Non-stationary log-periodogram regression

Carlos Velasco*

Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe (Madrid), Spain

Received 1 July 1997; received in revised form 1 July 1998; accepted 12 August 1998

Abstract

We study asymptotic properties of the log-periodogram semiparametric estimate of the memory parameter d for non-stationary ($d \geqslant \frac{1}{2}$) time series with Gaussian increments, extending the results of Robinson (1995) for stationary and invertible Gaussian processes. We generalize the definition of the memory parameter d for non-stationary processes in terms of the (successively) differentiated series. We obtain that the log-periodogram estimate is asymptotically normal for $d \in [\frac{1}{2}, \frac{3}{4})$ and still consistent for $d \in [\frac{1}{2}, 1)$. We show that with adequate data tapers, a modified estimate is consistent and asymptotically normal distributed for any d, including both non-stationary and non-invertible processes. The estimates are invariant to the presence of certain deterministic trends, without any need of estimation.

JEL classification: C11; C22

Keywords: Non-stationary time series; Log-periodogram regression; Semiparametric inference; Tapering

1. Introduction

Statistical inference for stationary long range dependent time series is often based on semiparametric estimates that avoid parameterization of the short run behaviour. One of most popular semiparametric estimates in the frequency domain is the log-periodogram regression, proposed initially by Geweke and Porter-Hudak (1983). Robinson (1995) showed the consistency and asymptotic normality of a version of that estimate for stationary and invertible Gaussian

* E-mail: cavelas@est-econ.uc3m.es

vector time series. He assumed that the spectral density $f(\lambda)$ of the observed stationary sequence satisfies for one constant $0 < G < \infty$,

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

where $d \in (-\frac{1}{2}, \frac{1}{2})$ is the parameter that governs the memory of the series. This is the interval of values of d for which the process is stationary and invertible. If $d \in (0, \frac{1}{2})$ then we say that the series exhibits long memory or long range dependence. Expression (1) reflects a linear relationship between the spectral density and the frequency in log–log coordinates, with slope $-2d$. This, together with the fact that the periodogram ordinates at Fourier frequencies around the origin are still approximately independent and unbiased for the spectral density f in the long memory case (1), constitute the basis for the log-periodogram estimate.

There have been proposals to extend the applicability of the log-periodogram estimate for non-stationary ($d \geqslant \frac{1}{2}$) or non-invertible ($d \leqslant -\frac{1}{2}$) time series, and indeed log-periodogram regressions have been applied to non-stationary observations (e.g. Agiakloglou et al., 1993; Bloomfield, 1991), opening the question of analysing the properties of the estimates when they give values outside the interval of allowed values of d. For $d \geqslant \frac{1}{2}$, a function $f(\lambda)$ behaving like Eq. (1) can be defined in terms of the differenced series, but it is no longer a spectral density, since it is not integrable and the time series is non-stationary with infinite variance. Hassler (1992) used the log-periodogram estimate to construct a unit root test ($d = 1$), but he gave no theoretical justification for his asymptotic theory in the non-stationary case. Hurvich and Ray (1995) studied the behaviour of the expectation of the periodogram at low Fourier frequencies for Gaussian non-stationary and non-invertible fractionally integrated processes. They showed that the normalized periodogram has bounded expectation for $d \in [\frac{1}{2}, \frac{3}{2})$ but it is biased (for the function f) in this case, and they proposed to taper the data with the full cosine window in order to reduce this bias.

Robinson (1995) advocated an initial differentiation (integration) of the observed time series when non-stationarity (non-invertibility) is suspected to obtain a value of d in the stationary and invertible interval $(-\frac{1}{2}, \frac{1}{2})$ and then perform the periodogram regression on the transformed series, adjusting the estimate with the number of differences (integrations) taken. However, the simulation work of Hassler (1992) and Hurvich and Ray (1995) suggests that, at least for values $d \in [\frac{1}{2}, 1)$, the estimation procedure using the original series can be consistent, although it will not coincide in general with the pre-differenced estimate.

Using Hurvich and Ray's definitions we extend Robinson's (1995) results to cover the non-stationary case. We find that in the Gaussian case the log-periodogram estimate is asymptotically normal for $d < \frac{3}{4}$ and still consistent for

$d < 1$. Here we are trying to approximate a different function than in the stationary situation, explaining the discrepancy with respect to the estimates which use previously differentiated observations. When we taper the periodogram with the cosine window, as suggested by Hurvich and Ray (1995), we show that the estimate is asymptotically normal even for $d < \frac{3}{2}$.

We also consider a general non-stationary model for any $d \geqslant \frac{1}{2}$, where the presence of deterministic time trends is allowed, and show that it is possible to design data tapers which deliver asymptotic normally distributed estimates of d under Gaussianity. The main idea is the same as in, e.g. Zhurbenko (1979, 1980, 1982), Robinson (1986) or Dahlhaus (1988), who showed that certain tapers or data windows allow statistical inference in the presence of non-stationary properties at certain frequencies. Their analyses used the improved convergence properties of the spectral window of some tapers and we will require those and some other special features to deal with the stochastic trends of non-stationary processes. The same principle will make the estimates robust to deterministic time trends up to certain order, avoiding any trend specification, testing or estimation as in most of non-stationary inference literature, both with the autoregressive approach (e.g. Durlauf and Phillips, 1988) or in the fractional differencing framework (Robinson, 1994b). Related ideas allow also the estimation of $d \leqslant -\frac{1}{2}$ for Gaussian non-invertible processes that may arise in overdifferencing to eliminate stochastic and deterministic trends. These properties enable us to abstract from deterministic behaviours and concentrate on the stochastic trends and their implications on the non-invertibility ($d \leqslant -\frac{1}{2}$), non-stationarity ($d \geqslant \frac{1}{2}$), mean reversion ($d < 1$), etc., of the observed time series.

Finally, we analyse empirically the performance of the estimates for finite sample sizes. We show how to base a choice of the degree of tapering, identifying when it produces biased estimates for all possible choices of a bandwidth parameter.

The paper is organized as follows. First we give the main assumptions and definitions. In Section 3 we study the non-tapered situation and in Section 4 we analyse the cosine bell window taper. Then we consider in Section 5 a general model for non-stationary time series and suitable data windows for their analysis and in Section 6 we apply the same methods to non-invertible processes. In Section 7 we analyse the performance of the estimates proposed for simulated data. Then we conclude and give some proofs and technical lemmas in three appendices.

2. Assumptions and definitions

Following Hurvich and Ray (1995), we say that the non-stationary process $\{X_t\}$ has memory parameter d ($\frac{1}{2} \leqslant d < \frac{3}{2}$) if the zero mean stationary process

$\varepsilon_t = \Delta X_t$ has spectral density

$$f_\varepsilon(\lambda) = |1 - \exp(\mathrm{i}\lambda)|^{-2(d-1)} f^*(\lambda),$$

where $f^*(\lambda)$ is a positive, integrable, even function on $[-\pi, \pi]$ which is bounded above and away from zero and is continuous at $\lambda = 0$. We will relax this assumption later, and consider a more general non-stationary process. Then, we can write, for any $t \geqslant 1$,

$$X_t = X_0 + \sum_{k=1}^{t} \varepsilon_k,$$

where X_0 is a random variable not depending on time t. Define the function

$$f(\lambda) = |1 - \exp(\mathrm{i}\lambda)|^{-2} f_\varepsilon(\lambda) = |1 - \exp(\mathrm{i}\lambda)|^{-2d} f^*(\lambda) = |2\sin(\lambda/2)|^{-2d} f^*(\lambda), \quad (2)$$

so $f(\lambda)$ satisfies Eq. (1). Note that $2d \geqslant 1$, so f is not integrable in $[-\pi, \pi]$ and is not a spectral density. We do not assume that f^* is the spectral density of a stationary and invertible ARMA process as would be the case if ε_t followed a fractional ARIMA model. Here f^* may have (integrable) poles or zeroes at frequencies beyond the origin.

We introduce now the following assumptions about the behaviour of the spectral density $f_\varepsilon(\lambda)$ (and thus of the functions $f(\lambda)$ and $f^*(\lambda)$) at the origin:

Assumption 1. The spectral density $f_\varepsilon(\lambda)$ satisfies for numbers $0 < \alpha \leqslant 2$, $0 < G < \infty$, $d \in [\frac{1}{2}, \frac{3}{2})$,

$$f_\varepsilon(\lambda) = G\lambda^{-2(d-1)} + \mathrm{O}(\lambda^{-2(d-1)+\alpha}) \quad \text{as } \lambda \to 0^+.$$

Under Assumption 1 we write, defining the function $g(\lambda) = G|\lambda|^{-2d}$, $0 < \alpha \leqslant 2$,

$$f(\lambda)/g(\lambda) = 1 + \mathrm{O}(\lambda^\alpha) \quad \text{as } \lambda \to 0^+. \quad (3)$$

This is equivalent to Assumption 1 in Robinson (1995) when f is the spectral density of X_t (stationary) and $d \in (-\frac{1}{2}, \frac{1}{2})$. See also Remark 3.1 in Giraitis et al. (1997).

Assumption 2. The spectral density $f_\varepsilon(\lambda)$ satisfies for numbers $0 < \alpha \leqslant 2$, $0 < G, E_\alpha < \infty$, $d \in [\frac{1}{2}, \frac{3}{2})$,

$$f_\varepsilon(\lambda) = G\lambda^{-2(d-1)} + G\,E_\alpha\lambda^{-2(d-1)+\alpha} + \mathrm{o}(\lambda^{-2(d-1)+\alpha}) \quad \text{as } \lambda \to 0^+.$$

This assumption implies obviously Assumption 1 and holds if $f_\varepsilon(\lambda) = g(\lambda)h(\lambda)$, $h(0) = 1$, with $h(\lambda)$ even, satisfying either a Lipschitz property around the origin of order α, for $0 < \alpha \leqslant 1$, or it is differentiable with derivative in $\mathrm{Lip}(\alpha - 1)$, for $1 < \alpha \leqslant 2$. Then, under Assumption 2 we can write, with the same definitions as before that, $0 < \alpha \leqslant 2$,

$$\frac{f(\lambda)}{g(\lambda)} = 1 + \mathrm{E}_\alpha \lambda^\alpha + \mathrm{o}(\lambda^\alpha) \quad \text{as } \lambda \to 0^+. \tag{4}$$

This last expression is now equivalent to Assumption 3 in Robinson (1994a) and was used also by Velasco (1997) to study the behaviour of the tapered periodogram for stationary long memory time series. Both assumptions are satisfied with $\alpha = 2$ if f_ε is the spectral density of a stationary, invertible fractional ARIMA process or fractional Gaussian noise, when $d > \frac{1}{2}$, so $d - 1 \in (-\frac{1}{2}, \frac{1}{2})$. With $d = \frac{1}{2}$, ε_t is not invertible but stationary.

Also, both Assumptions 1 and 2 imply that $f^*(\lambda)$ is bounded above and away from zero and is continuous in an interval $(0, \varepsilon)$, $\varepsilon > 0$. Finally we introduce

Assumption 3. In a neighbourhood $(0, \varepsilon)$ of the origin, $f_\varepsilon(\lambda)$ is differentiable and

$$\left| \frac{\mathrm{d}}{\mathrm{d}\lambda} f_\varepsilon(\lambda) \right| = \mathrm{O}(\lambda^{-1-2(d-1)}) \quad \text{as } \lambda \to 0^+.$$

Then $f(\lambda)$ has first derivative satisfying (cf. Assumption 2 of Robinson (1995) in the stationary case $d < \frac{1}{2}$),

$$\left| \frac{\mathrm{d}}{\mathrm{d}\lambda} f(\lambda) \right| = \mathrm{O}(\lambda^{-1-2d}) \quad \text{as } \lambda \to 0^+. \tag{5}$$

These assumptions could have been formulated in terms of the functions f and/or f^*, since we are precisely interested in the implications they have on the function f, Eqs. (3)–(5). However, we did not find appropriate to make assumptions directly on f or f^*, since these functions have no immediate and clear statistical interpretation as f_ε has.

Define the discrete Fourier transform of X_t for n observations, $t = 1, \ldots, n$, at Fourier frequencies $\lambda_j = 2\pi j/n$, j is an integer,

$$w(\lambda_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t \exp(\mathrm{i}\lambda_j t) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \sum_{k=1}^{t} \varepsilon_k \exp(\mathrm{i}\lambda_j t), \tag{6}$$

so $w(\lambda_j)$ is a complex linear combination of the stationary variables ε_k. The Fourier transform at any frequency λ_j, $0 < j < n$, of the sequence X_t allows the elimination of the random variable X_0, because $\sum_{t=1}^{n}\exp(\mathrm{i}\lambda_j t) = 0$ for $j \neq 0 \,(\mathrm{mod}\, n)$, so $w(\lambda_j)$ is not depending on the past values of ε_k for $k \leqslant 0$ possibly contained in X_0.

Defining the periodogram of X_t as

$$I(\lambda_j) = |w(\lambda_j)|^2,$$

we can consider Eq. (1) as an approximate semiparametric model under stationarity for the spectral density f at the first m Fourier frequencies λ_j, $j = 1, \ldots, m$, closest to the origin, so taking logarithms we obtain

$$\log f(\lambda_j) \approx \log G - 2d \log \lambda_j,$$

and adding $\log I(\lambda_j)$ to both sides, rearranging terms,

$$\log I(\lambda_j) \approx \log G - 2d \log \lambda_j + \log \frac{I(\lambda_j)}{f(\lambda_j)}. \tag{7}$$

In the weak dependence case, $d = 0$, the normalized random variables $I(\lambda_j)/f(\lambda_j)$ are asymptotically i.i.d. distributed (with $\chi_2^2/2$ distribution), so Eq. (7) is an approximate linear regression model with response variable $\log I(\lambda_j)$, regressor $-2\log \lambda_j$, and slope d, and standard methods could be used to estimate d if the above properties of the normalized periodogram extend to long memory ($d > 0$) or antipersistent ($d < 0$) processes. The log-periodogram regression estimate is just the ordinary least squares (OLS) solution as proposed by Geweke and Porter-Hudak (1983) with λ_j substituted by the asymptotically equivalent quantity $2 \sin \lambda_j/2$.

Among other issues, Robinson (1995) modified the OLS estimate considering the logs of a pooled periodogram of $J = 1, 2, \ldots$ (a fixed number of) periodogram ordinates,

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

(assuming $(m - l)/J$ integer) and showed that for Gaussian stationary and invertible time series the estimate

$$\hat{d} = \left(\sum_k \Lambda_k^2 \right)^{-1} \left(\sum_k \Lambda_k Y_k^{(J)} \right), \tag{8}$$

is consistent and asymptotically normal. Here $\Lambda_k = z_k - \bar{z}$, where $z_k = -2\log \lambda_k$ and $\bar{z} = \{J/(m-l)\}\sum_k z_k$ and the pooling $J > 1$ increases efficiency. The bandwidth number m is an integer smaller than n, and l is a user-chosen trimming number to avoid the very first periodogram ordinates, which have undesirable large sample properties. In the asymptotics both numbers tend to infinity with the sample size n, but more slowly.

The main idea to show that the previous Robinson's (1995) results go through in the non-stationary case ($d \geqslant \frac{1}{2}$) is to analyse the asymptotic behaviour of the discrete Fourier transform of X_t for frequencies λ_j, $l < j \leqslant m$. We will show that under some assumptions this behaviour is equivalent to the stationary case with respect to the function f defined in Eq. (2). Therefore, assuming Gaussianity for the ε_k's, we could repeat the steps in Robinson (1995) to obtain the consistency and asymptotic distribution of the log-periodogram estimate of the parameter d for non-stationary processes. This is possible, because the proof of Theorem 3 in Robinson (1995) only uses the error in the estimation of the covariance matrix of the discrete Fourier transforms at low frequencies and the Gaussianity of the discrete Fourier transform of X_t (implied by Eq. (6)).

The covariance matrix of $w(\lambda_j)$ can be studied in a similar way as in the stationary framework, extending Hurvich and Ray's (1995) analysis of the expectation of the periodogram. However, due to a bias problem, the same results as in Robinson (1995) can only be obtained for $d < \frac{3}{4}$ (consistency holds for $d < 1$). This problem can be overcome, as Hurvich and Ray (1995) suggested, with tapering. For example, tapering the data with the full cosine bell, allows the asymptotic normality of the estimate of d for any $d < \frac{3}{2}$, since it alleviates slightly the global bias problem for these values of d but will not be operative for bigger values (see discussion in Section 5).

3. Non-tapered periodogram

In this section we analyse the asymptotic properties of $\hat{d}$ as defined previously in Eq. (8), in terms of the raw (non-tapered) periodogram for non-stationary time series. We analyse the univariate case for simplicity, but the multivariate model does not involve new ideas and can be dealt with as in Robinson (1995), since the relationships between the elements of the spectral density matrix of ε_t go through for a matrix function $f(\lambda)$, although the interpretation is different.

Under Assumptions 1 and 3, the conditions on the behaviour of the function $f(\lambda)$ at the origin of Theorem 2 in Robinson (1995) hold, now for $d \in [\frac{1}{2}, \frac{3}{2})$. If the bar $\bar{\ }$ stands for complex conjugation and denoting the discrete Fourier transform of (Eq. (6)) by $w_j = w(\lambda_j)$, we have to analyse the covariances between the normalized versions of $[w_j, \overline{w_j}]$, $[w_j, w_j]$, $[w_j, w_k]$ and $[w_j, \overline{w_k}]$, for $k < j$, corresponding to parts (a)–(d) of Theorem 2 of Robinson (1995). Defining $v(\lambda) = w(\lambda)/(G^{1/2}\lambda^{-d})$, our first result is

Theorem 1. Under Assumptions 1 $[0 < \alpha \leqslant 2]$ *and* 3, $d \in [\frac{1}{2}, 1)$, *for any sequences of positive integers* $j = j(n)$ *and* $k = k(n)$ *such that* $0 < k < j$ *and* $j/n \to 0$ *as* $n \to \infty$, *defining* $\gamma_{k,j} = (jk)^{d-1}\log(k+1)$,

(a) $\mathrm{E}[v(\lambda_j)\overline{v}(\lambda_j)] = 1 + \mathrm{O}(\gamma_{j,j} + [j/n]^{\alpha})$,
(b) $\mathrm{E}[v(\lambda_j)v(\lambda_j)] = \mathrm{O}(\gamma_{j,j})$,
(c) $\mathrm{E}[v(\lambda_j)\overline{v}(\lambda_k)] = \mathrm{O}(k^{-1}\log(j+1) + \gamma_{k,j})$,
(d) $\mathrm{E}[v(\lambda_j)v(\lambda_k)] = \mathrm{O}(k^{-1}\log(j+1) + \gamma_{k,j})$.

Proof. See Appendix A. □

This result is weaker with respect to Robinson's (1995) stationary version for the extra term $\gamma_{k,j}$, and it is valid only for $d < 1$, making sense with Hurvich and Ray (1995) observation that the bias of the periodogram decreases as j grows only for $d < 1$ (but otherwise increases). Note that the theorem is valid for $j = 1$ in the sense that the expectations are bounded O(1) in n.

The intuition why the normalized periodogram is unbiased (and the discrete Fourier transforms at different frequencies are asymptotically uncorrelated) for non-stationary time series and increasing indices j and k is the following. It is possible to show that the expectation of the periodogram can be written like in the stationary case as

$$\mathrm{E}[I(\lambda_j)] = \int_{-\pi}^{\pi} f(\alpha)K(\lambda_j - \alpha)\,\mathrm{d}\alpha,$$

a convolution of f and the Fejér kernel

$$K(\lambda) = \frac{1}{2\pi n}|D(\lambda)|^2 = \frac{1}{2\pi n}\frac{\sin^2[n\lambda/2]}{\sin^2[\lambda/2]}, \tag{9}$$

where

$$D(\lambda) = \sum_{t=1}^{n} \mathrm{e}^{\mathrm{i}\lambda t} \tag{10}$$

is Dirichlet kernel and now f is a non-integrable function (so it is not a spectral density). However, Fejér kernel $K(\lambda)$ has zeroes of order 2 for all Fourier frequencies λ_j, $j \neq 0 \pmod n$, and this compensates for any pole in $f(\lambda)$ at the origin of order less than 3, i.e. $d < \frac{3}{2}$, just using the integrability of f outside the origin, implied by the integrability of the spectral density f_ε. This implies a bounded expectation for the normalized periodogram for $d < \frac{3}{2}$ at λ_j, but only unbiasedness for $d < 1$ when j is increasing with n.

Now we can show the consistency of $\hat{d}$ when $d < 1$:

Theorem 2. Under the assumptions of Theorem 1, ε_t Gaussian and

$$\frac{\log m}{l^{2(1-d)}} + \frac{1}{m-l} + \frac{(\log n)^2}{m} + \frac{m}{n} \to 0 \quad as\ n \to \infty, \tag{11}$$

the estimate $\hat{d}$ defined in Eq. (8) is consistent for d.

Proof. From Theorem 1, for $d < 1$ and frequencies λ_j, $j = l, \ldots, m$, with l increasing slowly with n, from the first condition in Eq. (11), the normalized discrete Fourier transforms of X_t have exactly the same first two moments structure as in the stationary and invertible case $(-\frac{1}{2} < d < \frac{1}{2})$. Then, given the Gaussianity assumption for ε_t, the Fourier transforms are also Gaussian distributed because they are a linear combination of Gaussian variables from Eq. (6). Then, following Remark 8 of Robinson (1992), the estimate of $d < 1$ will be consistent with condition (11). □

We observe that the trimming has to be more important (i.e. l increasing faster) as d approaches 1, since the function f is steeper. For values $d \geqslant 1$, the periodogram is not unbiased for the function f as j increases, and therefore the log-periodogram estimator $\hat{d}$ cannot be consistent. The asymptotic normality of $\hat{d}$ needs stronger assumptions on the trimming and bandwidth numbers to control the bias and can only be obtained for $d < \frac{3}{4}$:

Theorem 3. Under the assumptions of Theorem 1, with $d \in [\frac{1}{2}, \frac{3}{4})$, ε_t Gaussian and

$$\frac{m^{1/2} \log m}{l^{2(1-d)}} + \frac{l(\log n)^2}{m} + \frac{m^{1+1/2\alpha}}{n} \to 0 \quad as\ n \to \infty, \tag{12}$$

we obtain for $\hat{d}$ defined in Eq. (8) that

$$m^{1/2}(\hat{d} - d) \underset{\mathrm{d}}{\to} \mathrm{N}\left(0, \frac{J}{4}\psi'(J)\right),$$

where ψ' is the digamma function $\psi'(x) = (\mathrm{d}/\mathrm{d}x)\log \Gamma(x)$.

Proof. Further to the comments in the proof of the previous theorem, from Eq. (5.17) in Robinson (1995), we need the error terms in the covariance matrix of the discrete Fourier transform of Theorem 1 to be $o(m^{-1/2})$ and that l is tending to infinity slower than m to obtain $lm^{-1}(\log m)^2 = o(1)$ (see formulae after expression (5.4) in Robinson, 1995, p. 1066)). From Theorem 1, Eq. (12) is sufficient for that, but the choice of l and m is only possible for $d < \frac{3}{4}$ considering the first condition in Eq. (12). □

Note that this result for $d < \frac{3}{4}$ is exactly the same as in the stationary case, and that the asymptotic distribution of $\hat{d}$ does not depend on any unknown parameter. However when d is very close to the boundary $\frac{3}{4}$ the choice of the number l is very limited by the first condition in Eq. (12), and will depend on the true value of d. For example if $m = n^{4/5}$, in the boundary of the last condition of Eq. (12), and $d = 0.7$, we need simultaneously for l that $ln^{-0.8}(\log n)^2 \to 0$ and $n^{0.67} l^{-1} \to 0$. The limitations in the asymptotics are due to the extra bias in the estimation of the elements of the covariance matrix of the discrete Fourier transform because of the behaviour of f when $d \geqslant \frac{1}{2}$. Basically, the periodogram is asymptotically unbiased at λ_j as j increases only when $d < 1$, and the order magnitude of the bias depends on the value of d, unlike in the stationary case. Furthermore, the bounds for the biases of the covariance matrix of the Fourier transforms are not sufficient for the asymptotic normality for $d \geqslant \frac{3}{4}$.

One possible solution, as pointed out by Hurvich and Ray (1995), is the use of tapering. We will show that tapering allows a reduction of the order of magnitude of the bounds in Theorem 1, so we can estimate bigger values of d. Thus, with the cosine bell taper all the results go through for any $d < \frac{3}{2}$, since this data taper achieves a reduction of the overall bias from Robinson's (1995) results if f is smooth enough. This was observed by Velasco (1997) for a related problem with non-Gaussian stationary time series. However, as we will see in next section, the full advantage of the tapering improvement in the convergence in the tails of the spectral kernel, only shows up when we use Assumption 2 with $\alpha \geqslant 1$, increasing the smoothness of the function f near the origin. In Section 5 we find that other tapers reduce the bias even more and allow the consideration of series with $d \geqslant \frac{3}{2}$ and stationary increments with mean different from zero.

4. Cosine bell tapered periodogram

We consider in this section the full cosine bell taper, as suggested by Hurvich and Ray (1995). The tapered discrete Fourier transform for any taper sequence $\{h_t\}_{t=1}^n$ is defined as

$$w^T(\lambda_j) = \frac{1}{\sqrt{2\pi \sum h_t^2}} \sum_{t=1}^{n} h_t X_t \exp(\mathrm{i}\lambda_j t).$$

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

The benefits of tapering derive from the following properties of the cosine bell taper. We have (Bloomfield, 1976, pp. 80–84) or Percival and Walden, 1993, pp. 325–326)) that for $2 \leqslant j \leqslant n-2$ the tapered Fourier transform at λ_j is a linear combination of the usual Fourier transform at the frequencies λ_j, λ_{j-1} and λ_{j+1},

$$w^T(\lambda_j) = \frac{1}{\sqrt{6}}[-w(\lambda_{j-1}) + 2w(\lambda_j) - w(\lambda_{j+1})]. \tag{13}$$

Note that $\sum_{t=1}^{n} h_t \exp(\mathrm{i}\lambda_j t) = 0$. Then, the spectral kernel for the tapered periodogram, corresponding to Fejér kernel $K(\lambda)$ for the periodogram is

$$K^T(\lambda_j - \lambda) = \frac{1}{2\pi \sum h_t^2} |D^T(\lambda_j - \lambda)|^2 = \frac{1}{2\pi \sum h_t^2} \sin^2[n(\lambda_j - \lambda)/2]\, H_j^2(\lambda), \tag{14}$$

where

$$H_j(\lambda) = \frac{1}{\sqrt{6}} \left\{ \frac{2}{\sin[(\lambda_j - \lambda)/2]} - \frac{1}{\sin[(\lambda_{j-1} - \lambda)/2]} - \frac{1}{\sin[(\lambda_{j+1} - \lambda)/2]} \right\}, \tag{15}$$

and

$$D^T(\lambda) = \sum_{t=1}^{n} h_t \exp\{\mathrm{i}t\lambda\} \tag{16}$$

is the equivalent of the Dirichlet kernel $D(\lambda)$, (Eq. (10)), in the non-tapered case, from Eq. (13) equal to

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

Then $K^T(\lambda)$ is even, positive, integrates to one and satisfies (see, e.g., Bloomfield (1976) or Hannan, 1970, p. 265)) $\sup_{\lambda,n}|K^T(\lambda)| = O(\min\{n, n^{-5}|\lambda|^{-6}\})$. This property derives from the fact that $\sup_{\lambda,n}|D^T(\lambda)| = O(\min\{n, n^{-2}|\lambda|^{-3}\})$, so the tapered periodogram $I^T(\lambda_j) = |w^T(\lambda_j)|^2$ has improved asymptotic properties with respect to the usual periodogram, because the tails of the kernel $K^T(\lambda)$ decrease much faster with the frequency and with the sample size than the tails of Fejér kernel K. Therefore, we will be able to reduce the bias of the periodogram on the tails, even for frequencies close to a singularity and for non-integrable functions, if they are smooth enough.

As both functions, $K(\lambda)$ and $K^T(\lambda)$, integrate to one, there has to be a trade off between the behaviour of the kernels at the origin and at the tails, i.e. the tails of K^T are less thick than those of K, but the central lobe is much wider. This is the reason why we only can consider tapered periodogram ordinates or discrete Fourier transforms that are at least three basic frequencies $\lambda_1 = 2\pi/n$ away. Furthermore, the order of the zero of K^T at λ_j, $j = 1, 2, \ldots, n-1$, given by the function $\sin^2[n\lambda/2]$, is of the same order, 2, as in the case of Fejér kernel, so we cannot consider functions f with $d \geqslant \frac{3}{2}$, as the expectation of the periodogram will always diverge. The covariance structure of the normalized tapered Fourier transform $v^T(\lambda) = w^T(\lambda)/(G^{1/2}\lambda^{-d})$ is given by

Theorem 4. Under Assumptions 2 and 3 $[0 < \alpha \leqslant 2]$, $d \in [\frac{1}{2}, \frac{3}{2})$ *for any sequences of positive integers* $j = j(n)$ *and* $k = k(n)$, $1 < k$, j *and* $k + 2 < j$, *such that* $j/n \to 0$ *and defining* $\gamma_{j,k} \equiv (jk)^{d-3} \log k$,

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

Proof. See Appendix B. □

This result confirms Hurvich and Ray (1995) observation that the tapered periodogram is unbiased for f, even for values of d close to $\frac{3}{2}$. If $\alpha \leqslant 1$, it would be enough to consider Assumption 1, instead of the stronger Assumption 2. Comparing with Theorem 1 and forgetting about the term $\gamma_{j,k}$ due to the non-integrability of f, we obtain here a substantial improvement in parts (a) (when $\alpha \geqslant 1$) and (b), reducing the bounds, at most, to $\mathrm{O}(j^{-2})$ (for $\alpha = 2$) and to $\mathrm{O}(j^{-3} \log j)$ (for $d = \frac{3}{2}$), respectively. However, in parts (c) and (d) we only manage to eliminate the log factor. This is due to the reason pointed out before: K^T has better behaviour on the tails, but not in its central lobe, so in parts (c) and (d) we cannot improve too much if the numbers j and k can be arbitrarily close, satisfying only $j > k + 2$.

Defining $\hat{d}^T$ now as

$$\hat{d}^T = \left(\sum_k \Lambda_k^2 \right)^{-1} \left(\sum_k \Lambda_k Y_k^{(T,J)} \right), \tag{17}$$

with Λ_k as in (Eq. (8)) and with the pooled tapered periodogram ordinates

$$Y_k^{(T,J)} = \log\left(\sum_{j=1}^{J} I^T(\lambda_{k+3(j-J)}) \right), \quad k = l + 3J, l + 6J, \ldots, m,$$

and using Theorem 4, we can obtain, similarly to Theorem 3, *Theorem* 5. *Under the assumptions of Theorem* 4, ε_t *Gaussian and*

$$\frac{m^{1/2}}{l}+\frac{l(\log n)^2}{m}+\frac{m^{1+1/2\alpha}}{n}\to 0 \quad \text{as } n\to\infty, \tag{18}$$

we obtain for $\hat{d}^T$ *defined in Eq. (17),*

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

Proof. As in the proof of Theorem 3, we first observe that for $d\in[\frac{1}{2},\frac{3}{2})$ the uniform bound for the bias errors in the covariance matrix of the tapered discrete Fourier transforms at the frequencies considered in the definition of $\hat{d}^T$, is $o(m^{-1/2})$, using the last condition in (Eq. (18)). Hence, under (Eq. (18)) the asymptotic uncorrelatedness and then independence of w^T are enough to make valid all the asymptotic results of Robinson (1995) for $\hat{d}^T$ and non-stationary processes with memory parameter $d\in[\frac{1}{2},\frac{3}{2})$. □

Note that the conditions on the bandwidths are now slightly milder with respect to Theorem 3, since we do not have the term in $\log m$ thanks to tapering from Theorem 4. This result is again in line with Hurvich and Ray (1995) empirical findings for $\hat{d}^T$ and $d\geqslant 1$. In this case the choice of bandwidth and trimming numbers does not depend on the value of d, even when it is arbitrarily close to $\frac{3}{2}$. Also it tells us that, for any value of d, although tapering might reduce the bias of the periodogram and therefore of the estimate of d, it increases the variance by a factor of 3, due to the modification in the definition of $\hat{d}^T$ with respect to $\hat{d}$. We conjecture that this modification could be avoided, using all the Fourier frequencies, resulting in an increment of the variance of the estimate due to the autocorrelation between adjacent tapered Fourier transforms $w^T(\lambda_j)$, which are moving averages of the approximately independent $w(\lambda_j)$ in view of Eq. (13). However in this case Robinson's (1995) results cannot be applied directly since they are based on the asymptotic independence of those transforms.

5. General non-stationary processes and data tapers

In this section we propose a general model for non-stationary time series $d\geqslant\frac{1}{2}$ and show how to extend the previous ideas to the estimation of the memory parameter d when we use appropriate data tapers. The consideration of processes with, e.g. one or two unit roots in the classical sense will also lead us to the discussion of polynomic trends of time and how to discriminate between these deterministic trends and the stochastic trends produced by the integration of (zero mean) processes.

We say that the observed sequence X_t, $t = 1, \ldots, n$, has memory parameter $d > -\frac{1}{2}$ if $\Delta^s X_t = \varepsilon_t^{(s)}$, $s = \lfloor d + \frac{1}{2} \rfloor$, is stationary with mean μ possibly different from zero, and spectral density $f_{\varepsilon^{(s)}}(\lambda)$ behaving as $\lambda^{-2(d-s)}$ around the origin, $d - s \in [-\frac{1}{2}, \frac{1}{2})$. In Section 2 we have considered the case $s = 1$ which only covers $d < \frac{3}{2}$ and $\mu = 0$. Denote for $r = 1, 2, \ldots, s$, $\Delta^r X_t = \varepsilon_t^{(r)}$, so the function

$$f(\lambda) = |1 - \exp(i\lambda)|^{-2s} f_{\varepsilon^{(s)}}(\lambda) = |2 \sin(\lambda/2)|^{-2d} f^*(\lambda) \tag{19}$$

is defined as before in terms of the spectral density of the stationary sequence $\varepsilon_t^{(s)}$, $f_{\varepsilon^{(s)}}$, and the unit root transfer functions.

Extending the discussion in Hurvich and Ray (1995), we can write for random variables X_0, $\varepsilon_0^{(r)}$, $r = 1, \ldots, s-1$ which do not depend on time,

$$\begin{aligned} X_t &= X_0 + \sum_{j_1=1}^{t} \varepsilon_{j_1}^{(1)} \\ &= X_0 + \sum_{j_1=1}^{t} \left(\varepsilon_0^{(1)} + \sum_{j_2=1}^{j_1} \varepsilon_{j_2}^{(2)} \right) \\ &= X_0 + \sum_{r=1}^{s-1} \varepsilon_0^{(r)} p^{(r)}(t) + \mu p_\mu(t) + \sum_{j_1=1}^{t} \sum_{j_2=1}^{j_1} \cdots \sum_{j_s=1}^{j_{s-1}} \varepsilon_{j_s}^{(\star)}, \end{aligned}$$

where $p^{(r)}(t)$ are polynomials in t of order r (e.g. $p^{(1)}(t) = t$), $p_\mu(t)$ is a polynomial in t of order s and $\varepsilon_t^{(\star)} = \varepsilon_t^{(s)} - \mu$ has zero mean and the same spectral density as $\varepsilon_t^{(s)}$.

We consider now the discrete Fourier transform of the tapered series $h_t X_t$,

$$w^T(\lambda_j) = \frac{1}{\sqrt{2\pi \sum h_t^2}} \sum_{t=1}^{n} h_t X_t \exp(i\lambda_j t)$$

$$= \frac{1}{\sqrt{2\pi \sum h_t^2}} \sum_{t=1}^{n} h_t \left(X_0 + \sum_{r=1}^{s-1} \varepsilon_0^{(r)} p^{(r)}(t) + \mu p_\mu(t) \right) \exp(i\lambda_j t) \tag{20}$$

$$+ \frac{1}{\sqrt{2\pi \sum h_t^2}} \sum_{t=1}^{n} h_t \sum_{j_1=1}^{t} \sum_{j_2=1}^{j_1} \cdots \sum_{j_s=1}^{j_{s-1}} \varepsilon_{j_s}^{(\star)} \exp(i\lambda_j t). \tag{21}$$

We think of the term (20) as a nuisance term which comprises the information in $\{X_t\}_1^n$ from the past prior to $t = 1$. To make inferences about d we need to eliminate this dependence on the past or initial conditions as we did when $s = 1$ where only X_0 appeared, at least for some frequencies λ_j, by means of certain orthogonality properties of the weights h_t, like

$$\sum_{t=1}^{n} h_t(1 + t + t^2 + \cdots + t^s)\exp(\mathrm{i}\lambda_j t) = 0, \tag{22}$$

which is sufficient to cancel the contribution to the Fourier transform $w^T(\lambda_j)$ of the polynomials in t with unknown coefficients in the term (20). Observe that in the case $s = 1$ we have only required that $\sum_{t=1}^{n} h_t \exp(\mathrm{i}\lambda_j t) = 0$, because we were assuming $\mu = 0$ so the polynomial $p_\mu(t)$ did not show up, and we only needed to eliminate the influence from X_0 (constant with respect to t) and both the raw and cosine bell-tapered Fourier transforms satisfy condition (22) with $s = 0$ (but not for any $s \geqslant 1$). Condition (22) holds when for a particular frequency λ_j, the tapered Dirichlet kernel $D^T(\lambda_j)$ in (Eq. (16)) has all derivatives in λ up to order s equal to zero. The Dirichlet kernel $D(\lambda_j)$ in (Eq. (10)) is zero for all Fourier frequencies λ_j, $0 < j < n$, but its derivative is not zero. The same holds for the cosine bell taper using (Eq. (13)).

Next, we define a general class of data tapers which satisfy the orthogonality condition (22). We will only consider positive tapers symmetric around $n/2$, with $\max_t h_t = 1$. We say then that a sequence of data tapers $\{h_t\}_1^n$ is of *order p* if the following two conditions are satisfied:

- For a function $b = b(n)$, $0 < b < \infty$, $\forall n > 0$,

$$\sum_{t=1}^{n} h_t^2 = bn. \tag{23}$$

- For $N = n/p$ (which we assume as an integer), the Dirichlet kernel D_p^T satisfies

$$\mathrm{D}_p^T(\lambda) \equiv \sum_{t=1}^{n} h_t \exp\{\mathrm{i}\lambda t\} = \frac{a(\lambda)}{n^{p-1}}\left(\frac{\sin[n\lambda/2p]}{\sin[\lambda/2]}\right)^p, \tag{24}$$

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 $\lambda \in [-\pi, \pi]$.

Then, it is immediate to obtain from (Eq. (24)) that, $|\lambda| \leqslant \pi$,

$$|D_p^T(\lambda)| \leqslant \text{const.} \min\{n, n^{1-p}|\lambda|^{-p}\}. \tag{25}$$

This property will permit us to analyse nonparametrically functions $f(\lambda)$ with higher order poles at $\lambda = 0$. Also we have from (Eq. (24)) that $D_p^T(\lambda_{jp})$ has zeroes of order p and condition (22) is satisfied for $s \leqslant p-1$ at frequencies λ_{jp}, $0 < j < N$, allowing in our set-up the inclusion of deterministic time trends up to order $p-1$, which are removed in the calculation of $w^T(\lambda_j)$ without the need of estimating them by any means.

These definitions apply directly to the Fourier transform with $p = 1$. However the cosine bell taper does not belong to this class, though it has property (22) as if $p = 1$, and has the improved convergence property (25) with respect to the non-tapered case, corresponding to tapers of order $p = 3$, which are the basis of bias reduction, even for inference with stationary time series. In previous analysis of tapering properties only condition (25) has been required (see for instance Condition C1 in Robinson (1986) or Dahlhaus (1988) assumptions), but to deal with a general form of non-stationarity, condition (22) is essential, though Robinson, 1986, p. 246) pointed out that not only $D(\lambda)$, but also its derivatives should be small away from the origin if we want to control the trending behaviour due to non-random smooth functions in t.

We consider two examples of higher order, $p > 1$, data tapers. For sample size $n = 4N$, where N is an integer, the weights given by the Parzen window

$$h_t^{\text{Parzen}} =$$

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

satisfy (Eq. (22)) for $j = 4, 8, \ldots, n-4$ and $s = 3$. We can obtain (see, e.g. Percival and Walden, 1993)

$$D_{\text{Parzen}}^T(\lambda) = \sum_{t=1}^{n} h_t^{\text{Parzen}} \exp\{\mathrm{i}\lambda t\} = \frac{32}{n^3}(3 - 2\sin^2\lambda/2)\left(\frac{\sin n\lambda/8}{\sin\lambda/2}\right)^4 \exp\{\mathrm{i}n\lambda/2\}$$

and $\sum_{t=1}^{n}(h_t^{\text{Parzen}})^2 \sim \text{const.}\ n$, so they are of order $p = 4$. Zhurbenko (1979) defined a general class of data tapers suggested by Kolmogorov of orders $p = 1, 2, \ldots$. When $p = 4$, these weights are very close to Parzen's ones and both have the same asymptotic properties. Kolmogorov weights correspond to the pth convolution of the uniform density, so for $p = 1$ they are equivalent to the raw Fourier transform and with $p = 2$ they are equal to Barlett's or the triangular window. See Section 7 for some plots and Alekseev (1996) for a recent discussion and some explicit formulae.

We now analyse the covariance matrix of the (normalized) tapered Fourier transform $v_p^T(\lambda) = w_p^T(\lambda)/(G^{1/2}\lambda^{-d})$ with tapers of order p.

Theorem 6. Under Assumptions 2 and 3 $[d > -\frac{1}{2},\ 0 < \alpha \leqslant 2]$ *for* $f_{\varepsilon^{(s)}}$, *a data taper of order* $p = 2, 3, \ldots$, *with* $p \geqslant s+1$ [*or just* $p > d$ *if* $\mu = 0$], *for any sequence of* *positive integers $k = k(n)$ and $j = j(n)$, $0 < k < j$, such that $j/n \to 0$, and defining* $\gamma_{j,k} \equiv (jk)^{d-p} \log(k+1)$,

(a) $\mathrm{E}[v_p^T(\lambda_{jp})\overline{v_p^T}(\lambda_{jp})] = 1 + \mathrm{O}(\min\{j^{-\alpha}, j^{-1}\} + [j/n]^{\alpha} + \gamma_{j,j})$,
(b) $\mathrm{E}[v_p^T(\lambda_{jp})v_p^T(\lambda_{jp})] = \mathrm{O}(j^{-p} + j^{-1-p}\log n + \gamma_{j,j})$,
(c) $\mathrm{E}[v_p^T(\lambda_{jp})\overline{v_p^T}(\lambda_{kp})] = \mathrm{O}(|j-k|^{-p} + k^{-1}|j-k|^{1-p} + k^{-1}|j-k|^{-p}\log n + \gamma_{k,j})$,
(d) $\mathrm{E}[v_p^T(\lambda_{jp})v_p^T(\lambda_{kp})] = \mathrm{O}(|j-k|^{-p} + k^{-1}|j-k|^{1-p} + k^{-1}|j-k|^{-p}\log n + \gamma_{k,j})$.

Proof. See Appendix A. □

We obtain that the periodogram is unbiased for any $d < p$ if $\mu = 0$. The main problem here are the covariance terms, whose bounds depend on the distance between the Fourier transforms considered because tapering destroys the orthogonality of the sine and cosine functions. Therefore, in the log-periodogram regression we are led to consider frequencies which are moving closer somewhat slower than n^{-1}. We adapt consequently the definition of $\hat{d}^T$, taking $J = 1$ for simplicity,

$$\hat{d}_p^T = \left(\sum_k \Lambda_{kp}^2\right)^{-1}\left(\sum_k \Lambda_{kp} Y_{kp}^{(T,1)}\right), \tag{26}$$

with Λ_k as in (Eq. (8)) and

$$Y_{kp}^{(T,1)} = \log I_p^T(\lambda_{kp}), \quad k = l, l+\eta, l+2\eta, \ldots, m\eta,$$

in such a way that for $\eta = 1, 2, \ldots$ we are still using about m *observations* in the regression (ignoring the trimming), so the variance of $\hat{d}_p^T$ can be of order m^{-1} if $\eta > 1$. For the asymptotic distribution we need in the definition of $\hat{d}_p^T$ that η increases with n to obtain the approximate independence of the tapered periodogram ordinates used in the estimate.

Theorem 7. Under the assumptions of Theorem 6, $p \geqslant s+1$, $p > 1$, $\varepsilon_t^{(s)}$ Gaussian and

$$\frac{(m\eta)^{1/2}}{l^{\max\{1,\alpha\}}} + \frac{m^{1/(2p-1)}}{\eta} + \frac{(m\eta)^{1+1/2\alpha}}{n} + \frac{l(\log n)^2}{m\eta} \to 0 \quad \text{as } n \to \infty, \tag{27}$$

we obtain for the estimate $\hat{d}_p^T$ defined in (Eq. (26)),

$$m^{1/2}(\hat{d}_p^T - d) \underset{\mathrm{d}}{\to} \mathrm{N}(0, \pi^2/24).$$

If $\mu = 0$ this is valid for $p > d + \frac{1}{4}$ with the extra condition $(m\eta)^{1/2} l^{2(d-p)} \log m \to 0$ as $n \to \infty$.

Proof. First, we observe that the uniform bounds for the bias errors in the covariance matrix of the tapered discrete Fourier transforms at the frequencies considered in the definition of $\hat{d}^{(p)}$ are now $o((m\eta)^{-1/2})$, using the first three conditions in (Eq. (27)) and $p \geqslant s+1$, so $p > d + \frac{1}{2}$. Hence, under (Eq. (27)) the asymptotic uncorrelatedness and Gaussianity of w_p^T is enough to make valid all the asymptotic results of Robinson (1995) for $\hat{d}_p^T$ and non-stationary processes with $d \geqslant \frac{1}{2}$. The same argument applies when $\mu = 0$ and $p > d + \frac{1}{4}$. □

Note that the lower growth rate required in (Eq. (27)) for l can be significantly larger than for η because the improved convergence properties of tapering are used to keep the bias under control for the covariance terms in Theorem 6. However, for any d fixed, to increase p will not reduce significantly the bias for the variance of the Fourier transform unless we increment at the same time α (i.e. the smoothness of f near the origin in Assumption 2). For example with $\alpha = 2$ and $p > d+1$, we need $l^{-1}m^{p/2(2p-1)+\varepsilon} \to 0$ as $n \to \infty$, for $\varepsilon > 0$ arbitrarily small, where the exponent of m is tending to $\frac{1}{4}$ as p increases, so the trimming required is not specially significant.

Then, if we expect $d \leqslant 2$ in practical applications, choosing $p = 3, 4$, say, we can base consistently on $\hat{d}_p^T$ an initial decision about the number of (integer) differences to take, under very general specification assumptions and independently of the presence of trends or of the value of d. Once we are certain to have returned to the interval $(-\frac{1}{2}, \frac{1}{2})$ after (integer) differentiation, we can perform more efficient, semiparametric or full parametric, analyses of the memory d and of any deterministic trend. We finally point out that Theorem 7 is valid when no trend removal is required (e.g. $d < \frac{3}{2}$ and $\mu = 0$) setting $p = 1$ in (Eq. (26)), even if tapers with higher order $p > 1$ are used for bias reduction purposes.

6. Non-invertible processes

Differencing the observed time series is an effective way of reducing the magnitude of the memory parameter d and the maximum order of any polynomial deterministic trend. However, differencing to remove deterministic or stochastic trends may lead to non-invertible stationary time series satisfying (Eq. (1)) with $d \leqslant -\frac{1}{2}$. Otherwise we will not find the non-invertible ($d \leqslant -\frac{1}{2}$) situation very often in practical applications.

Hurvich and Ray (1995) considered the limit of the expectation of the periodogram when $d < -\frac{1}{2}$ and of the tapered periodogram with the full cosine window when $d \in (-2.5, 1.5)$. They found that the (normalized) periodogram's expectation diverges with n so the log-periodogram estimate will have negative bias, and that tapering reduces this bias, allowing the log-periodogram regression estimate to work well in simulations when $d \in [-1, -\frac{1}{2}]$.

In this section we analyse if tapering with higher-order ($p > 1$) tapers may be fruitful to estimate the memory d of non-invertible time series satisfying the semiparametric model (1). We shall obtain, using the techniques of Theorems 6 and 7, that with p big enough (for d fixed), $\hat{d}_p^T$ is consistent and asymptotically normal for any $d \leqslant -\frac{1}{2}$. We first consider the covariance matrix of the tapered Fourier transforms.

Theorem 8. Under Assumptions 2 and 3 $[0 < \alpha \leqslant 2]$ *for* $f(\lambda)$, $d \leqslant -\frac{1}{2}$, *we can choose a data taper of order* $p = 2, 3, \ldots$, *such that for any sequences of positive integers* $j = j(n)$ *and* $k = k(n)$, $0 < k < j$, *with* $j/n \to 0$, *and with p big enough* ($p > |d| + \frac{1}{2}$), *defining* $\gamma_j \equiv n^a j^b$, $a = 4d^2/(2(p-d)-1)$, $b = 2d - (2p-1)^2/(2(p-d)-1)$,

(a) $\mathrm{E}[v_p^T(\lambda_{jp})\overline{v_p^T(\lambda_{jp})}] = 1 + \mathrm{O}(\min\{j^{-1}, j^{-\alpha}\} + [j/n]^\alpha + \gamma_j)$,
(b) $\mathrm{E}[v_p^T(\lambda_{jp})v_p^T(\lambda_{jp})] = \mathrm{O}(j^{-p} + j^{-1-p}\log n + \gamma_j)$,
(c) $\mathrm{E}[v_p^T(\lambda_{jp})\overline{v_p^T(\lambda_{kp})}] = \mathrm{O}(|j-k|^{-p} + k^{-1}|j-k|^{-p}\log n + k^{-1}|j-k|^{1-p} + \gamma_j)$,
(d) $\mathrm{E}[v_p^T(\lambda_{jp})v_p^T(\lambda_{kp})] = \mathrm{O}(|j-k|^{-p} + k^{-1}|j-k|^{-p}\log n + k^{-1}|j-k|^{1-p} + \gamma_j)$.

Proof. See Appendix C. □

For $d < 0$, $p > 1$, the exponent a of n in γ_j is positive, but for d fixed, it can be made arbitrarily small with p large enough, and the exponent b of j is negative, and can be made as big as we want in absolute value, increasing p as necessary, so to obtain $\gamma_j \to 0$ as n increases we will need j to grow at a certain rate. Otherwise this result is equivalent to Theorem 6. The intuition under this modification is the following. With $d < 0$ the process is stationary, so there are no problems with the definition of the spectral density $f(\lambda)$ or with its integrability. Here, given the required normalization for the moments of the discrete Fourier transform ($f(\lambda_j) = \mathrm{O}(\lambda_j^{-2d}) = \mathrm{o}(1)$ for $d < 0$ and $j/n \to 0$), the issue is how to avoid leakage from high frequencies (i.e. outside a neigbourhood of the origin, where we do not assume anything for f apart from integrability) to the zero frequency, where the spectral density f has a zero of order $-2d > 0$. This problem can be controlled by the fast uniform convergence (25) of the tails of $D_p^T(\lambda)$ with n and λ when p is chosen suitably, resulting in the term γ_j.

Hence, with the definition of $\hat{d}_p^T$ as in the previous section, using exactly the same arguments as for Theorems 3, 5 or 7, we obtain

Theorem 9. Under the assumptions of Theorem 8, X_t *Gaussian and p big enough such that*

$$\frac{(m\eta)^{1/2}}{l^{\max\{1,\alpha\}}} + \frac{m^{1/(2p-1)}}{\eta} + \frac{(m\eta)^{1+1/2\alpha}}{n} + \frac{l(\log n)^2}{m\eta} + \frac{n^{8d^2} m^{2(p-d)-1}}{l^c} \to 0$$

$$as\ n \to \infty, \tag{28}$$

$c = -4d[2(p-d)-1] + 2(2p-1)^2$, *we obtain for* $\hat{d}_p^T$ *defined in* (*Eq.* (*26*))

$$m^{1/2}(\hat{d}_p^T - d) \underset{d}{\rightarrow} \mathrm{N}(0, \pi^2/24).$$

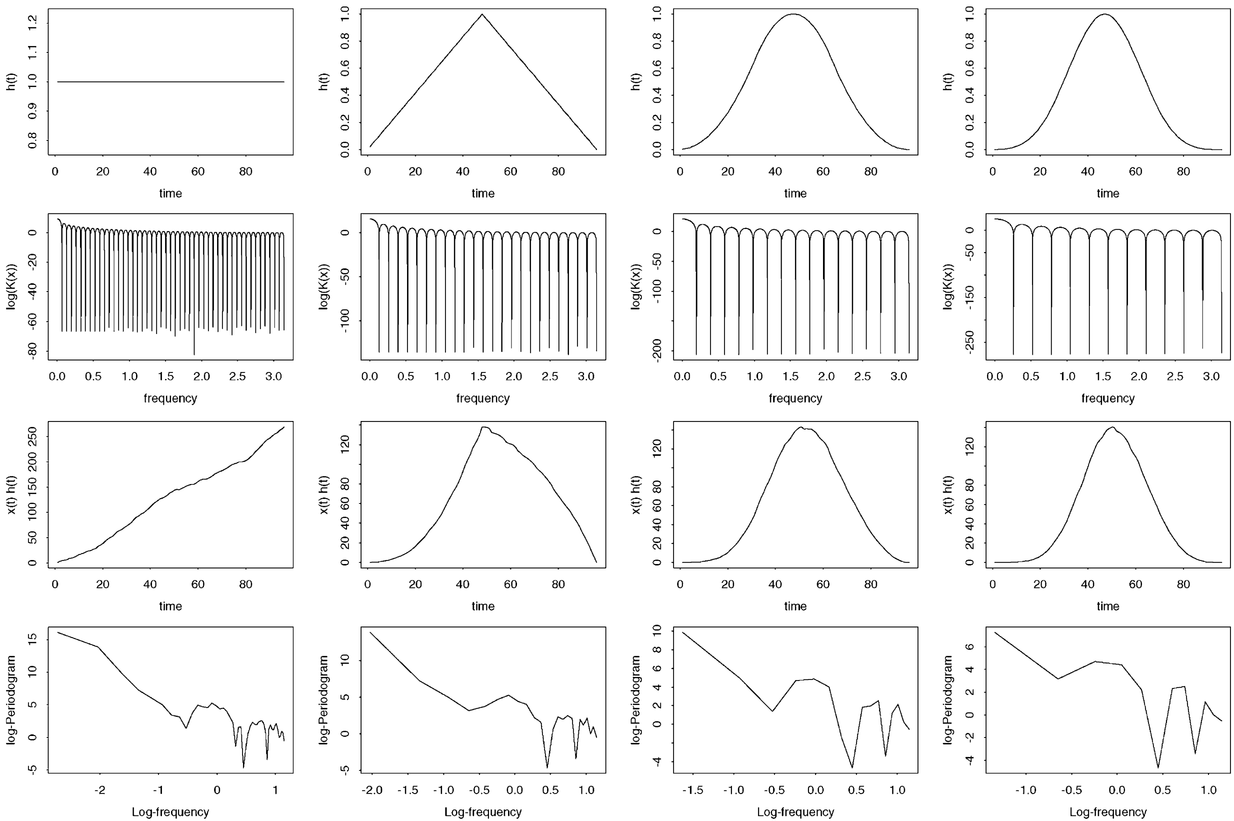
The last condition in (Eq. (28)) corresponds to $\gamma_l = o(m^{-1/2})$ to avoid the leakage from high frequencies. Here, the exponents of all the quantities are positive, the one for l growing very fast with p. For example, with $d = -2$ and $p = 4$ this is implied by $nm^{1/3}l^{-5.6} \to 0$, so if $m \sim n^{4/5}$ a choice of $l \sim n^{1/4}$ is sufficient. When $d = -2$ and $p = 3$ the condition is implied by $nm^{.29}l^{-3.82} \to 0$, so if $m \sim n^{4/5}$ again, a choice of $l \sim n^{0.31}$ would suffice. These conditions are in the same line with the ones required by, e.g. the log-periodogram regression estimate for stationary and invertible processes without tapering.

We could have considered all Fourier frequencies in the estimate (i.e. setting $p = 1$ in (Eq. (26)) even when higher-order tapers are used), but this will not improve in principle the estimation and will only complicate the inference, given the high correlation for adjacent periodogram ordinates when tapering. Finally we note that this theorem is valid as is stated for the cosine-bell taper for $d > -2.5$ when we fix $p = 3$ and $\eta = 1$ in the definition (26) of $\hat{d}_p^T$, since only the uniform bound (25) for the tails of the kernel $D_p^T(\lambda)$ is required, but not the property (22) of this taper at any particular Fourier frequency.

7. Simulation results

In this section we describe briefly the practical implementation of the previous estimates of the memory parameter d, with simulated non-stationary data. We will concentrate on Zhurbenko–Kolmogorov tapers with different values of p. We have plotted these data tapers for $p = 1$, 2, 3 and 4 and $n = 96$ on the first row of Fig. 1. We can observe that the larger the order p, the smoother is the transition in the extremes of the taper weights in the observed interval $1, \ldots, n$. The (logarithms of the) spectral windows $K_p^T(\lambda) = (2\pi \sum h_t^2)^{-1} |D_p^T(\lambda)|^2$ in the second row exhibit zeroes at different frequencies and central lobes with width increasing with p. In the third row we have the same simulated ARFIMA(0, 1.45, 0) tapered series for all the values of p considered. For any $p > 1$ the tapered series is hardly comparable with the original, $p = 1$, the shape of the tapering scheme is dominating. Finally, the last row of pictures corresponds to

Fig. 1. Zhurbenko–Kolmogorov tapers. The columns correspond to data tapers of orders $p = 1$, 2, 3 and 4, respectively, $n = 96$. In the first row, the plots correspond to the weights h_t^Z. In the second row we plot the logarithm of the spectral kernels $K_p(\lambda)$. In the third row appear the tapered series $h_t^Z X_t$, where X_t is a simulated Gaussian ARFIMA(0, 1.45, 0), and in the fourth row appear the log-periodogram of the above tapered series plotted against log-frequency.


h(t)
time
log(K(x))
frequency
x(t) h(t)
log-Periodogram
Log-frequency

the periodograms of the different tapered series in log–log coordinates, all of them being approximately linear, at least for the lower frequencies, though with different slopes as a consequence of the properties of the tapered periodogram for each p.

For $d \in [-1, 1.5]$ Hurvich and Ray (1995) provide an extensive simulation exercise for the log-periodogram estimate of d using the raw and the tapered (cosine bell) periodograms, confirming the results of our own simulations. Just for comparison purposes, we calculated the tapered log-periodogram estimate with the cosine bell and Zhurbenko taper with $p = 2$ for 1000 Gaussian ARFIMA(0, 1.8, 0), $\mu = 0$, simulated series and different bandwidth numbers m/p, covering the reasonable range of values for $n = 512$, without trimming ($l = 0$). For both data tapers we have tried two different versions of $\hat{d}^T$: first, for the cosine bell we considered all Fourier frequencies up to λ_m, except λ_1 and second, we considered only one frequency of every three, as in the definition of $\hat{d}^T$ in (Eq. (17)) (and therefore using a third of *observations* in each regression). For the Zhurbenko taper we calculated the estimates for $\eta = 1$ and $\eta = 2$, to check whether the central limit theorem of Theorem 7 is appropriate for such small values of η. Note that Theorem 5 does not hold under this set-up, but if (Eq. (20)) is negligible because $\mu = 0$ and $\varepsilon_0^{(1)} = 0$ the cosine bell tapered estimate could be consistent for $d < 3$ as if $p = 3$. The series were simulated with the S-Plus function `arima.fracdiff.sim` with $d = -0.2$ and then integrated twice.

The results of the simulation exercise are summarized in Table 1. In all cases we can observe that for small m the estimates have positive bias, which could be in part due to no trimming or due to the use of a very small number of frequencies. However, for big m, close to $n/2$, the bias is negative, and the variance is always decreasing with m, as we could expect. Part of the variability of the estimates can be due to the correlation between different periodogram ordinates when tapering, except for the cosine bell estimate (17), and perhaps for $\hat{d}_2^T$ if $\eta = 2$ is large enough to make adjacent periodogram ordinates almost independent. The standard deviations (s.d.) across simulations of $\hat{d}_2^T$ given by (Eq. (17)) never exceeded in more than 5% of the asymptotic s.d. given by Theorem 5 except for very small values of m, but when we take all possible frequencies in the regression, the s.d. are up to 128% of the asymptotic s.d. after taking into account that there are three times more points in the regressions. For Zhurbenko tapers, it seems that $\eta = 1$ is clearly insufficient for Theorem 7 to be a useful guide for inference, since the s.d. are about 17% higher than predicted, meanwhile for $\eta = 2$ they are about 10% higher.

The best results in terms of mean square error (MSE) were obtained for $m = 130$ or $m = 160$ and depended mainly on the number of frequencies used for the regression, so the first estimate considered, which uses m periodograms, gave the smallest MSE, except for very large m's. The estimates $\hat{d}_2^T$, $\eta = 1$ ($m/2$ frequencies), tended to give slightly better results than $\hat{d}^T$ with the cosine bell

Table 1
Tapered log-periodogram regression estimates, $\hat{d}_T$

m	Cosine bell, all freq.			Cosine bell, (17)			Zhurb., $p = 2$, $\eta = 1$			Zhurb., $p = 2$, $\eta = 2$		
	bias	s.d.	MSE	bias	s.d.	MSE	bias	s.d.	MSE	bias	s.d.	MSE
60	0.0525	0.1226	0.0178	0.0991	0.1612	0.0358	0.0710	0.1479	0.0269	– 0.0879	0.1800	0.0401
80	0.0297	0.1022	0.0113	0.0693	0.1322	0.0223	0.0508	0.1254	0.0183	0.0653	0.1542	0.0280
100	0.0087	0.0879	0.0078	0.0425	0.1145	0.0149	0.0321	0.1068	0.0124	0.0461	0.1341	0.0201
130	– 0.0220	0.0742	0.0060	0.0090	0.0984	0.0098	0.0032	0.0935	0.0088	0.0162	0.1201	0.0147
160	– 0.0556	0.0648	0.0073	– 0.0257	0.0884	0.0085	– 0.0279	0.0842	0.0079	– 0.0125	0.1092	0.0121
190	– 0.0943	0.0591	0.0124	– 0.0648	0.0813	0.0108	– 0.0644	0.0790	0.0104	– 0.0516	0.1022	0.0131
230	– 0.1509	0.0546	0.0258	– 0.1220	0.0764	0.0207	– 0.1222	0.0722	0.0202	– 0.1082	0.0934	0.0204

Bias, standard deviation (s.d) and mean square error (MSE) of $\hat{d}^T = \hat{d}^T(m)$, with the cosine bell and Zhurbenko–Kolmogorov tapers of order $p = 2$, $\eta = 1, 2$ for 1000 replications of a Gaussian ARFIMA(0, 1.8, 0), $n = 512$.

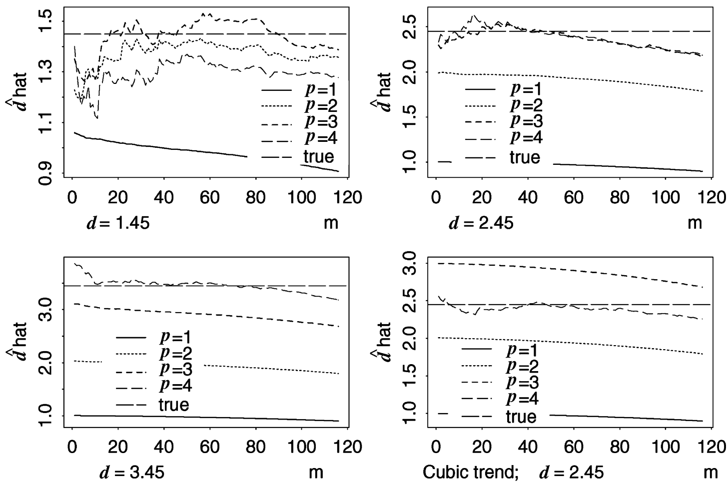


Fig. 2. Non-stationary and trending analysis. Plots of the true value of the parameter d and log-periodogram estimates $\hat{d}_p^T = \hat{d}_p^T(m)$ (for Zhurbenko tapers of orders $p = 1, 2, 3, 4$) against the value of the bandwidth number m calculated for the same simulated Gaussian ARFIMA(0, d, 0) series, $n = 1024$, integrated a different number of times, d, and with a cubic trend added in the last plot.

($m/3$ frequencies), but $\hat{d}_2^T$, $\eta = 2$, produced the worst simulations, as uses only $m/4$ frequencies.

The two main conclusions that we can draw from this and other related simulations we performed, and which will guide further analysis, are that except for very large sample sizes there seems to be no special advantage in taking $\eta > 1$, and that we would expect a positive bias for *small m*, and a negative bias for *large m*. Of course, this would be conditioned by the presence of other significant features in the dynamics of the process, like seasonal and cyclical components which may dominate the shape of $f(\lambda)$ at certain frequencies. It is important to note that model (1) is approximately valid for ARFIMA(0, d, 0) processes for all frequencies, so to increase m may reduce sometimes the bias, but this will not be the case for more general models.

Given the general class of estimates $\hat{d}_p^T$ defined by the Zhurbenko–Kolmogorov weights, it is interesting to study their different properties depending on the value of d and on the presence or not of deterministic trends in the observed time series. In Fig. 2 we show the typical behaviour of $\hat{d}_p^T$ for the same time series when integrated and/or added trends. The starting series is Gaussian ARFIMA(0, 0.45, 0), $\mu = 0$, $n = 1024$, and we integrate it once, twice and thrice and also when integrated twice, we added a cubic trend to it. Then we obtained

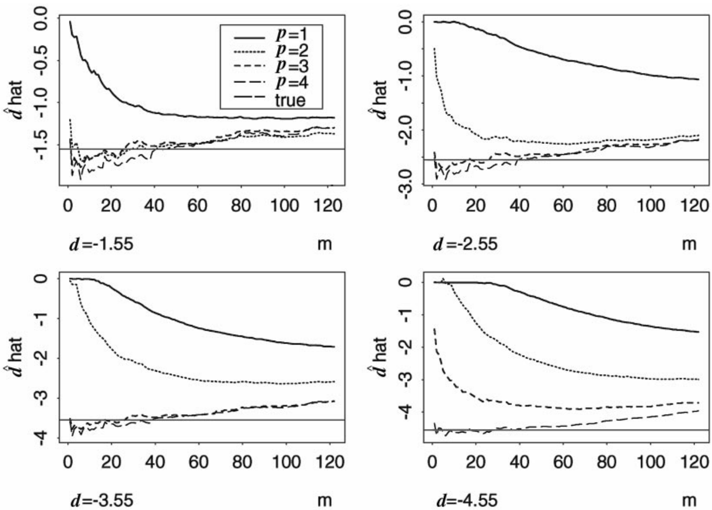


Fig. 3. Differencing analysis. Plots of the true value of the parameter $d < 0$ and log-periodogram estimates $\hat{d}_p^T = \hat{d}_p^T(m)$ (for Zhurbenko tapers of orders $p = 1, 2, 3, 4$) against the value of the bandwidth number m calculated for the same simulated Gaussian ARFIMA(0, d, 0) series, $n = 1024$, differentiated a different number of times, d.

the values of $\hat{d}_p^T$, $p = 1, \ldots, 4$, for a range of values of mp from about 25 to $n/2$, with increments of max$p = 4$.

When $d = 1.45$ the estimate $\hat{d}_1^T = \hat{d}$ does not work, as expected (in fact this is the usual log-periodogram estimate, valid only for $d < 1$, following Theorem 2). For $p > 2$ the results are much better and we can regard the estimates as consistent, the best results were obtained here for $p = 2, 3$. When $d = 2.45$ and $d = 3.45$ we can see that only with $p = 3, 4$ and $p = 4$, respectively, we capture the true features of the data, the estimates with $p < s + 1$ ($s = 2, 3$) converging invariantly to the value of p. This behaviour has been observed in all simulations and could be considered, among other problems, as an indication that the memory of the series is bigger that the value of p used, so we need to use higher-order tapers (and/or differentiate).

The fourth situation considered is an example of the confusion that the presence of deterministic trends may cause on the estimates for different p. We took the series with memory parameter $d = 2.45$ and added to it a cubic trend. Of the estimates considered, only $\hat{d}_4^T$ is resistant to that modification, as it is clear from the estimation results. Here $\hat{d}_3^T$ gives almost always 3, although for this series only $d = 2.45$: it takes wrongly the cubic trend estimating more memory than what actually is.

In conclusion, when apparently for a range of bandwidths m, an estimate $\hat{d}_p^T$ gives invariantly values about p, this indicates either that $s+1 \geqslant p$ (too much stochastic memory for that estimate) or that there is a deterministic trend of maximum exponent bigger or equal than p.

In Fig. 3 we repeat the same exercise as before, but now differencing the original series ($d = 0.45$) two to five times. In each case only the procedures with $p > |d| - 1$ give consistent estimates, taking into account that no deterministic trends are present. It can be observed that in all cases the leakage from high frequencies when m is big leads to positive biases. This suggests that non-invertibility presents as much difficulty for inference as non-stationary behaviours and stresses the risks of overdifferencing if deterministic trends are taken for (fractional) stochastic roots.

8. Conclusions

We have given a unified asymptotic theory for the log-periodogram estimate of the memory parameter d for Gaussian processes, including non-stationary ($d \geqslant \frac{1}{2}$) and non-invertible ($d \leqslant -\frac{1}{2}$) time series, with possibly deterministic trends, making of this semiparametric estimate a convenient tool for the analysis of the memory structure of a general class of processes under weak assumptions.

We have described the effects of tapering in terms of bias reduction, trend removal and estimation of non-standard values of d, showing why certain tapering schemes are resistant to particular non-stationary behaviours, but not to all. As Robinson (1986, p. 242) and Zhurbenko (1979) remark, the benefits of tapering only show up for certain data windows but not by tapering the data with any general smooth function. The results of this paper can be applied directly to obtain the asymptotic properties of non-parametric *spectral* estimates (of discrete average type) for functions f at fixed (Fourier) frequencies away from the origin, showing why traditional spectral non-parametric methods work in non-stationary situations for which they were not designed in first instance, justifying the conjecture of Robinson (1986, p. 246). This also confirms the observation of Granger (1966) about the shape of the (pseudo) spectral density of possibly non-stationary economic time series estimated from the original data.

The bounds for the moments of the discrete tapered Fourier transform for non-stationary processes obtained in this paper are only valid when evaluated at some particular Fourier frequencies λ_{jp}, $0 < j < N$, since it is only there where the spectral kernel of the Fourier transform has special properties. Thus, they do not extend for any continuously smoothed estimate of f or tapered auto-covariances, and only to non-stationarity at other frequencies different from zero if they coincide with a suitable Fourier frequency. It is very likely that the results of this paper about the asymptotic properties of the tapered periodogram can be adapted to carry out statistical inference with other semiparametric and parametric models of non-stationary (or non-invertible) observations without explicit specification of the degree of non-stationarity (or non-invertibility).

The generalization to multivariate time series follows immediately as in Robinson (1995), adapting his assumptions for the differenced stationary time series ε_t. The extension of the asymptotic theory given for the log-periodogram estimate to non-Gaussian time series could be tried under related conditions to those used in Velasco (1997).

Acknowledgements

I am grateful to P.M. Robinson for helpful discussions and suggestions. I also thank an Associate Editor and two Referees whose comments led to a great improvement of the paper. I also wish to thank P. Zaffaroni, C. Michelacci, F.J. Hidalgo, L. Giraitis and L.A. Gil Alaña for valuable comments. The first version of this paper was written while the author was at the London School of Economics and Political Science. Financial support from the Fundación Ramón Areces (Spain), the Economic and Social Research Council (ESRC) grant no. R000235892 and the Spanish Dirección General de Enseñanza Superior, Ref. no. PB95-0292 is gratefully acknowledged.

Appendix A. Proof of Theorems 1 and 6

Proof of Theorem 1. We can write the moments of the Fourier transform in terms of the function $f(\lambda)$, as if it were the spectral density of the non-stationary series X_t. Now the expectation of the periodogram $I(\lambda_j) = |w(\lambda_j)|^2$ is from Eq. (6)

$$\begin{aligned} \mathrm{E}[I(\lambda_j)] &= \frac{1}{2\pi n} \sum_{t_1=1}^{n} \sum_{k_1=1}^{t_1} \sum_{t_2=1}^{n} \sum_{k_2=1}^{t_2} \exp(\mathrm{i}\lambda_j t_1)\exp(-\mathrm{i}\lambda_j t_2)\mathrm{E}[\varepsilon_{k_1}\varepsilon_{k_2}] \\ &= \frac{1}{2\pi n} \sum_{t_1} \sum_{k_1} \sum_{t_2} \sum_{k_2} \exp\{\mathrm{i}\lambda_j (t_1 - t_2)\}\gamma_\varepsilon(k_1 - k_2) \\ &= \frac{1}{2\pi n} \int_{-\pi}^{\pi} \sum_{t_1} \sum_{k_1} \sum_{t_2} \sum_{k_2} \exp\{\mathrm{i}\lambda_j (t_1 - t_2)\}\exp\{-\mathrm{i}\lambda(k_1 - k_2)\} f_\varepsilon(\lambda)\, \mathrm{d}\lambda, \end{aligned}$$

where $\gamma_\varepsilon(k) = \mathrm{Cov}[\varepsilon_0, \varepsilon_k] = \int_{-\pi}^{\pi} f_\varepsilon(\lambda) \exp(\mathrm{i}\lambda k)\, \mathrm{d}\lambda$. Now

$$\sum_{t_1=1}^{n} \sum_{k_1=1}^{t_1} \exp\{\mathrm{i}\lambda_j t_1\}\exp\{-\mathrm{i}k_1\lambda\} = \sum_{t_1=1}^{n} \exp\{\mathrm{i}\lambda_j t_1\}\exp\left\{-\mathrm{i}\,\frac{\lambda(t_1+1)}{2}\right\} \frac{\sin t_1\lambda/2}{\sin\lambda/2}$$

$$= \sum_{t_1=1}^{n} \exp\{-\mathrm{i}\lambda/2 - \mathrm{i}t_1(\lambda/2 - \lambda_j)\}\frac{\sin t_1\lambda/2}{\sin\lambda/2}$$

$$= \frac{\exp\{-\mathrm{i}\lambda/2\}}{2\mathrm{i}\,\sin\lambda/2}\sum_{t_1=1}^{n}[\exp\{-\mathrm{i}t_1\lambda_j\} - \exp\{\mathrm{i}t_1(\lambda_j - \lambda)\}]$$

$$= -\frac{\exp\{-\mathrm{i}\lambda/2\}}{2\mathrm{i}\,\sin\lambda/2}\exp\left\{\mathrm{i}(\lambda_j - \lambda)\,\frac{n+1}{2}\right\}\frac{\sin n(\lambda_j - \lambda)/2}{\sin(\lambda_j - \lambda)/2}.$$

Repeating the same arguments for the sums in $\exp\{-\mathrm{i}\lambda_j t_2\}\exp\{\mathrm{i}k_2\lambda\}$, we get

$$\mathrm{E}[I(\lambda_j)] = \frac{1}{2\pi n}\int_{-\pi}^{\pi}\left[\frac{\sin n(\lambda_j - \lambda)/2}{\sin(\lambda_j - \lambda)/2}\right]^2\frac{f_\varepsilon(\lambda)}{4\sin^2\lambda/2}\,\mathrm{d}\lambda = \int_{-\pi}^{\pi}K(\lambda_j - \lambda)f(\lambda)\,\mathrm{d}\lambda, \tag{A.1}$$

where K is given by Eq. (9) and f by Eq. (2). For the other moments of the discrete Fourier transform,

$$\mathrm{E}[w^2(\lambda_j)] = \frac{1}{2\pi n}\sum_{t_1}\sum_{k_1}\sum_{t_2}\sum_{k_2}\exp\{\mathrm{i}\lambda_j(t_1 + t_2)\}\gamma_\varepsilon(k_1 - k_2)$$

$$= \frac{1}{2\pi n}\int_{-\pi}^{\pi}\sum_{t_1}\sum_{k_1}\sum_{t_2}\sum_{k_2}\exp\{\mathrm{i}\lambda_j(t_1 + t_2)\}\exp\{-\mathrm{i}\lambda(k_1 - k_2)\}f_\varepsilon(\lambda)\,\mathrm{d}\lambda$$

$$= \frac{1}{2\pi n}\int_{-\pi}^{\pi}\frac{\sin n(\lambda_j - \lambda)/2}{\sin(\lambda_j - \lambda)/2}\,\frac{\sin n(\lambda + \lambda_j)/2}{\sin(\lambda + \lambda_j)/2}\exp\{\mathrm{i}\lambda_j(n+1)\}\frac{f_\varepsilon(\lambda)}{4\sin^2\lambda/2}\,\mathrm{d}\lambda$$

$$= \frac{1}{2\pi n}\int_{-\pi}^{\pi}f(\lambda)D(\lambda_j - \lambda)D(\lambda + \lambda_j)\,\mathrm{d}\lambda,$$

where $D(\lambda)$ is Dirichlet kernel (10). Finally

$$\mathrm{E}[w(\lambda_j)\bar{w}(\lambda_k)] = \frac{1}{2\pi n}\sum_{t_1}\sum_{l_1}\sum_{t_2}\sum_{l_2}\exp\{\mathrm{i}\lambda_j t_1 - \mathrm{i}\lambda_k t_2)\}\gamma_\varepsilon(l_1 - l_2)$$

$$= \frac{1}{2\pi n}\int_{-\pi}^{\pi}\sum_{t_1}\sum_{l_1}\sum_{t_2}\sum_{l_2}\exp\{\mathrm{i}\lambda_j t_1 - \mathrm{i}\lambda_k t_2)\}\exp\{-\mathrm{i}\lambda(l_1 - l_2)\}f_\varepsilon(\lambda)\,\mathrm{d}\lambda$$

$$= \frac{1}{2\pi n} \int_{-\pi}^{\pi} \frac{\sin n(\lambda_j - \lambda)/2}{\sin(\lambda_j - \lambda)/2} \frac{\sin n(\lambda - \lambda_k)/2}{\sin(\lambda - \lambda_k)/2}$$

$$\times \exp\left\{ \mathrm{i}(\lambda_j - \lambda_k)\frac{n+1}{2} \right\} \frac{f_\varepsilon(\lambda)}{4\sin^2\lambda/2} \,\mathrm{d}\lambda$$

$$= \frac{1}{2\pi n} \int_{-\pi}^{\pi} f(\lambda) D(\lambda_j - \lambda) D(\lambda - \lambda_k) \,\mathrm{d}\lambda,$$

and

$$\mathrm{E}[w(\lambda_j) w(\lambda_k)] = \frac{1}{2\pi n} \int_{-\pi}^{\pi} f(\lambda) D(\lambda_j - \lambda) D(\lambda + \lambda_k) \,\mathrm{d}\lambda.$$

Then, the theorem follows from proof of Robinson's (1995) Theorem 2 where he considered the stationary and invertible case $d \in (-\frac{1}{2}, \frac{1}{2})$. For the interval around the origin, $[-\lambda_j/2, \lambda_j/2]$, where $f(\lambda)$ is no longer integrable when $d \geqslant \frac{1}{2}$, follow the method of the proof of Theorem 6 below used to bound (Eq. (A.12)) with $p = 1$, $|j - k| \geqslant 1$ and $d \in [\frac{1}{2}, 1)$, using the exact orthogonality of the sine and cosine components in the discrete Fourier transforms. □

Before giving the proof for Theorem 6 we prove two technical lemmas about tapering that will be required later.

Lemma A.1. For a data taper of order $p > 1$, and integers $j = j(n)$, $j = 1, 2, \ldots, n/2$,

$$n^{-1} \int_{-\pi}^{\pi} |D_p^T(\lambda_{jp} - \lambda) D_p^T(\lambda_{jp} + \lambda)| \,\mathrm{d}\lambda = \mathrm{O}(j^{-p}).$$

Proof. By symmetry we only need to consider $\lambda > 0$. Then, using Eq. (25)

$$\sup_{\lambda \geqslant 0} |D_p^T(\lambda_{jp} + \lambda)| = \mathrm{O}(n^{1-p} \lambda_{jp}^{-p}) = \mathrm{O}(n j^{-p}),$$

and the bound follows using $\int_{-\pi}^{\pi} |D_p^T(\lambda)| \,\mathrm{d}\lambda = \mathrm{O}(1)$, $p \geqslant 2$ for all n from Eq. (25). □

Lemma A.2. For a data taper of order $p > 1$, and integers $j = j(n)$, $k = k(n)$, $0 < k < j < n/2$,

$$n^{-1}\int_{-\pi}^{\pi} |D_p^T(\lambda_{jp} - \lambda)D_p^T(\lambda - \lambda_{kp})|\,\mathrm{d}\lambda = \mathrm{O}(|j-k|^{-p}).$$

Proof. Considering the intervals of integration $[-\pi, \lambda_{(k+j)p}/2]$ and $[\lambda_{(k+j)p}/2, \pi]$, and that $(\lambda_{jp} - \lambda_{kp})^{-1} = \mathrm{O}(n|j-k|^{-1})$, we have, for example,

$$\sup_{-\pi \leqslant \lambda \leqslant \lambda_{(k+j)p/2}} |D_p^T(\lambda_{jp} - \lambda)| = \mathrm{O}(n^{1-p}\lambda_{(j-k)p/2}^{-p}) = \mathrm{O}(n|j-k|^{-p}),$$

from (Eq. (25)) and the bound follows as before using the integrability of $|D_p^T|$, $p \geqslant 2$. □

Proof of Theorem 6. For part (a), we calculate the expectation of the periodogram $I_p^T(\lambda_{jp}) = |w_p^T(\lambda_{jp})|^2$ with respect to $f(\lambda_{jp})$. Proceeding as in the proof of Theorem 1

$$\begin{aligned}
\mathrm{E}[|w_p^T(\lambda_{jp})|^2] &= \frac{1}{2\pi b\, n^{2p-1}}\int_{-\pi}^{\pi} \frac{|a(\lambda - \lambda_{jp})|^2}{(2\sin[\lambda/2])^{2s}}\left(\frac{\sin^2[n(\lambda_{jp}-\lambda)/2p]}{\sin^2[(\lambda_{jp}-\lambda)/2]}\right)^p f_\varepsilon(\lambda)\,\mathrm{d}\lambda \\
&= \frac{1}{2\pi b\, n^{2p-1}}\int_{-\pi}^{\pi} |a(\lambda - \lambda_{jp})|^2 \frac{\sin^{2p}[n(\lambda_{jp}-\lambda)/2p]}{\sin^{2p}[(\lambda_{jp}-\lambda)/2]} f(\lambda)\,\mathrm{d}\lambda \\
&= \frac{1}{2\pi b\, n}\int_{-\pi}^{\pi} |D_p^T(\lambda_{jp}-\lambda)|^2 f(\lambda)\,\mathrm{d}\lambda = \int_{-\pi}^{\pi} K_p^T(\lambda_{jp}-\lambda) f(\lambda)\,\mathrm{d}\lambda,
\end{aligned}$$

similar to the stationary case, if $f(\lambda)$ defined in (Eq. (19)) were the (pseudo-) spectrum of X_t, on using the corresponding spectral kernel K_p^T with tapering, $K_p^T(\lambda) = (2\pi\sum h_t^2)^{-1}|D_p^T(\lambda)|^2$, satisfying with (Eq. (23)) the condition $|\lambda| \leqslant \pi$, then

$$|K_p^T(\lambda)| \leqslant \text{ const.} \min\{n, n^{1-2p}|\lambda|^{-2p}\}. \tag{A.2}$$

Now we generalize the proof in Theorem 2 of Robinson (1995), for $p > 1$, taking special care in the integration in the interval $[-\lambda_{jp}/2, \lambda_{jp}/2]$ where the integrability of $f(\lambda)$ can no longer be used when $d \geqslant \frac{1}{2}$. In the proof for the intervals $[-\pi, -\varepsilon]$ and $[\varepsilon, \pi]$ the integrability is used in that reference, but it is not necessary, restricting the integration in the bound to $|\lambda| > \varepsilon$. Note that we consider simultaneously the situations where $f(\lambda)$ diverges at the origin ($d > 0$), is a constant ($d = 0$) or tends to zero ($d < 0$).

The term in $[j/n]^\alpha$ comes from the normalization by $G^{1/2}\lambda^{-d}$, instead of by $f^{1/2}(\lambda)$ using Assumption 1.

Using the periodicity and integrability to 1 of K_p^T in $[-\pi, \pi]$, we consider the same intervals of integration to analyse the bias

$$\mathrm{E}[|w_p^T(\lambda_{jp})|^2] - f(\lambda_{jp}) = \int_{-\pi}^{\pi} [f(\lambda) - f(\lambda_{jp})] K_p^T(\lambda_{jp} - \lambda)\, \mathrm{d}\lambda,$$

as Robinson (1995). Consider a fixed $\varepsilon > 0$, such that $f(\lambda) \leqslant C_\varepsilon \lambda^{-2d}$, $|\lambda| \in (0, \varepsilon)$ for some positive constant C_ε, depending on ε, and n big enough such that $2\lambda_{jp}$, $2\lambda_{kp} < \varepsilon$. Then,

$$\left| \int_{-\pi}^{-\varepsilon} + \int_{\varepsilon}^{\pi} \right| \leqslant 2 \max_{|\lambda| \geqslant \varepsilon} |K_p^T(\lambda_{jp} - \lambda)| \int_{\varepsilon}^{\pi} |f(\lambda) - f(\lambda_{jp})|\, \mathrm{d}\lambda = \mathrm{O}([1 + f(\lambda_{jp})] n^{1-2p})$$
$$= \mathrm{O}(f(\lambda_{jp}) j^{-p}),$$

using the property (A.2) of $K_p^T(\lambda)$ and the integrability of f outside the origin. Next, using Eq. (A.2) and $f(\lambda) = \mathrm{O}(|\lambda|^{-2d})$,

$$\left| \int_{-\varepsilon}^{-\lambda_{jp}/2} \right| \leqslant f(\lambda_{jp}) \int_{-\varepsilon}^{-\lambda_{jp}/2} |K_p^T(\lambda_{jp} - \lambda)|\, \mathrm{d}\lambda + \int_{-\varepsilon}^{-\lambda_{jp}/2} f(\lambda) |K_p^T(\lambda_{jp} - \lambda)|\, \mathrm{d}\lambda$$
$$= \mathrm{O}\left(f(\lambda_{jp}) n^{1-2p} \int_{\lambda_{jp}/2}^{\infty} \lambda^{-2p}\, \mathrm{d}\lambda + n^{1-2p} \int_{\lambda_{jp}/2}^{\infty} \lambda^{-2d-2p}\, \mathrm{d}\lambda \right)$$
$$= \mathrm{O}(f(\lambda_{jp}) j^{1-2p}).$$

An identical bound can be obtained for the interval $[3\lambda_{jp}/2, \varepsilon]$. Now

$$\int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} \leqslant \left[\max_{-\lambda_{jp}/2 \leqslant \lambda \leqslant \lambda_{jp}/2} K_p^T(\lambda_{jp} - \lambda) \right] \int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} f(\lambda_{jp})\, \mathrm{d}\lambda \qquad \text{(A.3)}$$

$$+\int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} K_p^T(\lambda_{jp}-\lambda) f(\lambda)\,\mathrm{d}\lambda. \tag{A.4}$$

Now, with Eq. (A.2), the first term of Eq. (A.3) on the right-hand side is

$$\mathrm{O}(n^{1-2p}\lambda_{jp}^{-2p}\lambda_{jp}^{1-2d}) = \mathrm{O}(f(\lambda_{jp})j^{1-2p}).$$

If $d \in (-\frac{1}{2}, \frac{1}{2})$ the other contribution, Eq. (A.4), of the interval $[-\lambda_{jp}/2, \lambda_{jp}/2]$ is

$$\mathrm{O}\left(\sup_{-\lambda_{jp}/2 \leqslant \lambda \leqslant \lambda_{jp}/2} K_p^T(\lambda_{jp}-\lambda) \int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} f(\lambda)\,\mathrm{d}\lambda \right) = \mathrm{O}(n^{1-2p}\lambda_{jp}^{-2p}\lambda_{jp}^{1-2d})$$

$$= \mathrm{O}(f(\lambda_{jp})j^{1-2p}).$$

When $f(\lambda)$ is not integrable, $d \geqslant \frac{1}{2}$, to bound (Eq. (A.4)) we normalize by $1/f(\lambda_{jp})$ and in the definition of f in terms of $f^{\star}$, we substitute $2\sin[\lambda/2]$ by λ, since the terms $\mathrm{O}(|\lambda|^3)$ will cause negligible error in that interval. Then this contribution is of order

$$\frac{1}{2\pi b n^{2p-1}} \int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} |a(\lambda-\lambda_{jp})|^2 \frac{f^*(\lambda)}{f^*(\lambda_{jp})} \frac{\sin^{2p}[n(\lambda_{jp}-\lambda)/2p]}{\sin^{2p}[(\lambda_{jp}-\lambda)/2]} \lambda_{jp}^{2d} |\lambda|^{-2d}\,\mathrm{d}\lambda,$$

which making a change of variable is equal to

$$\frac{1}{2\pi b\, n^{2p}} \int_{-\pi jp}^{\pi jp} |a(\lambda/n-\lambda_{jp})|^2 \frac{f^*(\lambda/n)}{f^*(\lambda_{jp})} \frac{\sin^{2p}[(2\pi j-\lambda/p)/2]}{\sin^{2p}[(2\pi jp-\lambda)/2n]} \left(\frac{2\pi jp}{n}\right)^{2d} \left|\frac{\lambda}{n}\right|^{-2d} \mathrm{d}\lambda,$$

and this is not greater than a constant times

$$A \equiv \frac{1}{n^{2p}} \int_{-\pi jp}^{\pi jp} \frac{\sin^{2p}[\lambda/2p]}{\sin^{2p}[(2\pi jp-\lambda)/2n]} \left|\frac{\lambda}{2\pi jp}\right|^{-2d} \mathrm{d}\lambda,$$

since $\sin^{2p}[(2\pi j-\lambda/p)/2] = \sin^{2p}[\lambda/2p]$ for integer j, $|a(\lambda)|$ is bounded, b is bounded away from zero and $f^*(\lambda)$ is bounded above and away from zero around $\lambda = 0$. As $j/n \to 0$ and $\forall \lambda \in [-\pi jp, \pi jp]$ and $j = 1, 2, \ldots,$ we have $(2\pi jp - \lambda)/2n \to 0$. Bounding the sine function around 0 using $|\sin x| > \frac{1}{2}|x|$, $|x| \leqslant \pi/2$, we have

$$\sin^{-2p}[(2\pi jp - \lambda)/2n] \leqslant 2^{2p}\left(\frac{2\pi jp - \lambda}{2n}\right)^{-2p},$$

and

$$A \leqslant \frac{2^{2p}}{n^{2p}} \int_{-\pi jp}^{\pi jp} \frac{\sin^{2p}[\lambda/2p]}{[(2\pi jp - \lambda)/2n]^{2p}} \left|\frac{\lambda}{2\pi jp}\right|^{-2d} \mathrm{d}\lambda$$

$$\leqslant j^{2d} 2^{4p} (2\pi p)^{2d} \int_{-\pi jp}^{\pi jp} \frac{\sin^{2p}[\lambda/2p]}{(2\pi jp - \lambda)^{2p}} |\lambda|^{-2d} \, \mathrm{d}\lambda.$$

Now, using that $2(p-d) > -1$ from $p \geqslant s + 1$, so $\sin^{2p}[\lambda/2p]|\lambda|^{-2d}$ is integrable around the origin and $d \geqslant \frac{1}{2}$, we see that

$$\int_{-\pi jp}^{\pi jp} \sin^{2p}[\lambda/2p]|\lambda|^{-2d} \, \mathrm{d}\lambda = \mathrm{O}\left(\int_0^1 \lambda^{2(p-d)} \, \mathrm{d}\,\lambda + \int_1^{\pi jp} \lambda^{-2d} \, \mathrm{d}\lambda\right)$$
$$= \mathrm{O}(\log(j+1)), \tag{A.5}$$

(just O(1) if $d > 1/2$) and that, uniformly for $\lambda \in [-\pi jp, \pi jp]$,

$$(2\pi jp - \lambda)^{-2p} \leqslant 4(2\pi jp)^{-2p} = \mathrm{O}(j^{-2p}),$$

obtaining, with $j/n \to 0$,

$$A = \mathrm{O}([j^{2(d-p)} \log(j+1)).$$

Next, if $\alpha \in (1, 2]$, using the discussion after Assumption 2,

$$\left| \int_{\lambda_{jp}/2}^{3\lambda_{jp}/2} [f(\lambda) - f(\lambda_{jp})] K_p^T(\lambda_{jp} - \lambda) \, \mathrm{d}\lambda \right|$$

$$= \left| \int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} [f(\lambda_{jp} - \lambda) - f(\lambda_{jp})] K_p^T(\lambda) \, \mathrm{d}\lambda \right|$$

$$= \left| \int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} [\lambda f'(\lambda_{jp}) + \mathrm{O}(\lambda_{jp}^{-\alpha-2d} |\lambda|^{\alpha})] K_p^T(\lambda) \, \mathrm{d}\lambda \right|$$

$$= \mathrm{O}\left(\lambda_{jp}^{-\alpha-2d} \int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} |\lambda|^{\alpha} K_p^T(\lambda) \, \mathrm{d}\lambda \right), \tag{A.6}$$

since K_p^T is positive, even and we are integrating in a symmetric interval around 0. Now, with both bounds in (Eq. (A.2)), $1 \leqslant \alpha \leqslant 2$, $p \geqslant 2$,

$$\int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} |\lambda|^{\alpha} K_p^T(\lambda) = 2 \left\{ \int_0^{n^{-1}} + \int_{n^{-1}}^{\lambda_{jp}/2} \right\} \lambda^{\alpha} K_p^T(\lambda) \, \mathrm{d}\lambda$$

$$= \mathrm{O}\left(n \int_0^{n^{-1}} \lambda^{\alpha} \, \mathrm{d}\lambda + n^{1-2p} \int_{n^{-1}}^{\lambda_{jp}/2} \lambda^{\alpha-2p} \, \mathrm{d}\lambda \right) = \mathrm{O}(n^{-\alpha}). \tag{A.7}$$

Therefore

$$\left| \int_{\lambda_{jp}/2}^{3\lambda_{jp}/2} \right| = \mathrm{O}(\lambda_{jp}^{-\alpha-2d} n^{-\alpha}) = \mathrm{O}(f(\lambda_{jp}) j^{-\alpha}).$$

When $\alpha \leqslant 1$ using similar methods, the bound is seen to be $\mathrm{O}(f(\lambda_{jp}) j^{-1})$ applying with Assumption 3, the mean value theorem (MVT) to f on the right-hand side of (Eq. (A.6)) and using (Eq. (A.7)) with $\alpha = 1$. The proof of (a) is now complete.

Let us consider now the covariance terms. First, for part (b),

$$\mathrm{E}[w(\lambda_{jp})^2] = \frac{1}{2\pi b n^{2p-1}} \int_{-\pi}^{\pi} \frac{a(\lambda - \lambda_{jp}) a(\lambda_{jp} + \lambda)}{(2\sin[\lambda/2])^{2s}}$$

$$\times \left(\frac{\sin[n(\lambda_{jp} - \lambda)/2p]}{\sin[(\lambda_{jp} - \lambda)/2]} \right)^p \left(\frac{\sin[n(\lambda_{jp} + \lambda)/2p]}{\sin[(\lambda_{jp} + \lambda)/2]} \right)^p f_{\varepsilon}(\lambda) \, \mathrm{d}\lambda$$

$$= \frac{1}{2\pi b n^{2p-1}} \int_{-\pi}^{\pi} a(\lambda - \lambda_{jp}) a(\lambda_{jp} + \lambda) \frac{\sin[n(\lambda_{jp} - \lambda)/2p]}{\sin[(\lambda_{jp} - \lambda)/2]}$$

$$\times \frac{\sin[n(\lambda_{jp} + \lambda)/2p]}{\sin[(\lambda_{jp} + \lambda)/2]} f(\lambda)\, \mathrm{d}\lambda.$$

$$= \frac{1}{2\pi \sum h_t^2} \int_{-\pi}^{\pi} D_p^T(\lambda_{jp} - \lambda) D_p^T(\lambda_{jp} + \lambda) f(\lambda)\, \mathrm{d}\lambda.$$

Again, the only step different from the stationary case of Theorem 2 of Robinson (1995), is the bound for the integral in the interval $[-\lambda_{jp}/2, \lambda_{jp}/2]$. The other problem is the destruction of the orthogonality between Fourier transforms and their real and imaginary parts. The last expression can be seen to be equal to

$$\frac{1}{2\pi \sum h_t^2} \int_{-\pi}^{\pi} [f(\lambda) - f(\lambda_{jp})] D_p^T(\lambda_{jp} - \lambda) D_p^T(\lambda_{jp} + \lambda)\, \mathrm{d}\lambda + \mathrm{O}(f(\lambda_{jp}) j^{-p}) \qquad \text{(A.8)}$$

where the last term follows from the approximate orthogonality for frequencies that are moving apart (Lemma 1). Now, we can study the integral in (Eq. (A.8)) splitting the range of integration in the following intervals:

$$\left| \int_{-\pi}^{-\varepsilon} + \int_{\varepsilon}^{\pi} \right| = \mathrm{O}\left(\frac{1}{n^{2p-1} \varepsilon^{2p}} \int_{\varepsilon}^{\pi} [f(\lambda) + f(\lambda_{jp})]\, \mathrm{d}\lambda \right) = \mathrm{O}([1 + f(\lambda_{jp})] n^{1-2p})$$

$$= \mathrm{O}(f(\lambda_{jp}) j^{-p}), \qquad \text{(A.9)}$$

$$\left| \int_{-\varepsilon}^{-2\lambda_{jp}} + \int_{2\lambda_{jp}}^{\varepsilon} \right| = \mathrm{O}\left(f(\lambda_{jp}) n^{-1} \int_{2\lambda_{jp}}^{\pi} |D_p^T(\lambda_{jp} - \lambda) D_p^T(\lambda_{jp} + \lambda)|\, \mathrm{d}\lambda \right)$$

$$+ \mathrm{O}\left(n^{-1} \int_{2\lambda_{jp}}^{\varepsilon} f(\lambda) |D_p^T(\lambda_{jp} - \lambda) D_p^T(\lambda_{jp} + \lambda)|\, \mathrm{d}\lambda \right)$$

$$= \mathrm{O}\left(f(\lambda_{jp}) n^{1-2p} \int_{2\lambda_j}^{\pi} \lambda^{-2p}\, \mathrm{d}\lambda + n^{1-2p} \int_{\lambda_j}^{\infty} \lambda^{-2d-2p}\, \mathrm{d}\lambda \right)$$

$$= \mathrm{O}(f(\lambda_{jp}) j^{1-2p}).$$

Now, using $f(\lambda_{jp}) = f(-\lambda_{jp})$,

$$\left| \int_{-2\lambda_{jp}}^{-\lambda_{jp}/2} + \int_{\lambda_{jp}/2}^{2\lambda_{jp}} \right| = \mathrm{O}\Bigg(n^{-1} \sup_{\lambda_{jp}/2 \leqslant \lambda \leqslant 2\lambda_{jp}} |f'(\lambda) D_p^T(\lambda_{jp} + \lambda)|$$

$$\times \int_{\lambda_{jp}/2}^{2\lambda_{jp}} |\lambda_{jp} - \lambda| |D_p^T(\lambda_{jp} - \lambda)| \, \mathrm{d}\lambda \Bigg)$$

$$= \mathrm{O}\left(n^{-1} f(\lambda_{jp}) \lambda_{jp}^{-1} n^{1-p} \lambda_{jp}^{-p} \int_0^{2\lambda_{jp}} \lambda |D_p^T(\lambda)| \, \mathrm{d}\lambda \right)$$

$$= \mathrm{O}(f(\lambda_{jp}) j^{-1-p} \log n),$$

because

$$\int_0^{2\lambda_{jp}} \lambda |D_p^T(\lambda)| \mathrm{d}\lambda = \mathrm{O}\left(n \int_0^{n^{-1}} \lambda \, \mathrm{d}\lambda + n^{1-p} \int_{n^{-1}}^{2\lambda_{jp}} \lambda^{1-p} \, \mathrm{d}\lambda \right) = \mathrm{O}(n^{-1} \log n), \quad \text{(A.10)}$$

using the property (25) of $D_p^T(\lambda)$, the term $\log n$ appearing when $p = 2$ only. Finally

$$\left| \int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} \right| = \mathrm{O}\Bigg(n^{-1} \max_{-\lambda_{jp}/2 \leqslant \lambda \leqslant \lambda_{jp}/2} |D_p^T(\lambda_{jp} - \lambda) D_p^T(\lambda_{jp} + \lambda)|$$

$$\times \int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} f(\lambda_{jp}) \, \mathrm{d}\lambda \Bigg) \quad \text{(A.11)}$$

$$+ \mathrm{O}\left(n^{-1} \int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} D_p^T(\lambda_{jp} - \lambda) D_p^T(\lambda_{jp} + \lambda) f(\lambda) \, \mathrm{d}\lambda \right) \quad \text{(A.12)}$$

where the first term, (Eq. (A.11))'s right-hand side, is

$$\mathrm{O}(n^{1-2p} \lambda_{jp}^{-2p} \lambda_{jp} f(\lambda_{jp})) = \mathrm{O}(f(\lambda_{jp}) j^{1-2p}).$$

The second term (Eq. (A.12)) is also $O(f(\lambda_{jp})j^{1-2p})$ when f is integrable (see the bound for (Eq. (A.4))), and when not, making a change of variable similar as before, normalizing by $1/f(\lambda_{jp})$ and substituting $2\sin[\lambda/2]$ by λ, the contribution of the integral in the interval $[-\lambda_{jp}/2, \lambda_{jp}/2]$ is of the same order of magnitude as

$$\int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} \frac{|a(\lambda_{jp}-\lambda)||a(\lambda_{jp}+\lambda)|}{2\pi b n^{2p-1}} \frac{f^*(\lambda)}{f^*(\lambda_{jp})}$$

$$\times \left| \frac{\sin[n(\lambda_{jp}-\lambda)/2p]}{\sin[(\lambda_{jp}-\lambda)/2]} \frac{\sin[n(\lambda_{jp}+\lambda)/2p]}{\sin[(\lambda_{jp}+\lambda)/2]} \right|^p \lambda_{jp}^{2d} |\lambda|^{-2d} \, d\lambda$$

$$\leqslant \frac{\|a\|_\infty^2}{2\pi b n^{2p-1}} \int_{-\pi jp}^{\pi jp} \frac{f^*(\lambda/n)}{f^*(\lambda_{jp})} \left| \frac{\sin^2[\lambda/2p]}{\sin[(2\pi jp-\lambda)/2n]\sin[(2\pi jp+\lambda)/2n]} \right|^p$$

$$\times \left(\frac{2\pi jp}{n}\right)^{2d} \left|\frac{\lambda}{n}\right|^{-2d} d\lambda,$$

and exactly the same bound holds as before for A, since the two sine functions behave asymptotically in a similar way in this range of values of λ, i.e.

$$\sup_{-\pi jp \leqslant \lambda \leqslant \pi jp} |\sin[(2\pi jp \pm \lambda)/2n]|^{-1} = O\left(\left(\frac{j}{n}\right)^{-1}\right),$$

as $j/n \to 0$. Therefore the bound for part (b) follows.

Let us now study the covariance term, $0 < k < j$, $j/n \to 0$,

$$E[w_p^T(\lambda_{jp})\overline{w_p^T(\lambda_{kp})}]$$

$$= \frac{1}{2\pi b n^{2p-1}} \int_{-\pi}^{\pi} \frac{a(\lambda_{jp}-\lambda)a(\lambda-\lambda_{kp})}{(2\sin[\lambda/2])^{2s}}$$

$$\times \left(\frac{\sin[n(\lambda_{jp}-\lambda)/2p]}{\sin[\lambda/2]\sin[(\lambda_{jp}-\lambda)/2]} \frac{\sin[n(\lambda_{kp}-\lambda)/2p]}{\sin[\lambda/2]\sin[(\lambda_{kp}-\lambda)/2]} \right)^p f_\varepsilon(\lambda) \, d\lambda$$

$$= \frac{1}{2\pi b n^{2p-1}} \int_{-\pi}^{\pi} a(\lambda_{jp}-\lambda)a(\lambda-\lambda_{kp})$$

$$\times\left(\frac{\sin[n(\lambda_{jp}-\lambda)/2p]}{\sin[(\lambda_{jp}-\lambda)/2]}\frac{\sin[n(\lambda_{kp}-\lambda)/2p]}{\sin[(\lambda_{kp}-\lambda)/2]}\right)^p f(\lambda)\,\mathrm{d}\lambda$$

$$=\frac{1}{2\pi\sum h_t^2}\int_{-\pi}^{\pi} D_p^T(\lambda_{jp}-\lambda)D_p^T(\lambda-\lambda_{kp})f(\lambda)\,\mathrm{d}\lambda.$$

This covariance can be expanded, with error $\mathrm{O}(|j-k|^{-p})$ from Lemma A.2 due to the loss of orthogonality of data tapers, as

$$\frac{1}{2\pi\sum}h_t^2\Bigg[\int_{(\lambda_{kp}+\lambda_{jp})/2}^{2\lambda_{jp}} \{f(\lambda)-f(\lambda_{jp})\}D_p^T(\lambda_{jp}-\lambda)D_p^T(\lambda-\lambda_{kp})\,\mathrm{d}\lambda \tag{A.13}$$

$$+\int_{\lambda_{kp}/2}^{(\lambda_{kp}+\lambda_{jp})/2} \{f(\lambda)-f(\lambda_{kp})\}D_p^T(\lambda_{jp}-\lambda)D_p^T(\lambda-\lambda_{kp})\,\mathrm{d}\lambda \tag{A.14}$$

$$-\int_{\lambda_{kp}/2}^{(\lambda_{kp}+\lambda_{jp})/2} \{f(\lambda_{jp})-f(\lambda_{kp})\}D_p^T(\lambda_{jp}-\lambda)D_p^T(\lambda-\lambda_{kp})\,\mathrm{d}\lambda \tag{A.15}$$

$$+\left\{\int_{2\lambda_{jp}}^{\pi}+\int_{-\pi}^{\lambda_{kp}/2}\right\}\{f(\lambda)-f(\lambda_{jp})\}D_p^T(\lambda_{jp}-\lambda)D_p^T(\lambda-\lambda_{kp})\,\mathrm{d}\lambda\Bigg]. \tag{A.16}$$

Define $f_{jk}=(\lambda_k\lambda_j)^{-d}$. Now (Eq. (A.13)) can be bounded by

$$\mathrm{O}\left(n^{-1}\sup_{(\lambda_{kp}+\lambda_{jp})/2\leqslant\lambda\leqslant 2\lambda_{jp}}|f'(\lambda)D_p^T(\lambda-\lambda_{kp})|\int_{(\lambda_{kp}+\lambda_{jp})/2}^{2\lambda_{jp}}|\lambda_{jp}-\lambda||D_p^T(\lambda_{jp}-\lambda)|\,\mathrm{d}\lambda\right)$$

$$=\mathrm{O}\left(n^{-1}f(\lambda_{jp})\lambda_{jp}^{-1}n^{1-p}\lambda_{(j-k)p/2}^{-p}\int_0^{\lambda_{jp}}\lambda|D_p^T(\lambda)|\,\mathrm{d}\lambda\right)$$

$$=\mathrm{O}(f(\lambda_{jp})j^{-1}|j-k|^{-p}\log n)=\mathrm{O}(f_{jk}\,k^{-1}\,|j-k|^{-p}\log n),$$

because $\int_0^{\lambda_{jp}}\lambda|D_p^T(\lambda)|\,\mathrm{d}\lambda=\mathrm{O}(n^{-1}\log n)$ for $p\geqslant 2$ from (Eq. (A.10)). Next, for $k\geqslant j/2$, (Eq. (A.14)) is

$$\mathrm{O}\left(n^{-1}\sup_{\lambda_{kp}/2\leqslant\lambda\leqslant(\lambda_{kp}+\lambda_{jp})/2}|f'(\lambda)D_p^T(\lambda_{jp}-\lambda)|\int_{\lambda_{kp}/2}^{(\lambda_{kp}+\lambda_{jp})/2}|\lambda-\lambda_{kp}||D_p^T(\lambda-\lambda_{kp})|\,\mathrm{d}\lambda\right)$$

$$= \mathrm{O}\left(n^{-1} f(\lambda_{kp}) \lambda_{kp}^{-1} n |j-k|^{-p} \int_{0}^{\lambda_{jp}} \lambda |D_p^T(\lambda)| \, \mathrm{d}\lambda \right)$$

$$= \mathrm{O}(f(\lambda_{kp}) k^{-1} |j-k|^{-p} \log n) = \mathrm{O}(f_{jk} \, k^{-1} |j-k|^{-p} \log n),$$

since $f(\lambda_{kp}) = \mathrm{O}(f_{jk})$ if $k \geqslant j/2$, and when $k < j/2$, (Eq. (A.14)) is

$$\mathrm{O}\left(n^{-1} \sup_{\lambda_{kp}/2 \leqslant \lambda \leqslant (\lambda_{kp} + \lambda_{jp})/2} |f(\lambda) + f(\lambda_{kp})| |D_p^T(\lambda_{jp} - \lambda)| \int_{\lambda_{kp}/2}^{(\lambda_{kp} + \lambda_{jp})/2} |D_p^T(\lambda - \lambda_{kp})| \mathrm{d}\lambda \right)$$

$$= \mathrm{O}\left(n^{-1} [f(\lambda_{kp}) + f(\lambda_{jp})] n^{1-p} \lambda_{(j-k)p/2}^{-p} \int_{0}^{\lambda_{jp}} |D_p^T(\lambda)| \, \mathrm{d}\lambda \right)$$

$$= \mathrm{O}([f(\lambda_{kp}) + f(\lambda_{jp})] j^{-p}) = \mathrm{O}(f_{jk} \, k^{-p}),$$

using that $j - k > j/2$, $p > d$, and $\int_0^{\lambda_{jp}} |D_p^T(\lambda)| \, \mathrm{d}\lambda = \mathrm{O}(1)$ for $p \geqslant 2$ (which can be shown as (Eq. (A.10))).

For $k \geqslant j/2$ and using the MVT, Assumption 3 and (Eq. (25)), (Eq. (A.15)) is

$$\mathrm{O}\left(n^{-1} (\lambda_{jp} - \lambda_{kp}) \sup_{\lambda_{kp}/2 \leqslant \lambda \leqslant \lambda_{jp}} |f'(\lambda)| \sup_{\lambda_{kp}/2 \leqslant \lambda \leqslant (\lambda_{kp} + \lambda_{jp})/2} |D_p^T(\lambda_{jp} - \lambda)| \right.$$
$$\left. \int_{\lambda_{kp}/2}^{(\lambda_{kp} + \lambda_{jp})/2} |D_p^T(\lambda - \lambda_{kp})| \, \mathrm{d}\lambda \right)$$

$$= \mathrm{O}\left(n^{-1} |j-k| f(\lambda_{kp}) \lambda_{kp}^{-1} |j-k|^{-p} \int_{0}^{\lambda_{jp}} |D_p^T(\lambda)| \, \mathrm{d}\lambda \right)$$

$$= \mathrm{O}(f(\lambda_{kp}) k^{-1} |j-k|^{1-p}) = \mathrm{O}(f_{jk} \, k^{-1} |j-k|^{1-p}),$$

and when $k < j/2$

$$\mathrm{O}\left(n^{-1} [f(\lambda_{jp}) + f(\lambda_{kp})] \sup_{\lambda_{kp}/2 \leqslant \lambda \leqslant (\lambda_{kp} + \lambda_{jp})/2} |D_p^T(\lambda_{jp} - \lambda)| \right.$$

$$\int_{\lambda_{kp}/2}^{(\lambda_{kp}+\lambda_{jp})/2} |D_p^T(\lambda - \lambda_{kp})|\,\mathrm{d}\lambda\Bigg)$$

$$= \mathrm{O}\Bigg(n^{-1}[f(\lambda_{kp}) + f(\lambda_{jp})]n^{1-p}\lambda_{j-k}^{-p}\int_0^{\lambda_{jp}} |D_p^T(\lambda)|\mathrm{d}\lambda\Bigg)$$

$$= \mathrm{O}([f(\lambda_{kp}) + f(\lambda_{jp})]j^{-p}) = \mathrm{O}(f_{jk}\,k^{-p}),$$

with $j - k > j/2$ and $p > d$ again. Then, for n and ε chosen as before, for the following intervals in (Eq. (A.16)) we have

$$\left|\int_{-\pi}^{-\varepsilon} + \int_{\varepsilon}^{\pi}\right| = \mathrm{O}\left(\frac{1}{n^{1-2p}\varepsilon^{2p}}\int_{\varepsilon}^{\pi}[f(\lambda) + f(\lambda_{jp})]\,\mathrm{d}\lambda\right)$$

$$= \mathrm{O}([1 + f(\lambda_{jp})]n^{1-2p}) = \mathrm{O}(f_{jk}\,k^{-p}),$$

and

$$\left|\int_{-\varepsilon}^{-\lambda_{jp}} + \int_{2\lambda_{jp}}^{\varepsilon}\right| = \mathrm{O}\left(f(\lambda_{jp})n^{-1}\left\{\int_{-\varepsilon}^{-\lambda_{jp}} + \int_{2\lambda_{jp}}^{\varepsilon}\right\}|D_p^T(\lambda_{jp} - \lambda)D_p^T(\lambda - \lambda_{kp})|\,\mathrm{d}\lambda\right)$$

$$+ \mathrm{O}\left(n^{-1}\left\{\int_{-\varepsilon}^{-\lambda_{jp}} + \int_{2\lambda_{jp}}^{\varepsilon}\right\}f(\lambda)|D_p^T(\lambda_{jp} - \lambda)D_p^T(\lambda - \lambda_{kp})|\,\mathrm{d}\lambda\right)$$

$$= \mathrm{O}\left(f(\lambda_{jp})n^{1-2p}\int_{\lambda_{jp}}^{\pi} \lambda^{-2p}\,\mathrm{d}\,\lambda + n^{1-2p}\int_{\lambda_{jp}}^{\pi} \lambda^{-2d-2p}\,\mathrm{d}\lambda\right)$$

$$= \mathrm{O}(f(\lambda_{jp})j^{1-2p}) = \mathrm{O}(f_{jk}k^{1-2p}).$$

Next,

$$\left|\int_{-\lambda_{jp}}^{-\lambda_{kp}/2}\right| = \mathrm{O}\Bigg(n^{-1}\sup_{-\lambda_{jp}\leqslant\lambda\leqslant-\lambda_{kp}/2}[f(\lambda) + f(\lambda_{jp})]|D_p^T(\lambda_{jp} - \lambda)|$$

$$\times \int_{-\lambda_{jp}}^{-\lambda_{kp}/2} |D_p^T(\lambda - \lambda_{kp})| \, \mathrm{d}\lambda \Big)$$

$$= \mathrm{O}(n^{-1}[f(\lambda_{kp}) + f(\lambda_{jp})] n^{2-2p} \lambda_{jp}^{-p} \lambda_{kp}^{1-p})$$

$$= \mathrm{O}([f(\lambda_{kp}) + f(\lambda_{jp})] k^{1-p} j^{-p}) = \mathrm{O}(f_{jk}\, k^{-p}),$$

because $p > d$. Finally,

$$\left| \int_{-\lambda_{kp}/2}^{\lambda_{kp}/2} \right| = \mathrm{O}\left(n^{-1} \max_{-\lambda_{kp}/2 \leqslant \lambda \leqslant \lambda_{kp}/2} |D_p^T(\lambda - \lambda_{kp}) D_p^T(\lambda_{jp} - \lambda)| \int_{-\lambda_{kp}/2}^{\lambda_{kp}/2} f(\lambda_{jp}) \, \mathrm{d}\lambda \right) \tag{A.17}$$

$$+ \mathrm{O}\left(n^{-1} \int_{-\lambda_{kp}/2}^{\lambda_{kp}/2} |D_p^T(\lambda - \lambda_{kp}) D_p^T(\lambda_{jp} - \lambda)| f(\lambda) \, \mathrm{d}\lambda \right) \tag{A.18}$$

The first term (Eq. (A.17)) is

$$\mathrm{O}(n^{1-2p} \lambda_{kp}^{1-p} \lambda_{(j-k/2)p}^{-p} f(\lambda_{jp})) = \mathrm{O}(f(\lambda_{jp}) k^{1-p} j^{-p}) = \mathrm{O}(f_{jk}\, k^{1-2p})$$

and the second term (Eq. (A.18)) is $\mathrm{O}(f_{jk}\, k^{1-2p})$ if f is integrable, and otherwise, making a change of variable similar as before, the contribution from this interval, after normalization by f_{jk} and with the obvious notation, is of the same order as

$$\frac{\|a\|_\infty^2}{2\pi b n^{2p-1}} \int_{-\lambda_{kp}/2}^{\lambda_{kp}/2} \frac{f^*(\lambda)}{[f^*(\lambda_{jp}) f^*(\lambda_{kp})]^{1/2}}$$

$$\times \left| \frac{\sin[n(\lambda_{jp} - \lambda)/2p]}{\sin[(\lambda_{jp} - \lambda)/2]} \frac{\sin[n(\lambda_{kp} - \lambda)/2p]}{\sin[(\lambda_{kp} - \lambda)/2]} \right|^p \lambda_{jp}^d \lambda_{kp}^d |\lambda|^{-2d} \, \mathrm{d}\lambda$$

$$= \frac{\|a\|_\infty^2}{2\pi b n^{2p}} \int_{-\pi k p}^{\pi k p} \frac{f^*(\lambda/n)}{[f^*(\lambda_{jp}) f^*(\lambda_{kp})]^{1/2}} \left| \frac{\sin^2[\lambda/2p]}{\sin[(2\pi j p - \lambda)/2n] \sin[(2\pi k p - \lambda)/2n]} \right|^p$$

$$\times \left(\frac{2\pi k p}{n} \right)^d \left(\frac{2\pi j p}{n} \right)^d \left| \frac{\lambda}{n} \right|^{-2d} \mathrm{d}\lambda,$$

and exactly the same bound holds as before for each sine function in the denominator, in terms of k and j, since they behave asymptotically in a similar way in this range of values of λ. Because $0 < k < j$,

$$\sup_{-\pi kp \leqslant \lambda \leqslant \pi kp} |\sin[(2\pi jp - \lambda)/2n]|^{-1} = \mathrm{O}\left(\left(\frac{j}{n}\right)^{-1}\right),$$

as $j/n \to 0$, and

$$\sup_{-\pi kp \leqslant \lambda \leqslant \pi kp} |\sin[(2\pi kp - \lambda)/2n]|^{-1} = \mathrm{O}\left(\left(\frac{k}{n}\right)^{-1}\right).$$

Therefore, the bound for (Eq. (A.18)) is $\mathrm{O}((jk)^{d-p}\log(k+1))$, following part (c) of the theorem.

A similar procedure for the last covariance term in the theorem corresponding to (d),

$$\mathrm{E}[w_p^T(\lambda_{jp})w_p^T(\lambda_{kp})] = \frac{1}{2\pi\sum h_t^2}\int_{-\pi}^{\pi} D_p^T(\lambda_{jp} - \lambda)D_p^T(\lambda + \lambda_{kp})f(\lambda)\,\mathrm{d}\lambda,$$

can be followed easily, obtaining the same bound as for (c), since we do not need to distinguish between frequencies λ_{jp} and λ_{kp} close and far apart. □

Appendix B. Proof of Theorem 4

The proof of this theorem follows the lines of the previous one with $p = 3$, though the cosine bell is not of order 3, but shares properties (25) and (A.2) with $p = 3$, except for the integration of the convolutions around the origin of f. Alternatively, we can use the proof of Theorem 2 in Velasco (1997) for the tapered Fourier transform using the cosine bell for stationary processes, taking special care of those intervals.

For part (a), we consider the normalized expectation of the tapered periodogram which is now

$$\frac{\mathrm{E}[|w^T(\lambda_j)|^2]}{f(\lambda_j)} = \frac{|\sin[\lambda_j/2]|^{2d}}{2\pi\sum h_t^2}\int_{-\pi}^{\pi}\frac{f^*(\lambda)}{f^*(\lambda_j)}\sin^2[n\lambda/2]|\sin(\lambda/2)|^{-2d}H_j^2(\lambda)\,\mathrm{d}\lambda,$$

where H_j is given in (Eq. (15)). Again we only need to concentrate in the interval $[-\lambda_j/2, \lambda_j/2]$. All the other intervals contribute $O(\min\{j^{-1}, j^{-\alpha}\})$ and the term in $[j/n]^\alpha$ is the bias term due to the normalization using Assumption 1.

Then, we can obtain, making a change of variable as in A of the proof of Theorem 1, that the contribution from this interval is bounded by

$$\mathrm{B} \equiv \frac{|\sin[\lambda_j/2]|^{2d}}{n2\pi \sum h_t^2} \int_{-\pi j}^{\pi j} \frac{f^*(\lambda/n)}{f^*(\lambda_j)} \sin^2[\lambda/2] |\sin[\lambda/2n]|^{-2d} H_j^2(\lambda/n)\, \mathrm{d}\lambda$$

$$= O\left(n^{-2} j^{2d} \int_{-\pi j}^{\pi j} \sin^2[\lambda/2] |\lambda|^{-2d} H_j^2(\lambda/n)\, \mathrm{d}\lambda \right).$$

So now it remains to bound $H_j(\lambda/n)$ uniformly for values of λ in $[-\pi j, \pi j]$. Much as before, for $a \equiv (2\pi j - \lambda)/2n \to 0$, as $n \to \infty$, $\lambda \in [-\pi j, \pi j]$,

$$\sin a = a - \frac{1}{3!} a^3 + \frac{1}{5!} a^5 - \frac{1}{7!} a^7 + O(a^9),$$

so

$$(\sin a)^{-1} = a^{-1} + \frac{1}{6} a + \frac{7}{360} a^3 + c a^5 + O(a^7),$$

for some constant c. Similarly, since $a^{-1} = O(n/j)$,

$$\sin^{-1}\left[\frac{2\pi(j+1) - \lambda}{2n}\right] = a^{-1}\left[1 - \frac{(+2\pi)}{2\pi j - \lambda}\right] + \frac{1}{6}\left[a + \frac{(+2\pi)}{2n}\right] + O(n/j^3)$$
$$+ \frac{7}{360}\left[a^3 + 3\, \frac{(+2\pi)(2\pi j - \lambda)^2}{(2n)^3}\right] + O(j/n^3)$$
$$+ c\left[a^5 + 5\, \frac{(+2\pi)(2\pi j - \lambda)^4}{(2n)^5}\right] + O(j^3/n^5)$$

and we can obtain an equivalent expression for $\sin[(2\pi(j-1) - \lambda)/2n]$, substituting the $(+2\pi)$ terms by (-2π). Therefore, in $H_j(\lambda)$ all the terms up to order $O(nj^{-3} + j/n^3)$ cancel out, since $j^3/n^5 = o(j/n^3)$, obtaining with $j/n \to 0$,

$$H_j^2(\lambda/n) = O(n^2[j^{-6} + j^2 n^{-8}]) = O(n^2 j^{-6}).$$

This result corresponds to the well known fact that the tails of the cosine Hanning window are decreasing at the rate $|\lambda|^{-6}$. Finally we get,

$$B = \mathrm{O}(j^{2(d-3)} \log j).$$

For the covariances between tapered Fourier transforms of X_t, as before, we are led to the following expression in the case of the analysis of (b)

$$\mathrm{E}[w^T(\lambda_j)^2] = \frac{1}{8\pi \sum h_t^2} \int_{-\pi}^{\pi} \exp\left(\frac{-\mathrm{i}(n+1)(\lambda_j - \lambda)}{2} - \frac{\mathrm{i}(n+1)(\lambda_j + \lambda)}{2} \right)$$

$$\times \frac{\sin[n(\lambda_j - \lambda)/2]}{\sin[\lambda/2]} \frac{\sin[n(\lambda_j + \lambda)/2]}{\sin[\lambda/2]} H_j(\lambda) H_j(-\lambda) f_\varepsilon(\lambda)\, \mathrm{d}\lambda$$

$$= \frac{1}{2\pi \sum h_t^2} \int_{-\pi}^{\pi} \exp(-\mathrm{i}(n+1)\lambda_j)$$

$$\times \sin[n(\lambda_j - \lambda)/2] \sin[n(\lambda_j + \lambda)/2] H_j(\lambda) H_j(-\lambda) f(\lambda)\, \mathrm{d}\lambda.$$

Therefore, the only difference with respect to the previous integral for (a) in the restricted interval $[-\pi j, \pi j]$ is the cross product $|H_j(\lambda) H_j(-\lambda)|$. However, the same bounds as before hold for each of the H_j functions, so we obtain the equivalent result

$$|H_j(\lambda/n) H_j(-\lambda/n)| = \mathrm{O}(n^2 j^{-6}), \quad \lambda \in [-\pi j, \pi j],$$

and the same bound for that integral as for B. The contribution from the other intervals is, from Theorem 6, $\mathrm{O}(j^{-4})$, because the $\log n$ term does not show up in (Eq. (A.10)) as $p = 3$ and the bound for (Eq. (A.9)) can be improved to $\mathrm{O}(f(\lambda_{jp}) j^{-1-p})$ when $p > 2$ and $d > -\frac{1}{2}$.

The obvious modifications apply for the cross terms at the frequencies λ_j and λ_k and the function H_k, exactly as in the general tapered case, now using parts (c) and (d) of Theorem 6. □

Appendix C. Proof of Theorem 8

Here we can proceed exactly as in the proof of Theorem 2 of Robinson (1995), since the time series is not invertible, but stationary, so the spectral density is well defined, and for example for the expectation of the periodogram we want to study the difference

$$\mathrm{E}[I_p^T(\lambda_{jp})] - f(\lambda_{jp}) = \int_{-\pi}^{\pi} [f(\lambda) - f(\lambda_{jp})] K_p^T(\lambda_{jp} - \lambda)\, \mathrm{d}\lambda.$$

We can consider the following intervals of integration as in the proof of Theorem 6, for the same choice of ε,

$$\left| \int_{-\pi}^{-\varepsilon} + \int_{\varepsilon}^{\pi} \right| \leqslant \max_{|\lambda| \geqslant \varepsilon} |K_p^T(\lambda_{jp} - \lambda)| \int_{-\varepsilon}^{\pi} [f(\lambda) + f(\lambda_{jp})]\, \mathrm{d}\lambda = \mathrm{O}(n^{1-2p}),$$

using the property (Eq. (A.2)) of $K_p^T(\lambda)$ and the integrability of f outside the origin. Note that here we cannot include in the bound a term like $f(\lambda_{jp})$, because this is tending to zero with $j/n \to 0$. Then this is

$$\mathrm{O}(f(\lambda_j) n^{1-2(p+d)} j^{2d}) = \mathrm{O}(f(\lambda_j) j^{1-2p}),$$

since, e.g. $p > |d| + \frac{1}{2}$ and $d < 0$. Next, for a sequence $\delta_n = \delta(n,j)$ with $\lambda_{jp}/\delta_n \to 0$ as $n \to \infty$, to be chosen optimally later,

$$\left| \int_{-\delta_n}^{-\lambda_{jp}/2} \right| \leqslant \left[\max_{\lambda_{jp}/2 \leqslant \lambda \leqslant \delta_n} f(\lambda) + f(\lambda_{jp}) \right] \int_{-\delta_n}^{-\lambda j/2} K_p^T(\lambda_{jp} - \lambda)\, \mathrm{d}\lambda$$

$$= \mathrm{O}\left(\delta_n^{-2d} n^{1-2p} \int_{\lambda_{jp}/2}^{\pi} \lambda^{-2p}\, \mathrm{d}\lambda \right) = \mathrm{O}(\delta_n^{-2d} j^{1-2p})$$

$$= \mathrm{O}(f(\lambda_{jp})(j/n)^{2d} \delta_n^{-2d} j^{1-2p}). \tag{C.1}$$

and,

$$\left| \int_{-\varepsilon}^{-\delta_n} \right| \leqslant \left[\max_{\delta_n \leqslant \lambda \leqslant \varepsilon} f(\lambda) + f(\lambda_{jp}) \right] \int_{-\pi}^{-\delta_n} K_p^T(\lambda_{jp} - \lambda)\, \mathrm{d}\lambda$$

$$= \mathrm{O}(n^{1-2p} \delta_n^{1-2p}) = \mathrm{O}(f(\lambda_{jp})(j/n)^{2d} n^{1-2p} \delta_n^{1-2p}). \tag{C.2}$$

Identical bounds can be obtained when we split the interval $[3\lambda_{jp}/2, \varepsilon]$ using a sequence δ_n. Next

$$\left|\int_{-\lambda_{jp}/2}^{\lambda_{jp}/2}\right| \leqslant \left[\max_{-\lambda_{jp}/2\leqslant\lambda\leqslant\lambda_{jp}/2} K_p^T(\lambda_{jp}-\lambda)\right]\int_{-\lambda_{jp}/2}^{\lambda_{jp}/2}[f(\lambda_{jp})+f(\lambda)]\,\mathrm{d}\lambda$$

$$= \mathrm{O}(f(\lambda_{jp})j^{1-2p}) = \mathrm{O}(f(\lambda_{jp})j^{-1}).$$

Finally, using Assumptions 2 and 3 so as to bound the left-hand side of (Eq. (6)),

$$\left|\int_{\lambda_{jp}/2}^{3\lambda_{jp}/2}[f(\lambda)-f(\lambda_{jp})]K_p^T(\lambda_{jp}-\lambda)\,\mathrm{d}\lambda\right| = \mathrm{O}(f(\lambda_{jp})j^{-\max\{1,\alpha\}}).$$

Now, it remains to find the optimal choice of δ_n. This is given when $\delta_n^{-2d}j^{1-2p} = \delta_n^{1-2p}n^{1-2p}$, so

$$\delta_n = \left(\frac{j}{n}\right)^{(2p-1)/(2(p-d)-1)},$$

and the bounds (C.1) and (C.2) are

$$\mathrm{O}(f(\lambda_{jp})[n^{4d^2/(2(p-d)-1)}j^{2d-(2p-1)^2/(2(p-d)-1)}+j^{-1}]).$$

For the covariance terms the reasoning is exactly the same as in Theorem 6, using Lemmas 1 and 2, and considering the intervals with the optimal sequence δ_n to control the leakage from high frequencies, beyond $|\lambda_j|$, $j > k$. □

References

Agiakloglou, C., Newbold, P., Wohar, M., 1993. Bias in an estimator of the fractional difference parameter. Journal of Time Series Analysis 14, 235–246.

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

Bloomfield, P., 1976. Fourier Analysis of Time Series: an Introduction. Wiley, New York.

Bloomfield P., 1991. Time series methods. In: Hinkley, D.V., Reid, N., Snell, E.J. (Eds.), Statistical Theory and Modeling. Chapman & Hall, London, pp. 152–176.

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

Durlauf, S., Phillips, P.C.B., 1988. Trends versus random walks in time series analysis. Econometrica 56, 1333–1354.
Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4, 221–238.
Giraitis, L., Robinson, P.M., Samarov, A., 1997. Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long range dependence. Journal of Time Series Analysis 18, 49–60.
Granger, C.W.J., 1966. The typical spectral shape of an economic variable. Econometrica 34, 150–161.
Hannan, E.J., 1970. Multiple Time Series. Wiley, New York.
Hassler, U., 1992. Unit root test: the autoregressive approach in comparison with the periodogram regression. Statistical Papers 34, 67–82.
Hurvich, C.M., Ray, B.K., 1995. Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. Journal of Time Series Analysis 16, 17–42.
Percival, D.B., Walden, A.T., 1993. Spectral Analysis for Physical Applications. Cambridge University Press, Cambridge.
Robinson, P.M., 1986. On the errors-in-variables problem for time series. Journal of Multivariate Analysis 19, 240–250.
Robinson P.M., 1992. Log-periodogram regression of time series with long range dependence. Preprint London School of Economics.
Robinson, P.M., 1994a. Rates of convergence and optimal spectral bandwidth for long range dependence. Probability Theory and Related Fields 99, 443–473.
Robinson, P.M., 1994b. Efficient tests of nonstationary hypotheses. Journal of the American Statistical Association 89, 1420–1437.
Robinson, P.M., 1995. Log-periodogram regression of time series with long range dependence. Annals of Statistics 23, 1048–1072.
Velasco, C., 1997. Non-Gaussian log-periodogram regression. Preprint.
Zhurbenko, I.G., 1979. On the efficiency of estimates of a spectral density. Scandinavian Journal of Statistics 6, 49–56.
Zhurbenko, I.G., 1980. On the efficiency of spectral estimates density estimators of a stationary process I. Theory of Probability and its Applications 6, 466–480.
Zhurbenko, I.G., 1982. On the efficiency of spectral estimates density estimators of a stationary process II. Theory of Probability and its Applications 8, 409–419.