	0.190	0.343	0.186	0.387	0.220
	18	Mean		0.488	0.354	0.412	0.320	0.492	0.351	0.451	0.323	0.416	0.286
		sd (0.35)	(0.31)	0.323	0.307	0.316	0.312	0.418	0.302	0.411	0.293	0.415	0.301
		mse		0.112	0.097	0.100	0.103	0.183	0.094	0.171	0.092	0.173	0.104

We report the simulation results for the ARFIMA$(1, d, 0)$ cointegrating series in Tables 4–6 for $\phi_1 = 0.3, 0.6$. Here the estimation is more difficult, since the signal/noise ratio at low frequencies is smaller than in the previous model. The results for $\phi_1 = 0.3$ are similar than before, though the best results correspond always to residual-based

Table 3
Model 3, $\alpha = 0.7$, CI$(1.1, 0.4)$, $\phi_1 = 0.34$, $\phi_2 = -0.9$

p	m			$q = 2$		$q = 1$		$X_t, \hat{Z}_t$		$\Delta X_t, \Delta\hat{Z}_t$		$\Delta X_t, \hat{Z}_t$	
				$\hat{\alpha}$	$\hat{\alpha}_\Delta$	$\hat{\alpha}$	$\hat{\alpha}_\Delta$	$\hat{\alpha}_L$	$\hat{\alpha}_G$	$\hat{\alpha}_L$	$\hat{\alpha}_G$	$\hat{\alpha}_L$	$\hat{\alpha}_G$
$n = 256$													
1	12	Mean		0.818	0.392	0.663	0.351	0.715	0.502	0.402	0.287	0.740	0.602
		sd (0.29)	(0.24)	0.333	0.329	0.303	0.328	0.364	0.250	0.386	0.246	0.359	0.252
		mse		0.125	0.203	0.093	0.230	0.132	0.101	0.238	0.231	0.130	0.073
	24	Mean		0.773	0.434	0.682	0.405	0.729	0.511	0.485	0.351	0.744	0.581
		sd (0.19)	(0.16)	0.199	0.237	0.196	0.235	0.206	0.146	0.266	0.179	0.201	0.149
		mse		0.045	0.127	0.039	0.142	0.043	0.057	0.117	0.154	0.042	0.036
	36	Mean		0.773	0.455	0.701	0.427	0.776	0.566	0.557	0.427	0.766	0.621
		sd (0.15)	(0.13)	0.160	0.180	0.158	0.191	0.165	0.124	0.219	0.164	0.162	0.123
		mse		0.031	0.092	0.025	0.111	0.033	0.033	0.068	0.101	0.030	0.021
2	12	Mean		0.741	0.571	0.635	0.544	0.728	0.512	0.649	0.458	0.651	0.453
		sd (0.47)	(0.42)	0.509	0.475	0.464	0.432	0.659	0.493	0.649	0.480	0.645	0.482
		mse		0.260	0.242	0.220	0.211	0.435	0.278	0.424	0.289	0.418	0.293
	24	Mean		0.701	0.593	0.638	0.578	0.727	0.524	0.672	0.482	0.671	0.478
		sd (0.29)	(0.26)	0.289	0.290	0.282	0.269	0.325	0.249	0.325	0.233	0.322	0.237
		mse		0.083	0.096	0.084	0.087	0.106	0.093	0.107	0.102	0.105	0.105
	36	Mean		0.669	0.591	0.626	0.582	0.766	0.583	0.73	0.552	0.726	0.544
		sd (0.22)	(0.20)	0.215	0.217	0.215	0.208	0.247	0.196	0.248	0.186	0.245	0.187
		mse		0.047	0.059	0.052	0.057	0.065	0.052	0.062	0.057	0.061	0.059
$n = 128$													
1	6	Mean		0.818	0.306	0.596	0.244	0.745	0.539	0.48	0.396	0.773	0.662
		sd (0.37)	(0.35)	0.570	0.422	0.485	0.439	0.466	0.296	0.497	0.295	0.479	0.333
		mse		0.339	0.334	0.246	0.401	0.220	0.113	0.296	0.180	0.235	0.112
	12	Mean		0.841	0.381	0.677	0.334	0.749	0.526	0.408	0.305	0.747	0.606
		sd (0.29)	(0.24)	0.336	0.310	0.304	0.302	0.378	0.250	0.359	0.232	0.355	0.266
		mse		0.133	0.198	0.093	0.226	0.145	0.093	0.214	0.210	0.128	0.080
	18	Mean		0.829	0.411	0.698	0.366	0.793	0.570	0.507	0.384	0.783	0.636
		sd (0.23)	(0.19)	0.249	0.247	0.234	0.245	0.268	0.192	0.29	0.198	0.255	0.196
		mse		0.079	0.144	0.055	0.172	0.081	0.054	0.121	0.139	0.072	0.042
2	6	Mean		0.678	0.500	0.499	0.361	0.735	0.553	0.640	0.500	0.648	0.500
		sd (0.60)	(0.68)	0.645	0.594	0.754	0.712	1.072	0.474	0.899	0.463	0.900	0.479
		mse		0.417	0.393	0.609	0.622	1.151	0.247	0.812	0.254	0.813	0.269
	12	Mean		0.711	0.545	0.621	0.513	0.756	0.552	0.665	0.492	0.678	0.499
		sd (0.47)	(0.42)	0.464	0.454	0.446	0.437	0.647	0.466	0.619	0.453	0.634	0.472
		mse		0.215	0.230	0.205	0.226	0.422	0.239	0.384	0.248	0.402	0.264
	18	Mean		0.690	0.560	0.622	0.536	0.800	0.577	0.732	0.528	0.739	0.528
		sd (0.35)	(0.32)	0.313	0.310	0.315	0.305	0.435	0.330	0.438	0.323	0.429	0.324
		mse		0.098	0.115	0.106	0.120	0.200	0.124	0.193	0.134	0.186	0.135

estimates: the log-periodogram regression for Models 1 and 3 and Gaussian estimation for Model 2. In this last case, coherence-based estimates have large biases, usually growing with m, but the sds in all cases are in line with approximation (19). For

Table 4
Model 1′, $\alpha = 1$, CI(1, 0)

$n = 256$				$q = 2$		$q = 1$		$X_t, \hat{Z}_t$		$\Delta X_t, \Delta\hat{Z}_t$		$\Delta X_t, \hat{Z}_t$	
p	m			$\hat{\alpha}$	$\hat{\alpha}_\Delta$	$\hat{\alpha}$	$\hat{\alpha}_\Delta$	$\hat{\alpha}_L$	$\hat{\alpha}_G$	$\hat{\alpha}_L$	$\hat{\alpha}_G$	$\hat{\alpha}_L$	$\hat{\alpha}_G$
$\phi_1 = 0.3$													
1	12	Mean		1.024	0.539	0.877	0.507	0.936	0.683	0.594	0.430	0.924	0.736
		sd (0.29)	(0.24)	0.330	0.346	0.297	0.350	0.361	0.262	0.433	0.287	0.355	0.252
		mse		0.109	0.332	0.103	0.366	0.134	0.169	0.353	0.408	0.132	0.133
	24	Mean		0.829	0.488	0.778	0.491	0.889	0.694	0.656	0.519	0.866	0.716
		sd (0.19)	(0.16)	0.211	0.215	0.198	0.227	0.214	0.157	0.285	0.207	0.213	0.153
		mse		0.074	0.308	0.089	0.311	0.058	0.118	0.200	0.274	0.063	0.104
	36	Mean		0.696	0.415	0.682	0.434	0.827	0.690	0.634	0.553	0.791	0.705
		sd (0.15)	(0.13)	0.175	0.163	0.170	0.177	0.186	0.125	0.219	0.169	0.173	0.122
		mse		0.123	0.369	0.130	0.351	0.065	0.112	0.182	0.229	0.074	0.102
$\phi_1 = 0.6$													
1	12	Mean		0.837	0.412	0.737	0.429	0.866	0.666	0.621	0.486	0.841	0.696
		sd (0.29)	(0.24)	0.329	0.300	0.298	0.294	0.368	0.259	0.396	0.271	0.353	0.251
		mse		0.134	0.436	0.158	0.412	0.154	0.178	0.300	0.337	0.150	0.156
	24	Mean		0.603	0.321	0.588	0.358	0.754	0.629	0.582	0.524	0.718	0.646
		sd (0.19)	(0.16)	0.213	0.187	0.203	0.190	0.222	0.153	0.256	0.180	0.215	0.153
		mse		0.203	0.496	0.211	0.448	0.110	0.161	0.241	0.259	0.126	0.149
	36	Mean		0.473	0.252	0.485	0.294	0.654	0.576	0.514	0.507	0.614	0.592
		sd (0.14)	(0.13)	0.172	0.145	0.169	0.151	0.186	0.121	0.192	0.137	0.170	0.120
		mse		0.307	0.581	0.294	0.521	0.155	0.195	0.273	0.262	0.178	0.180

larger ϕ_1 the performance of all estimates deteriorates, especially that of $\hat{\alpha}$ and $\hat{\alpha}_\Delta$ for Models 2 and 3.

In conclusion, $\hat{\alpha}$ seems a simple competitive alternative to residual-based estimates, which may be affected by the combination of memory estimates for observed series and cointegrating residuals.

6. Empirical example

Dueker and Startz (1998) analysed 120 monthly observations from January 1987 to December 1996 on 10-year government bond rates from the United States and Canada. We analyse here the same log series, denoted as X_t and Y_t, respectively (see Fig. 1). Standard procedures used by these authors do not reject the hypothesis of a unit root ($d = 1$) for both series nor the hypothesis of no cointegration, but the visual evidence is in favour of a long-run relationship, probably different from the CI(1, 0) paradigm.

Dueker and Startz (1998) also fit a bivariate ARFIMA model with two orders of integration, one for the differenced US series ΔX_t (d) and one for the cointegration error ($d - \alpha$). They find that $\hat{d} = 0.674$ (0.25) and $\widehat{d - \alpha} = 0.2$ (0.10), so $\hat{\alpha} = 0.474$, while a joint Wald test rejects $d = 1$ and $\alpha = 1$. They also estimate the memory of the observed residuals with Lobato and Robinson's (1996) average periodogram semiparametric

Table 5
Model 2′, $\alpha = 0.4$, CI(1.3, 0.9)

$n = 256$			$q = 2$		$q = 1$		$X_t, \hat{Z}_t$		$\Delta X_t, \Delta\hat{Z}_t$		$\Delta X_t, \hat{Z}_t$	
p	m		$\hat{\alpha}$	$\hat{\alpha}_\Delta$	$\hat{\alpha}$	$\hat{\alpha}_\Delta$	$\hat{\alpha}_L$	$\hat{\alpha}_G$	$\hat{\alpha}_L$	$\hat{\alpha}_G$	$\hat{\alpha}_L$	$\hat{\alpha}_G$
$\phi_1 = 0.3$												
1	12	Mean	0.251	0.153	0.185	0.150	0.265	0.225	0.302	0.269	0.390	0.338
		sd (0.29) (0.24)	0.329	0.271	0.289	0.250	0.344	0.222	0.308	0.225	0.346	0.253
		mse	0.131	0.135	0.130	0.125	0.137	0.080	0.105	0.068	0.120	0.068
	24	Mean	0.172	0.134	0.143	0.137	0.183	0.152	0.278	0.251	0.327	0.286
		sd (0.19) (0.16)	0.205	0.172	0.194	0.168	0.219	0.140	0.187	0.140	0.216	0.154
		mse	0.094	0.100	0.104	0.098	0.095	0.081	0.050	0.042	0.052	0.037
	36	Mean	0.122	0.108	0.107	0.116	0.133	0.118	0.243	0.234	0.284	0.260
		sd (0.15) (0.13)	0.156	0.134	0.152	0.133	0.176	0.113	0.148	0.108	0.170	0.119
		mse	0.101	0.103	0.109	0.099	0.102	0.093	0.046	0.039	0.042	0.034
$\phi_1 = 0.6$												
1	12	Mean	0.168	0.069	0.122	0.074	0.225	0.207	0.248	0.251	0.339	0.323
		sd (0.29) (0.24)	0.310	0.251	0.274	0.230	0.335	0.222	0.302	0.223	0.328	0.253
		mse	0.150	0.172	0.152	0.159	0.143	0.086	0.114	0.072	0.111	0.070
	24	Mean	0.061	0.053	0.050	0.059	0.096	0.102	0.190	0.189	0.254	0.237
		sd (0.19) (0.16)	0.192	0.160	0.182	0.157	0.216	0.140	0.175	0.135	0.205	0.150
		mse	0.152	0.146	0.155	0.141	0.139	0.108	0.075	0.062	0.063	0.049
	36	Mean	−0.013	0.029	−0.011	0.039	0.012	0.042	0.128	0.128	0.198	0.178
		sd (0.14) (0.13)	0.156	0.125	0.150	0.125	0.181	0.116	0.136	0.101	0.163	0.117
		mse	0.195	0.154	0.192	0.146	0.183	0.141	0.092	0.084	0.068	0.063

estimator, obtaining significantly different from zero values of $\hat{\alpha}$, ranging from 0.2 to 0.28 for small bandwidths ($m < 10$ in similar notation to our coherence-based $\hat{\alpha}$) and about 0.4 for larger values of m. This is an alternative procedure to the one justified by Hassler et al. (2001) and Velasco (2001) for other semiparametric estimates.

We reanalyse this data set first using the techniques summarized in Lobato and Velasco (2000) using a multivariate generalization of Robinson's (1995b) Gaussian semiparametric estimate of the memory d. We use the increments of the original series without tapering and with a taper of order $p = 2$ and bandwidths $m = 6, 12, 18$. A semiparametric Wald test of equal memory for both bond rates series is performed in first place, with p-values equal to:

p-value Wald Test	$m = 6$	$m = 12$	$m = 18$
$p = 1$ $(\Delta X, \Delta Y)$	0.102	0.131	0.119
$p = 2$ (X, Y)	0.644	0.211	0.245

and not rejecting the equal memory hypothesis, though by a small margin using nontapered differenced data, as memory estimates for X are slightly larger than those

Table 6
Model 3′, $\alpha = 0.7$, CI(1.1, 0.4)

$n = 256$				$q = 2$		$q = 1$		$X_t, \hat{Z}_t$		$\Delta X_t, \Delta\hat{Z}_t$		$\Delta X_t, \hat{Z}_t$	
p	m			$\hat{\alpha}$	$\hat{\alpha}_\Delta$	$\hat{\alpha}$	$\hat{\alpha}_\Delta$	$\hat{\alpha}_L$	$\hat{\alpha}_G$	$\hat{\alpha}_L$	$\hat{\alpha}_G$	$\hat{\alpha}_L$	$\hat{\alpha}_G$
$\phi_1 = 0.3$													
1	12	Mean		0.660	0.354	0.544	0.357	0.663	0.490	0.543	0.424	0.664	0.529
		sd (0.29)	(0.24)	0.333	0.289	0.307	0.274	0.370	0.247	0.357	0.250	0.356	0.254
		mse		0.112	0.203	0.119	0.193	0.138	0.105	0.152	0.139	0.128	0.094
	24	Mean		0.524	0.305	0.488	0.331	0.598	0.456	0.517	0.440	0.590	0.499
		sd (0.19)	(0.16)	0.214	0.181	0.200	0.183	0.217	0.148	0.223	0.158	0.216	0.154
		mse		0.077	0.189	0.085	0.170	0.058	0.081	0.083	0.092	0.059	0.064
	36	Mean		0.438	0.254	0.417	0.278	0.542	0.433	0.474	0.437	0.526	0.480
		sd (0.15)	(0.13)	0.172	0.141	0.169	0.141	0.172	0.115	0.170	0.121	0.169	0.119
		mse		0.098	0.219	0.108	0.198	0.054	0.085	0.080	0.084	0.059	0.063
$\phi_1 = 0.6$													
1	12	Mean		0.492	0.211	0.415	0.236	0.593	0.471	0.467	0.421	0.572	0.510
		sd (0.29)	(0.24)	0.328	0.266	0.303	0.253	0.366	0.245	0.340	0.243	0.351	0.253
		mse		0.151	0.309	0.173	0.280	0.146	0.113	0.170	0.137	0.140	0.100
	24	Mean		0.346	0.162	0.325	0.189	0.480	0.391	0.407	0.391	0.469	0.438
		sd (0.19)	(0.16)	0.208	0.169	0.199	0.167	0.213	0.144	0.202	0.149	0.207	0.152
		mse		0.169	0.317	0.180	0.289	0.094	0.116	0.127	0.118	0.096	0.092
	36	Mean		0.260	0.121	0.259	0.149	0.387	0.321	0.334	0.340	0.381	0.371
		sd (0.14)	(0.13)	0.161	0.133	0.160	0.134	0.164	0.111	0.153	0.113	0.158	0.115
		mse		0.219	0.353	0.220	0.322	0.125	0.156	0.157	0.142	0.127	0.121

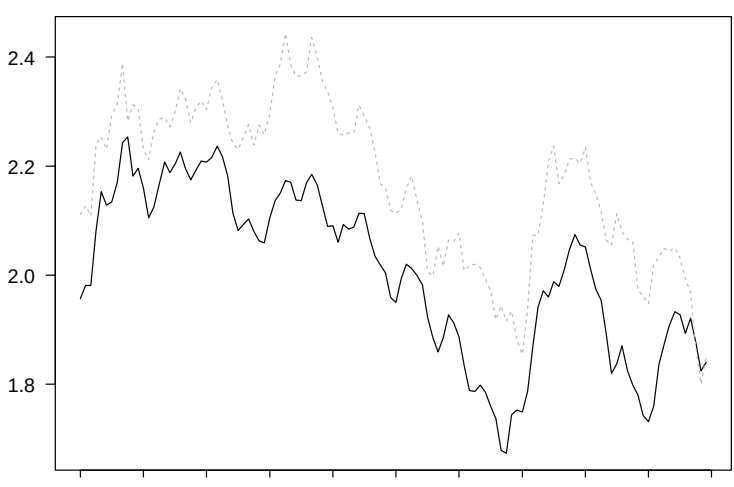

Fig. 1. Logarithm of bond rates, US (solid) and Canada (dotted).

for Y. Then the common memory parameter Gaussian estimates $\hat{d}$ are:

Common $\hat{d}$	$m=6$	$m=12$	$m=18$
$p=1$ $(\Delta X, \Delta Y)$	0.982 (0.09)	1.010 (0.10)	0.990 (0.08)
$p=2$ (X, Y)	0.818 (0.12)	1.146 (0.14)	0.855 (0.12)

These are noticeable larger that ML estimates obtained by Dueker and Startz (1998). We finally compute zero frequency coherence estimates $|\hat{H}_{xy}(0)|^2 = |\hat{G}_{xy}|^2(\hat{G}_{xx}\hat{G}_{yy})^{-1}$ obtained from the previous semiparametric estimation:

$\lvert\hat{H}_{xy}(0)\rvert^2$	$m=6$	$m=12$	$m=18$
$p=1$ (X, Y)	0.795 (0.11)	0.752 (0.09)	0.743 (0.07)
$p=2$ (X, Y)	0.800 (0.15)	0.747 (0.13)	0.759 (0.10)

which are inconclusive of coherence smaller than 1 given the sample size and the bandwidths employed. Notice that the previous procedures have been only justified under the hypothesis of no cointegration, i.e. with **G** being nonsingular.

We now estimate nonparametrically the coherence with $|\hat{H}_{xy}(\lambda_j)|^2$ for $q=1,2,3$ and $p=1,2$. We plot the estimates in Fig. 2 for $\lambda_\ell - \lambda_{60} = \pi$, where $\ell = \lfloor (2q+1)/2 \rfloor$. Standard errors can be approximated by (23). In all plots is evident the effect of increasing the smoothing in nonparametric estimates and it can be observed that $|H_{xy}(\lambda)|^2 \approx 1 - G_{\mathrm{H}}\lambda^{2\alpha}$ is a reasonable approximation. In the plots of $\log(1 - |\hat{H}_{xy}(\lambda_j)|^2)$ against $2\log\lambda_j$, for $j = \ell, \ldots, 30$, see Fig. 3, the linear relationship becomes more clear as q increases, though this is not valid for all the range of frequencies plotted. The OLS estimates of $\hat{\alpha}$ obtained from (X, Y) are

Log-coherence $\hat{\alpha}$		$m=6$	$m=12$	$m=18$
$q=1$	$p=1$	0.371 (0.35)	0.292 (0.24)	0.177 (0.19)
	$p=2$	0.574 (0.68)	0.549 (0.42)	0.545 (0.32)
$q=2$	$p=1$	0.650 (0.37)	0.299 (0.29)	0.226 (0.23)
	$p=2$	0.757 (0.64)	0.670 (0.47)	0.665 (0.37)
$q=3$	$p=1$	0.704 (0.51)	0.305 (0.31)	0.299 (0.25)
	$p=2$	1.023 (0.86)	0.681 (0.49)	0.768 (0.39)

As in the simulations, the results with differenced data were not interpretable and are not reported. For $p=1$ and $m=18$ we are including in the regression the high coherence points around frequency λ_{15}, explaining the low value of $\hat{\alpha}$ obtained in this

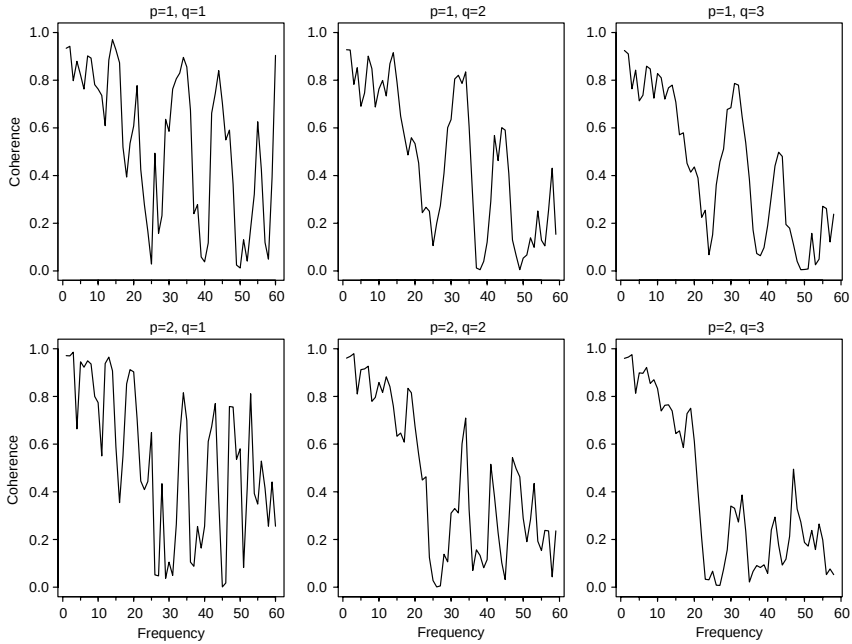

Fig. 2. Coherence estimates $|\hat{H}_{xy}(\lambda_j)|^2$ of bond rates series, against log-frequency $\log \lambda_j$, $j = \ell, \ell+1, \ldots, n/2$. Upper row, no tapering $p = 1$; lower row, tapering $p = 2$. Uniform weights: $q = 1, 2, 3$.

case. The estimates for $m = 6$ are rather unstable due to the small number of points in the regression. Then, since for nontapered series the estimates were quite uniform across values of q, we prefer estimates with $q = 1$ and $m = 12$ which have smallest standard errors and should be also less biased. This gives $\hat{\alpha} = 0.292$ (0.24), which is lower than the value given by Dueker and Startz (1998), 0.474. However, for tapered series the estimates are more smoothing dependent, and we obtained from $\hat{\alpha} = 0.55$ for $q = 1$ to $\hat{\alpha} = 0.68$ for $q = 3$, more in agreement with that paper.

Finally, we used residual estimation with a multivariate two-step Gaussian semiparametric estimate (Lobato, 1999; Lobato and Velasco, 2000), which remains consistent if $\alpha > 0$ and has the usual asymptotic distribution if $\alpha > 0.5$ (Velasco, 2001). We applied joint estimation between the OLS residual series $\tilde{Z}$ and ΔX to obtain standard errors for $\tilde{\alpha} = \tilde{d} - \tilde{d}_Z$.

Residual-based $\tilde{\alpha}$	$m = 6, 8$	$m = 12$	$m = 18$
$p = 1$ $(\Delta X, \tilde{Z})$	0.186 (0.29)	0.094 (0.20)	0.068 (0.17)
$p = 2$ $(\Delta X, \tilde{Z})$	0.159 (0.34)	0.002 (0.37)	0.079 (0.31)

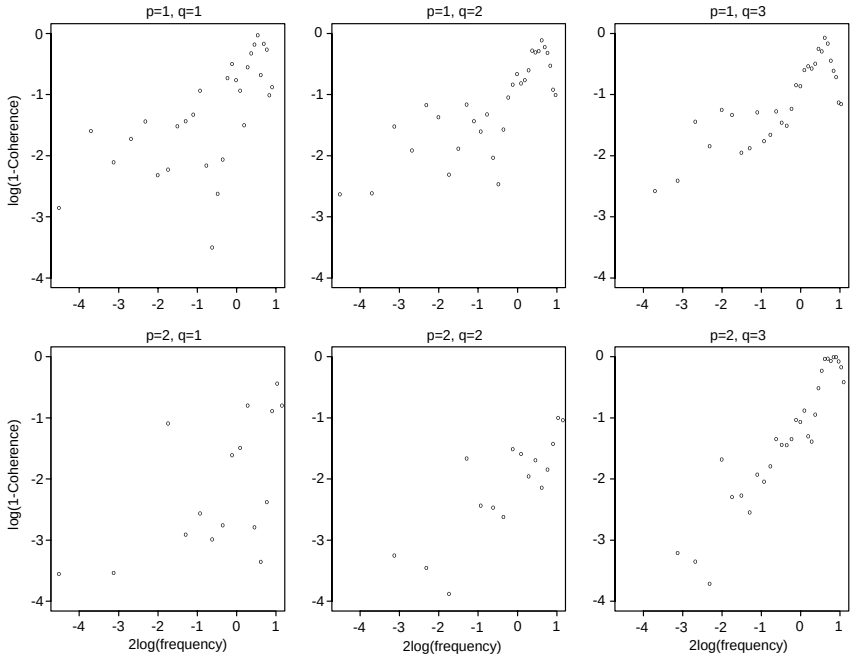

Fig. 3. Log-coherence estimates $\log(1 - |\hat{H}_{xy}(\lambda_j)|^2)$ of bond rates series against log-frequency $2\log\lambda_j$, $j = \ell, \ell+1, \ldots, n/4$. Upper row, no tapering $p = 1$; lower row, tapering $p = 2$. Uniform weights: $q = 1, 2, 3$.

The residual-based procedure obtained values of $\tilde{\alpha}$ noticeably much smaller than coherence-based $\hat{\alpha}$ for all combinations of m and tapering, casting some doubts about the reliability of linear OLS-based inference for the present data set.

Acknowledgements

I am grateful to U. Hassler, C.S. Kim and the participants at the Cowles Foundation conference on New Developments in Time Series Econometrics (1999), the associate editor and the referee for helpful comments. Research funded by the Spanish Dirección General de Enseñanza Superior, Ref. no. BEC2001-1270.

Appendix

Proof of Theorem 3. We start approximating the cross-periodogram of the observed vector series by that of the linear innovations, $I^{\epsilon,p}(\lambda_{jp})$, times the transfer function,

including the unit root filters of the integer differences. Define

$$\hat{f}_{rs}^{\epsilon,M}(v) = \frac{2\pi p}{n} \sum_j K_M(v - \lambda_{jp}) B_r(\lambda_{jp}) I^{\epsilon,p}(\lambda_{jp}) B_s^*(\lambda_{jp}),$$

where the index j runs for $|\lambda_{jp} - v| \leqslant M^{-1}\pi$.

No tapering [$p = 1$]. We consider the case with $d_* = \max\{d_r, d_s\} \in (-\frac{1}{2}, 1)$, and $\mu_r = \mu_s = 0$ and $D_r, D_s = 0, 1$. Using Lemma 2 below and the arguments of the proof of Theorem 1 of Robinson (1995b) (see also the proofs of Theorem 2 and Lemma 1 of Velasco (1999b) and Appendix C of Lobato and Velasco (2000)),

$$\begin{aligned}\hat{f}_{rs}^{M}(v) - \hat{f}_{rs}^{\epsilon,M}(v) &= \frac{2\pi}{n} \sum_j K_M(v - \lambda_j)[I_{rs}(\lambda_j) - B_r(\lambda_j) I^{\epsilon}(\lambda_j) B_s^*(\lambda_j)] \\ &= \mathrm{O_p}\left(\frac{M}{n} \sum_j [j^{d_*-1} + j^{-1/2}](\log n)^{1/2}\right) \\ &= \mathrm{O_p}([n^{d_*-1} + n^{-1/2}](\log n)^{1/2}), \end{aligned} \tag{A.1}$$

which is $\mathrm{o_p}(1)$ if $d_* < 1$. Notice that for $\lambda_j \in [v - \pi M^{-1}, v + \pi M^{-1}]$, $\max_j \lambda_j^{-1} = \mathrm{O}(1)$ and $\max j^{-1} = \mathrm{O}(n^{-1})$ as $n \to \infty$, and from Assumptions 1 and 4, $\max_j |B_a(\lambda_j)|$, $a = r, s$, are bounded if $|v| > 0$. Now the theorem follows as when $p > 1$ below, using the exact orthogonality of the sine and cosine instead of Lemma A.4(A).

Tapering [$p > 1$]. From Theorem 2 and using the same argument as in the proof of Theorem 3 in Velasco (1999b) or Lemma A.1,

$$\hat{f}_{rs}^{M}(v) - \hat{f}_{rs}^{\epsilon,M}(v) = \frac{2\pi p}{n} \sum_j^n K_M(v - \lambda_{jp})[I_{rs}^p(\lambda_j) - B_r(\lambda_j) I^{\epsilon,p}(\lambda_j) B_s^*(\lambda_j)] \tag{A.2}$$

$$\begin{aligned} &= \mathrm{O_p}\left(\frac{M}{n} \sum_j [j^{d_*-p}(\log n)^{1/2} + j^{-1/2}]\right) \\ &= \mathrm{O_p}(n^{d_*-p}(\log n)^{1/2} + n^{-1/2}), \end{aligned} \tag{A.3}$$

which is $\mathrm{o_p}(1)$ because $p > d_*$.

Now, using the differentiability of $f_{rs}(\lambda)$ around v and the Lipschitz property of $K_M(\lambda)$, $p > 1$,

$$E[\hat{f}_{r,s}^{\epsilon,M}(v)] = \frac{2\pi p}{n} \sum_j K_M(v - \lambda_{jp}) f_{rs}(\lambda_{jp}) = f_{rs}(v) + \mathrm{O}(n^{-1}M + M^{-1}), \tag{A.4}$$

where the error is o(1) using (6). For the variance we first obtain for all j,k, and $r,r',s,s' \in \{1,\ldots,R\}$, $N = 2\pi \sum_{t=1}^{n} h_t^2$,

$$
\begin{aligned}
&\mathrm{Cov}[B_r(\lambda_{jp})I^{\epsilon}(\lambda_{jp})B_s^*(\lambda_{jp}), B_{r'}(\lambda_{kp})I^{\epsilon}(\lambda_{kp})B_{s'}^*(\lambda_{kp})] \\
&\quad = \sum_{a=1}^{R}\sum_{b=1}^{R}\sum_{c=1}^{R}\sum_{d=1}^{R} B_{ra}(\lambda_{jp})\overline{B_{sb}(\lambda_{jp})B_{r'c}(\lambda_{kp})}B_{s'd}(\lambda_{kp})\,\mathrm{Cov}[I_{ab}^{\epsilon,p}(\lambda_{jp}), I_{cd}^{\epsilon,p}(\lambda_{kp})] \\
&\quad = N^{-2}\sum_{a=1}^{R}\sum_{b=1}^{R}\sum_{c=1}^{R}\sum_{d=1}^{R} B_{ra}(\lambda_{jp})\overline{B_{sb}(\lambda_{jp})B_{r'c}(\lambda_{kp})}B_{s'd}(\lambda_{kp}) \\
&\qquad \times \left\{ \Sigma_{ac}\Sigma_{bd}\left[\sum_{t}^{n} h_t^2 \cos t(\lambda_{jp}-\lambda_{kp})\right]^2 + \Sigma_{ad}\Sigma_{cb}\left[\sum_{t}^{n} h_t^2 \cos t(\lambda_{jp}+\lambda_{kp})\right]^2 \right. \\
&\qquad \left. + \sum_{t}^{n} \kappa_{abcd} h_t^4 \right\} \\
&\quad = N^{-2} B_r(\lambda_{jp})\Sigma B_{r'}^*(\lambda_{kp})B_s(-\lambda_{jp})\Sigma B_{s'}^*(-\lambda_{kp})\left[\sum_{t}^{n} h_t^2 \cos t(\lambda_{jp}-\lambda_{kp})\right]^2 \\
&\qquad + N^{-2} B_r(\lambda_{jp})\Sigma B_{s'}^*(-\lambda_{kp})B_{r'}(-\lambda_{kp})\Sigma B_s^*(\lambda_{jp})\left[\sum_{t}^{n} h_t^2 \cos t(\lambda_{jp}+\lambda_{kp})\right]^2 \\
&\qquad + N^{-2}\sum_{1}^{n} h_t^4 \sum_{a=1}^{R}\sum_{b=1}^{R}\sum_{c=1}^{R}\sum_{d=1}^{R} B_{ra}(\lambda_{jp})\overline{B_{sb}(\lambda_{jp})B_{r'c}(\lambda_{kp})}B_{s'd}(\lambda_{kp})\kappa_{abcd},
\end{aligned}
$$

where κ_{abcd} is the joint fourth-order cumulant of the a,b,c,dth components of ϵ_t. Using Lemma A.4(A), $\mathrm{Var}[\hat{f}_{rs}^{\epsilon,M}(v)]$ is

$$
\begin{aligned}
&\left(\frac{2\pi p}{n}\right)^2 \sum_{j} K_M^2(v-\lambda_{jp})\,\mathrm{Var}[B_r(\lambda_{jp})I^{\epsilon,p}(\lambda_{jp})B_s^*(\lambda_{jp})] \\
&\quad + \left(\frac{2\pi p}{n}\right)^2 \sum_{j}\sum_{k\neq j} K_M(v-\lambda_{jp})K_M(v-\lambda_{kp}) \\
&\quad \times \mathrm{Cov}[B_r(\lambda_{jp})I^{\epsilon}(\lambda_{jp})B_s^*(\lambda_{jp}), B_r(\lambda_{kp})I^{\epsilon}(\lambda_{kp})B_s^*(\lambda_{kp})] \\
&= n^{-2}\mathrm{O}\Bigg(\sum_{j} K_M^2(v-\lambda_{jp}) + \sum_{j}\sum_{k>j} K_M(v-\lambda_{jp})K_M(v-\lambda_{kp})
\end{aligned}
$$

$$\times[|j-k|^{-2p}+|j+k|^{-2p}+n^{-1}]\Bigg)$$

$$=\mathrm{O}\left(n^{-1}M+n^{-2}M^2\sum_j\sum_{k>j}\{|j-k|^{-2p}+|j+k|^{-2p}+n^{-1}\}\right)$$

$$=\mathrm{O}(n^{-1}M)=\mathrm{o}(1)$$

as $n\to\infty$, so $\hat{f}_{rs}^{M}(v)-f_{rs}(v)\to_p 0$ follows using (6).

For the proof of the central limit theorem we assume for simplicity only positive frequencies $v_a>0\ \forall a$. This entails no restriction, since it is always possible to write for positive v that $\hat{f}_{rs}(-v)=\overline{\hat{f}_{rs}(v)}=\hat{f}_{sr}(v)$ and deduce the variances and covariances for the conjugate estimates from those with positive argument and reversed indexes. We follow the same procedure as in the proof of consistency, but employ Lemma A.2 when $p=1$ instead of Lemma A.1.

No tapering $[p=1]$. From Lemma A.2 and (8), $\|\hat{f}_{\mathbf{rs}}^{M}(v)-\hat{f}_{\mathbf{rs}}^{\epsilon,M}(v)\|=\mathrm{o_p}((n/M)^{-1/2})$. As in the proof of Theorem 3, with Assumption 6 and for all p,

$$E[\hat{f}_{rs}^{\epsilon,M}(v)]=f_{rs}(v)+\mathrm{O}(M^2n^{-1}+M^{-2}).$$

Tapering $[p>1]$. Now (A.3) is $\mathrm{o_p}((n/M)^{-1/2})$ with (9), if $\mu=0$, or with $p-d_*>0.5$ if $\mu\neq 0$.

For the central limit theorem we follow Hall and Heyde (1980, Section 3.2) and consider in detail only the case $p>1$. We have to consider linear combinations of the estimates, so for any $J\times 1$ vector ξ we have that $\xi'\hat{f}_{\mathbf{rs}}^{M}(v)=\xi'\hat{f}_{\mathbf{rs}}^{M,\epsilon}(v)+\mathrm{o_p}((n/M)^{-1/2})$. Now

$$\begin{aligned}
&\xi'\hat{f}_{\mathbf{rs}}^{\epsilon,M}(v)-E[\xi'\hat{f}_{\mathbf{rs}}^{\epsilon,M}(v)]\\
&=\sum_{a=1}^{J}\xi_a\left\{\frac{2\pi p}{n}\sum_j K_M(v_a-\lambda_j)B_{r(a)}(\lambda_j)I^{\epsilon}(\lambda_j)B^*_{s(a)}(\lambda_j)\right.\\
&\qquad\left.-\frac{p}{n}\sum_j K_M(v-\lambda_j)B_{r(a)}(\lambda_j)\Sigma B^*_{s(a)}(\lambda_j)\right\}\\
&=\sum_{a=1}^{J}\xi_a\frac{p}{n}\sum_j K_M(v_a-\lambda_j)B_{r(a)}(\lambda_j)[\nabla_{\epsilon}^{(1)}-\Sigma]B^*_{s(a)}(\lambda_j)\\
&\qquad+\sum_{a=1}^{J}\xi_a\frac{p}{n\sum h_t^2}\sum_j K_M(v_a-\lambda_j)B_{r(a)}(\lambda_j)\nabla_{\epsilon}^{(2)}B^*_{s(a)}(\lambda_j),
\end{aligned}$$

with equivalent notation as before, possibly now with data tapers, $\nabla_{\epsilon}^{(1)}=(\sum h_t^2)^{-1}\sum_{t=1}^{n}h_th_{t'}\epsilon_t\epsilon_t'$ and $\nabla_{\epsilon}^{(2)}=\sum_t\sum_{t'\neq t}h_th_{t'}\epsilon_t\epsilon_{t'}'\exp\{\mathrm{i}(t-t')\lambda_j\}$. The first term is negligible

$\mathrm{o_p}((Mn^{-1})^{1/2})$ since $\|\nabla_\epsilon^{(1)} - \Sigma\| = \mathrm{O_p}(n^{-1/2})$ with Assumption 2. Then we have

$$\sqrt{\frac{n}{Mp}}\{\xi' \hat{f}_{\mathbf{rs}}^M(v) - \xi' f_{\mathbf{rs}}(v)\} = \sum_{t=1}^n z_t + \mathrm{o_p}(1),$$

where

$$\begin{aligned} z_t &= \sum_{a=1}^J \xi_a \frac{p}{n\sum h_t^2} \sqrt{\frac{n}{Mp}} \sum_{j(a)} K_M(v_a - \lambda_{pj(a)}) B_{r(a)}(\lambda_{pj(a)}) \\ &\quad \times \sum_{s\neq t} h_t \epsilon_t h_s \epsilon_s' B_{s(a)}^*(\lambda_{pj(a)}) \exp\{\mathrm{i}(t-s)\lambda_{pj(a)}\} \\ &= h_t \epsilon_t' \sum_{s=1}^{t-1} \Lambda_{t-s} h_s \epsilon_s \end{aligned}$$

is a martingale difference sequence,

$$\Lambda_t = \frac{1}{\sum h_t^2} \sqrt{\frac{p}{nM}} \sum_{a=1}^J \sum_{j(a)} \Theta_{pj(a)}^a \cos t\lambda_{pj(a)},$$

$\Theta_{pj(a)}^a = 2\xi_a K_M(v_a - \lambda_{pj(a)}) B_{r(a)}'(\lambda_{pj(a)}) \overline{B_{s(a)}}(\lambda_{pj(a)})$, and the summation in $j(a)$ runs from $-n/2p$ to $n/2p+1$, with steps of size p, assuming $n/2p$ is integer for simplicity. We estimate first the asymptotic variance of $\hat{f}^M$,

$$\sum_{t=1}^n E[z_t \overline{z_t} | F_{t-1}] = \sum_{t=2}^n h_t^2 \sum_{s=1}^{t-1} h_s^2 \epsilon_s' \Lambda_{t-s}' \Sigma \overline{\Lambda_{t-s}} \epsilon_s \tag{A.5}$$

$$+ \sum_{t=1}^n h_t^2 \sum_{s=1}^{t-1} \sum_{r\neq s}^{t-1} h_s h_r \epsilon_s' \Lambda_{t-s}' \Sigma \overline{\Lambda_{t-r}} \epsilon_r. \tag{A.6}$$

The right-hand side of (A.5) is

$$\sum_{t=2}^{n-1} h_t^2 \sum_{s=1}^{t-1} h_s^2 \epsilon_s' \Lambda_{t-s} \Sigma \Lambda_{t-s}^* \epsilon_s = \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \operatorname{Trace}[(\epsilon_t \epsilon_t' - \Sigma) \Lambda_s' \Sigma \overline{\Lambda_s}] \tag{A.7}$$

$$+ \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \operatorname{Trace}[\Sigma \Lambda_s' \Sigma \overline{\Lambda_s}], \tag{A.8}$$

where the right-hand side of (A.7) is $\mathrm{o_p}(1)$, because it has zero mean and variance, $\|\Lambda_{t-s}\| = \mathrm{O}((Mn)^{-1/2})$,

$$\begin{aligned} &\mathrm{O}\left(\sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \{2 \operatorname{Trace}[\Sigma \Lambda_s' \Sigma \overline{\Lambda_s} \Sigma \Lambda_s' \Sigma \overline{\Lambda_s}] + \kappa_{abcd} \operatorname{Trace}[\Sigma \Lambda_s' \Sigma \overline{\Lambda_s}]^2\}\right) \\ &= \mathrm{O}\left(\sum_{t=1}^{n-1} \sum_{s=1}^{n-t} (Mn)^{-2}\right) = \mathrm{O}(M^{-2}) = \mathrm{o}(1). \end{aligned}$$

Term (A.8) is, using trigonometric identities (see Velasco (1999b, Lemma 6) for details),

$$\sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \operatorname{Trace}[\Sigma \Lambda_s' \Sigma \overline{\Lambda_s}]$$

$$= \frac{p}{nMH^2} \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2$$

$$\times \operatorname{Trace}\left[\Sigma \sum_a \xi_a \sum_{j(a)} \Theta^{a'}_{pj(a)} \cos s\lambda_{pj(a)} \Sigma \sum_b \xi_b \sum_{k(b)} \overline{\Theta^b_{pk(b)}} \cos s\lambda_{pk(b)}\right]$$

$$= \frac{p}{nMH^2} \sum_a \sum_b \sum_{j(a)} \xi_a \xi_b$$

$$\times \operatorname{Trace}[\Sigma \Theta^{a'}_{pj(a)} \Sigma \overline{\Theta^{b*}_{pj(a)}} + \Sigma \Theta^{a}_{pj(a)} \Sigma \overline{\Theta^{b}_{-pj(a)}}] \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \cos^2 s\lambda_{pj(a)} \tag{A.9}$$

$$+ \frac{p}{2nMH^2} \sum_a \sum_{j(a)} \sum_b \sum_{k(b) \neq \pm j(a)} \xi_a \xi_b \operatorname{Trace}[\Sigma \Theta^{a'}_{pj(a)} \Sigma \overline{\Theta^b_{pk(b)}}]$$

$$\times \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \{\cos(sp\lambda_{j(a)+k(b)}) + \cos(sp\lambda_{j(a)-k(b)})\}. \tag{A.10}$$

Using $\|\Theta^a_{j(a)}\| = \mathrm{O}(M)$ and Lemma A.4(B), (A.9) is equal to

$$\frac{p}{4nM} \sum_a \sum_b \sum_{j(a)} \xi_a \xi_b \operatorname{Trace}[\Sigma \Theta^{a'}_{pj(a)} \Sigma \overline{\Theta^b_{pj(a)}} + \Sigma \Theta^{a'}_{pj(a)} \Sigma \overline{\Theta^b_{-pj(a)}}] + \mathrm{O}(M/n).$$

If $k(b) \neq \pm j(a)$, using Lemma A.4(C), (A.10) is

$$\frac{p}{4nMH^2} \sum_a \sum_{j(a)} \sum_b \sum_{k(b) \neq \pm j(a)} \xi_a \xi_b \operatorname{Trace}[\Sigma \Theta^{a'}_{pj(a)} \Sigma \overline{\Theta^b_{pk(b)}}]$$

$$\times \left\{ \left(\sum_{t=1}^{n-1} h_t^2 \cos tp\lambda_{j(a)+k(b)} \right)^2 + \left(\sum_{t=1}^{n-1} h_t^2 \cos tp\lambda_{j(a)-k(b)} \right)^2 \right\} + \mathrm{O}(M/n),$$

where $\operatorname{Trace}[\Sigma \Theta^{a'}_{pj(a)} \Sigma \overline{\Theta^b_{pk(b)}}]$ is

$$4K_M(v_a - \lambda_{pj(a)}) K_M(v_b - \lambda_{pk(b)}) B_{r(a)}(\lambda_{pj(a)})$$

$$\Sigma B^*_{r(b)}(\lambda_{pk(b)}) B_{s(a)}(-\lambda_{pj(a)}) \Sigma B^*_{s(b)}(-\lambda_{pk(b)}).$$

Therefore, using Lemma A.4(A), the differentiability of $B(\lambda)$, the compact support of K and approximating sums by integrals, with the same method as in the proof of Lemma 7 in Velasco and Robinson (2000),

$$\frac{p}{4nMH^2}\sum_a\sum_{j(a)}\sum_b\sum_{k(b)}\xi_a\xi_b\,\mathrm{Trace}[\Sigma\Theta^{a'}_{pj(a)}\Sigma\overline{\Theta^{b}_{pk(b)}}]\left(\sum_{t=1}^{n-1}h_t^2\cos tp\lambda_{j(a)-k(b)}\right)^2$$

$$=2\pi\Phi_p\|K\|_2^2\sum_a\sum_b\xi_a\xi_b\delta(v_a-v_b)f_{r(a)r(b)}(v_a)f_{s(a)s(b)}(-v_a)+\mathrm{o}(1),$$

and

$$\frac{p}{4nMH^2}\sum_a\sum_{j(a)}\sum_b\sum_{k(b)}\xi_a\xi_b\,\mathrm{Trace}[\Sigma\Theta^{a'}_{pj(a)}\Sigma\overline{\Theta^{b*}_{pk(b)}}]\left(\sum_{t=1}^{n-1}h_t^2\cos tp\lambda_{j(a)+k(b)}\right)^2$$

$$=\mathrm{o}(1),$$

since the frequencies covered by $K_M(v_a-\lambda_{pj(a)})K_M(v_b-\lambda_{pk(b)})$ are such that $j(a)+k(b)>(v_a+v_b)n/2\pi p-n/Mp>Cn$, $\bmod n$, if $\pi>v_a,v_b>0$ and n and M are big enough, so $H^{-2}(\sum h_t^2\cos tp\lambda_{j(a)+k(b)})^2=\mathrm{O}(|j(a)+k(b)|^{-2p})$ from Lemma A.4(A).

The second term (A.6) is $\mathrm{o_p}(1)$ because it has zero mean and variance equal to

$$2\sum_{t=2}^{n}h_t^2\sum_{u=2}^{n}h_u^2\sum_{s}^{\min\{t-1,u-1\}}\sum_{r\neq s}h_s^2h_r^2\,\mathrm{Trace}[\Lambda'_{t-s}\Sigma\overline{\Lambda_{t-r}}\Sigma(\Lambda'_{u-s}\Sigma\overline{\Lambda_{u-r}})^*]$$

$$=2\sum_{t=2}^{n}h_t^4\sum_s\sum_{r\neq s}h_s^2h_r^2\,\mathrm{Trace}[\Lambda'_{t-s}\Sigma\overline{\Lambda_{t-r}}\Sigma\Lambda'_{t-r}\Sigma\overline{\Lambda_{t-s}}] \tag{A.11}$$

$$+\,4\sum_{t=3}^{n}h_t^2\sum_{u=2}^{t-1}h_u^2\sum_{s}^{u-1}\sum_{r\neq s}^{u-1}h_s^2h_r^2\,\mathrm{Trace}\,[\Lambda'_{t-s}\Sigma\overline{\Lambda_{t-r}}\Sigma\Lambda'_{u-r}\Sigma\overline{\Lambda_{u-s}}], \tag{A.12}$$

since the weights $\{h_t\}$ are symmetric around $\lfloor n/2\rfloor$. By summation by parts we find that $\|\Lambda_t\|=\mathrm{O}(M^{1/2}n^{-1/2}t^{-1})$, $t<n/2$, $\|\Lambda_t\|=\|\Lambda_{n-t}\|$, so (A.11) is

$$\mathrm{O}\left(\sum_{t=2}^{n}\sum_s\sum_{r\neq s}\|\Lambda_{t-r}\|^2\|\Lambda_{t-s}\|^2\right)=\mathrm{O}\left(M^2n^{-2}\sum_{t=2}^{n}\sum_s\sum_{r\neq s}|t-r|^{-2}|t-s|^{-2}\right)$$

$$=\mathrm{O}(M^2n^{-1})=\mathrm{o}(1),$$

and, following Robinson (1995b, p. 1646), (A.2) has absolute value bounded by

$$C\sum_{t=3}^{n}\sum_{u=2}^{t-1}\left(\sum_{s}^{u-1}\|\Lambda_{t-r}\|^2\sum_{r\neq s}^{u-1}\|\Lambda_{u-r}\|^2\right)\leqslant C\left(\sum_{1}^{n}\|\Lambda_t\|^2\right)\left(\sum_{t=3}^{n}\sum_{u=2}^{t-1}\sum_{r=t-u+1}^{t-1}\|\Lambda_r\|^2\right),$$

since $\max_t |h_t| \leqslant 1$, and using the same arguments as in that reference the last bracketed factor is

$$\sum_{j=1}^{n-2} j(n-j-1)\|\Lambda_j\|^2 \leqslant 2n \sum_{1}^{[n/2]} j\|\Lambda_j\|^2 = \mathrm{O}(M \log n),$$

and $\sum_1^n \|\Lambda_t\|^2 = \mathrm{O}(M^2 n^{-1})$ so (A.6) is $\mathrm{O_p}([n^{-1}M^3 \log n]^{1/2}) = \mathrm{o_p}(1)$ with (7).

Finally, it remains to show that Lindeberg's conditions holds,

$$\sum_{1}^{n} E[z_t \overline{z}_t I(|z_t| > \rho)] = \sum_{1}^{n} E[|z_t|^2 I(|z_t| > \rho)] \to 0 \quad \text{for all } \rho > 0.$$

Proceeding as in Robinson (1995b), we check the sufficient condition $\sum_1^n E|z_t|^4 \to 0$ as $n \to \infty$. Following his arguments we have also in our case

$$\begin{aligned}\sum_{1}^{n} E|z_t|^4 &\leqslant C \sum_{1}^{n} \left(\sum_{1}^{n} \|\Lambda_{t-s}\|^4 \right) + C \sum_{1}^{n} \sum_{1}^{t-1} \sum_{1}^{t-1} \|\Lambda_{t-s}\|^2 \|\Lambda_{t-r}\|^2 \\ &= \mathrm{O}(n^{-1}M^2) = \mathrm{o}(1),\end{aligned}$$

using the previous bound for $\|\Lambda_{t-s}\|$, completing the proof of the theorem. □

Proof of Theorem 4. For $|v| \to 0$ as $n \to \infty$ we only stress the main differences with respect to the proof of Theorem 3. First notice that since $(|vM|)^{-1} \to 0$ as $n \to \infty$, we have that for all $\lambda_j \in [v - \pi/M, v + \pi/M]$:

$$\inf_{[v-\pi/M, v+\pi/M]} \lambda_j = v[1 + \mathrm{O}((vM)^{-1}] = v[1 + \mathrm{o}(1)] \sim v \quad \text{as } n \to \infty, \tag{A.13}$$

and $f_{rr}(\lambda_j) \sim f_{rr}(v)$. Then all results are valid if we normalize all the quantities by $[f_{rr}(v) f_{ss}(v)]^{1/2}$ since, using the differentiability of B_r and f_{rr}, for $\lambda_j \in [v - \pi/M, v + \pi/M]$,

$$\max_{\lambda_j} |f_{rr}^{-1}(v) - f_{rr}^{-1}(\lambda_j)| = \mathrm{O}(f_{rr}^{-1}(v)(|v|M)^{-1}) = \mathrm{o}(f_{rr}^{-1}(v)),$$

applying the mean value theorem.

Therefore, when no tapering is applied, the left-hand side of (A.1) is

$$\mathrm{O_p}([f_{rr}(v) f_{ss}(v)]^{1/2}[(n|v|)^{d_*-1} + (n|v|)^{-1/2}](\log n)^{1/2}) = \mathrm{o_p}([f_{rr}(v) f_{ss}(v)]^{1/2}),$$

as the summation is running for integers j between $vn/(2\pi p) \pm n/(2M)$. When tapering is applied, (A.2) is

$$\mathrm{O_p}([f_{rr}(v) f_{ss}(v)]^{1/2}[|nv|^{d_*-p}(\log n)^{1/2} + |nv|^{-1/2}]) = \mathrm{o_p}([f_{rr}(v) f_{ss}(v)]^{1/2}),$$

using (6) and (10). Finally, the left-hand side of (A.4) is

$$f_{rs}(v) + \mathrm{O}([f_{rr}(v) f_{ss}(v)]^{1/2}[n^{-1}M + (|v|M)^{-1}]) = f_{rs}(v) + \mathrm{o}([f_{rr}(v) f_{ss}(v)]^{1/2}),$$

while the bound for the variance of $\hat{f}_{rs}^{\epsilon,M}(v)$ follows as for fixed v due to the normalization by $f_{rs}(v)$.

For the analysis of the asymptotic distribution for positive $v \to 0$ as $n \to \infty$, we first obtain for $p = 1$ that

$$
\begin{aligned}
&[f_{rr}(v)f_{ss}(v)]^{-1/2}(nM^{-1})^{1/2}|\hat{f}_{rs}^{M}(v) - \hat{f}_{rs}^{\epsilon,M}(v)| \\
&= \mathrm{O_p}\left(\begin{array}{l} |nv|^{d_*-1}\log^{1/2} n + |nv|^{-1/2}\log^{1/2} n \\ \quad + (nM^{-1})^{1/2}[|nv|^{2(d_*-1)}\log n + |nv|^{-3/4} \\ \quad + |nv|^{(d_*-2)/2}\log^{3/2} n + |nv|^{d_*-5/4}\log^{1/2} n] \end{array}\right) \\
&= \mathrm{O_p}((nM^{-1})^{1/2}|nv|^{d_*-5/4}\log^{1/2} n) + \mathrm{o_p}(1) = \mathrm{o_p}(1),
\end{aligned}
$$

using (7), (11) and $d_* < \frac{3}{4}$. When we apply tapering, $p > 1$, we find that

$$
\begin{aligned}
&[f_{rr}(v)f_{ss}(v)]^{-1/2}(M^{-1}n)^{1/2}|\hat{f}_{rs}^{M}(v) - \hat{f}_{rs}^{\epsilon,M}(v)| \\
&= \mathrm{O_p}((nM^{-1})^{1/2}[|nv|^{-1/2} + |nv|^{d_*-p}\log^{1/2} n]) + \mathrm{o_p}(1),
\end{aligned}
$$

which is $\mathrm{o_p}(1)$ by (12). Finally, with Assumption 6:

$$
\begin{aligned}
E[\hat{f}_{rs}^{\epsilon,M}(v)] &= \frac{2\pi p}{n}\sum_j K_M(v - \lambda_{jp})f_{rs}(\lambda_{jp}) \\
&\quad + \mathrm{O}([f_{rr}(v)f_{ss}(v)]^{1/2}[n^{-1}M + n^{-1}|v|^{-1}]) \\
&= f_{rs}(v) + \mathrm{O}([f_{rr}(v)f_{ss}(v)]^{1/2}[n^{-1}M + |vM|^{-2}]),
\end{aligned}
$$

where the error term is $\mathrm{o}([f_{rr}(v)f_{ss}(v)]^{1/2}(Mn^{-1})^{1/2})$ using (7) and (11), and the theorem follows as for fixed v. □

Lemma A.1. *Under* (6) *for* $|v| > 0$ *or* (10) *for* $|v| \to 0$ *and* (i) *or* (ii) *of Theorem* 3, *for* $\lambda_{jp} \in [v - \pi(1/M), v + \pi(1/M)]$, $p \geqslant 1$,

$$
f_{rs}^{-1}(v)|I_{rs}(\lambda_{jp}) - B_r(\lambda_{jp})I^{\epsilon}(\lambda_{jp})B_s^*(\lambda_{jp})| = \mathrm{O_p}(e_{p,d_s}^{1/2} + e_{p,d_r}^{1/2}),
$$

where e_{p,d_r} *is the error term in the part* (a) *of Theorem* 1 *or* 2, *for each* p, *depending on the values of* d_r, $\beta = 1$.

Proof. We write for $|v| > 0$, following Robinson (1995b), proof of Theorem 1, suppressing in the notation the frequency λ_{jp},

$$
\begin{aligned}
I_{rs} - B_r I^{\epsilon} B_s^* &= w_r\bar{w}_s - B_r w^{\epsilon} w^{\epsilon *} B_s^* \\
&= \tfrac{1}{2}\{(w_r - B_r w^{\epsilon})(\bar{w}_s + w^{\epsilon *}B_s^*) + (w_r + B_r w^{\epsilon})(\bar{w}_s - w^{\epsilon *}B_s^*)\}.
\end{aligned}
$$

Then $E|I_{rs} - BI^{\epsilon}B_s^*| = \frac{1}{2}E[|(w_r - B_r w^{\epsilon})(\bar{w}_s + w^{\epsilon *}B_s^*)|] + \frac{1}{2}E[|(w_r + B_r w^{\epsilon})(\bar{w}_s - w^{\epsilon *}B_s^*)|]$, where

$$\begin{aligned}
&E[|(w_r - B_r w^{\epsilon})(\bar{w}_s + w^{\epsilon *}B_s^*)|] \\
&\quad \leqslant E[|w_r - B_r w^{\epsilon}||\bar{w}_s + w^{\epsilon *}B_s^*|] \\
&\quad \leqslant (EI_r - B_r E w^{\epsilon}\bar{w}_r - E w_r \bar{w}^{\epsilon}B_r^* + B_r E w^{\epsilon} w^{\epsilon *}B_r^*)^{1/2} \\
&\quad \times (EI_s + B_s E w^{\epsilon}\bar{w}_s + E w_s w^{\epsilon *}B_s^* + B_s E w^{\epsilon} w^{\epsilon *}B_s^*)^{1/2},
\end{aligned}$$

and since $B_r E[w^{\epsilon}w^{\epsilon *}]B_r = f_{rr}$, and from the proof of Theorem 1 in Velasco (1999a) it follows that $B_r E[w^{\epsilon}\bar{w}_r] = B_r(1/2\pi)\Sigma B_r^* + \mathrm{O}(\lambda_j^{-2d_r}e_{p,d_r}) = f_{rr} + \mathrm{O}(\lambda_j^{-2d_r}e_{p,d_r})$, $E[w_r\bar{w}^{\epsilon *}]$ $B_r^* = B_r(1/2\pi)\Sigma B_r^* + \mathrm{O}(\lambda_j^{-2d_r}e_{p,d_r}) = f_{rr} + \mathrm{O}(\lambda_j^{-2d_r}e_{p,d_r})$, this completes the proof for $|v| > 0$. When $|v| \to 0$ as $n \to \infty$ the argument follows using (A.3) because $(|v|M)^{-1} \to 0$ as $n \to \infty$ from (10). □

Lemma A.2. *Under the assumptions of Theorem* 3 ($|v| > 0$) *or Theorem* 4 ($|v| \to 0$), $p = 1$,

$$\frac{2\pi}{n}\sum_j K_M(v - \lambda_j)[I_{rs}(\lambda_j) - B_r(\lambda_j)I^{\epsilon}(\lambda_j)B_s^*(\lambda_j)] = \mathrm{o_p}((n/M)^{-1/2}). \tag{A.14}$$

Proof. Using the second moments of the periodogram as in the argument in p. 1648 of Robinson (1995b) and in Velasco (1999b), $d_a \in (-\frac{1}{2}, 1)$, we have with $K_{Mj} = K_M(v - \lambda_j)$ that the expectation of the square of the left-hand side of (A.14) is

$$\begin{aligned}
&\sum_j \left(\frac{2\pi K_{Mj}}{n}\right)^2 E[(w_r(\lambda_j)\overline{w_s}(\lambda_j) - B_r(\lambda_j)w^{\epsilon}(\lambda_j)w^{\epsilon *}(\lambda_j)B_s^*(\lambda_j))(w_r(\lambda_j)\overline{w_s}(\lambda_j) \\
&\quad - B_r(\lambda_j)w^{\epsilon}(\lambda_j)w^{\epsilon *}(\lambda_j)B_s^*(\lambda_j))^*] \\
&\quad + \sum_j \sum_{k \neq j} \frac{2\pi K_{Mj}}{n}\,\frac{2\pi K_{Mk}}{n} E[(w_r(\lambda j)\overline{w_s}(\lambda_j) - B_r(\lambda_j)w^{\epsilon}(\lambda_j)w^{\epsilon *}(\lambda_j) \\
&\quad \times B_s^*(\lambda_j))(w_r(\lambda_k)\overline{w_s}(\lambda_k) - B_r(\lambda_k)w^{\epsilon}(\lambda_k)w^{\epsilon *}(\lambda_k)B_s^*(\lambda_k))^*].
\end{aligned}$$

Then, using the same procedure as in Robinson (1995b, proof of Theorem 2), calculating the expectations in terms of the second moments and fourth cumulants, we obtain with Theorem 1 above and Lemma A.3, that the left-hand side of (A.14) is

$$\begin{aligned}
&\mathrm{O_p}\left(Mn^{-1}\left[\sum_j \{j^{-1} + j^{2(d_*-1)}\}\log n + \sum_j \sum_{k>j} \{j^{-2} + j^{4(d_*-1)}\}\log^2 n\right]^{1/2}\right) \\
&\quad + \mathrm{O_p}\left(Mn^{-1}\left[\sum_j \{j^{-3/2} + j^{3(d_*-1)} + n^{-1/2}(j^{-1} + j^{2(d_*-1)})\}\right]^{1/2} \log n\right)
\end{aligned}$$

$$+ \mathrm{O_p}\left(Mn^{-1} \left[\sum_j \sum_{k>j} \{ j^{-3/2} + (jk)^{2(d_*-1)} \log^2 n \right.\right.$$

$$\left.\left. + n^{-1} k^{d_*-1} \log^3 n + n^{-1/2}((jk)^{-1/2} + (jk)^{d_*-1}) \log n \} \right]^{1/2} \right)$$

$$= \mathrm{O_p}(M^{1/2} n^{-1/2} [n^{d_*-1} \log^{1/2} n + n^{-1/2} \log^{1/2} n$$

$$+ n^{2d_*-3/2} M^{-1/2} \log n + M^{-1/2} n^{-1/2} \log n$$

$$+ n^{-3/4} + (n^{3(d_*-1)/2} + n^{d_*-5/4} + n^{-3/4}) \log n$$

$$+ M^{-1/2} n^{-1/4} + M^{-1/2} n^{(d_*-1)/2} \log^{3/2} n$$

$$+ M^{-1/2} (n^{d_*-3/4} + n^{-1/4}) \log^{1/2} n]),$$

which is $\mathrm{o_p}(M^{1/2} n^{-1/2})$ with (8). □

Lemma A.3. *Under Assumptions* 1,2, 4′, 5′, p=1, d=max$\{d_a, d_b\} < 1$, j=1, 2 . . . , $n/2$,

$$E|I_{ab}(\lambda_j) - B_a(\lambda_j) I^{\epsilon}(\lambda_j) B_b^*(\lambda_j)|^2$$

$$= \mathrm{O}(f_{aa}(\lambda_j) f_{bb}(\lambda_j)[(j^{-1} + j^{2(d-1)}) \log n$$

$$+ j^{-3/2} + j^{3(d-1)} + n^{-1/2}(j^{-1} + j^{2(d-1)} \log n)]),$$

and for $j < k$,

$$E[I_{ab}(\lambda_j) - B_a(\lambda_j) I^{\epsilon}(\lambda_j) B_b^*(\lambda_j)] \overline{[I_{ab}(\lambda_k) - B_a(\lambda_k) I^{\epsilon}(\lambda_k) B_b^*(\lambda_k)]}$$

$$= \mathrm{O}\left(|f_{ab}(\lambda_j) f_{ab}(\lambda_k)| \begin{array}{c} [j^{-2} \log^2 n + j^{4(d-1)} \log^2 n + j^{-3/2} + (jk)^{2(d-1)} \log^2 n \\ + n^{-1} k^{d-1} \log^3 n + n^{-1/2}((jk)^{-1/2} + (jk)^{d-1}) \log n] \end{array} \right).$$

Proof. It follows from Lobato (1999) multivariate treatment, adapting Robinson (1995b, Theorem 2), and Velasco (1999b, Lemmas 1–3) for $d \geqslant 0.5$. □

The following lemma is Lemma 8 of Velasco and Robinson (2000) and Lemma 7 of Velasco (1999b).

Lemma A.4. *If the sequence* $\{h_j\}$ *is a data taper of order* p *as defined previously*,

(A) $(\sum_{t=1}^n h_t^2)^{-1} |\sum_{t=1}^n h_t^2 \cos t\lambda_j| = \mathrm{O}(|j|^{-p})$, $0 < |j| < n/2$;
(B) $\sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \cos^2 s\lambda_j = \frac{1}{4}(\sum_{t=1}^n h_t^2)^2 + \mathrm{O}(n^2 |j|^{-2p} + n)$, $0 < |j| < n/2$;
(C) $\sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \cos s\lambda_j = \frac{1}{2}(\sum_{t=1}^n h_t^2 \cos t\lambda_j)^2 + \mathrm{O}(n)$, $0 < |j| < n$.

Proof of Lemma 1. We do the proof for $\hat{f}^M_{yy}(\lambda_j)$, for the cross-spectral estimate it follows similarly. Denote $D_j = f_{yy}^{-1}(\lambda_j)[1 - |H_{xy}(\lambda_j)|^2]^{-1}$,

$$\max_{\ell \leqslant j \leqslant m} D_j |\hat{f}^M_{yy}(\lambda_j) - f_{yy}(\lambda_j)| \leqslant \max_{\ell \leqslant j \leqslant m} D_j |\hat{f}^M_{yy}(\lambda_j) - E[\hat{f}^M_{yy}(\lambda_j)]| \tag{A.15}$$

$$+ \max_{\ell \leqslant j \leqslant m} D_j |E[\hat{f}^M_{yy}(\lambda_j)] - f_{yy}(\lambda_j)|. \tag{A.16}$$

First (A.16) is

$$O\left(\max_{\ell \leqslant j \leqslant m} \lambda_j^{-2\alpha}\left[\frac{M}{n} + \lambda_j^{-2} M^{-2}\right]\right) = \mathrm{O}(n^{2\alpha-1}\ell^{-2\alpha}M + n^{2(1+\alpha)}\ell^{-2(1+\alpha)}M^{-2}).$$

Now we have that the right-hand side of (A.15) is, $\tau \geqslant 1$,

$$\max_{\ell \leqslant j \leqslant m} \left| [1 - |H_{xy}(\lambda_j)|^2]^{-1} \frac{2\pi}{n} \sum_r K_M(\lambda_r - \lambda_j) f_{yy}^{-1}(\lambda_j)[I_{yy}(\lambda_r) - E[I_{yy}(\lambda_r)]] \right|$$

$$\leqslant \left(\max_{\ell \leqslant j \leqslant m} [1 - |H_{xy}(\lambda_j)|^2]^{-\tau} \left| \frac{2\pi}{n} \sum_r \Xi^M_{r,j}[I_{yy}(\lambda_r) - E[I_{yy}(\lambda_r)]] \right|^\tau \right)^{1/\tau},$$

where $\Xi^M_{r,j} = K_M(\lambda_r - \lambda_j) f_{yy}^{-1}(\lambda_j)$, and $\max_j \max_r |\Xi_{r,j} f_{yy}(\lambda_r)| = \mathrm{O}(M)$ for r in the compact support of $K_M(\lambda_r - \lambda_j)$ for each λ_j. Thus, the supremum inside the brackets is less than or equal to

$$C \sum_{j=\ell}^{m} \lambda_j^{-2\alpha\tau} \left| \frac{2\pi}{n} \sum_r \Xi^M_{r,j}[I_{yy}(\lambda_r) - E[I_{yy}(\lambda_r)]] \right|^\tau$$

$$\leqslant C \left(\frac{M}{n}\right)^{\tau/2} n^{2\alpha\tau} \sum_{j=\ell}^{m} j^{-2\alpha\tau} \left| \frac{2\pi}{\sqrt{Mn}} \sum_r \Xi^M_{r,j}[I_{yy}(\lambda_r) - E[I_{yy}(\lambda_r)]] \right|^\tau$$

$$= \mathrm{O_p}\left(\left(\frac{M}{n}\right)^{\tau/2} n^{2\alpha\tau} \ell^{1-2\alpha\tau} \right),$$

taking expectations, $\tau > 1/(2\alpha)$, since the standardized quantity $(Mn)^{-1/2} \sum_r \Xi^M_{r,j}[I_{yy}(\lambda_r) - E[I_{yy}(\lambda_r)]]$ has zero mean and bounded variance and moments of any order for any j and $d < 0.75$, since higher order moments depend only on second-order properties, i.e. $f_{yy}(\lambda)$, by Gaussianity. □

References

Beran, J., 1994. Statistics for Long-Memory Processes. Chapman & Hall, London.

Brillinger, D.R., 1975. Time Series. Data Analysis and Theory. Holden-Day, San Francisco.

Chen, G., Abraham, B., Peiris, S., 1994. Lag window estimation of the degree of differencing in fractionally integrated time series models. J. Time Ser. Anal. 15, 473–487.
Dahlhaus, R., 1985. Asymptotic normality of spectral estimates. J. Multivariate Anal. 16, 412–431.
Dueker, M., Startz, R., 1998. Maximum-likelihood estimation of fractional cointegration with an application to U.S. and Canadian bond rates. Rev. Econom. Statist. 80, 420–426.
Engle, R., Granger, C., 1987. Cointegration and error correction: representation, estimation and testing Econometrica 55, 251–276.
Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. J. Time Ser. Anal. 4, 221–238.
Granger, C.W.J., 1981. Some properties of time series data and their use in econometric model specification. J. Econometrics 16, 121–130.
Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and its Application. Academic Press, New York.
Hannan, E.J., 1963. Regression for time series. In: Rosenblatt, M. (Ed.), Proceedings of the Symposium on Time Series Analysis. Wiley, New York, pp. 17–37.
Hannan, E.J., 1970. Multiple Time Series. Wiley, New York.
Hassler, U., 1993. Regression of spectral estimators with fractionally integrated time series. J. Time Ser. Anal. 14, 369–380.
Hassler, U., Marmol, F., Velasco, C., 2001. Residual log-periodogram inference for long run relationships. Preprint.
Hidalgo, F.J., 1996. Spectral analysis for bivariate long-memory time series. Econometric Theory 12, 773–792.
Hurvich, C.M., Ray, B.K., 1995. Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. J. Time Ser. Anal. 16, 17–42.
Künsch, H.R., 1987. Statistical aspects of self-similar processes. Proceedings of the First World Congress of the Bernoulli Society. VNU Science Press, Utrech, pp. 67–74.
Lobato, I N., 1997. Consistency of the averaged cross-periodogram in long memory series. J. Time Ser. Anal. 18, 137–155.
Lobato, I.N., 1999. A semiparametric two step estimator in a multivariate long memory model. J. Econometrics 90, 129–153.
Lobato, I N., Robinson, P.M., 1996. Averaged periodogram estimation of long memory. J. Econometrics 73, 303–324.
Lobato, I.N., Velasco, C., 2000. Long memory in stock market trading volume. J. Bus. Econom. Statist. 18, 410–427.
Marinucci, D., 2000. Spectral regression for cointegrated time series with long-memory innovations. J. Time Ser. Anal. 21, 685–705.
Reisen, V.A., 1994. Estimation of the fractional difference parameter in the ARIMA(p.d.q) model using the smoothed periodogram. J. Time Ser. Anal. 15, 335–350.
Robinson, P.M., 1986. On the errors-in-variables problem for time series. J. Multivariate Anal. 19, 240–250.
Robinson, P.M., 1994a. Time series with strong dependence. In: Sims, C.A. (Ed.), Advances in Econometrics. Sixth World Congress, Vol. I. Cambridge University Press, Cambridge, pp. 47–95.
Robinson, P M., 1994b. Semiparametric analysis of long-memory time series. Ann. Statist. 22, 515–539.
Robinson, P.M., 1995a. Log-periodogram regression of time series with long range dependence. Ann. Statist. 23, 1048–1072.
Robinson, P.M., 1995b. Gaussian semiparametric estimation of long range dependence. Ann. Statist. 23, 1630–1661.
Robinson, P.M., Hidalgo, J., 1998. Time series regression with long range dependence. Ann. Statist. 25, 77–104.
Robinson, P.M., Marinucci, D., 2001. Narrow-band analysis of nonstationary processes. Ann. Statist. 29, 947–986.
Solo, V., 1992. Intrinsic random functions and the paradox of 1/f noise. SIAM J. Appl. Math. 52, 270–291.
Tanaka, K., 1999. The nonstationary fractional unit root. Econometric Theory 15, 549–582.
Tukey, J.W., 1967. An introduction to the calculation of numerical spectrum analysis. In: Harris, B. (Ed.), Advanced Seminar on Spectral Analysis of Time Series. Wiley, New York, pp. 25–46.
Velasco, C., 1999a. Non-stationary log-periodogram regression. J. Econometrics 91, 325–371.

Velasco, C., 1999b. Gaussian semiparametric estimation for non-stationary time series. J. Time Ser. Anal. 20, 87–127.

Velasco, C., 2001. Gaussian semiparametric estimation of fractional cointegration. Preprint.

Velasco, C., Robinson, P.M., 2000. Whittle pseudo-maximum likelihood estimation of non-stationary fractional time series. J. Amer. Statist. Assoc. 95, 1229–1243.

Zhurbenko, I.G., 1979. On the efficiency of estimates of a spectral density. Scand. J. Statist. 6, 49–56.