Nonparametric Frequency Domain Analysis of Nonstationary Multivariate Time Series

Carlos Velasco*
Universidad Carlos III de Madrid

September 23, 1999

Abstract

We analyze the properties of standard 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 can be consistent for observations with any degree of nonstationarity, including polynomial trends. Then we propose an estimate of the degree of fractional cointegration for possibly nonstationary series based on coherence estimates around zero frequency and analyze its finite sample properties in comparison with residual based inference.

Key words: Coherence, semiparametric estimation, long memory, fractional cointegration.

1 Introduction

In many empirical studies it is found that spectral density estimates of observed time series display a peak at the origin. 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 make difficult the interpretation of such features, well-documented otherwise, since the nonstationarity may affect the properties of spectral estimates.

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,

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

where $d < \frac{1}{2}$ is the parameter that governs the degree of memory of the series, but it 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*

*Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Calle Madrid 126, E-28903 Getafe (Madrid), Spain. E-mail: `cavelas@est-econ.uc3m.es`. I am grateful to U. Hassler for helpful comments on an earlier version. Research funded by the Spanish Dirección General de Enseñanza Superior, Ref. no. PB98-0025.

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. These properties can be expressed equivalently in the time domain in terms of the autocovariance sequence. 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 $d \geq \frac{1}{2}$. Then it is possible to consider the transfer function of the difference operator to define a generalized or pseudo spectral density with power law behaviour at the origin as in (1), but adding to d the number of integer differences taken to achieve stationarity. This function $f(\lambda)$, though with similar shape to the spectral density of the stationary increments for frequencies away from the origin, is not integrable and can not 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 is well-known for stationary series. We show in this paper that this generalized spectral density is the quantity actually estimated in practice by smoothed spectral estimates, as has already been analyzed for full and semi-parametric estimates of long memory models without assumptions about the degree of possible nonstationarity (see Velasco and Robinson (1999) or Velasco (1999a, b)).

We analyze in a multivariate context 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. For estimates based on discrete averages of the periodogram ordinates closest to the frequency of interest, 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. Hidalgo (1996) analyzed the properties of spectral estimators based on autocovariances for stationary bivariate long-memory processes under time domain 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 too high, tapering the data (Tukey (1967)) might be necessary to reduce the bias in the estimation or to eliminate stochastic and deterministic trends, confirming the desirable resolution properties of such technique for stationary series found by e.g. Robinson (1986), Zhurbenko (1979) and Dahlhaus (1985).

For the analysis of multivariate long memory 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 the reduction of the memory. 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 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 relationship, mostly through residuals in an estimated regression model. Here we use nonparametric estimates of the coherence between two series at frequencies close to the origin to propose narrow band estimates of the order of cointegration in the spirit of Robinson and Marinucci's (1998) slope estimates. We discuss inference based of such estimates and compare its finite sample performance with residual-based semiparametric estimates.

The paper is organized as follows. We first present the main definitions for nonstationary long memory time series and data taper sequences and introduce the technical assumptions used throughout. In Section 3 we define the nonparametric estimates of the (pseudo) spectral density 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 analyzed and it is applied to economic time series. In Appendix A we review some asymptotics for the discrete (tapered) Fourier transform for multivariate long memory time series, possibly nonstationary, while the proofs of the main results are given in Appendix B.

2 Nonstationary time series and data tapers

Following Hurvich and Ray (1995) in a univariate context, we propose a general model for possibly nonstationary 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, $-\frac{1}{2} \leq d_r - D_r < \frac{1}{2}$. Here $\Delta = 1 - L$, where L is the lag operator. However using partial sums of stationary long memory processes as we do here is not the only possibility to define long memory or fractional nonstationary models (see e.g. Robinson and Marinucci (1998)).

Define the (pseudo) spectral density (PSD) of X_{rt} as

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

$0 < G_{rr} < \infty$. When $2d_r \geq 1$, $f_{rr}(\lambda)$ 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. Set the zero mean stationary process $U_{rt}^{(*)} = U_{rt} - \mu_r$. Note that if $\mu_r \neq 0$ the observed time series has a deterministic component and if $D_r \geq 1$ this is a polynomial trend. Similarly we can define the (pseudo) cross-spectral density of a pair of series (X_{rt}, X_{st}) as

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

$0 \leq |G_{rs}| < \infty$, where g_{rs} 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.

Define the tapered discrete Fourier transform (DFT) of X_{rt} and a deterministic taper sequence h_t, for n observations $t = 1, \ldots, n$ as $r = 1, \ldots, R$, $(\lambda_j = 2\pi j/n)$,

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

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, helping to control leakage problems in spectral estimation when nonstationarity is suspected in the observed time series.

The usual DFT $w_r(\lambda_j)$ 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 $n = 4N$, N integer, the weights of the Parzen window are given by

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

Zhurbenko (1979) used a class of data tapers $\{h_t^Z\}$ 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 (see Alekseev (1996) for more examples). The raw DFT weights, the identity, are of order $p = 1$, and from now on when we say that $p = 1$ we will imply the usual DFT, without tapering. 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.

As discussed in Velasco (1999a), 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. 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 of the deterministic trends (see Lobato and Velasco (1999) for an empirical application of this property). Thus summation by parts yields for a differentiable taper which vanishes at the boundaries, with derivative h_t',

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

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

$$E[I_{rr}^p(\lambda_{jp})] = \left(2\pi \sum_{t=1}^{n} h_t^2\right)^{-1} \int_{-\pi}^{\pi} |D_h(\lambda - \lambda_{jp})|^2 f_{rr}(\lambda) d\lambda \to f_{rr}(\lambda_{jp}), \tag{3}$$

as $n \to \infty$, where $D_h(\lambda) = \sum_{t=1}^n h_t \exp\{i\lambda t\}$. Then the tapered periodogram can be asymptotically unbiased for the PSD f_{rr} of nonstationary series at Fourier frequencies λ_{jp}, $j \neq 0 (\operatorname{mod} N)$, not too close to the origin, though with increased correlation between adjacent ordinates (see Theorems 3 and 4 in Appendix A). Furthermore, for a data taper of order p

$$w(t^\ell, h_t, \lambda_{jp}) = 0, \qquad \ell = 0, 1, \ldots, p-1, \tag{4}$$

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

We review in Appendix A in a multivariate context some results obtained in Robinson (1995a), Velasco (1999a) for the (tapered) discrete Fourier transform of possibly nonstationary time series as defined before, extending some results of Robinson (1995a) for nontapered stationary observations. Here we are only concerned with Fourier frequencies $0 < \lambda_j < n/2$, since we can analyze negative ones by complex conjugation and symmetry. These results assume the following conditions concerning the behaviour of $g_{rr}(\lambda)$ for frequencies around the frequency of interest $0 \leq \nu < \pi$, such that $\lambda \to \nu$ as $n \to \infty$. The case $\nu = 0$ is of interest in order to analyze the persistence properties of the observed vector series, including its possible cointegration properties. These conditions have obvious implications on the PSD f_{rr} for nonstationary series.

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

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

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

Assumption 2 *[$\boldsymbol{\nu} = \mathbf{0}$] $g_{rs}(\lambda)$, $r, s = 1, \ldots, R$, is differentiable in an interval of the origin $(0, \varepsilon)$, $\varepsilon > 0$, and*

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

[$\boldsymbol{\nu} \neq \mathbf{0}$] $|g_{rs}(\lambda)| > 0$, $r, s = 1, \ldots, R$, and $g_{rs}(\lambda)$ is boundedly differentiable in an interval $(\nu - \varepsilon, \nu + \varepsilon)$, $\varepsilon > 0$.

Assumption 3 *[$\boldsymbol{\nu} = \mathbf{0}$] $g_{rs}(\lambda)$, $r, s = 1, \ldots, R$, is twice differentiable in an interval of the origin (0,ε), $\varepsilon > 0$, with*

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

[$\boldsymbol{\nu} \neq \mathbf{0}$] *$g_{rs}(\lambda)$, $r, s = 1, \ldots, R$, is twice boundedly differentiable in an interval $(\nu - \varepsilon, \nu + \varepsilon)$, $\varepsilon > 0$.*

The first assumption deals with the possible long memory or nonstationarity of each univariate observed series, while Assumptions 2 and 3 impose some smoothness on the spectral density g_{rs} around the frequency of interest. Only the first two are needed for the analysis of the covariance of the usual DFT in Theorem 3 while Assumption 3 is used to control smoothing bias and to fully use tapering properties when $p > 1$ in Theorem 4 in Appendix A. Similar conditions are used for parametric (e.g. Fox and Taqqu (1986)) and semiparametric inference on long memory processes, imposing only local conditions around the frequency of interest. This allows for the possibility of PSDs with (integrable) poles or zeroes at remote frequencies.

We introduce the following linear process assumptions about the (differenced) stationary zero mean series $\mathbf{U}_t^{(\star)} = \mathbf{U}_t - \boldsymbol{\mu}$, $\mathbf{U}_t = (U_{1t}, \ldots, U_{Rt})'$. They were introduced by Robinson (1995b) and Lobato (1997, 1999) to analyze semiparametric estimates of d for stationary long memory processes. They do not restrict the form of $f_U(\lambda) = g(\lambda)$ in any way and are only restrictive in the linearity they impose.

Assumption 4 $\mathbf{U}_t = \boldsymbol{\mu} + \sum_{j=0}^{\infty} \mathbf{A}_j \boldsymbol{\epsilon}_{t-j}$ *with* $\sum_{j=0}^{\infty} \|\mathbf{A}_j\|^2 < \infty$, *where $\|\cdot\|$ denotes the supremum norm and $\boldsymbol{\epsilon}_t$ satisfies a.s. $E(\boldsymbol{\epsilon}_t |\Im_{t-1}) = 0$; $E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' |\Im_{t-1}) = \boldsymbol{\Sigma}$, $\Sigma_{rr} = 1$; $E(\epsilon_a(t)\epsilon_b(t)\epsilon_c(t) |\Im_{t-1}) = \mu_{abc}$ with $|\mu_{abc}| < \infty$ for $a, b, c = 1, ..., R$; $E(\epsilon_a(t)\epsilon_b(t)\epsilon_c(t)\epsilon_d(t) |\Im_{t-1}) = \mu_{abcd}$, where $|\mu_{abcd}| < \infty$ for $a, b, c, d = 1, ..., R$ where $\Im_{t-1}$ is the σ-field of events generated by $\{\boldsymbol{\epsilon}_s, s \leq t-1\}$.*

Note that the variance of the components of $\boldsymbol{\epsilon}_t$ is set to one for identifiability. 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 $\boldsymbol{\epsilon}_t$ is $\mathbf{f}_\epsilon(\lambda) = (2\pi)^{-1}\boldsymbol{\Sigma}$, so $\mathbf{g}(\lambda) = (2\pi)^{-1}\mathbf{A}(\lambda)\boldsymbol{\Sigma}\mathbf{A}^*(\lambda)$, and each element is

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

where $*$ stands for simultaneous transposition and complex conjugation. Denote $\mathbf{B}_r(\lambda) = \left(1 - e^{i\lambda}\right)^{-D_r} \mathbf{A}_r(\lambda)$, $r = 1, \ldots, R$, so $f_{rs}(\lambda) = (2\pi)^{-1}\mathbf{B}_r(\lambda)\boldsymbol{\Sigma}\mathbf{B}_s^*(\lambda)$. We also impose the following condition on the components of $\mathbf{A}$ (equivalently $\mathbf{B}$) around the frequency of interest, which implies Assumption 2.

Assumption 5 *[$\boldsymbol{\nu} = \mathbf{0}$] $\mathbf{A}_r(\lambda)$ is differentiable in an interval $(0, \varepsilon)$, $\varepsilon > 0$, $r = 1, \ldots, R$, and*

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

[$\boldsymbol{\nu} \neq \mathbf{0}$] $\mathbf{A}_{rs}(\lambda)$ is continuously differentiable in an interval $(\nu - \varepsilon, \nu + \varepsilon)$, $\varepsilon > 0$, $r, s = 1, \ldots, R$.

3 Nonparametric Estimates of the Pseudo-Spectral Density Matrix

We analyze in this section the properties of traditional nonparametric kernel spectral estimates at frequencies ν, $0 < \nu < \pi$, fixed in the asymptotics, and at frequencies in a degenerating band around the origin with $\nu \to 0$ as the sample size increases, as e.g. $\nu = \lambda_j$ 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 ν:

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

where $K_M(x) = M\,K(Mx)$ 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, $j = v - \lfloor \frac{n}{2Mp} \rfloor, v - \lfloor \frac{n}{2Mp} \rfloor + 1, \ldots, v-1, v, v+1, \ldots, v + \lfloor \frac{n}{2Mp} \rfloor$, if the support of K is exactly $[-\pi, \pi]$, where λ_{pv} is the closest frequency to ν for each n, v integer). Appendix A analyzes the asymptotic behaviour of the periodogram ordinates of nonstationary observations at these frequencies under Assumptions 1 to 3.

When the series is not stationary, the frequency domain estimates $\widehat{f}^M$ 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.

Assumption 6 *The function K is even, has compact support inside $[-\pi, \pi]$, satisfies a Lipschitz condition and*

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

This condition is standard in nonparametric kernel estimation and it 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 of the estimates of $f_{rs}(\nu)$.

Theorem 1 *Under Assumptions 1, 2, 4, 5, 6, $p \geq \max\{D_r, D_s\} + 1$ [only $p > \max\{d_r, d_s\}$ if $\mu_r = \mu_s = 0$], and*

$$Mn^{-1} + M^{-1} \to 0 \qquad as \quad n \to \infty, \tag{5}$$

- *For $|\nu| > 0$ as $n \to \infty$,*

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

- *For $|\nu| \to 0$ and*

$$(|\nu|M)^{-1} + (n|\nu|)^{2(d-p)} \log n + (n|\nu|)^{-1} \log n \to 0 \qquad as \quad n \to \infty \tag{6}$$

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

Proof of Theorem 1. See Appendix B. •

A condition like (5) is also minimal for nonparametric estimates $\widehat{f}_{rs}^M(\nu)$ of smooth spectral densities. Note that tapering allows the consistent estimation of f with trending observations without need of initial detrending, and that without any kind of tapering it is possible to estimate the function f for nonstationary but *mean reverting* processes ($d < 1$). The condition $(|\nu|M)^{-1} \to 0$ as $n \to \infty$ in (6) is needed for spectral estimation at frequencies in a degenerating band around zero frequency, to avoid periodogram ordinates too close to the singularity of the PSD at the origin. The second condition in (6) has the same purpouse when the degree of nonstationary d is close to the degree of tapering applied, p. When $\nu = \lambda_j$ then j may grow as n^{ε} for any $\varepsilon > 0$.

The next theorem gives the asymptotic distribution of the nonparametric estimates. To centre them in the actual value of $f_{rs}(\nu)$, we need to undersmooth the nonparametric estimates in order to obtain a bias of smaller order of magnitude, $O(M^{-2})$, than the standard deviation, $O((M/n)^{1/2})$. We assume a quite smooth PSD matrix here, with Assumption 3 for bias control. We recall that for complex quantities, the covariance is defined conjugating the second term, 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 m-dimensional complex normal distribution, denoted as N_m^C), m fixed. Set $d_* = \max_{r=1,\dots,R} d_r$ and for $r_i, s_j \in \{1,\dots,R\}$,

$$F_{\mathbf{rs}}(\boldsymbol{\nu}) = \left(f_{r_1 s_1}(\nu_1), \dots, f_{r_m s_m}(\nu_m)\right)'.$$

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

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

Theorem 2 *Under Assumptions 1, 2, 3, 4, 5, 6 with*

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

- *if* $p = 1$, $\boldsymbol{\mu} = 0$ *and*

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

  (so $d_* < \frac{3}{4}$ *is sufficient for any choice of* M*).*

- *If* $p > 1$ *(any* $\boldsymbol{\mu}$*), we need* $p \geq \lfloor d_* + \frac{1}{2}\rfloor + 1$, *[and if* $\boldsymbol{\mu} = 0$ *only* $p > d_*$ *and*

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

For $\nu_i \in (-\pi,\pi) - \{0\}$, $i = 1,\dots,m$, *fixed with* n,

$$\sqrt{\frac{n}{M}}\left\{\widehat{F}_{\mathbf{rs}}^M(\boldsymbol{\nu}) - F_{\mathbf{rs}}(\boldsymbol{\nu})\right\} \to_d N_m^C\left(0, 2\pi p\Phi_p\|K\|^2\Omega(\boldsymbol{\nu})\right), \qquad as \quad n\to\infty,$$

where $\Omega(\boldsymbol{\nu}) = [\sigma_{ij}(\boldsymbol{\nu})]$, $\sigma_{ij}(\boldsymbol{\nu}) = \delta(\nu_i - \nu_j) f_{r_i r_j}(\nu_i) f_{s_i s_j}(-\nu_i) + \delta(\nu_i + \nu_j) f_{r_i s_j}(\nu_i) f_{r_j s_i}(\nu_i)$, $i,j \in \{1,\dots,R\}$.

For $|\nu| \to 0$ *as* $n\to\infty$ *if additionally*

$$M^{-5}|\nu|^{-4} n \to 0 \qquad as\ n\to\infty, \tag{10}$$

- *if* $p = 1$, $d_* < \frac{3}{4}$;
- *if* $p > 1$,

$$M^{-1} n (n|\nu|)^{2(d_*-p)}\log n \to 0 \qquad as\ n\to\infty, \tag{11}$$

  (so $p > d_* + \frac{1}{2}$ *is enough for all* M*);*

we obtain

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

Proof of Theorem 2. See Appendix B. •

We decided for simplicity not to provide the result for $\nu_i = \pm\pi$ (but the standard result holds) and to consider Assumption 3 and a univariate central limit theorem for degenerating $|\nu| \to 0$. Note that (7) and (10) imply (6) in Theorem 1. Condition (10) further restricts the range of low frequencies ν for which we can obtain the asymptotic normality of $\widehat{f}^M(\nu)$, though if $\nu = \lambda_j$ 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 in the same environment by Velasco (1999a, b) and by Velasco and Robinson (1999) for parametric estimates. When $p = 1$ and all ν_i are fixed, it is possible to find sequences M which lead to asymptotically normal estimates if $d_* < \frac{5}{6}$. Some weaker conditions on the smoothing bandwidth M would be sufficient for the result if we substitute $F_{\mathbf{rs}}(\boldsymbol{\nu})$ by $E[\widehat{F}_{\mathbf{rs}}(\boldsymbol{\nu})]$ or if we employ higher order kernels or estimate higher order bias terms.

The taper variance inflation factor Φ_p is smaller than 1.05 for Zhurbenko kernels with $p > 1$ ($\Phi_1 \equiv 1$), implying increments in the asymptotic variance of the estimates of less than 5% (apart from the p factor due to the reduced number of frequencies used in $\widehat{f}^M$). Note that $\Phi_p = \Upsilon_p$, where Υ_p is the usual tapering variance correction (see e.g. Dahlhaus (1985)) if the sums 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. This approach can be applied to the cosine bell when $-\frac{1}{2} < d_* < \frac{3}{2}$, $\boldsymbol{\mu} = 0$, considering all the possible frequencies in the summation, like if $p = 1$ (with $|j| > 1$) and no extra sampling of frequencies, obtaining exactly $\Phi_c = \Upsilon_c = \frac{35}{18}$ (see the discussion in Velasco (1999b)). Also notice that spectral estimates at different frequencies are asymptotically independent for $|\nu_h| > 0$ as in the weak dependence case with bounded spectral density.

Finally note that the asymptotic variance changes the sign of the frequency when $\nu_i = -\nu_j$ in $f_{r_j s_i}(\nu_i)$ with respect to Brillinger's (1975, Theorem and Corollary 7.4.3), otherwise when $\nu_i = \nu_j$ and considering $\widehat{f}_{r_j s_j}(-\nu_1) = \widehat{f}_{s_j r_j}(\nu_1)$ we would obtain a contradiction. However this is correctly stated in his equation (7.2.14), but there is a typo in the second line of (7.2.13). Thus all the standard results for linear and nonlinear functions of the PSD (real and imaginary parts, moduli, coherency, phase, transfer function) can be deduced from Theorem 2. See e.g. Hannan (1970, Section V.5).

These spectral estimates can be used for long memory estimation (e.g Hassler (1993), Chen et al. (1994) and Reisen (1994)) and for efficient Hannan's (1963) regression for long memory and nonstationary series in the spirit of Robinson and Hidalgo (1998), apart from nonparametric descriptive analysis. In the next section we take this further and analyze the behaviour of coherence measures for cointegrated series, estimating a semiparametric model for them.

4 Spectral analysis of cointegrated time series

Let the observable bivariate time series (Y_t, X_t) be $I(d)$ (i.e. $d_Y = d_X = d > 0$ in Assumption 1), satisfy

$$Y_t = bX_t + Z_t, \tag{12}$$

for some $b \neq 0$, where the cointegrating error Z_t is $I(d - \alpha)$, $0 < \alpha \leq d$, and may be correlated with X_t at some frequencies (all $\mu = 0$). We strengthen Assumption 1 and suppose that the PSD matrix $\mathbf{f}$ of X and Z satisfies

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

for some constants $0 < |G_{ab}| < \infty$, $a, b \in \{x, z\}$, and the matrix $\mathbf{G} = \{G_{ab}\}$ is hermitian and nonsingular. Define the (squared) coherence $H_{ab}(\lambda)$ between two time series a_t and b_t at frequency λ as

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

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

Then, employing model (12), and pretending that the series are stationary to calculate the covariances (otherwise, integer difference (12) a sufficient number of times and then multiply by the unit root filters) the PSD of Y is

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

and the cross PSD of X and Y satisfies

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

Therefore X and Y have coherence equal to one at zero frequency, $H_{xy}(0) = 1$ and the PSD matrix of (X, Y) 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 frequencies as in the stationary case, independently of whether the PSDs are unbounded or zero at some frequencies.

After straightforward manipulations using the model (12) we can write that

$$|H_{xy}(\lambda)|^2 = 1 - \frac{f_{zz}(\lambda)}{f_{yy}(\lambda)} + \frac{|f_{zx}(\lambda)|^2}{f_{yy}(\lambda)f_{xx}(\lambda)}, \tag{13}$$

Substituting in (13) the approximations of f_{zx} and f_{yy} as $\lambda \to 0^+$,

$$\begin{aligned} |H_{xy}(\lambda)|^2 &= 1 - \left(\frac{G_{zz}}{G_{xx}} - \frac{|G_{zx}|^2}{G_{xx}^2}\right)\left(1 - 2\lambda^{\alpha}\frac{ReG_{zx}}{G_{xx}}\right)\lambda^{2\alpha} + O(\lambda^{4\alpha} + \lambda^{2+2\alpha}) \\ &\sim 1 - G_H\lambda^{2\alpha} \end{aligned}$$

for a real constant $0 < G_H < \infty$,

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

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

$$\log(1 - |H_{xy}(\lambda)|^2) \sim \log G_H + 2\alpha \log \lambda \quad \text{as } \lambda \to 0^+, \tag{14}$$

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 α the worse is the above approximation for $|H_{xy}(\lambda)|$ based on the leading terms of the expansion, $1 - G_H\lambda^{2\alpha}$, but in this case also estimates of b have slower rates of convergence (see e.g. Robinson and Marinucci (1998, p. 15)). This approach is valid for both stationary and nonstationary series (tapering might be used to eliminate some intercept or polynomial trend in (12) or to cover very nonstationary situations, $d \geq 1$) and it is neither affected asymptotically by the endogeneity of the residuals ($H_{zx}(\lambda) \neq 0$) because of its semiparametric nature. However if X and Z are incoherent at zero frequency, $H_{zx}(0) = 0$, so $G_{zx} = 0$ and $G_H = G_{zz}G_{xx}^{-1}$, then $|H_{xy}(\lambda)|^2 = 1 - G_H\lambda^{2\alpha} + O(\lambda^{4\alpha})$ reducing the bias in the semiparametric model. 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 $\mathbf{H}(\lambda)$.

We take (14) as the basis to estimate α. Denote by $\widehat{\alpha}$ the least squares estimate of α based on the regression of $\log(1 - |\widehat{H}_{xy}(\lambda_j)|^2)$ on $W_j = 2\log\lambda_j$, for frequencies λ_j, $j = \ell, \ldots, m$,

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

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

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

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 have not very desirable asymptotic properties. Asymptotic properties of $\widehat{\alpha}$ are complicated with respect to the log-periodogram regression estimate of d due to the nonlinear and nonparametric nature of sample coherences $|\widehat{H}_{xy}(\lambda_j)|^2$. We show first the consistency of $\widehat{\alpha}$ under conditions similar to those of Theorem 1 and then approximate its distribution for large samples.

We can write

$$\widehat{\alpha} - \alpha = \left(\sum_j \widetilde{W}_j^2\right)^{-1} \sum_j \widetilde{W}_j \log \frac{1 - |\widehat{H}_{xy}(\lambda_j)|^2}{1 - |H_{xy}(\lambda_j)|^2} + \Lambda,$$

where the bias term Λ is as in Robinson (1995a) $\Lambda = O\left(m^{-1}\sum_j |\widetilde{W}_j|\lambda_j^\alpha\right) = O\left(\left(\frac{m}{n}\right)^\alpha\right)$. We can obtain then

$$\widehat{\alpha} - \alpha = \left(\sum_j \widetilde{W}_j^2\right)^{-1} \sum_j \widetilde{W}_j \widehat{S}_j + \left(\sum_j \widetilde{W}_j^2\right)^{-1} \sum_j \widetilde{W}_j \widehat{C}_j + O\left(\left(\frac{m}{n}\right)^\alpha\right)$$

where $\widehat{S}_j = \widehat{A}_j - \widehat{B}_j$,

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

and

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

We now analyze the uniformity properties of the spectral estimates included in $\widehat{\alpha}$ under some stronger conditions than before.

Lemma 1 *Under Assumptions 1, 2, 4, 5, 6, ($p = 1$, $\mu = 0$) for Gaussian X_t, Y_t, $\alpha < \frac{1}{2}$, $d < \frac{3}{4}$, $M^{-1} + Mn^{-1} + (m - \ell)^{-1} + mn^{-1} \to 0$ as $n \to \infty$, and if for some $\tau \geq 1$,*

$$\left(n^{2\alpha-1}\ell^{-2\alpha}M + n^{2(1+\alpha)}\ell^{-2(1+\alpha)}M^{-2}\right)\log n + M^{\tau/2}n^{\tau(2\alpha-0.5)}\ell^{1-2\alpha\tau}\log^\tau n \to 0, \tag{15}$$

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

$$\max_{\ell \leq j \leq m} (1 - |H_{xy}(\lambda_j)|^2)^{-1} \left(f_{aa}(\lambda_j)f_{bb}(\lambda_j)\right)^{-1/2} \left|\widehat{f}^M_{ab}(\lambda_j) - f_{ab}(\lambda_j)\right| = o_p(\log^{-1} n). \tag{16}$$

We impose $d < \frac{3}{4}$ and Gaussianity to simplify proofs and avoid conditions on the moments of the linear innovations. Note that under (15), the condition for the consistency of $\widehat{f}^M_{ab}(\nu)$ (6) holds for $\nu = \lambda_j$, $\ell \leq j \leq m$. 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 used for the log-coherence regression, though in practical applications this may not be needed as long as zero frequency periodograms are avoided. Thus if $\ell \sim An^a$, $m \sim Bn^b$, $M \sim Cn^c$, $0 < a \leq b < 1$, $0 < c < 1$, (15) holds if

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

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

Then under the assumptions of Lemma 1, $\widehat{\alpha} - \alpha = o_p(1)$ as $n \to \infty$, because we obtain that $\max_j |\widetilde{W}_j| = O(\log n)$ and $\max_j \left|\widehat{A}_j - 1\right| = o_p(\log^{-1} n)$, $\max_j \left|\widehat{B}_j - 1\right| = o_p(\log^{-1} n)$ by (16), which imply that $\max_j \left|\widehat{S}_j\right| = o_p(\log^{-1} n)$ and $\max_j \left|\widehat{C}_j\right| \leq \max_j |\widehat{A}_j - 1|^2 + \max_j |\widehat{B}_j - 1|^2 = o_p(\log^{-1} n)$.

Now we proceed heuristically. Since the spectral estimates are approximately independent if they include periodogram ordinates at not 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 $\mathrm{Var}\left[\log(1 - |\widehat{H}_{xy}(\lambda_j)|^2)\right]$ by $4\mathrm{Var}\left[\tanh^{-1}(|\widehat{H}_{xy}(\lambda_j)|)\right]$ using that $\log(1 - x^2) + 2\tanh^{-1}(x) = \log(1 + x) \to 2\log 2$ as $x \to 1^-$, so

$$\mathrm{Var}[\widehat{\alpha}] \approx \left(\sum \widetilde{W}_j^2\right)^{-2} 4 \sum_j \sum_k \widetilde{W}_j \widetilde{W}_k \,\mathrm{Cov}\left[\tanh^{-1}(|\widehat{H}_{xy}(\lambda_j)|), \tanh^{-1}(|\widehat{H}_{xy}(\lambda_k)|)\right]. \tag{17}$$

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

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

We can also approximate for $|t| \leq 2q$

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

and if the estimates $\widehat{H}_{xy}$ are evaluated at frequencies sufficiently far apart we suppose they are asymptotically uncorrelated. Plugging the last two expressions in (17) we can estimate the sampling variance of $\widehat{\alpha}$ for each m and q. For tapered series this approximation to the variance can be adjusted by Φ_p and p as for $\widehat{f}$.

We can also justify those variance estimates using the previous approximations of $\widehat{\alpha}$. Thus

$$\begin{aligned}
&\mathrm{Var}\left[\widehat{S}_j\right] \left\{f_{xx}(\lambda_j) f_{yy}(\lambda_j)(1 - |H_{xy}(\lambda_j)|^2)\right\}^2 \\
&= |H_{xy}(\lambda_j)|^4 \mathrm{Var}\left[\widehat{f}_{xx}^M(\lambda_j)\widehat{f}_{yy}^M(\lambda_j)\right] - 2|H_{xy}(\lambda_j)|^2 \mathrm{Cov}\left[\widehat{f}_{xx}^M(\lambda_j)\widehat{f}_{yy}^M(\lambda_j), |\widehat{f}_{xy}^M(\lambda_j)|^2\right] + \mathrm{Var}\left[|\widehat{f}_{xy}^M(\lambda_j)|^2\right],
\end{aligned}$$

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

$$\begin{aligned}
&|H_{xy}(\lambda_j)|^4 \left\{2 f_{xx}(\lambda_j) f_{yy}(\lambda_j) |f_{xy}(\lambda_j)|^2 + 8 f_{xx}^2(\lambda_j) f_{yy}^2(\lambda_j)\right\} - 20 |H_{xy}(\lambda_j)|^2 f_{xx}(\lambda_j) f_{yy}(\lambda_j) |f_{xy}(\lambda_j)|^2 \\
&\quad + 2 f_{xx}(\lambda_j) f_{yy}(\lambda_j) |f_{xy}(\lambda_j)|^2 + 8 |f_{xy}(\lambda_j)|^4,
\end{aligned}$$

under appropriate conditions on the higher order cumulants of ϵ_t or assuming Gaussianity. Thus, simplifying terms, for small λ_j,

$$\mathrm{Var}\left[\widehat{S}_j\right] \approx \frac{2}{2q+1} |H_{xy}(\lambda_j)|^2 \approx \frac{2}{2q+1}. \tag{18}$$

Alternatively we can write that

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

and then use the fact that

$$\mathrm{Var}\left[|\widehat{H}_{xy}(\lambda_j)|^2\right] \approx \frac{2}{2q+1} |H_{xy}(\lambda_j)|^2 (1 - |H_{xy}(\lambda_j)|^2)^2, \tag{19}$$

(e.g. Brillinger, 1975, p. 309) to obtain again the approximation (18). Similar approximations can be used for the covariances between $S(\lambda_j)$ and $S(\lambda_k)$, $j \neq k$.

5 Simulation results

In this section we simulate the performance of the estimate $\widehat{\alpha}$ 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), Velasco (1999a)) and an estimate based on a local Gaussian or Whittle likelihood (Künsch (1987), Robinson (1995b), Velasco (1999b)). These estimates are consistent for nonstationary series when $d < 1$, or $d < p$ if tapering of order p is applied ($\mu = 0$). We use the Zhurbenko taper of order $p = 2$, which is valid for $d < 1.5$ for memory estimation.

We have simulated cointegrated Gaussian series (X_t, Y_t) of lengths $n = 128$ and 256 according to (12) with three pairs of cointegration values, $\mathrm{CI}(1,0)$ ($\alpha = 1$), $\mathrm{CI}(1.3, 0.9)$ ($\alpha = 0.4$), and $\mathrm{CI}(1.1, 0.4)$ ($\alpha = 0.7$). 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 $\mathrm{ARFIMA}(0, d, 0)$, while Z_t are $\mathrm{ARFIMA}(2, d - \alpha, 0)$ with autoregressive coefficients $\phi_1 = 0.34$ and $\phi_2 = -0.9$, guarantying that the PSDs of Z_t shows a peak at $\lambda = 4\pi/9$, or $\mathrm{ARFIMA}(1, d - \alpha, 0)$, with $\phi_1 = 0.3$ and 0.6. The innovations are zero mean Gaussian independent sequences η_X, η_Z with standard deviations $\sigma_X = 1$, $\sigma_Z = 2$ 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 $\mathrm{CI}(0, 1)$ case by Robinson and Marinucci (1998).

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 $\widehat{\alpha}$ are calculated from the original data and from tapered data with Zhurbenko taper of order $p = 2$. Note that if no taper is used $\widehat{f}^M$ is not consistent for our simulated series ($d_X \geq 1$). We also construct similar estimates $\widehat{\alpha}_\Delta$ based on the increments $(\Delta X_t, \Delta Y_t)$. For AR(1) series we only report the estimates based on nontapered series with length $N = 256$.

For comparison purpouses we consider alternative estimates of α based on OLS residuals. Notice that for these series the OLS estimate satisfies $\widehat{b} - b = O_p(n^{-\alpha})$ (cases II, IV and III respectively of Robinson and Marinucci (1998)). We consider two semiparametric estimation procedures with the same bandwidths as for $\widehat{\alpha}$: the log-periodogram regression ($\widehat{\alpha}_L$) and the local Gaussian semiparametric estimate ($\widehat{\alpha}_G$). These estimates are implemented with three different input series. We first estimate d starting with the vector (X_t, Y_t) and the restriction $d = d_X = d_Y$ (see Lobato (1998) and Lobato and Velasco (1999) for a two-step modification of the possibly tapered Gaussian estimate) and with the OLS residuals $\widehat{Z}_t$ we estimate the order of integration of Z_t and set the estimate of α as $\widehat{d} - \widehat{d}_Z$. We also substitute (X_t, Y_t) by $(\Delta X_t, \Delta Y_t)$ and $\widehat{Z}_t$ by $\Delta\widehat{Z}_t$ 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, and that some systematic biases may cancel out.

We report the mean, standard deviation (sd) and mean square error (mse) of the estimates across 500 replications. We also give in parenthesis the approximations of the standard deviation of $\widehat{\alpha}$ based on (17) 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 to 3 for $N = 256$ and $N = 128$ and Models 1 to 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 standard deviation is smaller than predicted by (17). The estimates $\widehat{\alpha}_\Delta$ 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 (17) gives a good indication of the sample variability of $\widehat{\alpha}$ 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 (17) 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 where attained for the largest values of m, both in terms of sample bias and standard deviation.

Comparing with residual-based inference, coherence estimates have similar properties 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 one is to use differenced data and original residuals, though, as expected the one using both differenced data and residuals work better for Model 2, while the one with both original data and residuals seems to have no advantage in any case. Gaussian semiparametric estimates have less variability than log-periodogram ones, but are in general more biased. Here again tapering increases standard deviations and mse's.

We report the simulation results for the ARFIMA(1,d,0) cointegrating series in Tables 4 to 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 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 standard deviations in all cases are in line with the approximation in (17). For larger ϕ_1 the performance of all estimates deteriorates, especially that of $\widehat{\alpha}$ and $\widehat{\alpha}_\Delta$ for Models 2 and 3.

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

6 Empirical application

We analyze briefly the same data set as in Dueker and Startz (1998). These authors analyze 120 monthly observations from January 1987 to December 1996 on 10-year government bond rates from the United States and Canada. As they, we analyze the log series, denoted as X_t and Y_t respectively (see Figure 1). Standard procedures used by these authors do not reject the hypothesis of a unit root 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 $\widehat{d} = .674$ (.25) and $\widehat{d-\alpha} = .2$ (.10), so $\widehat{\alpha} = .474$, and that 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 estimator, obtaining significantly different from zero values of $\widehat{\alpha}$, ranging from .2 to .28 for small bandwidths ($m < 10$ in similar notation to our coherence-based $\widehat{\alpha}$) and about .4 for larger values of m. This is an alternative procedure to the one justified by Hassler et al. (1999) and Velasco (1999c) for other semiparametric estimates.

We reanalyze this data set first using the techniques summarized in Lobato and Velasco (1999) 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-v Wald Test	$m=6$	$m=12$	$m=18$
$p=1$ $(\Delta X, \Delta Y)$	.102	.131	.119
$p=2$ (X,Y)	.644	.211	.245

and not rejecting the equal variance hypothesis, though by a small margin using not tapered differenced data, as memory estimates for X are slightly larger than those for Y. Then the common memory parameter Gaussian estimates $\widehat{d}$ are

Common $\widehat{d}$	$m=6$	$m=12$	$m=18$
$p=1$ $(\Delta X, \Delta Y)$	.982 (.09)	1.010 (.10)	.990 (.08)
$p=2$ (X,Y)	.818 (.12)	1.146 (.14)	.855 (.12)

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

$\lvert\widehat{H}_{xy}(0)\rvert^2$	$m=6$	$m=12$	$m=18$
$p=1$ (X,Y)	.795 (.11)	.752 (.09)	.743 (.07)
$p=2$ (X,Y)	.800 (.15)	.747 (.13)	.759 (.10)

which are inconclusive of coherence smaller than 1 given the sample size and the bandwidths employed.

We now estimate nonparametrically the coherence with $|\widehat{H}_{xy}(\lambda_j)|^2$ for $q=1,2,3$ and $p=1,2$. We plot the estimates in Figure 2 for λ_ℓ to $\lambda_{60}=\pi$, where $\ell = \lfloor (2q+1)/2 \rfloor$. Standard errors can be approximated by (19). In all six plots is evident the effect of increasing smoothing in nonparametric estimates and it can be observed that $|H_{xy}(\lambda)|^2 \approx 1 - G_H \lambda^{2\alpha}$ approximately holds. In the plots of $\log(1-|\widehat{H}_{xy}(\lambda_j)|^2)$ against $2\log\lambda_j$, for $j=\ell,\ldots,30$,, see Figure 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 $\widehat{\alpha}$ obtained from (X,Y) are

Log-coherence $\widehat{\alpha}$		$m=6$	$m=12$	$m=18$
$q=1$	$p=1$	.371 (.35)	.292 (.24)	.177 (.19)
	$p=2$	.574 (.68)	.549 (.42)	.545 (.32)
$q=2$	$p=1$	.650 (.37)	.299 (.29)	.226 (.23)
	$p=2$	.757 (.64)	.670 (.47)	.665 (.37)
$q=3$	$p=1$	.704 (.51)	.305 (.31)	.299 (.25)
	$p=2$	1.023 (.86)	.681 (.49)	.768 (.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 $\widehat{\alpha}$ obtained in this case. The estimates for $m=6$ are rather unstable due to the small number of points in the regression. Then, since for untapered series the estimates were quite uniform across values of q, we prefer estimates with $q=1$ which have smallest standard errors and should be also less biased. This gives $\widehat{\alpha} = .292$ (.24), which is lower that the value given by Dueker and Startz (1998), .474. However for tapered series the estimates are more smoothing-dependent, and we obtained from $\widehat{\alpha} = .55$ for $q=1$ to $\widehat{\alpha} = .68$ for $q=3$, more in agreement with that paper.

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

Residual-based $\widetilde{\alpha}$	$m = 6, 8$	$m = 12$	$m = 18$
$p = 1$ $(\Delta X, \widetilde{Z})$	.186 (.29)	.094 (.20)	.068 (.17)
$p = 2$ $(\Delta X, \widetilde{Z})$	.159 (.34)	.002 (.37)	.079 (.31)

The residual-based procedure obtained values of $\widetilde{\alpha}$ noticeably much smaller than coherence-based $\widehat{\alpha}$ for all combinations of m and tapering, casting some doubts about the reliability of OLS-based inference for such α.

7 Appendix A: DFT Asymptotics for Non-Stationary Time Series

We say that a sequence of positive data tapers $\{h_t\}_1^n$ symmetric around $\lfloor n/2 \rfloor$ is of *order* p if the following two conditions are satisfied:

- The weights are normalized to $\max_t h_t = 1$, and for some $0 < b < \infty$, $\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^n h_t^2 = b$.
- For $N = n/p$ (which we assume integer),
$$\sum_{t=1}^{n} h_t \exp\{it\lambda\} = \frac{a(\lambda)}{n^{p-1}} \left(\frac{\sin[n\lambda/2p]}{\sin[\lambda/2]}\right)^p, \tag{20}$$
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 $\lambda \in [-\pi, \pi]$.

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

Theorem 3 (p=1) *Under Assumptions 1 and 2, $d_r \in (-\frac{1}{2}, 1)$, $r = 1, \ldots, R$, $\boldsymbol{\mu} = 0$, for sequences of positive integers $j = j(n)$ and $k = k(n)$ such that $0 < k < j < n/2$, λ_j,$\lambda_k \to \nu \in [0, \pi]$ as $n \to \infty$ and defining $\gamma_{j,k} = j^{d_r-1}k^{d_s-1}\log(1 + j)$,*
(a) $E[v_r(\lambda_j)\overline{v}_s(\lambda_j)] = H_{rs}(\lambda_j) + O\left(j^{-1}\log(1 + j) + \gamma_{j,j}\right)$,
(b) $E[v_r(\lambda_j)v_s(\lambda_j)] = O\left(j^{-1}\log(1 + j) + \gamma_{j,j}\right)$,
(c) $E[v_r(\lambda_j)\overline{v}_s(\lambda_k)] = O\left(k^{-1}\log(j + 1) + \gamma_{j,k}\right)$,
(d) $E[v_r(\lambda_j)v_s(\lambda_k)] = O\left(k^{-1}\log(j + 1) + \gamma_{j,k}\right)$.

The term $\gamma_{j,k}$ is the nonstationarity bias. For values $d_r \geq 1$ the periodogram is not unbiased for the function f_{rr}, due to the impossibility to compensate for the non-integrable pole of $f_{rr}(\lambda)$ at the origin, though its expectation is finite for $d < 1.5$. Tapering helps to control this bias using the property (20) 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 (or g) in Assumption 3 with two derivatives. Denote by $\beta = 1, 2$ the number of derivatives of f around ν. Thus the tapered periodogram (with a taper of order p) is unbiased at certain Fourier frequencies for any $d_r < p$ if $\mu_r = 0$ or with some extra tapering if there are deterministic polynomial trends in time ($\mu_r \neq 0$). Define as before the normalized tapered Fourier transform $v_r^T(\lambda) = w_r^T(\lambda)/f_{rr}^{1/2}(\lambda)$.

Theorem 4 ($p \geq 2$) *Under Assumptions 1, 2 or 3, $d_r > -\frac{1}{2}$, a data taper of order $p = 2, 3, \ldots$, with $p \geq \max\{D_r, D_s\} + 1$ [or just $p > \max\{d_r, d_s\}$ if $\mu_r = \mu_s = 0$], for sequences of positive integers $k = k(n)$ and $j = j(n)$, $k < j$, and $\eta \equiv j - k$, $1 \leq k < j \leq n/(2p)$, λ_{jp},$\lambda_{kp} \to \nu \in [0, \pi]$ as $n \to \infty$, $\gamma_{j,k} \equiv j^{d_r-p}k^{d_s-p}\log(1 + j)$,*
(a) $E[v_r^{T,p}(\lambda_{jp})\overline{v_s^{T,p}}(\lambda_{jp})] = H_{rs}(\lambda_j) + O\left(j^{-\beta} + \gamma_{j,j}\right)$,
(b) $E[v_r^{T,p}(\lambda_{jp})v_s^{T,p}(\lambda_{jp})] = O\left(j^{-p} + \gamma_{j,j}\right)$,
(c) $E[v_r^{T,p}(\lambda_{jp})\overline{v_s^{T,p}}(\lambda_{kp})] = O\left(k^{-1}\eta^{1-p} + k^{-1}\eta^{-p}\{\log n\}_{[p=2]} + \eta^{-p} + \gamma_{j,k}\right)$,
(d) $E[v_r^{T,p}(\lambda_{jp})v_s^{T,p}(\lambda_{kp})] = O\left(k^{-1}\eta^{1-p} + k^{-1}\eta^{-p}\{\log n\}_{[p=2]} + \eta^{-p} + \gamma_{j,k}\right)$.

In parts (c) and (d) the $\log n$ factor only appears when $p=2$ but not otherwise.

8 Appendix B: Proofs

Proof of Theorem 1. We first prove the theorem for fixed $|\nu|>0$ in the asymptotics. 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

$$\widehat{f}_{rs}^{\epsilon,M}(\nu)=\frac{2\pi p}{n}\sum_{j}K_M(\nu-\lambda_{jp})B_r(\lambda_{jp})I^{\epsilon,p}(\lambda_{jp})B_s^*(\lambda_{jp}),$$

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

No tapering $[p=1]$. We consider the case with $d_*=\max\{d_r,d_s\}\in\left(-\frac{1}{2},1\right)$, 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 (1999)),

$$\begin{aligned}\widehat{f}_{rs}^{M}(\nu)-\widehat{f}_{rs}^{\epsilon,M}(\nu) &= \frac{2\pi}{n}\sum_{j}K_M(\nu-\lambda_j)\left[I_{rs}(\lambda_j)-B_r(\lambda_j)I^{\epsilon}(\lambda_j)B_s^*(\lambda_j)\right] \qquad (21)\\ &= O_p\left(\frac{M}{n}\sum_{j}\left[j^{d_*-1}+j^{-1/2}\right](\log n)^{1/2}\right)=O_p\left(\left[n^{d_*-1}+n^{-1/2}\right](\log n)^{1/2}\right),\end{aligned}$$

which is $o_p(1)$ if $d_*<1$. Notice that for $\lambda_j\in\left[\nu-\pi M^{-1},\nu+\pi M^{-1}\right]$, $\max_j\lambda_j^{-1}=O(1)$ and $\max j^{-1}=O(n^{-1})$ as $n\to\infty$, and from Assumptions 1 and 2, $\max_j|B_a(\lambda_j)|$, $a=r,s$, are bounded if $|\nu|>0$. Now the theorem follows as when $p>1$ below, using the exact orthogonality of the sine and cosine instead of Lemma 5.

Tapering $[p>1]$. From Theorem 4 and using the same argument as in the proof of Theorem 3 in Velasco (1999b) or Lemma 2,

$$\widehat{f}_{rs}^{M}(\nu)-\widehat{f}_{rs}^{\epsilon,M}(\nu) = \frac{2\pi p}{n}\sum_{j}^{n}K_M(\nu-\lambda_{jp})\left[I_{rs}^{p}(\lambda_j)-B_r(\lambda_j)I^{\epsilon,p}(\lambda_j)B_s^*(\lambda_j)\right] \qquad (22)$$

$$= O_p\left(\frac{M}{n}\sum_{j}\left[j^{d_*-p}(\log n)^{1/2}+j^{-1/2}\right]\right) = O_p\left(n^{d_*-p}(\log n)^{1/2}+n^{-1/2}\right), \qquad (23)$$

which is $o_p(1)$ because $p>d_*$.

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

$$E\left[\widehat{f}_{r,s}^{\epsilon,M}(\nu)\right]=\frac{2\pi p}{n}\sum_{j}K_M(\nu-\lambda_{jp})f_{rs}(\lambda_{jp}) = f_{rs}(\nu)+O(n^{-1}M+M^{-1}), \qquad (24)$$

where the error is $o(1)$ using (5). For the variance we first obtain for all j,k, and $r,r',s,s'\in\{1,\ldots,R\}$, $N=2\pi\sum_{t=1}^{n}h_t^2$,

$$\begin{aligned}&\mathrm{Cov}\left[B_r(\lambda_{jp})I^{\epsilon}(\lambda_{jp})B_s^*(\lambda_{jp}),B_{r'}(\lambda_{kp})I^{\epsilon}(\lambda_{kp})B_{s'}^*(\lambda_{kp})\right]\\ &= \sum_{a=1}^{R}\sum_{b=1}^{R}\sum_{c=1}^{R}\sum_{d=1}^{R}B_{ra}(\lambda_{jp})\overline{B_{sb}(\lambda_{jp})B_{r'c}(\lambda_{kp})}B_{s'd}(\lambda_{kp})\mathrm{Cov}\left[I_{ab}^{\epsilon,p}(\lambda_{jp}),I_{cd}^{\epsilon,p}(\lambda_{kp})\right]\\ &= N^{-2}\sum_{a=1}^{R}\sum_{b=1}^{R}\sum_{c=1}^{R}\sum_{d=1}^{R}B_{ra}(\lambda_{jp})\overline{B_{sb}(\lambda_{jp})B_{r'c}(\lambda_{kp})}B_{s'd}(\lambda_{kp})\end{aligned}$$

$$\times \left\{ \Sigma_{ac}\Sigma_{bd} \left[\sum_t^n h_t^2 \cos t(\lambda_{jp} - \lambda_{kp}) \right]^2 + \Sigma_{ad}\Sigma_{cb} \left[\sum_t^n h_t^2 \cos t(\lambda_{jp} + \lambda_{kp}) \right]^2 + \sum_t^n \kappa_{abcd} h_t^4 \right\}$$

$$= N^{-2} B_r(\lambda_{jp})\Sigma B_{r'}^*(\lambda_{kp}) B_s(-\lambda_{jp})\Sigma B_{s'}^*(-\lambda_{kp}) \left[\sum_t^n h_t^2 \cos t(\lambda_{jp} - \lambda_{kp}) \right]^2$$

$$+ N^{-2} B_r(\lambda_{jp})\Sigma B_{s'}^*(-\lambda_{kp}) B_{r'}(-\lambda_{kp})\Sigma B_s^*(\lambda_{jp}) \left[\sum_t^n h_t^2 \cos t(\lambda_{jp} + \lambda_{kp}) \right]^2$$

$$+ N^{-2} \sum_1^n h_t^4 \sum_{a=1}^R \sum_{b=1}^R \sum_{c=1}^R \sum_{d=1}^R B_{ra}(\lambda_{jp}) \overline{B_{sb}(\lambda_{jp}) B_{r'c}(\lambda_{kp})} B_{s'd}(\lambda_{kp}) \kappa_{abcd},$$

where κ_{abcd} is the joint fourth order cumulant of the a, b, c, d-th components of $\boldsymbol{\epsilon}_\mathbf{t}$. Using Lemma 5,

$$\begin{aligned}
&\mathrm{Var}\left[\widehat{f}_{rs}^{\epsilon,M}(\nu) \right] \\
&= \left(\frac{2\pi p}{n} \right)^2 \sum_j K_M^2(\nu - \lambda_{jp}) \mathrm{Var}\left[B_r(\lambda_{jp}) I^{\epsilon,p}(\lambda_{jp}) B_s^*(\lambda_{jp}) \right] \\
&\quad + \left(\frac{2\pi p}{n} \right)^2 \sum_j \sum_{k \neq j} K_M(\nu - \lambda_{jp}) K_M(\nu - \lambda_{kp}) \mathrm{Cov}\left[B_r(\lambda_{jp}) I^{\epsilon}(\lambda_{jp}) B_s^*(\lambda_{jp}), B_r(\lambda_{kp}) I^{\epsilon}(\lambda_{kp}) B_s^*(\lambda_{kp}) \right] \\
&= O\left(n^{-2} \sum_j K_M^2(\nu - \lambda_{jp}) \right) \\
&\quad + O\left(n^{-2} \sum_j \sum_{k>j} K_M(\nu - \lambda_{jp}) K_M(\nu - \lambda_{kp}) \left[|j-k|^{-2p} + |j+k|^{-2p} + n^{-1} \right] \right) \\
&= O\left(n^{-1}M + n^{-2}M^2 \sum_j \sum_{k>j} \left\{ |j-k|^{-2p} + |j+k|^{-2p} + n^{-1} \right\} \right) = O\big(n^{-1}M\big) = o(1)
\end{aligned}$$

as $n \to \infty$, and the theorem follows for fixed ν using (5).

For $|\nu| \to 0$ as $n \to \infty$ we only stress the main differences. First notice that since $(|\nu M|)^{-1} \to 0$ as $n \to \infty$, we have that for all $\lambda_j \in [\nu - \frac{\pi}{M}, \nu + \frac{\pi}{M}]$

$$\inf_{[\nu - \pi/M, \nu + \pi/M]} \lambda_j = \nu[1 + O((\nu M)^{-1}] = \nu[1 + o(1)] \sim \nu \quad \text{as } n \to \infty, \tag{25}$$

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

$$\max_{\lambda_j} \left| f_{rr}^{-1}(\nu) - f_{rr}^{-1}(\lambda_j) \right| = O({f_{rr}}^{-1}(\nu)(|\nu| M)^{-1}) = o(f_{rr}^{-1}(\nu)),$$

applying the mean value theorem.

Therefore, when no tapering is applied, the left hand side of (21) is

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

as the summation is running for integers j between $\nu n/(2\pi p) \pm n/(2M)$. When tapering is applied (22) is

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

using (5) and (6). Finally the left hand side of (24) is

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

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

Proof of Theorem 2. We assume for simplicity only positive frequencies $\nu_h>0$, $\forall h$. This entails no restriction, since it is always possible to write for positive ν, that $\widehat{f}_{rs}(-\nu)=\overline{\widehat{f}_{rs}(\nu)}=\widehat{f}_{sr}(\nu)$ and deduce the variances and covariances for the conjugate estimates from those with positive argument and reversed indexes. We considerate later frequencies in a degenerating band around zero frequency. We follow the same procedure as in the proof of Theorem 1 but employ Lemma 3 when $p=1$ instead of Lemma 2.

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

$$E\left[\widehat{f}_{rs}^{\epsilon,M}(\nu)\right]=f_{rs}(\nu)+O(M^2n^{-1}+M^{-2}).$$

Tapering $[p>1]$. Now (23) is $o_p((n/M)^{-1/2})$ 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 $m\times 1$ vector ξ we have that $\xi'\widehat{f}_{\mathbf{rs}}^{M}(\nu)=\xi'\widehat{f}_{\mathbf{rs}}^{M,\epsilon}(\nu)+o_p((n/M)^{-1/2})$. Now

$$\begin{aligned}
&\xi'\widehat{f}_{\mathbf{rs}}^{\epsilon,M}(\nu)-E\left[\xi'\widehat{f}_{\mathbf{rs}}^{\epsilon,M}(\nu)\right]\\
&=\sum_{h=1}^{m}\xi_h\left\{\frac{2\pi p}{n}\sum_j K_M(\nu_h-\lambda_j)B_{r_h}(\lambda_j)I^{\epsilon}(\lambda_j)B_{s_h}^{*}(\lambda_j)-\frac{p}{n}\sum_j K_M(\nu-\lambda_j)B_{r_h}(\lambda_j)\Sigma B_{s_h}^{*}(\lambda_j)\right\}\\
&=\sum_{h=1}^{m}\xi_h\frac{p}{n}\sum_j K_M(\nu_h-\lambda_j)B_{r_h}(\lambda_j)\left[\nabla_{\epsilon}^{(1)}-\Sigma\right]B_{s_h}^{*}(\lambda_j)\\
&\quad+\sum_{h=1}^{m}\xi_h\frac{p}{n\sum h_t^2}\sum_j K_M(\nu_h-\lambda_j)B_{r_h}(\lambda_j)\nabla_{\epsilon}^{(2)}B_{s_h}^{*}(\lambda_j),
\end{aligned}$$

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 $\nabla_{\epsilon}^{(2)}=\sum_t\sum_{t'\neq t}h_th_{t'}\epsilon_t\epsilon_{t'}'\exp\{i(t-t')\lambda_j\}$. The first term is negligible $o_p\left((Mn^{-1})^{1/2}\right)$ since $\left\|\nabla_{\epsilon}^{(1)}-\Sigma\right\|=O_p(n^{-1/2})$, with Assumption 4. Then we have,

$$\sqrt{\frac{n}{Mp}}\left\{\xi'\widehat{f}_{\mathbf{rs}}^{M}(\nu)-\xi'f_{\mathbf{rs}}(\nu)\right\}=\sum_{t=1}^{n}z_t+o_p(1)$$

where

$$\begin{aligned}
z_t &= \sum_{h=1}^{m}\xi_h\frac{p}{n\sum h_t^2}\sqrt{\frac{n}{Mp}}\sum_{j_h}K_M(\nu_h-\lambda_{pj_h})B_{r_h}(\lambda_{j_h})\sum_{s\neq t}h_t\epsilon_th_s\epsilon_s'B_{s_h}^{*}(\lambda_{pj_h})\exp\{i(t-s)\lambda_{pj_h}\}\\
&= h_t\epsilon_t'\sum_{s=1}^{t-1}\Lambda_{t-s}h_s\epsilon_s,
\end{aligned}$$

is a martingale difference sequence,

$$\Lambda_t=\frac{1}{\sum h_t^2}\sqrt{\frac{p}{nM}}\sum_{h=1}^{m}\sum_{j_h}\Theta_{pj_h}^{h}\cos t\lambda_{pj_h},$$

$\Theta^h_{pj_h} = 2\xi_h K_M(\nu_h - \lambda_{pj_h}) B'_{r_h}(\lambda_{pj_h})\overline{B_{s_h}}(\lambda_{pj_h})$, and the summation in j_h runs from $-n/2p$ to $n/2p+1$, with p steps, assuming $n/2p$ is integer for simplicity. We estimate first the asymptotic variance of $\widehat{f}^M$,

$$\sum_1^n E\left[z_t \overline{z_t} | F_{t-1}\right] = \sum_{t=2}^n h_t^2 \sum_{s=1}^{t-1} h_s^2 \epsilon_s' \Lambda_{t-s}' \Sigma \overline{\Lambda_{t-s}} \epsilon_s \quad (26)$$

$$+ \sum_{t=1}^n h_t^2 \sum_{s=1}^{t-1} \sum_{r \neq s}^{t-1} h_s h_r \epsilon_s' \Lambda_{t-s}' \Sigma \overline{\Lambda_{t-r}} \epsilon_r. \quad (27)$$

The right hand side of (26) is

$$\sum_{t=2}^{n-1} h_t^2 \sum_{s=1}^{t-1} h_s^2 \epsilon_s' \Lambda_{t-s} \Sigma \Lambda_{t-s}^* \epsilon_s = \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \text{ Trace}\left[(\epsilon_t\epsilon_t' - \Sigma)\Lambda_s' \Sigma \overline{\Lambda_s}\right] \quad (28)$$

$$+ \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \text{ Trace}[\Sigma \Lambda_s' \Sigma \overline{\Lambda_s}] \quad (29)$$

where the right hand side of (28) is $o_p(1)$, because it has zero mean and variance, $\|\Lambda_{t-s}\| = O((Mn)^{-1/2})$,

$$O\left(\sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \left\{2 \text{ Trace}\left[\Sigma\Lambda_s'\Sigma\overline{\Lambda_s}\Sigma\Lambda_s'\Sigma\overline{\Lambda_s}\right] + \kappa_{abcd} \text{ Trace}\left[\Sigma\Lambda_s'\Sigma\overline{\Lambda_s}\right]^2\right\}\right)$$

$$= O\left(\sum_{t=1}^{n-1}\sum_{s=1}^{n-t}(Mn)^{-2}\right) = O(M^{-2}) = o(1).$$

The term (29) is, using trigonometric identities (see Velasco (1999b), Lemma 6 for details),

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

$$= \left(\sum h_t^2\right)^{-2} \frac{p}{nM} \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \text{ Trace}\left[\Sigma \sum_h \xi_h \sum_{j_h} \Theta^{h'}_{pj_h} \cos s\lambda_{pj_h} \Sigma \sum_\ell \xi_\ell \sum_{k_\ell} \overline{\Theta^\ell_{pk_\ell}} \cos s\lambda_{pk_\ell}\right]$$

$$= \left(\sum h_t^2\right)^{-2} \frac{p}{nM} \sum_h \sum_\ell \sum_{j_h} \xi_h \xi_\ell \text{ Trace}\left[\Sigma\Theta^{h'}_{pj_h}\Sigma\overline{\Theta^{\ell *}_{pj_h}} + \Sigma\Theta^h_{pj_h}\Sigma\overline{\Theta^\ell_{-pj_h}}\right] \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \cos^2 s\lambda_{pj_h} \quad (30)$$

$$+\left(\sum h_t^2\right)^{-2} \frac{p}{nM} \sum_h \sum_{j_h} \sum_\ell \sum_{k_\ell \neq \pm j_h} \xi_h \xi_\ell \text{ Trace}\left[\Sigma\Theta^{h'}_{pj_h}\Sigma\overline{\Theta^\ell_{pk_\ell}}\right] \frac{1}{2} \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \{\cos(s\lambda_{pj_h+pk_\ell}) + \cos(s\lambda_{pj_h-pk_\ell})\} \quad (31)$$

Using $\left\|\Theta^h_{j_h}\right\| = O(M)$ and Lemma 6A, (30) is equal to

$$\frac{p}{4nM} \sum_h \sum_\ell \sum_{j_h} \xi_h \xi_\ell \text{ Trace}\left[\Sigma\Theta^{h'}_{pj_h}\Sigma\overline{\Theta^\ell_{pj_h}} + \Sigma\Theta^{h'}_{pj_h}\Sigma\overline{\Theta^\ell_{-pj_h}}\right] + O(M/n).$$

If $k_\ell \neq \pm j_h$, using, Lemma 6B, (31) is

$$\left(\sum h_t^2\right)^{-2} \frac{p}{4nM} \sum_h \sum_{j_h} \sum_\ell \sum_{k_\ell \neq \pm j_h} \xi_h \xi_\ell \text{ Trace}\left[\Sigma\Theta^{h'}_{pj_h}\Sigma\overline{\Theta^\ell_{pk_\ell}}\right]$$

$$\times \left\{\left(\sum_{t=1}^{n-1} h_t^2 \cos t\lambda_{pj_h+pk_\ell}\right)^2 + \left(\sum_{t=1}^{n-1} h_t^2 \cos t\lambda_{pj_h-pk_\ell}\right)^2\right\} + O(M/n).$$

Then

$$\text{Trace}\left[\Sigma\Theta^{h'}_{pj_h}\Sigma\overline{\Theta^{\ell}_{pk_\ell}}\right] = 4K_M(\nu_h-\lambda_{pj_h})K_M(\nu_\ell-\lambda_{pk_\ell})B_{r_h}(\lambda_{pj_h})\Sigma B^*_{r_\ell}(\lambda_{pk_\ell})B_{s_h}(-\lambda_{pj_h})\Sigma B^*_{s_\ell}(-\lambda_{pk_\ell}).$$

Therefore, using Lemma 5, 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 (1999),

$$\begin{aligned}&\left(\sum h_t^2\right)^{-2}\frac{p}{4nM}\sum_h\sum_{j_h}\sum_\ell\sum_{k_\ell}\xi_h\xi_\ell\ \text{Trace}\left[\Sigma\Theta^{h'}_{pj_h}\Sigma\overline{\Theta^{\ell}_{pk_\ell}}\right]\left(\sum_{t=1}^{n-1}h_t^2\cos t\lambda_{pj_h-pk_\ell}\right)^2\\ =\ & 2\pi\Phi_p\left\|K\right\|_2^2\sum_h\sum_\ell\xi_h\xi_\ell\delta(\nu_h-\nu_\ell)f_{r_hr_\ell}(\nu_h)f_{s_hs_\ell}(-\nu_h)+o(1),\end{aligned}$$

and

$$\left(\sum h_t^2\right)^{-2}\frac{p}{4nM}\sum_h\sum_{j_h}\sum_\ell\sum_{k_\ell}\xi_h\xi_\ell\ \text{Trace}\left[\Sigma\Theta^{h'}_{pj_h}\Sigma\overline{\Theta^{\ell*}_{pk_\ell}}\right]\left(\sum_{t=1}^{n-1}h_t^2\cos t\lambda_{pj_h+pk_\ell}\right)^2 = o(1),$$

since the frequencies covered by $K_M(\nu_h-\lambda_{pj_h})K_M(\nu_\ell-\lambda_{pk_\ell})$ are such that $j_h+k_\ell > \frac{\nu_h+\nu_\ell}{2\pi}\frac{n}{p}-\frac{n}{Mp} > Cn$, $\operatorname{mod} n$, if $\pi > \nu_h, \nu_\ell > 0$ and n and M are big enough, so $\left(\sum h_t^2\right)^{-2}\left(\sum h_t^2\cos t\lambda_{pj_h+pk_\ell}\right)^2 = O(|j_h+k_\ell|^{-2p})$ from Lemma 5.

The second term (27) is $o_p(1)$ because it has zero mean and variance equal to

$$\begin{aligned}&2\sum_{t=2}^n h_t^2\sum_{u=2}^n h_u^2\sum_s^{\min\{t-1,u-1\}}\sum_{r\neq s}h_s^2h_r^2\ \text{Trace}\left[\Lambda'_{t-s}\Sigma\overline{\Lambda_{t-r}}\Sigma\left(\Lambda'_{u-s}\Sigma\overline{\Lambda_{u-r}}\right)^*\right]\\ =\ & 2\sum_{t=2}^n h_t^4\sum_s\sum_{r\neq s}h_s^2h_r^2\ \text{Trace}\left[\Lambda'_{t-s}\Sigma\overline{\Lambda_{t-r}}\Sigma\Lambda'_{t-r}\Sigma\overline{\Lambda_{t-s}}\right] \qquad (32)\\ & +4\sum_{t=3}^n h_t^2\sum_{u=2}^{t-1}h_u^2\sum_s^{u-1}\sum_{r\neq s}^{u-1}h_s^2h_r^2\ \text{Trace}\left[\Lambda'_{t-s}\Sigma\overline{\Lambda_{t-r}}\Sigma\Lambda'_{u-r}\Sigma\overline{\Lambda_{u-s}}\right], \qquad (33)\end{aligned}$$

since the weights $\{h_t\}$ are symmetric around $\lfloor n/2\rfloor$. By summation by parts we find that $\|\Lambda_t\| = O(M^{1/2}n^{-1/2}t^{-1})$, $t < n/2$, $\|\Lambda_t\| = \|\Lambda_{n-t}\|$, so (32) is

$$\begin{aligned}O\left(\sum_{t=2}^n\sum_s\sum_{r\neq s}\left\|\Lambda_{t-r}\right\|^2\left\|\Lambda_{t-s}\right\|^2\right) &= O\left(M^2n^{-2}\sum_{t=2}^n\sum_s\sum_{r\neq s}|t-r|^{-2}|t-s|^{-2}\right)\\ &= O\big(M^2n^{-1}\big) = o(1),\end{aligned}$$

and, following Robinson (1995b), p. 1646, (33) has absolute value bounded by

$$C\sum_{t=3}^n\sum_{u=2}^{t-1}\left(\sum_s^{u-1}\left\|\Lambda_{t-r}\right\|^2\sum_{r\neq s}^{u-1}\left\|\Lambda_{u-r}\right\|^2\right) \le C\left(\sum_1^n\left\|\Lambda_t\right\|^2\right)\left(\sum_{t=3}^n\sum_{u=2}^{t-1}\sum_{r=t-u+1}^{t-1}\left\|\Lambda_r\right\|^2\right),$$

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

$$\sum_{j=1}^{n-2} j(n-j-1)\left\|\Lambda_j\right\|^2 \le 2n\sum_1^{[n/2]} j\left\|\Lambda_j\right\|^2 = O(M\log n),$$

and $\sum_1^n\left\|\Lambda_t\right\|^2 = O\big(M^2n^{-1}\big)$ so (27) is $O_p\left(\left[n^{-1}M^3\log n\right]^{1/2}\right) = o_p(1)$ with (7).

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

$$\sum_1^n E[z_t \overline{z_t} I(|z_t| > \rho)] = \sum_1^n E[|z_t|^2 I(|z_t| > \rho)] \to 0, \qquad \text{for all } \ \rho > 0.$$

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

$$\sum_1^n E\,|z_t|^4 \le C \sum_1^n \left(\sum_1^n \|\Lambda_{t-s}\|^4 \right) + C \sum_1^n \sum_1^{t-1} \sum_1^{t-1} \|\Lambda_{t-s}\|^2 \|\Lambda_{t-r}\|^2 = O\big(n^{-1}M^2\big) = o(1),$$

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

For positive $\nu \to 0$ as $n \to \infty$, we first obtain for $p = 1$ that

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

using (7), (10) and $d_* < \frac{3}{4}$. When we apply tapering, $p > 1$, we find that

$$[f_{rr}(\nu) f_{ss}(\nu)]^{-1/2} \left(M^{-1}n\right)^{1/2} \left| \widehat{f}_{rs}^{M}(\nu) - \widehat{f}_{rs}^{\epsilon,M}(\nu) \right| = O_p((nM^{-1})^{1/2} [|n\nu|^{-1/2} + |n\nu|^{d_*-p} \log^{1/2} n]) + o_p(1),$$

which is $o_p(1)$ by (11). Finally with Assumption 3

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

where the error term is $o([f_{rr}(\nu) f_{ss}(\nu)]^{1/2} (Mn^{-1})^{1/2})$ using (7) and (10), and the theorem follows as for fixed ν. •

Lemma 2 *Under the Assumptions of Theorem 1, for* $\lambda_{jp} \in \left[\nu - \pi \frac{1}{M}, \nu + \pi \frac{1}{M}\right]$, $p \ge 1$,

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

where e_{p,d_r} *is the error term in the part (a) of Theorems 3 or 4 in Appendix A, for each* p*, depending on the values of* d_r*,* $\beta = 1$*.*

Proof of Lemma 2. We write for $|\nu| > 0$, following Robinson (1995b), proof of Theorem 1, suppressing in the notation the frequency λ_{jp},

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

Then

$$E\left| I_{rs} - B I^{\epsilon} B_s^* \right| = \frac{1}{2} E\left[\left| \left(w_r - B_r w^{\epsilon}\right) \left(\bar{w}_s + w^{\epsilon *} B_s^*\right) \right| \right] + \frac{1}{2} E\left[\left| \left(w_r + B_r w^{\epsilon}\right) \left(\bar{w}_s - w^{\epsilon *} B_s^*\right) \right| \right]$$

and

$$E\left[\left|(w_r - B_r w^\epsilon)\left(\bar{w}_s + w^{\epsilon *} B_s^*\right)\right|\right] \le E\left[|w_r - B_r w^\epsilon|\,|\bar{w}_s + w^{\epsilon *} B_s^*|\right]$$
$$\le \left(EI_r - B_r E w^\epsilon \bar{w}_r - E w_r \bar{w}^\epsilon B_r^* + B_r E w^\epsilon w^{\epsilon *} B_r^*\right)^{1/2}$$
$$\times \left(EI_s + B_s E w^\epsilon \bar{w}_s + E w_s w^{\epsilon *} B_s^* + B_s E w^\epsilon w^{\epsilon *} B_s^*\right)^{1/2}$$

and since $B_r E\left[w^\epsilon w^{\epsilon *}\right] B_r = f_{rr}$, and from the proof of Theorem 3 in Velasco (1999a) it follows that $B_r E\left[w^\epsilon \bar{w}_r\right] = B_r \frac{1}{2\pi} \Sigma B_r^* + O(\lambda_j^{-2d_r} e_{p,d_r}) = f_{rr} + O(\lambda_j^{-2d_r} e_{p,d_r})$, $E\left[w_r \bar{w}^{\epsilon *}\right] B_r^* = B_r \frac{1}{2\pi} \Sigma B_r^* + O(\lambda_j^{-2d_r} e_{p,d_r}) = f_{rr} + O(\lambda_j^{-2d_r} e_{p,d_r})$, this completes the proof for $|\nu| > 0$. When $|\nu| \to 0$ as $n \to \infty$ the argument follows using (25) because $(|\nu| M)^{-1} \to 0$ as $n \to \infty$ from (6). •

Lemma 3 *Under the Assumptions of Theorem* 2, $p = 1$,

$$\frac{2\pi}{n} \sum_j K_M(\nu - \lambda_j)\left[I_{rs}(\lambda_j) - B_r(\lambda_j) I^\epsilon(\lambda_j) B_s^*(\lambda_j)\right] = o_p\left((n/M)^{-1/2}\right).$$

Proof of Lemma 3. Using the second moments of the periodogram as in the argument in page 1648 of Robinson (1995b) and in Velasco (1999b), $d_h \in \left(-\frac{1}{2}, 1\right)$, we have with $K_{Mj} = K_M(\nu - \lambda_j)$, that

$$E\left(\sum_j \frac{2\pi K_{Mj}}{n}\left[I_{rs}(\lambda_j) - B_r(\lambda_j) I^\epsilon(\lambda_j) B_s^*(\lambda_j)\right]\right)^2$$
$$= \sum_j \left(\frac{2\pi K_{Mj}}{n}\right)^2$$
$$\times E\left[\left(w_r(\lambda_j)\overline{w_s}(\lambda_j) - B_r(\lambda_j) w^\epsilon(\lambda_j) w^{\epsilon *}(\lambda_j) B_s^*(\lambda_j)\right)\left(w_r(\lambda_j)\overline{w_s}(\lambda_j) - B_r(\lambda_j) w^\epsilon(\lambda_j) w^{\epsilon *}(\lambda_j) B_s^*(\lambda_j)\right)^*\right]$$
$$+ \sum_j \sum_{k \ne j} \frac{2\pi K_{Mj}}{n} \frac{2\pi K_{Mk}}{n}$$
$$\times E\left[\left(w_r(\lambda_j)\overline{w_s}(\lambda_j) - B_r(\lambda_j) w^\epsilon(\lambda_j) w^{\epsilon *}(\lambda_j) B_s^*(\lambda_j)\right)\left(w_r(\lambda_k)\overline{w_s}(\lambda_k) - B_r(\lambda_k) w^\epsilon(\lambda_k) w^{\epsilon *}(\lambda_k) B_s^*(\lambda_k)\right)^*\right].$$

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 3 above and Lemma 4,

$$\frac{2\pi}{n} \sum_j K_M(\nu - \lambda_j)\left[I_{rs}(\lambda_j) - B_r(\lambda_j) I^\epsilon(\lambda_j) B_s^*(\lambda_j)\right]$$
$$= O_p\left(M n^{-1}\left[\sum_j \left\{j^{-1} + j^{2(d_*-1)}\right\} \log n\right]^{1/2}\right) + O_p\left(M n^{-1}\left[\sum_j \sum_{k>j} \left\{j^{-2} + j^{4(d_*-1)}\right\} \log^2 n\right]^{1/2}\right)$$
$$+ O_p\left(M n^{-1}\left[\sum_j \left\{j^{-3/2} + j^{3(d_*-1)} + n^{-1/2}\left(j^{-1} + j^{2(d_*-1)}\right)\right\}\right]^{1/2} \log n\right)$$
$$+ O_p\left(M n^{-1}\left[\sum_j \sum_{k>j} \left\{j^{-3/2} + (jk)^{2(d_*-1)} \log^2 n + n^{-1} k^{d_*-1} \log^3 n + n^{-1/2}\left((jk)^{-1/2} + (jk)^{d_*-1}\right) \log n\right\}\right]^{1/2}\right)$$
$$= O_p\Big(M^{1/2} n^{-1/2}\Big[n^{d_*-1} \log^{1/2} n + n^{-1/2} \log^{1/2} n + n^{2d_*-3/2} M^{-1/2} \log n + M^{-1/2} n^{-1/2} \log n$$
$$+ n^{-3/4} + \left(n^{3(d_*-1)/2} + n^{d_*-5/4} + n^{-3/4}\right) \log n$$
$$+ M^{-1/2} n^{-1/4} + M^{-1/2} n^{(d_*-1)/2} \log^{3/2} n + M^{-1/2}\left(n^{d_*-3/4} + n^{-1/4}\right) \log^{1/2} n\Big]\Big),$$

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

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

$$E\left|I_{ab}(\lambda_j)-B_a(\lambda_j)I^{\epsilon}(\lambda_j)B_b^*(\lambda_j)\right|^2$$
$$=O\Big(f_{aa}(\lambda_j)f_{bb}(\lambda_j)\left[\left(j^{-1}+j^{2(d-1)}\right)\log n+j^{-3/2}+j^{3(d-1)}+n^{-1/2}\left(j^{-1}+j^{2(d-1)}\log n\right)\right]\Big),$$

and for $j<k$,

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

Proof of Lemma 4. It follows from Lobato (1999) multivariate treatment, adapting Robinson (1995b, Theorem 2), and Velasco (1999b, Lemmas 1 to 3) for $d\geq 0.5$. •

The following two lemmas are Lemma 8 of Velasco and Robinson (1999) and Lemma 7 of Velasco (1999b), respectively.

Lemma 5 *If the sequence $\{h_j\}$ is a data taper of order p as defined previously,* $0<|j|<n/2$,

$$\left(\sum_{t=1}^n h_t^2\right)^{-1}\left|\sum_{t=1}^n h_t^2\cos t\lambda_j\right|=O(|j|^{-p}).$$

Lemma 6 *If the sequence $\{h_j\}$ is a data taper of order p as defined previously,*

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

(B) $$\sum_{t=1}^{n-1}h_t^2\sum_{s=1}^{n-t}h_{s+t}^2\cos s\lambda_j=\frac{1}{2}\left(\sum_{t=1}^n h_t^2\cos t\lambda_j\right)^2+O(n),\ 0<|j|<n.$$

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

$$\max_{\ell\leq j\leq m}D_j\left|\widehat{f}^M_{yy}(\lambda_j)-f_{yy}(\lambda_j)\right|\leq\max_{\ell\leq j\leq m}D_j\left|\widehat{f}^M_{yy}(\lambda_j)-E\left[\widehat{f}^M_{yy}(\lambda_j)\right]\right| \tag{34}$$
$$+\max_{\ell\leq j\leq m}D_j\left|E\left[\widehat{f}^M_{yy}(\lambda_j)\right]-f_{yy}(\lambda_j)\right|. \tag{35}$$

First (35) is

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

Now we have that (34) is, $\tau\geq 1$,

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

where $\Xi_{r,j}^M = K_M(\lambda_r - \lambda_j) f_{yy}^{-1}(\lambda_j)$, and $\max_j \max_r |\Xi_{r,j} f_{yy}(\lambda_r)| = O(M)$ 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 or equal than

$$\begin{aligned} & C \sum_{j=\ell}^{m} \lambda_j{}^{-2\alpha\tau} \left| \frac{2\pi}{n} \sum_r \Xi_{r,j}^M \left[I_{yy}(\lambda_r) - E[I_{yy}(\lambda_r)] \right] \right|^\tau \\ \leq \; & C \left(\frac{M}{n} \right)^{\tau/2} n^{2\alpha\tau} \sum_{j=\ell}^{m} j^{-2\alpha\tau} \left| \frac{2\pi}{\sqrt{Mn}} \sum_r \Xi_{r,j}^M \left[I_{yy}(\lambda_r) - E[I_{yy}(\lambda_r)] \right] \right|^\tau \\ = \; & O_p \left(\left(\frac{M}{n} \right)^{\tau/2} n^{2\alpha\tau} \ell^{1-2\alpha\tau} \right), \end{aligned}$$

taking expectations, since the standardized quantity $(Mn)^{-1/2} \sum_r \Xi_{r,j}^M \left[I_{yy}(\lambda_r) - E[I_{yy}(\lambda_r)] \right]$ 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

Alekseev, V.G. (1996). Jackson- and Jackson-Vallée Poussin-type kernels and their probability applications. *Theory Probability and its Applications* 41, 137-143.

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

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

Chen, G., B. Abraham and S. Peiris (1994). Lag window estimation of the degree of differencing in fractionally integrated time series models. *Journal of Time Series Analysis* 15, 473-487.

Dahlhaus, R. (1985). Asymptotic normality of spectral estimates. *Journal of Multivariate Analysis* 16, 412-431.

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

Dueker, M. and R. Startz (1998). Maximum-likelihood estimation of fractional cointegration with an application to U.S. and Canadian bond rates. *Review of Economics and Statistics* 80, 420-426.

Engle, R. and C. Granger (1987). Cointegration and error correction: representation, estimation and testing. *Econometrica* 55, 251-276.

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

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

Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and its Application.* Academic Press, New York.

Hannan, E.J. (1963). Regression for time series. (Proc. Symp. on) *Time Series Analysis*, (ed. M. Rosenblatt), 17-37. Wiley, New York.

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

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

Hassler, U., F. Marmol and C. Velasco (1999). 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. and B.K. Ray (1995). Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *Journal of Time Series Analysis* 16, 17-42.

Künsch, H.R. (1987) Statistical aspects of self-similar processes. *Proceedings 1st World Congress of the Bernoulli Society,* 67-74. VNU Science Press.

Lobato, I.N. (1997). Consistency of the averaged cross-periodogram in long memory series. *Journal of Time Series Analysis* 18, 137-155.

Lobato, I.N. (1999). A semiparametric two step estimator in a multivariate long memory model. *Journal of Econometrics* 90, 129-153.

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

Lobato, I.N. and C. Velasco (1999). Long memory in stock market trading volume. Preprint.

Reisen, V.A. (1994). Estimation of the fractional difference parameter in the ARIMA(p.d.q) model using the smoothed periodogram. *Journal of Time Series Analysis* 15, 335-350.

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

Robinson, P.M. (1994a). Time series with strong dependence. In *Advances in Econometrics. Sixth World Congress, Vol. I*, C.A. Sims ed., pp. 47-95. Cambridge University Press, Cambridge.

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

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

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

Robinson, P.M. and J. Hidalgo (1998). Time series regression with long range depedence. *Annals of Statistics* 25, 77-104.

Robinson, P.M. and D. Marinucci (1998). Semiparametric frequency domain analysis of fractional cointegration. STICERD Discussion Paper EM/98/348, London School of Economics and Political Science.

Solo, V. (1992). Intrinsic random functions and the paradox of 1/f noise. *SIAM Journal of Applied Mathematics* 52, 270-291.

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

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

Velasco, C. (1999b). Gaussian semiparametric estimation for non-stationary time series. *Journal of Time Series Analysis* 20, 87-127.

Velasco, C. (1999c). Gaussian semiparametric estimation of fractional cointegration. Preprint.

Velasco, C. and P.M. Robinson (1999). Whittle pseudo-maximum likelihood estimation of non-stationary fractional time series. Preprint.

Zhurbenko, I.G. (1979). On the efficiency of estimates of a spectral density. *Scandinavian Journal of Statistics* 6, 49-56.

Table 1 **MODEL 1, ® = 1, CI(1,0), Á$_1$ = 0.34, Á$_2$ = −0.9**

n=256				$q=2$		$q=1$		$\mathbf{X}, \widehat{Z}_t$		$\Delta\mathbf{X}_t, \Delta\widehat{Z}_t$		$\Delta\mathbf{X}_t, \widehat{Z}_t$	
p	m			$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$
1	12	mean		1.167	0.335	0.976	0.282	0.980	0.683	0.249	0.175	1.000	0.890
		sd (.29)	(.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 (.19)	(.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 (.15)	(.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 (.47)	(.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 (.29)	(.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 (.22)	(.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				$q=2$		$q=1$		$\mathbf{X}_t, \widehat{Z}_t$		$\Delta\mathbf{X}_t, \Delta\widehat{Z}_t$		$\Delta\mathbf{X}_t, \widehat{Z}_t$	
p	m			$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$
1	6	mean		1.141	0.305	0.869	0.221	0.995	0.686	0.461	0.411	1.042	0.955
		sd (.37)	(.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 (.29)	(.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 (.23)	(.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 (.60)	(.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 (.47)	(.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 (.36)	(.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

Table 2 **MODEL 2, ® = 0.4, CI(1.3, 0.9), Á$_1$ = 0.34, Á$_2$ = −0.9**

n=256				$q=2$		$q=1$		$\mathbf{X}_t, \widehat{Z}_t$		$\Delta\mathbf{X}_t, \Delta\widehat{Z}_t$		$\Delta\mathbf{X}_t, \widehat{Z}_t$	
p	m			$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$
1	12	mean		0.344	0.304	0.254	0.279	0.316	0.242	0.370	0.274	0.465	0.346
		sd (.29)	(.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 (.19)	(.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 (.15)	(.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 (.47)	(.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 (.29)	(.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 (.22)	(.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				$q=2$		$q=1$		$\mathbf{X}_t, \widehat{Z}_t$		$\Delta\mathbf{X}_t, \Delta\widehat{Z}_t$		$\Delta\mathbf{X}_t, \widehat{Z}_t$	
p	m			$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$
1	6	mean		0.403	0.241	0.269	0.208	0.359	0.308	0.404	0.318	0.473	0.376
		sd (.37)	(.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 (.29)	(.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 (.23)	(.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 (.60)	(.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 (.47)	(.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 (.35)	(.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

Table 3 **MODEL 3, ® = 0.7, CI(1.1, 0.4), Á$_1$ = 0.34, Á$_2$ = −0.9**

n=256					$q=2$		$q=1$		$\mathbf{X}_t, \widehat{Z}_t$		$\Delta\mathbf{X}_t, \Delta\widehat{Z}_t$		$\Delta\mathbf{X}_t, \widehat{Z}_t$	
p	m				$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$
1	12	mean			0.818	0.392	0.663	0.351	0.715	0.502	0.402	0.287	0.740	0.602
		sd (.29)		(.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 (.19)		(.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 (.15)		(.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 (.47)		(.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 (.29)		(.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 (.22)		(.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					$q=2$		$q=1$		$\mathbf{X}_t, \widehat{Z}_t$		$\Delta\mathbf{X}_t, \Delta\widehat{Z}_t$		$\Delta\mathbf{X}_t, \widehat{Z}_t$	
p	m				$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$
1	6	mean			0.818	0.306	0.596	0.244	0.745	0.539	0.48	0.396	0.773	0.662
		sd (.37)		(.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 (.29)		(.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 (.23)		(.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 (.60)		(.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 (.47)		(.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 (.35)		(.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

Table 4 **MODEL 1' ® $= \mathbf{1}$, CI$(\mathbf{1}, \mathbf{0})$**

Á$_{\mathbf{1}}$ $= \mathbf{0.3}$

n=256				$q=2$		$q=1$		$\mathbf{X}_t, \widehat{Z}_t$		$\Delta\mathbf{X}_t, \Delta\widehat{Z}_t$		$\Delta\mathbf{X}_t, \widehat{Z}_t$	
p	m			$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$
1	12	mean		1.024	0.539	0.877	0.507	0.936	0.683	0.594	0.430	0.924	0.736
		sd (.29)	(.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 (.19)	(.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 (.15)	(.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

Á$_{\mathbf{1}}$ $= \mathbf{0.6}$

n=256				$q=2$		$q=1$		$\mathbf{X}_t, \widehat{Z}_t$		$\Delta\mathbf{X}_t, \Delta\widehat{Z}_t$		$\Delta\mathbf{X}_t, \widehat{Z}_t$	
p	m			$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$
1	12	mean		0.837	0.412	0.737	0.429	0.866	0.666	0.621	0.486	0.841	0.696
		sd (.29)	(.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 (.19)	(.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 (.14)	(.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

Table 5 **MODEL 2' ® $= \mathbf{0.4}$, CI$(\mathbf{1.3}, \mathbf{0.9})$**

Á$_{\mathbf{1}}$ $= \mathbf{0.3}$

n=256				$q=2$		$q=1$		$\mathbf{X}_t, \widehat{Z}_t$		$\Delta\mathbf{X}_t, \Delta\widehat{Z}_t$		$\Delta\mathbf{X}_t, \widehat{Z}_t$	
p	m			$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$
1	12	mean		0.251	0.153	0.185	0.150	0.265	0.225	0.302	0.269	0.390	0.338
		sd (.29)	(.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 (.19)	(.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 (.15)	(.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

Á$_{\mathbf{1}}$ $= \mathbf{0.6}$

n=256				$q=2$		$q=1$		$\mathbf{X}_t, \widehat{Z}_t$		$\Delta\mathbf{X}_t, \Delta\widehat{Z}_t$		$\Delta\mathbf{X}_t, \widehat{Z}_t$	
p	m			$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$
1	12	mean		0.168	0.069	0.122	0.074	0.225	0.207	0.248	0.251	0.339	0.323
		sd (.29)	(.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 (.19)	(.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 (.14)	(.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

Table 6 **MODEL 3' ® $= \mathbf{0.7}$, CI$(\mathbf{1.1, 0.4})$**

Á$_1$ $= \mathbf{0.3}$

n=256				$q=2$		$q=1$		$\mathbf{X}_t, \widehat{Z}_t$		$\Delta\mathbf{X}_t, \Delta\widehat{Z}_t$		$\Delta\mathbf{X}_t, \widehat{Z}_t$	
p	m			$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$
1	12	mean		0.660	0.354	0.544	0.357	0.663	0.490	0.543	0.424	0.664	0.529
		sd (.29)	(.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 (.19)	(.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 (.15)	(.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

Á$_1$ $= \mathbf{0.6}$

n=256				$q=2$		$q=1$		$\mathbf{X}_t, \widehat{Z}_t$		$\Delta\mathbf{X}_t, \Delta\widehat{Z}_t$		$\Delta\mathbf{X}_t, \widehat{Z}_t$	
p	m			$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}$	$\widehat{\alpha}_\Delta$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$	$\widehat{\alpha}_L$	$\widehat{\alpha}_G$
1	12	mean		0.492	0.211	0.415	0.236	0.593	0.471	0.467	0.421	0.572	0.510
		sd (.29)	(.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 (.19)	(.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 (.14)	(.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

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

Figure 2: Coherence of bond rates series

Figure 3: Log-coherence of bond rates series