Working Paper 98-20
Statistics and Econometrics Series 13
January 1998

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

GAUSSIAN SEMIPARAMETRIC ESTIMATION OF NON-STATIONARY TIME SERIES

Carlos Velasco*

Abstract

Generalizing the definition of the memory parameter d in terms of the differentiated series, we showed in Velasco (1997a) that it is possible to estimate consistently the memory of non-stationary processes using methods designed for stationary long range dependent time series. In this paper we consider the Gaussian semi-parametyric estimate anlyzed in Robinson (1995b) for stationary processes. Without a priori knowledge about the possible non-stationarity of the observed process, we obtain that this estimate is consistent for $d \in (-1/2, 1)$ and asymptotically normal for $d \in (-1/2, 3/4)$ under a similar set of assumptions to Robinson´s paper. Tapering the observations, we can estimate any degree of non-stationarity, even in the presence of deterministic polynomial trends of time. The semi-parametric efficiency of this estimate for stationary sequences also extends to the non-stationary framework.

Keywords:
Non-stationary time series, semiparametric inference, tapering.

*Departamento de Estadística y Econometría, Universidad Carlos III de Madrid. C/ Madrid, 126 28903 Madrid. Spain. Ph: 34-1-624.98.87, Fax: 34-1-624.98.49, e-mail: cavelas@est-econ.uc3m.es. Research funded by the Spanish Dirección General de Enseñanza Superior, Ref. n. PB95-0292.

Gaussian Semiparametric Estimation of Non-Stationary Time Series

Carlos Velasco

Department of Statistics and Econometrics

Universidad Carlos III de Madrid

28903 Getafe (Madrid)

Spain

January 22, 1998

Gaussian Semiparametric Estimation of Non-Stationary Time Series

Carlos Velasco

Abstract

Generalizing the definition of the memory parameter d in terms of the differentiated series, we showed in Velasco (1997a) that it is possible to estimate consistently the memory of non-stationary processes using methods designed for stationary long range dependent time series. In this paper we consider the Gaussian semi-parametric estimate analyzed in Robinson (1995b) for stationary processes. Without a priori knowledge about the possible non-stationarity of the observed process, we obtain that this estimate is consistent for $d \in (-\frac{1}{2}, 1)$ and asymptotically normal for $d \in (-\frac{1}{2}, \frac{3}{4})$ under a similar set of assumptions to Robinson's paper. Tapering the observations, we can estimate any degree of non-stationarity, even in the presence of deterministic polynomial trends of time. The semi-parametric efficiency of this estimate for stationary sequences also extends to the non-stationary framework.

Key words: Non-stationary time series; 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. Frequently, it is assumed that the spectral density $f(\lambda)$ of the observed stationary sequence satisfies for one $0 < G < \infty$,

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

where $d \in (-\frac{1}{2}, \frac{1}{2})$ is the parameter that governs the degree of 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. When $d < 0$ the spectral density satisfies $f(0) = 0$ and if $d \leq -\frac{1}{2}$ the process is not invertible.Many non-stationary time series are transformed into stationary ones after taking enough number of differences. In this case it is straightforward to generalize the definition of the memory parameter d in terms of the properties of the spectral density of the stationary increments of the observed process and the unit root filter(s). Robinson (1995a) recommended an initial, possibly repeated, 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 stationary procedures on the transformed series, adjusting the estimate with the number of differences (integrations) taken.

However in many empirical applications values of d outside the stationary range are found when the estimates are not constrained to the stationary range, $d < \frac{1}{2}$, as is the case of explicit form estimates, like the log-periodogram regression (e.g. Agiakloglou et al. (1993), Bloomfield (1991)). In Velasco (1997a) we considered the application of the log-periodogram regression estimate (see Robinson (1995a) and Geweke and Porter-Hudak (1983)) to the raw non-stationary processes, following some previous ideas in Hassler (1993) and Hurvich and Ray (1995). The last reference considered the expectation of the periodogram at low Fourier frequencies for 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 a function f satisfying (1)) in this case.

Robinson (1995b) found that in the stationary and invertible case an estimate of d minimizing an approximation to a Gaussian likelihood for frequencies close to the origin had better efficiency properties than rival semiparametric estimates, in the sense of having smaller asymptotic variance after proper normalization when using the same amount of sample information. Using Velasco's (1997a) results for the periodogram of non-stationary time series, we address in this paper whether it is possible to extend the range of allowed values of d in this implicitly defined estimate to cover some non-stationary situations and what are the properties of the estimates when the series is non stationary, including some possible efficiency gains.

Under similar conditions to those assumed by Robinson we find that the Gaussian semiparametric estimate is consistent for $d \in (-\frac{1}{2}, 1)$, asymptotically normal for $d < \frac{3}{4}$, with the same variance as

in the stationary situation, being more efficient that the log-periodogram regression estimator. If we taper the observations adequately we can estimate higher degrees of nonstationarity, as was found for the log-periodogram estimate in Velasco (1997a). Finally, we perform a limited numerical study with simulated and real data of these theoretical results. We give all the proofs at the end of the paper in several appendices together with some technical lemmas.

We do not discuss the non-invertible case here, $d \leq -\frac{1}{2}$, but this could be done using similar methods to those of Velasco (1997a) for the log-periodogram estimate (see Theorems 9 and 10 in that paper).

2 Assumptions and definitions

In the first two sections we consider the original estimate analyzed by Robinson (1995b) and concentrate in the interval $-\frac{1}{2} < d < \frac{3}{2}$. When the observed time series is stationary with spectral density $f_X(\lambda)$ satisfying (1), $d < \frac{1}{2}$, we say that the process has memory d and we define the function f, as

$$f(\lambda) = f_X(\lambda).$$

When $\{X_t\}$ is a non-stationary process, we say that it has memory parameter d ($\frac{1}{2} \leq d < \frac{3}{2}$) if the zero mean stationary process $U_t = \Delta X_t$ has spectral density

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

where $f^*(\lambda)$ is a spectral density on $[-\pi, \pi]$ which is bounded above and away from zero and is continuous at $\lambda = 0$. Thus $f_U(\lambda)$ satisfies (1) with some $-\frac{1}{2} \leq d_U < \frac{1}{2}$, but we do not restrict its form for frequencies away the origin. Then we assume, following Hurvich and Ray (1995), that for any $t \geq 1$,

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

where X_0 is a random variable not depending on time t. Next, define the function $f(\lambda)$ for $d \geq \frac{1}{2}$,

$$f(\lambda) = |1 - \exp(i\lambda)|^{-2} f_U(\lambda) = |1 - \exp(i\lambda)|^{-2d} f^*(\lambda) = |2\sin(\lambda/2)|^{-2d} f^*(\lambda).$$

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

We want to give a unified theory for semiparametric estimates of $d \in (-\frac{1}{2}, 1)$, including stationary (with $f_X(0)$ equal to zero, a constant or infinity) and non-stationary processes. We introduce now the following assumptions about the behaviour of the spectral densities $f_X(\lambda)$ ($d < \frac{1}{2}$) and $f_U(\lambda)$ ($d \geq \frac{1}{2}$) (and thus of the functions $f(\lambda)$ and $f^*(\lambda)$) at the origin:

Assumption 1 *When $d \in (-\frac{1}{2}, \frac{1}{2})$, the spectral density $f_X(\lambda)$ satisfies, for $0 < G < \infty$,*

$$f_X(\lambda) \sim G \cdot \lambda^{-2d} \qquad \text{as } \lambda \to 0^+$$

and when $d \in [\frac{1}{2}, \frac{3}{2})$, the spectral density $f_U(\lambda)$ satisfies,

$$f_U(\lambda) \sim G \cdot \lambda^{-2(d-1)} \qquad \text{as } \lambda \to 0^+.$$

A slightly stronger version of this assumption, and that we will use to obtain the asymptotic normality of our estimates is

Assumption 2 *When $d \in (-\frac{1}{2}, \frac{1}{2})$, the spectral density $f_X(\lambda)$ satisfies for numbers $0 < \beta \leq 2$, $0 < G < \infty$,*

$$f_X(\lambda) = G \cdot \lambda^{-2d} + O(\lambda^{-2d+\beta}) \qquad \text{as } \lambda \to 0^+,$$

and when $d \in [\frac{1}{2}, \frac{3}{2})$, the spectral density $f_U(\lambda)$ satisfies

$$f_U(\lambda) = G \cdot \lambda^{-2(d-1)} + O(\lambda^{-2(d-1)+\beta}) \qquad \text{as } \lambda \to 0^+.$$

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

$$\frac{f(\lambda)}{g(\lambda)} = 1 + O(\lambda^{\beta}) \qquad \text{as } \lambda \to 0^+. \tag{2}$$

This is equivalent to Assumption 1 in Robinson (1995a) 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. (1995).

Also, Assumption 2 implies that $f^*(\lambda)$ is bounded above and away from zero and is continuous in an interval $(0, \varepsilon)$, $\varepsilon > 0$.

Assumption 3 *In a neighbourhood $(0, \varepsilon)$ of the origin, if $d \in (-\frac{1}{2}, \frac{1}{2})$, $f_X(\lambda)$ is differentiable and*

$$\left| \frac{d}{d\lambda} f_X(\lambda) \right| = O(\lambda^{-1-2d}) \qquad \text{as } \lambda \to 0^+,$$

and if $d \geq \frac{1}{2}$, $f_U(\lambda)$ is differentiable and

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

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

$$\left| \frac{d}{d\lambda} f(\lambda) \right| = O(\lambda^{-1-2d}) \qquad \text{as } \lambda \to 0^+. \tag{3}$$

These assumptions could have been formulated in terms of the functions f^* and/or f, since we are interested in the implications they have on the function f, (2) and (3). However, we did not find appropriate to make assumptions directly on f or f^*, since these functions have not immediate and clear statistical interpretation as f_U or f_X have.

Now we make the following assumptions about the series U_t when $d \geq \frac{1}{2}$, or for X_t when $d < \frac{1}{2}$, paralleling Robinson (1995b),

Assumption 4 *We have, for $-\frac{1}{2} < d < \frac{1}{2}$, $y_t = X_t$ or for $\frac{1}{2} \leq d < 1$, $y_t = U_t$, with*

$$y_t = \sum_{\ell=0}^{\infty} \alpha_\ell \epsilon_{t-\ell}, \qquad \sum_{\ell=0}^{\infty} \alpha_\ell^2 < \infty,$$

where

$$E[\epsilon_t | F_{t-1}] = 0, \quad E[\epsilon_t^2 | F_{t-1}] = 1, \quad a.s., \quad t = 0, \pm 1, \ldots,$$

in which F_t is the σ-field of events generated by ϵ_t, $s \leq t$, and there exists a random variable ϵ, such that $E\epsilon^2 < \infty$ and for all $\eta > 0$ and some $C > 0$, $P(|\epsilon_t| > \eta) \leq CP(|\epsilon| > \eta)$.

Then we obtain that, $d \geq \frac{1}{2}$,

$$f(\lambda) = |1 - \exp(i\lambda)|^{-2} f_U(\lambda) = |1 - \exp(i\lambda)|^{-2} \frac{|\alpha(\lambda)|^2}{2\pi},$$

where

$$\alpha(\lambda) = \sum_{\ell=0}^{\infty} \alpha_\ell e^{i\ell\lambda}$$

and $|\alpha(\lambda)|^2/2\pi = f_U(\lambda)$, the spectral density of U_t.

Define the discrete Fourier transform of X_t, $t = 1, \ldots, n$, $\lambda_j = 2\pi j/n$, j integer,

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

and when $d \geq \frac{1}{2}$, we obtain,

$$w(\lambda_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \sum_{k=1}^{t} U_k \exp(i\lambda_j t),$$

so $w(\lambda_j)$ is a complex linear combination of the (non observable) stationary variables U_k. The Fourier transform at any frequency λ_j, $0 < j < n$, of a non-stationary sequence X_t allows the elimination of the random variable X_0, so $w(\lambda_j)$ is not depending on the values of U_k for $k < 1$. Define the periodogram of X_t as

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

Because the estimate is not defined in closed form, we denote by G_o and d_o the true parameter values, and by G and d any admissible values. Consider the objective function (see Robinson (1995b) and Künsch (1987)),

$$Q(G, d) = \frac{1}{m} \sum_{j=1}^{m} \left\{ \log G\lambda_j^{-2d} + \frac{I(\lambda_j)}{G\lambda_j^{-2d}} \right\},$$

and define the closed interval of admissible estimates of d_o, $\Theta = [\nabla_1, \nabla_2]$, where ∇_1 and ∇_2 are numbers such that $-\frac{1}{2} < \nabla_1 < \nabla_2 < 1$. Note that we cover part of the range of values of d for which X_t is non-stationary. As in Robinson (1995b) ∇_1 and ∇_2 can be chosen arbitrarily close to $-\frac{1}{2}$ and 1 ($\frac{1}{2}$ in his case), respectively, or reflecting some prior knowledge on d_o. When $d_o \in (-\frac{1}{2}, \frac{1}{2})$ the asymptotics for $I(\lambda_j)$ are exactly the same as in Robinson's discussion, but when $d_o \geq \frac{1}{2}$, we have to resort to the results of Velasco (1997a), weaker in general. Robinson used notation in terms of the parameter $H = d + \frac{1}{2}$,

but we find more natural to use the number of differences parameter d in a possibly non-stationary context. We define the estimates

$$(\widehat{G}, \widehat{d}) = \arg \min_{0<G<\infty, d\in\Theta} Q(G, d),$$

which always exist and also

$$\widehat{d} = \arg \min_{d\in\Theta} R(d),$$

where

$$R(d) = \log \widehat{G}(d) - 2d\frac{1}{m}\sum_{j=1}^{m} \log \lambda_j, \quad \widehat{G}(d) = \frac{1}{m}\sum_{1}^{m} \lambda_j^{2d} I(\lambda_j).$$

Using the discussion in Velasco (1997a), the main idea to show that Robinson (1995b) results go through in the non-stationary case ($d_o \geq \frac{1}{2}$) is to analyse the asymptotic behaviour of the discrete Fourier transform of X_t for frequencies λ_j, $1 \leq j \leq m$, with $1/m + m/n \to 0$ as $n \to \infty$. Therefore, assuming the same conditions for ϵ_k's, we could repeat the steps in Robinson (1995b) to obtain the consistency and asymptotic distribution of the estimate of the parameter d for non-stationary processes. However, due to a bias problem, the same results as in Robinson (1995b) can only be obtained for $d_o < \frac{3}{4}$, consistency holding for $d_o < 1$.

We stress the point that the discrete sum in the previous definitions *cannot* be substituted by an integral form as is considered for related estimates in a full parametric context (see Fox and Taqqu (1986), and Giraitis and Surgailis (1990)), since the properties of the periodogram for non-stationary processes are only equivalent to the stationary case when evaluated at frequencies λ_j, $1 \leq j \leq n-1$.

3 Consistency

In this section we obtain the consistency of $\widehat{d}$ as defined previously for values $d_o \in (-\frac{1}{2}, 1)$. Under Assumptions 2 and 3, the conditions on the behaviour of the function $f(\lambda)$ at the origin of Theorem 1 in Robinson (1995b) hold now also for $d_o \in [\frac{1}{2}, \frac{3}{2})$ (we do not need the integrability of f).

In the stationary case, the analysis of the asymptotic properties of $w(\lambda_j)$ is done in Robinson (1995a). For the non-stationary situation, $d \geq \frac{1}{2}$, we can obtain following some ideas of Hurvich and Ray (1995) that

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

where $K(\lambda) = (2\pi n)^{-1} |\sum_1^n \exp\{i\lambda t\}|^2$ is Fejér kernel. From this expression it is possible to see that when X_t is non stationary, $f(\lambda)$ plays exactly the same role as a spectral density in the asymptotics for the discrete Fourier transform at frequencies λ_j, $j \neq 0 \bmod(n)$, and Velasco (1997a) showed that the periodogram is (asymptotically) unbiased for f if j is growing slowly with n and $d < 1$. This is done in the next theorem, which is Theorem 1 in Velasco (1997a). Defining $v(\lambda) = w(\lambda)/f^{1/2}(\lambda)$,

Theorem 1 *Under Assumptions 1 and 3, $d \in [\frac{1}{2}, 1)$, for any sequences of positive integers $j = j(n)$ and $k = k(n)$ such that $1 \leq k < j$ and $j/n \to 0$ as $n \to \infty$, defining*

$$\delta_{k,j} = (jk)^{d-1} \log(j+1),$$

(a) $E[v(\lambda_j)\overline{v}(\lambda_j)] = 1 + O(\delta_{j,j})$,

(b) $E[v(\lambda_j)v(\lambda_j)] = O(\delta_{j,j})$,

(c) $E[v(\lambda_j)\overline{v}(\lambda_k)] = O\left(k^{-1}\log(j) + \delta_{k,j}\right)$,

(d) $E[v(\lambda_j)v(\lambda_k)] = O\left(k^{-1}\log(j) + \delta_{k,j}\right)$.

The next two results hold in a similar way for the log periodogram estimate of d for non-stationary Gaussian time series. Here we do not need to assume Gaussianity in any form. First we show the consistency of $\widehat{d}$ when $d < 1$:

Theorem 2 *Under Assumptions 1 ($d_o \in \left(-\frac{1}{2}, 1\right)$), 3, 4 and*

$$\frac{1}{m} + \frac{m}{n} \to 0 \qquad \text{as } n \to \infty,$$

we obtain $\widehat{d} \to_p d$.

4 Asymptotic Normality

For values $d_o \geq 1$ the periodogram at frequencies λ_j is not unbiased for the function f as j increases, and therefore $\widehat{d}$ can not be consistent. Unlike for stationary processes, we can only obtain the asymptotic distribution for $\widehat{d}$ in the non-stationary case for a smaller range of values of d_o, ($d_o < \frac{3}{4}$) than the interval where the estimate is consistent, $d_o < 1$. This is due to the fact that the properties of the periodogram depend on convolutions of the function $f(\lambda)$, which deteriorate rapidly as f becomes more "non-integrable", i.e. as d_o increases (see Theorem 1 above and Theorem 1 in Velasco (1997a), and the subsequent discussion).

We introduce two new assumptions that will be needed in the proofs.

Assumption 5 *In a neighbourhood $(0, \varepsilon)$ of the origin, $\alpha(\lambda)$ is differentiable and*

$$\left|\frac{d}{d\lambda}\alpha(\lambda)\right| = O\left(\frac{|\alpha(\lambda)|}{\lambda}\right) \qquad \text{as } \lambda \to 0^+.$$

Clearly Assumption 5 implies Assumption 3, since $f(\lambda) = |\alpha(\lambda)|^2/2\pi$ when $-\frac{1}{2} < d_o < \frac{1}{2}$ and $f(\lambda) = (2\sin\lambda/2)^{-2}|\alpha(\lambda)|^2/2\pi$ when $d_o \geq \frac{1}{2}$.

Assumption 6 *Assumption 4 holds and also*

$$E[\epsilon_t^3|F_{t-1}] = \mu_3 \quad a.s., \quad E[\epsilon_t^4|F_{t-1}] = \mu_4, \qquad t = 0, \pm 1, \ldots,$$

for finite constants μ_3 and μ_4.

Theorem 3 *Under Assumptions 2, 5, 6, with $d_o \in (-\frac{1}{2}, \frac{3}{4})$, and*

$$\frac{1}{m} + \frac{m^{1+2\beta}(\log m)^2}{n^{2\beta}} \to 0 \qquad as \;\; n \to \infty, \tag{4}$$

we obtain

$$m^{1/2}(\widehat{d} - d_o) \to_D N(0, \frac{1}{4}).$$

This theorem coincides, not surprisingly, with the results of Velasco (1997a) for the log-periodogram regression estimate of non-stationary time series with Gaussian increments. Beyond these values of d, the slow convergence of the expectation of the periodogram to the function f leads to a slower convergence of the estimates of d. In Velasco (1997a) this problem was overcome for the log-periodogram estimate using the bias reduction technique of tapering, as suggested by Hurvich and Ray (1995). We do not pursue this approach here, but the corresponding theory is similar to that obtained in the next section for general nonstationary processes and tapering schemes.

Another important point is that the efficiency property of this Gaussian estimate with respect to other comparable semiparametric estimates observed by Robinson (1995b) for stationary process, holds as well for non-stationary processes when the same number of periodogram ordinates, m, is used. Further, the asymptotic distribution of $\widehat{d}$ does not depend on any unknown constants, not even d_o, beyond the definition of the suitable range of valid values for the theorem, which is only limited by $d < \frac{3}{4}$.

5 General non-stationary time series

In this section we consider the estimation of the memory parameter for general non-stationary time series which after a finite number of differentiations are stationary. In general, a (possibly non-stationary) process $\{X_t\}$ has memory parameter $d > -\frac{1}{2}$ if the process $\Delta^s X_t = U_t^{(s)}$, $s = \lfloor d + \frac{1}{2} \rfloor$, is stationary with mean μ, possibly different from zero, and spectral density $f_{U^{(s)}}(\lambda)$ behaving as $G\lambda^{-2(d-s)}$, $-\frac{1}{2} \leq d - s < \frac{1}{2}$, around the origin for some positive constant G. Robinson (1995b) considered the case $s = 0$ and in Section 2 we have considered the case $s = 1$, $d < 1$, $\mu = 0$.

Define the function

$$f(\lambda) = |1 - \exp(i\lambda)|^{-2s} f_{U^{(s)}}(\lambda) = |2\sin(\lambda/2)|^{-2d} f^{\star}(\lambda),$$

in terms of the spectral density of the stationary sequence $U_t^{(s)}$ or the function $f^{\star}(\lambda)$. Following the discussion in Velasco (1997a), we can write for random variables $R^{(r)}$, $r = 1, \ldots, s$ which do not depend on time t

$$\begin{aligned} X_t &= R^{(1)} + \sum_{j_1=1}^{t} U_{j_1}^{(1)} \\ &= R^{(1)} + \sum_{j_1}^{t} \left(R^{(2)} + \sum_{j_2}^{j_1} U_{j_2}^{(2)} \right) \end{aligned}$$

$$
\begin{aligned}
&= R^{(1)} + t\,R^{(2)} + \sum_{j_1}^{t}\sum_{j_2}^{j_1}\left(R^{(3)} + \sum_{j_3}^{j_2} U_{j_3}^{(3)}\right) \\
&= R^{(1)} + t\,R^{(2)} + \frac{1}{2}(t+t^2)R^{(3)} + \sum_{j_1}^{t}\sum_{j_2}^{j_1}\sum_{j_3}^{j_2} U_{j_3}^{(3)} \\
&= \sum_{r=1}^{s} R^{(r)}p^{(r)}(t) + \mu\, p_\mu(t) + \sum_{j_1}^{t}\sum_{j_2}^{j_1}\cdots\sum_{j_s}^{j_{s-1}} U_{j_s}^{(\star)},
\end{aligned}
$$

where $p^{(r)}(t)$ are polynomials in t of order $r-1$, $p_\mu(t)$ is a polynomial of order s and $U_t^{(\star)} = U_t^{(s)} - \mu$ has zero mean and the same spectral density as $U_t^{(s)}$. These two polynomials can be regarded as the initial conditions of the observed non-stationary sequence and as a deterministic trend, respectively. In Velasco (1997a) we proposed to use, instead of the original series, a tapered version with a weight sequence $\{h_t\}_{t=1}^n$, symmetric around $\lfloor n/2 \rfloor$, such that $\max_t h_t = 1$. Hurvich and Ray (1995) used the cosine bell to analyze the expectation of the periodogram when $d < 1.5$. Others authors, Zhurbenko (1979), Robinson (1986), Dahlhaus (1988), have also shown that tapering allows inference in the presence of non-stationary distortions in the observed stationary time series.

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

$$
\begin{aligned}
w^T(\lambda_j) &= \frac{1}{\sqrt{2\pi\sum h_t^2}}\sum_{t=1}^{n} w_t X_t \exp(i\lambda_j t) \\
&= \frac{1}{\sqrt{2\pi\sum h_t^2}}\sum_{t=1}^{n} w_t \left(\sum_{r=1}^{s} R^{(r)}p^{(r)}(t) + \mu p_\mu(t)\right)\exp(i\lambda_j t) \qquad (5) \\
&\quad + \frac{1}{\sqrt{2\pi\sum h_t^2}}\sum_{t=1}^{n} w_t \sum_{j_1}^{t}\sum_{j_2}^{j_1}\cdots\sum_{j_s}^{j_{s-1}} U_{j_s}^{(\star)}\exp(i\lambda_j t). \qquad (6)
\end{aligned}
$$

The term (6) is reflects the accumulation of information in the non-stationary time series X_t, starting from $t = 1$, but the term (5) is a nuisance component of the discrete Fourier transform which comprises the information in $\{X_t\}_1^n$ from the past. To make inferences about d we make this expression (5) equal to zero for certain frequencies λ_j, using specific orthogonality properties of the weights h_t, i.e.

$$
\sum_{t=1}^{n} h_t\left(1 + t + t^2 + \cdots + t^s\right)\exp(i\lambda_j t) = 0. \qquad (7)
$$

Observe than in the case $s = 1$ we have only required that $\sum_{t=1}^n h_t \exp(i\lambda_j t) = 0$, because we were assuming $\mu = 0$ to eliminate the influence from the polynomial $p^{(1)}(t) = 1$ of order 0 (a constant with respect to t). The raw Fourier transform satisfies condition (7), $s = 0$ (but not any of higher order). In other words, without tapering we can consider $d < 1$ but always without drift.

Defining the equivalent to the Dirichlet kernel in the tapered case,

$$
D_p^T(\lambda) = \sum_{t=1}^{n} h_t e^{it\lambda},
$$

we say that a <u>sequence of data tapers $\{h_t\}_1^n$ is of order $p = 1, 2, \ldots$</u> if the following two conditions are satisfied:

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

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

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

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

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

Then, it is immediate to obtain that

$$|D_p^T(\lambda)| \leq \text{const.} \min \left\{ n,\ n^{1-p}|\lambda|^{-p} \right\}$$

and, with the equivalent to the Fejér kernel, $K_p^T(\lambda) = (2\pi \sum h_t^2)^{-1} |D_p^T(\lambda)|^2$,

$$|K_p^T(\lambda)| \leq \text{const.} \min \left\{ n,\ n^{1-2p}|\lambda|^{-2p} \right\}.$$

Also we have that $D_p^T(\lambda_{jp})$ has zeroes of order p and that thanks to

$$\left. \frac{d^{p-1}}{(d\lambda)^{p-1}} D_p^T(\lambda) \right|_{\lambda=\lambda_{jp}} = 0, \qquad 0 < j < N,$$

condition (7) is satisfied.

If condition (7) holds, deterministic time trends up to order s can be removed in the calculation of $w^T(\lambda_j)$ without need to estimate them by any means. The cosine bell taper is of order 1, so its utilization is only justified in the case $d < 1.5$ with $\mu = 0$, as was shown by Velasco (1997a) for the log-periodogram semiparametric estimate. Here we do not consider explicitly this tapering scheme, but given the asymptotic behaviour of tails of the kernel K_p^T in this case, the conclusions are equivalent to those with $p = 3$ and for $d < 1.5$.

Two examples of data tapers satisfying the above conditions are Parzen and Zhurbenko-Kolmogorov proposals (see also Alekseev (1996) for further examples and discussion). For sample size $n = 4N$, N integer, the weights given by the Parzen window

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

satisfy (7) for $j = 4, 8, \ldots, n-4$ and $s = 3$. We can obtain (see e.g. Percival and Walden (1993))

$$D^P(\lambda) = \frac{32}{n^3}(3 - 2\sin^2\lambda/2) \left(\frac{\sin n\lambda/8}{\sin \lambda/2} \right)^4 \exp\{in\lambda/2\}$$

and $\sum_{t=1}^{n} (h_t^P)^2 \sim \text{const.}\, n$. Zhurbenko (1979) used the data weights $\{h_t^Z\}$ suggested by Kolmogorov,

$$h_t^Z = \rho(p, N) \left(\frac{p(N^2-1)}{12\pi} \right)^{1/4} N^{-p}\, c_{p,N}(t),$$

where the coefficients $c_{p,N}(t)$ are given by

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

Then, it follows that

$$\sqrt{2\pi \sum h_t^2} D^Z(\lambda) = \rho \left(\frac{p(N^2-1)}{12\pi}\right)^{1/4} \left(\frac{1-e^{iN\lambda}}{N(1-e^{i\lambda})}\right)^p,$$

and hence

$$K^Z(\lambda) = \rho^2 \left(\frac{p(N^2-1)}{12\pi}\right)^{1/2} \left(\frac{\sin^2[n\lambda/2p]}{N^2 \sin^2[\lambda/2]}\right)^p,$$

where ρ is defined adequately to make K^Z integrate to one and it can be seen to be very close to 1 for p and N big enough (see Zhurbenko (1980)). Therefore, this class of taper weights for $p = 1, 2, \ldots$, fixed in the asymptotics, and $n = pN$ satisfies condition (7) with $s \leq p-1$ at frequencies λ_{jp}, $0 < j < N$.

6 Tapered estimates

In this section we obtain the consistency and asymptotic distribution of a modified version of $\widehat{d}$ when we use the previous data tapers for values $d_o > -\frac{1}{2}$. We introduce now the following assumptions about the behaviour of the spectral density $f_{U(s)}(\lambda)$ (and thus of the functions $f(\lambda)$ and $f^*(\lambda)$) at the origin:

Assumption 7 *The spectral density $f_{U(s)}(\lambda)$, $s = \lfloor d + \frac{1}{2} \rfloor$, satisfies, for some constant $0 < G < \infty$,*

$$f_{U(s)}(\lambda) \sim G \cdot \lambda^{-2(d-s)} \qquad as \;\; \lambda \to 0^+. \tag{8}$$

A slightly stronger version of Assumption 2 is the following condition, where we give more information about the behaviour of the spectral density $f_{U(s)}(\lambda)$ at the origin. This extra information will be used to reduce the bias of the tapered periodogram for f as was done in Velasco (1997b) in a related context (see also Assumption 3 in Robinson (1994b)).

Assumption 8 *When $d \in (-\frac{1}{2}, \frac{1}{2})$, the spectral density $f_{U(s)}(\lambda)$ satisfies for numbers $0 < \beta \leq 2$, $0 < G, E_\beta < \infty$,*

$$f_{U(s)}(\lambda) = G \cdot \lambda^{-2(d-s)} + E_\beta \lambda^{-2(d-s)+\beta} + o(\lambda^{-2(d-s)+\beta}) \qquad as \;\; \lambda \to 0^+.$$

As before, Assumption 8 implies that $f^*(\lambda)$ is bounded above and away from zero and is continuous in an interval $(0, \varepsilon)$, $\varepsilon > 0$.

We will need also the equivalent to Assumption 3

Assumption 9 *In a neighbourhood $(0, \varepsilon)$ of the origin, if $d \in (-\frac{1}{2}, \frac{1}{2})$, $f_{U(s)}(\lambda)$ is differentiable and*

$$\left| \frac{d}{d\lambda} f_{U(s)}(\lambda) \right| = O(\lambda^{-1-2(d-s)}) \qquad as \; \lambda \to 0^+.$$

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

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

Now we make the following assumption about the series $U_t^{(s)}$, equivalent to Assumption 6.

Assumption 10 *We have*

$$U_t^{(*)} = \sum_{\ell=0}^{\infty} \alpha_\ell \epsilon_{t-\ell}, \qquad \sum_{\ell=0}^{\infty} \alpha_\ell^2 < \infty,$$

where the ϵ_t's satisfy the conditions of Assumptions 4 and 6.

Then we obtain for any $d > -\frac{1}{2}$ that

$$f(\lambda) = |1 - \exp(i\lambda)|^{-2s} f_{U(s)}(\lambda) = |1 - \exp(i\lambda)|^{-2s} \frac{|\alpha(\lambda)|^2}{2\pi}.$$

Defining the (tapered) periodogram of X_t as

$$I_p^T(\lambda_j) = \left|w_p^T(\lambda_j)\right|^2$$

we consider now the objective function

$$Q_p(G, d) = \frac{p}{m} \sum_j^m \left\{ \log G\lambda_j^{-2d} + \frac{I_p^T(\lambda_j)}{G\lambda_j^{-2d}} \right\},$$

where all the summations run for $j = p, 2p, \ldots, m$, assuming m/p integer, unless otherwise stated. Define the closed interval of admissible estimates of d_o, $\Theta = [\nabla_1, \nabla_2]$, where ∇_1 and ∇_2 are numbers such that $-\frac{1}{2} < \nabla_1 < \nabla_2 < d^*$, where $p \geq \lfloor d^* + \frac{1}{2} \rfloor + 1$. This last condition is equivalent to $d^* < p + \frac{1}{2}$, where d^* is the maximum value of d we can estimate with tapers of order p. Note that we can cover part of the range of values of d for which X_t is non-stationary. As in Robinson (1995b), ∇_1 and ∇_2 can be chosen arbitrarily close to $-\frac{1}{2}$ and to a maximum value of d, d^*, restricted only by the order p of the taper weights used, or reflecting some prior knowledge on d_o. When $\mu = 0$ with $d^* < p$ is enough.

We define the estimates

$$(\widehat{G}_p, \widehat{d}_p) = \arg \min_{0<G<\infty, d\in\Theta} Q_p(G, d),$$

which always exist and also

$$\widehat{d}_p = \arg\min_{d\in\Theta} R_p(d),$$

where

$$R_p(d) = \log \widehat{G}_p(d) - 2d\frac{p}{m} \sum_j^m \log \lambda_j, \qquad \widehat{G}_p(d) = \frac{p}{m} \sum_j^m \lambda_j^{2d} I_p^T(\lambda_j).$$

The discrete sums in the previous definitions include only frequencies λ_j, $j = p, \ldots, m$, since the properties of the periodogram for non-stationary processes are only equivalent to the stationary case when evaluated at those frequencies.

When X_t is non stationary, $f(\lambda)$ plays exactly the same role as a spectral density in the asymptotics for the discrete Fourier transform at frequencies λ_j, $j \neq 0 \bmod(n)$, and Velasco (1997a) showed that the

periodogram is (asymptotically) unbiased for f if j is growing slowly with n and p is chosen adequately. This is done in the next theorem, which is essentially Theorem 6 in Velasco (1997a). Note that the non-tapered periodogram is an estimate with $p = 1$. Defining now $v_p^T(\lambda) = w^T(\lambda)/(G^{1/2}\lambda^{-d})$, for a taper of order p,

Theorem 4 ($p \geq 2$) *Under Assumptions 8 and 9 [$d > -\frac{1}{2}$, $0 < \beta \leq 2$] for $f_{U(\cdot)}$, a data taper of order $p = 2, 3, \ldots$, with $p \geq s + 1$ [or just $p > d$ if $\mu = 0$], for any sequences of positive integers $k = k(n)$ and $j = j(n)$, $1 \leq k < j$, and $\eta = j - k$, such that $j/n \to 0$,*

$$\gamma_{j,k} \equiv (jk)^{d-p} \log(j+1) \to 0$$

we get

(a) $E[v_p^T(\lambda_{jp})\overline{v_p^T}(\lambda_{jp})] = 1 + O\left(\min\{j^{-\beta}, j^{-1}\} + [j/n]^{\beta} + \gamma_{j,j}\right)$,

(b) $E[v_p^T(\lambda_{jp})v_p^T(\lambda_{jp})] = O\left(j^{-p} + \gamma_{j,j}\right)$,

(c) $E[v_p^T(\lambda_{jp})\overline{v_p^T}(\lambda_{kp})] = O\left(k^{-1}\eta^{1-p} + k^{-1}\eta^{-p}\log n + \eta^{-p} + \gamma_{k,j}\right)$,

(d) $E[v_p^T(\lambda_{jp})v_p^T(\lambda_{kp})] = O\left(k^{-1}\eta^{1-p} + k^{-1}\eta^{-p}\log n + \eta^{-p} + \gamma_{k,j}\right)$.

Then we obtain the consistency of $\widehat{d}_p$ in the following Theorem. Note that we only require for this result Assumption 7, but not 8, which will be used to derive the asymptotic distribution of $\widehat{d}$ in the next section and that was used in the previous theorem because we normalized the discrete Fourier transform by $(G\lambda^{-2d})^{1/2}$ and not by $(f(\lambda))^{1/2}$.

Theorem 5 *Under Assumptions 7, 9 and 10, with $\nabla_1 > -\frac{1}{2}$ and $p \geq \lfloor \nabla_2 + \frac{1}{2} \rfloor + 1$ such that $d_o \in [\nabla_1, \nabla_2]$, $p = 2, 3, \ldots$ and*

$$\frac{1}{m} + \frac{m}{n} \to 0 \qquad \text{as } n \to \infty,$$

we obtain $\widehat{d}_p \to_P d_o$.

If we assume $\mu = 0$ then we only need in fact $p > \nabla_2$ if there are only deterministic trends in X_t up to order $p - 1$. We do not consider here the case $p = 1$ because this is equivalent to the non-tapered situation, with $\nabla_2 < 1$ (and $\mu = 0$ necessarily). With respect to Theorem 2, the only extra condition we have used is the fourth moment of the innovations ϵ_t in Assumption 10.

Then we obtain the asymptotic normality of $\widehat{d}_p$,

Theorem 6 *Under Assumptions 5, 8 ($\beta > 1$, $\nabla_1 > -\frac{1}{2}$ and $p \geq \lfloor \nabla_2 + \frac{1}{2} \rfloor + 1$ such that $d_o \in [\nabla_1, \nabla_2]$, $p = 2, 3, \ldots$), 10 and*

$$\frac{1}{m} + \frac{m^{1+2\beta}(\log m)^2}{n^{2\beta}} \to 0 \qquad \text{as } n \to \infty, \tag{9}$$

we obtain

$$m^{1/2}(\widehat{d} - d_o) \to_D N(0, \tfrac{1}{4} p \Phi),$$

where

$$\Phi = \lim_{n\to\infty} \left(\sum_1^n h_t^2\right)^{-2} \sum_{k=0,p,2p,\ldots}^{n-p} \left[\sum_1^n h_t^2 \cos t\lambda_k\right]^2 . \tag{10}$$

This theorem is equivalent to the results of Velasco (1997a) for the log-periodogram regression estimate of non-stationary time series with Gaussian increments. There, we changed the definition of the estimate to adapt the proofs of Robinson (1995a), but here, even with the correlation between the tapered periodogram ordinates we do not need to modify the definition of the estimate. However, the variance of the estimate is increased slightly by a factor of Φ (generally bigger than 1), because of this correlation of the tapered periodogram, due to the lack of orthogonality of the taper weights. This Φ takes the values of 1.05000, 1.00354 and 1.00086 for the Zhurbenko kernels with $p = 2, 3, 4$ respectively, implying increments of the variance of 5%, 0.35% and 0.09% for each of the data tapers (apart from the factor p/m in the normalization of the estimate). When $\mu = 0$, then the theorem is valid with just $p > \nabla_2$. If we consider the full cosine window taper, $h_t = \frac{1}{2}(1 - \cos[2\pi t/n])$, then if we regard this taper as of order $p = 3$, with the same definitions as before, $\mu = 0$ and $d < \frac{3}{2}$, Theorem 6 holds with $\Phi = 1$, but if we use all the Fourier frequencies, from λ_2 to λ_m (i.e., without spacing), then $\Phi = \frac{35}{18}$ (see the discussion in Velasco (1997a, 1997b)). Note also that if we take in (10) the sum accross all frequencies we obtain with Parseval's identity,

$$\left(\sum_1^n h_t^2\right)^{-2} \sum_{k=0,1,2,\ldots}^{n-1} \left[\sum_1^n h_t^2 \cos t\lambda_k\right]^2 = n \left(\sum_1^n h_t^2\right)^{-2} \sum_{t=1}^n h_t^4,$$

the right hand side being the usual tapering variance adjustment (cf., e.g., Dahlhaus (1985), expression (3)).

The increased smoothness of the function $f(\lambda)$, $\beta > 1$, is used in conjunction with the tapering to approximate the periodogram of the observed time series by that of the innovations (see the proof of Theorem 6 in Velasco (1997a) and Theorem 2 in Velasco (1997b)). Here we cannot resort to the second moments of the tapered periodogram as was done in the non-tapered case, since the correlation problem just pointed out impedes further improvement of the approximations.

7 Empirical work

The aim of the first simulation exercise is to address the previous properties of $\widehat{d}$, specially in comparison with the log-periodogram regression estimate,

$$\widetilde{d} = -\frac{1}{2} \frac{\sum_{j=1}^m \log I(\lambda_j)[\log j - (1/m)\sum_{\ell=1}^m \log \ell]}{\sum_{j=1}^m \log j[\log j - (1/m)\sum_{\ell=1}^m \log \ell]}.$$

To that end we simulate 1000 Gaussian Fractional ARIMA$(0, d, 0)$ for each value of d in .45(.1)1.25, $n = 256$ and we choose a relatively small value for m, 32. We do not perform any trimming in the definition of $\widetilde{d}$. The series are simulated with the S-Plus function `arima.frac.diff` and the minimum of the objective function is found with the `nlmin` command. In the search for the minimum we use as

initial values for d and G those obtained with the log-periodogram regression, and we do not restrict the range of possible values for d. This procedure gave no problems for any value of d, indicating a relatively well-behaved objective function, even for values $d > 1$.

The box plots for the estimates are given in Figure 1, only up to $d = 1.05$. The main features of the plots are the invariance of the distributions of $\widehat{d}$ and $\widetilde{d}$ to the actual value of d_o and the efficiency and smaller bias of the Gaussian estimate with respect to the log-periodogram across all d_o. For $d_o = 1.05$ ($d_o \geq 1$) neither of the two estimates is consistent and this fact is reflected by the negative bias for both, just in the opposite direction of the biases when $d_o < 1$.

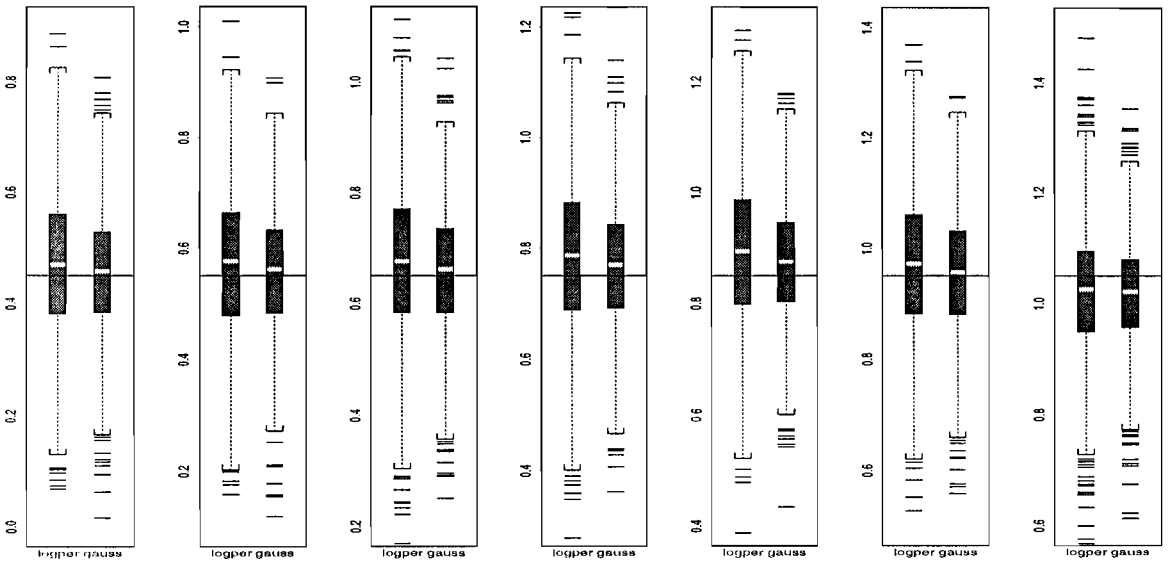

Figure 1: Gaussian semiparametric and Log-periodogram estimates, Gaussian ARFIMA$(0, d, 0)$, $n = 256$, $m = 32$, 1000 Replications.

The basic statistics summary is contained in the Table I, including the bias of the estimates, the standard deviation, the expected standard deviation from the corresponding central limit theorems and the mean square error (MSE) across replications. Note that for $d_o \geq \frac{3}{4}$, Theorem 3 does not hold.

Table I	Gaussian estimate				Log-Periodogram estimate			
d_o	bias	s.d.	th. s.d.	MSE	bias	s.d.	th. s.d.	MSE
0.45	0.0041	0.1151	0.0884	0.0132	0.0179	0.1383	0.1134	0.0194
0.55	0.0050	0.1115	0.0884	0.0124	0.0186	0.1330	0.1134	0.0180
0.65	0.0089	0.1155	0.0884	0.0134	0.0259	0.1446	0.1134	0.0215
0.75	0.0164	0.1168	-	0.0139	0.0324	0.1439	-	0.0217
0.85	0.0213	0.1116	-	0.0129	0.0398	0.1399	-	0.0211
0.95	0.0026	0.1108	-	0.0123	0.0160	0.1342	-	0.0182
1.05	-0.0309	0.1004	-	0.0110	-0.0286	0.1240	-	0.0161
1.15	-0.0837	0.0969	-	0.0164	-0.0867	0.1207	-	0.0221
1.25	-0.1638	0.1043	-	0.0377	-0.1776	0.1251	-	0.0472

In the second simulation we consider the estimation of values $d \geq 1$. The only modification with respect to the previous exercise is that now the series are of length $n = 512$ and $m = 100$. The values of d_o considered are .95 and 1.8, one close to the border line of the asymptotics presented in this paper for this estimate and the other well outside. The results for $\widehat{d}$ and $\widetilde{d}$ are given in Figure 2. In the top row of graphics we give the box plots and in the bottom row nonparametric smoothed estimates of the simulated probability density of the estimates of d. The two left most columns of plots, for $d = .95, 1.9$ indicate that the two semiparametric estimates considered work relatively well for values close to 1, but not for more non-stationary time series, for which the estimates converge extremely fast to values close to 1, except for a long tail towards the right value. The plots on the right are for the same estimates, but when we use a tapered periodogram with the triangular Barlett window taper (equivalent to Zhurbenko tapers with $p = 2$), and we define our estimates for frequencies $\lambda_2, \lambda_4, \ldots, \lambda_m$, assuming m is even. In this case it seems also that the Gaussian estimate is more efficient that the log-periodogram regression.

Now we consider a simple application with real data. Different parameterizations have been proposed in the literature to explain the persistence in the volatility of the returns found in many financial data sets. Robinson (1991) introduced a long memory generalized ARCH model which was retaken by Baillie at al. (1996) and Bollerslev and Mikkelsen (1996) to define the FIGARCH class,

$$\phi(L)(1-L)^d x_t^2 = \omega + b(L)\nu_t,$$

where all the roots of the polynomials ϕ and b in the lag operator L lie outside the unit circle and $\nu_t = x_t^2 - \tau_t^2$ are martingale differences, $E[\nu_t|\mathcal{F}_{t-1}] = 0$, $\tau_t^2 = Var[x_t|\mathcal{F}_{t-1}]$, a.s and $\mathcal{F}_t$ is the $\sigma-$ field of events generated by $\{x_s : s \leq t\}$.

These models allow persistence or long memory in the squares x_t^2 of martingale-difference levels x_t when $d > 0$ and are basically equivalent to the fractional ARIMA models for means, but in the variance, generalizing for any $0 \leq d \leq 1$ the full integrated IGARCH model, equivalent to a unit root

in the mean. Though our asymptotic theory for semiparametric estimation are not ready applicable for this situation (due to the linear process assumption) we investigate the possible utility of the tapered estimates proposed in exploratory analysis to detect the persistence in some crude approximations to the volatility (like the squares and absolute value of the levels) without need to model the short run dependence. The above models are strictly stationary for any $0 \leq d \leq 1$, but a further difficulty is that when $\omega > 0$ the squared process has a drift term so it is non-covariance stationary. We hope that with enough tapering (large p) we can alleviate the effect of this possible drift, which is a smooth function of the time t, and could be well approximated by polynomials of t.

We do this for two data sets corresponding to the returns (defined as the increment of the logarithm) of the exchange rates of the French Franc (FF) and the Deusche Mark (DM) against US dollar, using 2000 daily observations, running form November, 1972 to January, 1981. The plots of the relevant series are in Figure 3 and the results are in Figure 7. We employ bandwidth numbers $m = 15, 18, \ldots, 100$ and tapers with $p = 1, 2, 3$. We plot all the estimates obtained in that way, using the squares and the absolute value of the return series.

The main conclusions we can draw are as follows. The estimates with $p = 1$ usually obtain a lower range of values than the ones with higher values of p. In all cases, when we take m *too big*, the estimates produce much lower values of $\widehat{d}$ as a consequence of moving away from the origin, where we would not expect the model (1) to hold. For the significative range of values of m the estimates with $p = 2$ and 3 are almost always very close, indicating perhaps that with $p = 1$ we can not estimate appropriately high values of d. For the French Franc the persistence in the volatility is in general higher than for the Mark, obtaining values of d up to .9 with the absolute value while only of 0.7 for the DM. This agrees with the findings of the previous authors, who reported values for the DM between 0.6 and 0.8 depending on the parametric model assumed for the short run dynamics of the volatility.

8 Discussion

In this paper and in Velasco (1997a) we have shown that the semiparametric model (1) is valid to estimate the memory d of possibly non-stationary time series. If the observed process is non-stationary $f(\lambda)$ is no longer a spectral density, but is the limit of the expectation of the (tapered) periodogram, and therefore can be estimated non-parametrically. Both the log-periodogram and the Gaussian semi-parametric estimates compare the non-parametric estimate of $f(\lambda)$ given by the periodogram at the relevant frequencies with the model (1) and obtain the best estimate of d under different criteria. For that, it does not matter the integrability or not of the function f around the origin, only the accuracy with which we can estimate it by means of the periodogram ordinates. Of course, the steeper and more non-integrable f is, the more complicated this approximation will be, but the error can be controlled if enough degree of tapering is applied.

The same principle will undoubtedly work for full parametric models of functions f corresponding to nonstationary observations if tapered observations are used. Then, simultaneous estimation of d and the other short-run memory parameters is possible without a priori assumptions about the degree of (possible) non-stationarity of the observed sequence.

Nevertheless this approach will surely break down if we try to estimate the integral below $f(\lambda)$, $\int_0^\alpha f(\lambda)d\lambda$ for any $\alpha > 0$, instead of the function f itself, since this integral diverges for $d \geq \frac{1}{2}$. This problem arises for the semiparametric estimate of d considered by Robinson (1994a) and Lobato and Robinson (1996), based precisely on the estimation of the cumulative spectral distribution function. Simulations with this estimator $\bar{d}$ always result in estimates of d constrained to $\bar{d} < \frac{1}{2}$, for any $d \geq \frac{1}{2}$ and any order of data tapering.

A further approach to deal with long memory, non-stationarity and polynomial trends could be the use of wavelets and there are several recent references which deal with the estimation of d and related topics for fractional white noise inference using wavelets (e.g. Jensen (1995), McCoy and Walden (1996) and the references therein). Based on the wavelet decomposition of the variance at different scales, a variety of estimates of d are proposed, some close to the log-periodogram estimate and others related to Gaussian maximum likelihood, always using the information at all possible scales, being mainly then of full parametric nature. The lack of rigorous asymptotic theory for such estimates in a general case is related with some possible bias problems if the spectral density is not proportional to λ^{-2d} for all frequencies. Furthermore, the assumption of covariance stationarity of the filtered series makes difficult to predict how these procedures will deal with non-stationary observations.

9 Appendix: Proofs of Section 3

Proof of Theorem 2. We repeat the steps of the proof of Theorem 1 in Robinson (1995b), with the same definitions and with the notation in terms of $d = H - \frac{1}{2}$, readjusting accordingly the set of admissible values $[\nabla_1, \nabla_2]$. More details can be found in that reference or in the proof of Theorem 5. We will concentrate mainly in the asymptotics when $d_o \geq \frac{1}{2}$, since the case $d_o \in (-\frac{1}{2}, \frac{1}{2})$ is covered in Robinson's paper.

As in Robinson's proof we define $\nabla = \nabla_1$ when $d_o < \frac{1}{2} + \nabla_1$ and $d_o - \frac{1}{2} < \nabla \leq d_o$ otherwise. Then define $\Theta_1 = \{d : \nabla \leq d \leq \nabla_2\}$, and $\Theta_2 = \{d : \nabla_1 \leq d < \nabla\}$, possibly empty. We retake the proof after expression (3.12) in that reference. Given that now we can consider values of d arbitrarily close to 1, we obtain that for $r = 1, 2, \ldots, m$

$$\sup_{\Theta_1} \left| \left(1 + \frac{1}{r}\right)^{2(d-d_o)} - 1 \right| \leq \frac{12}{r},$$

so the bound is of the same order of magnitude as in the (exclusively) stationary case.

When the observed time series is stationary $d_o < \frac{1}{2}$, all Robinson's result apply, even if $\nabla_2 \geq \frac{1}{2}$. The differences arise when we have to consider the periodogram $I_j = I(\lambda_j)$ and $d_o \geq \frac{1}{2}$. When $d_o < \frac{1}{2}$,

we can use expression (3.14) in Robinson's paper,

$$\frac{I_j}{g_j} - 1 = \left(1 - \frac{g_j}{f_j}\right)\frac{I_j}{g_j} + \frac{1}{f_j}\left(I_j - |\alpha_j|^2 I_{\epsilon j}\right) + (2\pi I_{\epsilon j} - 1)$$

where $I_{\epsilon j} = I_\epsilon(\lambda_j)$ is the periodogram of $\{\epsilon_t\}_1^n$, $f_j = f(\lambda_j)$, $\alpha_j = \alpha(\lambda_j)$ and $g_j = G\lambda_j^{-2d_o}$. However, when $\frac{1}{2} \leq d_o < 1$ we have to consider the additional transfer function of the linear filter of first differences before writing down the previous decomposition in terms of the sequence ϵ_t,

$$\frac{I_j}{g_j} - 1 = \left(1 - \frac{g_j}{f_j}\right)\frac{I_j}{g_j} + \frac{1}{f_j}\left(I_j - |1 - e^{i\lambda_j}|^{-2}|\alpha_j|^2 I_{\epsilon j}\right) + (2\pi I_{\epsilon j} - 1).$$

Now, from Theorem 1 (see also Theorem 1 in Hurvich and Ray (1995)), $d_o \geq \frac{1}{2}$, for n sufficiently large,

$$E\left|\frac{I_j}{g_j}\right| \leq C, \qquad j = 1, \ldots, m, \tag{11}$$

for a generic positive finite constant C, in a similar way as when $d_o < \frac{1}{2}$.

Next, paralleling expression (3.17) in Robinson (1995b) for the stationary situation,

$$\begin{aligned}
& E\left|I_j - |1 - e^{i\lambda_j}|^{-2}|\alpha_j|^2 I_{\epsilon j}\right| \\
\leq\ & E[|w_j - (1 - e^{i\lambda_j})^{-1}\alpha_j w_{\epsilon j}||w_j + (1 - e^{i\lambda_j})^{-1}\alpha_j w_{\epsilon j}|] \\
\leq\ & \left(EI_j - (1 - e^{i\lambda_j})^{-1}\alpha_j E w_{\epsilon j}\bar{w}_j - \overline{(1 - e^{i\lambda_j})^{-1}\alpha_j} E\bar{w}_{\epsilon j} w_j + |(1 - e^{i\lambda_j})^{-1}\alpha_j|^2 EI_{\epsilon j}\right)^{1/2} \\
\times\ & \left(EI_j + (1 - e^{i\lambda_j})^{-1}\alpha_j E w_{\epsilon j}\bar{w}_j + \overline{(1 - e^{i\lambda_j})^{-1}\alpha_j} E\bar{w}_{\epsilon j} w_j + |(1 - e^{i\lambda_j})^{-1}\alpha_j|^2 EI_{\epsilon j}\right)^{1/2},
\end{aligned} \tag{12}$$

denoting as $w_{j\epsilon} = w_\epsilon(\lambda_j)$ the Fourier transform of ϵ_t. Then, from the proof of part (a) in Theorem 1 (see Velasco (1997a)) we can obtain, $\frac{1}{2} \leq d_o < 1$,

$$\begin{aligned}
EI_j &= f_j\left(1 + O(j^{2(d_o-1)}\log j)\right), \\
Ew_j\bar{w}_{\epsilon j} &= \frac{(1 - e^{i\lambda_j})^{-1}\alpha_j}{2\pi} + O\left(j^{2(d_o-1)}\lambda_j^{-d}\log j\right) \\
EI_{\epsilon j} &= \frac{1}{2\pi} + O\left(j^{2(d_o-1)}\log j\right)
\end{aligned}$$

uniformly in $j = 1, \ldots, m$. Thus (12) is $O\left(j^{d_o-1}(\log j)^{1/2}\right)$, and following with Robinson's proof, when $d_o \geq \frac{1}{2}$

$$\begin{aligned}
& E\left[\sum_1^{m-1}\left(\frac{r}{m}\right)^{2(\nabla-d_o)+1}\frac{1}{r^2}\left|\sum_1^r \frac{1}{f_j}(I_j - |(1 - e^{i\lambda_j})^{-1}\alpha_j|^2 I_{\epsilon j})\right|\right] \\
\leq\ & C\sum_1^m\left(\frac{r}{m}\right)^{2(\nabla-d_o)+1}\frac{1}{r^2}\sum_1^r\left(j^{d_o-1}(\log j)^{1/2}\right) \\
\leq\ & Cm^{2(d_o-\nabla)-1}\sum_1^m r^{2(\nabla-d_o)-1+d_o}(\log r)^{1/2} \\
=\ & O\left(m^{2(d_o-\nabla)-1} + m^{d_o-1}(\log m)^{3/2}\right) = o(1),
\end{aligned}$$

where the last line follows from the separate consideration of the cases $2(\nabla - d_o) - 1 + d_o < -1$ and $2(\nabla - d_o) - 1 + d_o \geq -1$. Also we can check, using the same techniques, that, as $n \to \infty$, for arbitrarily

small η, $\frac{1}{2} \leq d_o < 1$,

$$\left|\frac{1}{m}\sum_1^m (\frac{I_j}{g_j} - 1)\right| = O_P\left(\eta + \frac{1}{m}\sum_1^m j^{d_o-1}(\log m)^{1/2}\right) + o_P(1) = o_P(1).$$

Using Robinson's definitions the next point that deserves attention when $d_o \geq \frac{1}{2} + \nabla_1$ is

$$\left|\frac{1}{m}\sum_1^m (a_j - 1)(1 - \frac{g_j}{f_j})\frac{I_j}{g_j}\right| = O_P\left(\frac{\eta}{m}\sum_1^m (a_j + 1)\right) = O_P(\eta),$$

with (11).

Observe that after equation (3.22) in Robinson (1995b) we need to choose in fact $\nabla < d_o - \frac{1}{2} + \frac{1}{4e}$ without loss of generality. Due to this modification, we have to proceed in a different way to bound the next expression, $\frac{1}{2} \leq d_o < 1$,

$$\left|\frac{1}{m}\sum_1^m \frac{(a_j - 1)}{f_j}\left(I_j - |(1 - e^{i\lambda_j})^{-1}\alpha_j|^2 I_{\epsilon j}\right)\right| \tag{13}$$

$$= O_P\left(\frac{1}{m}\sum_1^m (a_j + 1) j^{d_o-1}(\log m)^{1/2}\right)$$

$$= O_P\left(\frac{1}{m}\sum_1^m a_j j^{d_o-1}(\log m)^{1/2} + \frac{1}{m}\sum_1^m j^{d_o-1}(\log m)^{1/2}\right). \tag{14}$$

Next, since $p = \exp\left(m^{-1}\sum_1^m \log j\right) \sim m/e$,

$$\sum_1^p a_j j^{d_o-1} = p^{2(d_o-\nabla)}\sum_1^p j^{2(\nabla-d_o)+d_o-1} = O(m^{d_o}),$$

if $2\nabla - d_o > 0$, and $O(m^{2(d_0-\nabla)}\log m)$ if $2\nabla - d_o \leq 0$. Then, using $\sum_p^m a_j = O(m)$ and $\sup_{j>p} j^{d_o-1} = O(p^{d_o-1}) = O(m^{d_o-1})$, we obtain that (14) is

$$O_P\left(m^{-1}[m^{d_o} + m^{2(d_0-\nabla)}](\log m)^{3/2}\right) = o_P(1),$$

with $d_0 < 1$ and $d_0 - \frac{1}{2} < \nabla$, and the proof is completed. •

10 Appendix: Proofs of Section 4

Proof of Theorem 3. Again we retrace the steps in the proof of Theorem 2 in Robinson (1995b). The main step here is to obtain the equivalent to expression (4.7) in that proof bounding in probability the quantity

$$\sum_1^r \left(\frac{I_j}{g_j} - 2\pi I_{\epsilon j}\right)$$

for the general case $d_o \in (-\frac{1}{2}, \frac{3}{4})$. We will see that the bounds for the case $d_o \geq \frac{1}{2}$ are weaker in general than for the stationary case, so these will be the leading terms in the bounds.

First, we need the quantity (cf. equation (4.7) in Robinson (1995b)), for $0 < \delta < \frac{1}{2}$,

$$\sum_{r=1}^m \left(\frac{r}{m}\right)^{1-2\delta}\frac{1}{r^2}\left|\sum_1^r \left(\frac{I_j}{g_j} - 1\right)\right| + \frac{1}{m}\left|\sum_1^m \left(\frac{I_j}{g_j} - 1\right)\right| \tag{15}$$

to be $o_P((\log m)^{-6})$. From Lemma 1, the second term in (15) is, $d_o \geq \frac{1}{2}$,

$$\begin{aligned}
&O_P\left(m^{4(d_o-1)/(5-4d_o)}(\log m)^{2/(5-4d_o)} + m^{\beta}n^{-\beta}\right.\\
&\left.+m^{2(d_o-1)}\log m + n^{-1/2}m^{(d_o-1)/2}(\log n)^{5/4} + n^{-1/4}m^{d_o-1}(\log m)^{1/2}\right)\\
&= o_P((\log m)^{-6}),
\end{aligned}$$

if $d_o < 1$, with (4), and the first one is in order of probability,

$$\begin{aligned}
&m^{2\delta-1}\sum_{r=1}^{m} r^{-1-2\delta}\left(r^{1/(5-4d_o)}(\log r)^{2/(5-4d_o)} + r^{\beta+1}n^{-\beta}\right.\\
&\left.+r^{2d_o-1}\log r + n^{-1/2}r^{(1+d_o)/2}(\log n)^{5/4} + n^{-1/4}r^{d_o}(\log r)^{1/2}\right)\\
= \;& O\left(m^{2\delta-1}\left\{1 + m^{1-2\delta}\left(m^{4(d_o-1)/(5-4d_o)}(\log m)^{2/(5-4d_o)} + m^{\beta}n^{-\beta}\right.\right.\right.\\
&\left.\left.\left.+m^{2(d_o-1)}\log m + n^{-1/2}m^{(d_o-1)/2}(\log n)^{5/4} + n^{-1/4}m^{d_o-1}(\log m)^{1/2}\right)\right\}\right)\\
= \;& o_P\left((\log m)^{-6}\right).
\end{aligned}$$

From Lemma 1, we can see also, $\widehat{F}_k(d) = m^{-1}\sum_1^m (\log j)^k \lambda_j^{2d} I_j$,

$$\begin{aligned}
&\left|\hat{F}_k(d_o) - G_o\frac{1}{m}\sum_1^m (\log j)^k\right|\\
= \;& O_P\left([m^{4(d_o-1)/(5-4d_o)} + m^{2(d_o-1)} + n^{-1/2}m^{(d_o-1)/2}(\log n)^{5/4} + n^{-1/4}m^{d_o-1}](\log m)^2\right)\\
= \;& o_P(1),
\end{aligned}$$

if $\frac{1}{2} \leq d_o < 1$. Next, the error in probability after expression (4.11) in Robinson's proof is now with Lemma 1

$$\begin{aligned}
&O_P\left(\left[m^{(4d_o-3)/(10-8d_o)}(\log m)^{3/2} + m^{\beta+1/2}n^{-\beta}\right.\right.\\
&\left.\left.+m^{2d_o-3/2}\log m + n^{-1/2}m^{d_o/2}(\log n)^{5/4} + n^{-1/4}m^{d_o-1/2}(\log m)^{1/2}\right]\log m\right)\\
= \;& o_P(1)
\end{aligned}$$

if $\frac{1}{2} \leq d_o < \frac{3}{4}$, using (4). This completes the proof using the same central limit theorem. •

11 Appendix: Proofs of Section 6

Proof of Theorem 5. We repeat the steps of the proof of Theorem 1 in Robinson (1995b), with the same definitions and with the notation in terms of $d = H - \frac{1}{2}$, readjusting accordingly the set of admissible values $[\nabla_1, \nabla_2]$.

For $\frac{1}{2} > \delta > 0$ let $N_\delta = \{d : |d - d_o| < \delta\}$ and $\overline{N}_\delta = (-\infty, \infty) - N_\delta$. Then for $S_p(d) = R_p(d) - R_p(d_o)$,

$$\begin{aligned}
P\left(|\widehat{d} - d_o| \geq \delta\right) &= P\left(\widehat{d} \in \overline{N}_\delta \cap \Theta\right)\\
&= P\left(\inf_{\overline{N}_\delta \cap \Theta} R_p(d) \leq \inf_{N_\delta \cap \Theta} R_p(d)\right)\\
&\leq P\left(\inf_{\overline{N}_\delta \cap \Theta} S_p(d) \leq 0\right),
\end{aligned}$$

because $d_0 \in N_\delta \cap \Theta$. As in Robinson's proof, we define $\nabla = \nabla_1$ when $d_o < \frac{1}{2} + \nabla_1$ and $d_o - \frac{1}{2} < \nabla \leq d_o$ otherwise. Then $\Theta_1 = \{d : \nabla \leq d \leq \nabla_2\}$, and $\Theta_2 = \{d : \nabla_1 \leq d < \nabla\}$, possibly empty. It follows that

$$P\left(|\widehat{d} - d_o| \geq \delta\right) \leq P\left(\inf_{\overline{N}_\delta \cap \Theta_1} S_p(d) \leq 0\right) + P\left(\inf_{\Theta_2} S_p(d) \leq 0\right). \tag{16}$$

The sets Θ_1 and Θ_2 are treated separately because of the nonuniform behaviour of $R_p(d)$ around $d = d_o - \frac{1}{2}$. The first probability on the right of (16) is bounded by

$$P\left(\sup_{\Theta_1} |T_p(d)| \geq \inf_{\overline{N}_\delta \cap \Theta_1} U_p(d)\right), \tag{17}$$

where

$$\begin{aligned}
T_p(d) &= \log\left\{\frac{\widehat{G}(d)}{G_o}\right\} - \log\left\{\frac{\widehat{G}(d_o)}{G(d)}\right\} - \log\left\{\frac{2(d-d_o)+1}{m^{2(d-d_0)}}\frac{p}{m}\sum_j^m j^{2(d-d_0)}\right\} \\
&\quad +2(d-d_0)\left\{\frac{p}{m}\sum_j^m \log j - (\log m - 1)\right\} \\
U_p(d) &= 2(d-d_o) - \log\{2(d-d_o)+1\}, \\
G_p(d) &= G_o \frac{p}{m}\sum_j^m \lambda_j^{2(d-d_0)},
\end{aligned}$$

so that $S_p(d) = U_p(d) - T_p(d)$. As in Robinson (1995b),

$$\inf_{\overline{N}_\delta \cap \Theta_1} U_p(d) > \frac{1}{2}\delta^2, \tag{18}$$

and $\sup_{\overline{N}_\delta \cap \Theta_1} |T_p(d)| \to_p 0$ if

$$\sup_{\Theta_1} \left|\frac{\widehat{G}_p(d) - G_p(d)}{G_p(d)}\right| \tag{19}$$

is $o_P(1)$, while

$$\sup_{\Theta_1} \left|\frac{p[2(d-d_o)+1]}{m}\sum_j^m \left(\frac{j}{m}\right)^{2(d-d_o)} - 1\right| \tag{20}$$

and

$$\left|\frac{p}{m}\sum_j^m \log m - (\log m - 1)\right| \tag{21}$$

are both $o(1)$.

From Lemmas 4 and 5 bellow, (20) and (21) are $O(m^{-2(\nabla - d_o)-1}) = o(1)$ and $O(\log m/m) = o(1)$ as $m \to \infty$, respectively. We write

$$\frac{\widehat{G}_p(d) - G_p(d)}{G_p(d)} = \frac{A_p(d)}{B_p(d)},$$

where

$$\begin{aligned}
A_p(d) &= \frac{p[2(d-d_o)+1]}{m}\sum_j^m \left(\frac{j}{m}\right)^{2(d-d_o)}\left(\frac{I_j}{g_j} - 1\right), \\
B_p(d) &= \frac{p[2(d-d_o)+1]}{m}\sum_j^m \left(\frac{j}{m}\right)^{2(d-d_o)}
\end{aligned}$$

for $g_j = G_o\lambda_j^{-2d_0}$. Now

$$\inf_{\Theta_1} B_p(d) \geq 1 - \sup_{\Theta_1}\left|\frac{p[2(d-d_o)+1]}{m}\sum_j^m\left(\frac{j}{m}\right)^{2(d-d_o)} - 1\right| \geq \frac{1}{2}, \tag{22}$$

for all sufficiently large m, by Lemma 4. By summation by parts

$$|A_p(d)| \leq \frac{3p}{m}\left|\sum_r^{m-p}\left\{\left(\frac{r}{m}\right)^{2(d-d_o)} - \left(\frac{r+p}{m}\right)^{2(d-d_o)}\right\}\sum_j^r\left(\frac{I_j}{g_j}-1\right)\right| + \frac{3p}{m}\left|\sum_j^m\left(\frac{I_j}{g_j}-1\right).\right| \tag{23}$$

Because $|(1+1/r)^{2(d-d_o)} - 1| \leq C_{\nabla_2,p}/r$ on Θ_1 when $r > 0$, where $C_{\nabla_2,p}$ is a constant depending on ∇_2 and p, such that

$$C_{\nabla_2,p} \leq (2\nabla_2+1)\left(\frac{p+1}{p}\right)^{2\nabla_2},$$

the first term on the right of (23) has supremum on Θ_1 bounded by

$$3C_{\nabla_2,p}p\sup_{\Theta_1}\sum_r^{m-p}\left(\frac{r}{m}\right)^{2(d-d_o)+1}\frac{1}{r^2}\left|\sum_j^r\left(\frac{I_j}{g_j}-1\right)\right| \leq 3C_{\nabla_2,p}p\sum_r^{m-p}\left(\frac{r}{m}\right)^{2(\nabla-d_o)+1}\frac{1}{r^2}\left|\sum_j^r\left(\frac{I_j}{g_j}-1\right)\right|, \tag{24}$$

the inequality being due to $0 < 2(\nabla - d_o) + 1 \leq 2(d-d_o)+1$ on Θ_1.

Now we have to consider the periodogram $I_j^T = I_p^T(\lambda_j)$ in the decomposition

$$\frac{I_j^T}{g_j} - 1 = \left(1 - \frac{g_j}{f_j}\right)\frac{I_j^T}{g_j} + \frac{1}{f_j}\left(I_j^T - |1-e^{i\lambda_j}|^{-2s}|\alpha_j|^2 I_{\epsilon j}^T\right) + (2\pi I_{\epsilon j}^T - 1). \tag{25}$$

For any $\eta > 0$, Assumptions 7 and (8) imply that n can be chosen such that

$$\left|1 - \frac{g_j}{f_j}\right| \leq \eta, \qquad j = 1,\ldots,m. \tag{26}$$

Now, from the proof of Theorem 4 in Velasco (1997a), for n sufficiently large,

$$E\left|\frac{I_j^T}{g_j}\right| \leq C, \qquad j = 1,\ldots,m, \tag{27}$$

for a generic positive finite constant C, in a similar way as when $d_o \in (-\frac{1}{2},\frac{1}{2})$. Thus

$$E\left[\sum_r^{m-p}\left(\frac{r}{m}\right)^{2(\nabla-d_o)+1}\frac{1}{r^2}\left|\sum_j^r\left(1-\frac{g_j}{f_j}\right)\frac{I_j^T}{g_j}\right|\right] \leq \frac{C\eta}{2(\nabla-d_o)+1}.$$

Next, Next, generalizing expression (12),

$$\begin{aligned}
& E\left|I_j^T - |1-e^{i\lambda_j}|^{-2s}|\alpha_j|^2 I_{\epsilon j}^T\right| \\
\leq\ & E[|w_j^T - (1-e^{i\lambda_j})^{-s}\alpha_j w_{\epsilon j}^T||w_j^T + (1-e^{i\lambda_j})^{-s}\alpha_j w_{\epsilon j}^T|] \\
\leq\ & \left(EI_j^T - (1-e^{i\lambda_j})^{-s}\alpha_j E w_{\epsilon j}^T\bar{w}_j^T - \overline{(1-e^{i\lambda_j})^{-s}\alpha_j}E\bar{w}_{\epsilon j}^T w_j^T + |(1-e^{i\lambda_j})^{-s}\alpha_j|^2 EI_{\epsilon j}^T\right)^{1/2} \\
& \times\left(EI_j^T + (1-e^{i\lambda_j})^{-s}\alpha_j E w_{\epsilon j}^T\bar{w}_j^T + \overline{(1-e^{i\lambda_j})^{-s}\alpha_j}E\bar{w}_{\epsilon j}^T w_j^T + |(1-e^{i\lambda_j})^{-s}\alpha_j|^2 EI_{\epsilon j}^T\right)^{1/2},
\end{aligned} \tag{28}$$

denoting as $w_{j\epsilon}^T = w_\epsilon^T(\lambda_j)$ the (tapered) Fourier transform of ϵ_t. Then, from the proof of part (a) in Theorem 4 (see Velasco (1997a)) we can obtain that, as $n \to \infty$,

$$\begin{aligned} EI_j^T &= f_j\left(1 + O(j^{-1} + j^{2(d_o-p)}\log j)\right), \\ Ew_j^T\bar{w}_{\epsilon j}^T &= \frac{(1-e^{i\lambda_j})^{-s}\alpha_j}{2\pi} + O\left(j^{-1}\lambda_j^{-d_o} + j^{2(d_o-p)}\lambda_j^{-d_o}\log j\right) \\ EI_{\epsilon j}^T &= \frac{1}{2\pi} + O\left(j^{-1} + j^{2(d_o-p)}\log j\right) \end{aligned}$$

uniformly in $j = p, 2p, \ldots, m$. Thus (28) is $O\left(f_j\left\{j^{-1/2} + j^{d_o-p}(\log[j+1])^{1/2}\right\}\right)$, and following with Robinson's proof,

$$\begin{aligned} & E\left[\sum_1^{m-p}\left(\frac{r}{m}\right)^{2(\nabla-d_o)+1}\frac{1}{r^2}\left|\sum_1^r \frac{1}{f_j}(I_j - |(1-e^{i\lambda_j})^{-s}\alpha_j|^2 I_{\epsilon j})\right|\right] \\ & \leq C\sum_1^m\left(\frac{r}{m}\right)^{2(\nabla-d_o)+1}\frac{1}{r^2}\sum_1^r\left(j^{-1/2} + j^{d_o-p}(\log j)^{1/2}\right) \\ & \leq Cm^{2(d_o-\nabla)-1}\sum_1^m\left[r^{2(\nabla-d_o)-1/2} + r^{2(\nabla-d_o)-p+d_o}(\log j)^{1/2}\right] \\ & = O\left(m^{2(d_o-\nabla)-1} + m^{-1/2}\log m + m^{d_o-p}(\log m)^{3/2}\right) = o(1), \end{aligned}$$

with $d_o < p$, where the last line follows from the separate consideration of the cases $2(\nabla - d_o) - \frac{1}{2} < -1$ and $2(\nabla - d_o) - \frac{1}{2} \geq -1$ and $2(\nabla - d_o) - p + d_o < -1$ and $2(\nabla - d_o) - p + d_o \geq -1$.

To deal with the final contribution to (25) we need to consider the variance of $\sum_j^r(2\pi I_{\epsilon j}^T - 1)$, since it has zero mean. The variance and covariances of $I_{\epsilon j}^T$ have two components. The first is due to the fourth cumulant,

$$\left(\sum_1^n h_t^2\right)^{-2}\int_{[-\pi,\pi]^3} D_p^T(\omega_1+\lambda_j)D_p^T(\omega_2-\lambda_k)D_p^T(\omega_3-\lambda_j)D_p^T(\lambda_k - \sum_{1,2,3}\omega_i)f_\epsilon^{(4)}(\omega_1,\omega_2,\omega_3)d\omega,$$

which is of order n^{-1}, given the boundedness of $f_\epsilon^{(4)}$ and the properties of D_p^T, $\int_{-\pi}^{\pi}|D_p^T(\lambda)|d\lambda = O(1)$ and $\sup_\lambda|D_p^T(\lambda)| = O(n)$, all n and $p > 1$. The second component is due to the second moments. The variance of $I_{\epsilon j}^T$ is then of order $O(1)$ and for the covariance between $I_{\epsilon j}^T$ and $I_{\epsilon k}^T$, $k \neq j$, beside the $O(n^{-1})$ fourth cumulant term, we have to consider the following convolutions

$$\left(\sum_1^n h_t^2\right)^{-2}\int_{-\pi}^{\pi} D_p^T(\lambda+\lambda_j)D_p^T(\lambda-\lambda_k)d\lambda\int_{-\pi}^{\pi} D_p^T(\lambda-\lambda_j)D_p^T(\lambda+\lambda_k)d\lambda,$$

and

$$\left(\sum_1^n h_t^2\right)^{-2}\int_{-\pi}^{\pi} D_p^T(\lambda+\lambda_j)D_p^T(\lambda+\lambda_k)d\lambda\int_{-\pi}^{\pi} D_p^T(\lambda-\lambda_j)D_p^T(\lambda-\lambda_k)d\lambda,$$

since $f_\epsilon(\lambda)$ is constant. These terms, from Lemmas 1 and 2 of Velasco (1997a) are of order $O(|j-k|^{-2p})$ and $O(|j+k|^{-2p})$, respectively, if $j, k > 0$. Thus

$$\begin{aligned} \mathrm{Var}\left[\sum_j^r(2\pi I_{\epsilon j}^T - 1)\right] &= \sum_j^r \mathrm{Var}\left[2\pi I_{\epsilon j}^T\right] + \sum_{j\neq}^r\sum_k^r \mathrm{Cov}\left[2\pi I_{\epsilon j}^T, 2\pi I_{\epsilon k}^T\right] \\ &= O(r) + O\left(\sum_{j\neq}^r\sum_k^r\{|j-k|^{-2p} + |j+k|^{-2p} + n^{-1}\}\right) = O(r). \end{aligned}$$

We have obtained that $\sum_j^r (2\pi I_{\epsilon j}^T - 1) = O_P(r^{1/2})$ so that

$$\begin{aligned}\sum_1^m \left(\frac{r}{m}\right)^{2(\nabla - d_o)+1} \frac{1}{r^2} \left|\sum_1^r (2\pi I_{\epsilon j} - 1)\right| &= O_P\left(\sum_1^m \left(\frac{r}{m}\right)^{2(\nabla - d_o)+1} r^{-3/2}\right) \\ &= O_P\left(m^{2(d_o - \nabla)-1} \sum_1^m r^{2(\nabla - d_o)-1/2}\right) \\ &= O_P\left(m^{-1/2} + m^{2(d_o-\nabla)-1} \log m\right) = o_P(1)\end{aligned}$$

as $n \to \infty$ because $2(d_o - \nabla) < 1$.

Also we can check, using the same techniques, that, as $n \to \infty$, for arbitrarily small η, since $d_o < p$,

$$\left|\frac{1}{m}\sum_1^m (\frac{I_j}{g_j} - 1)\right| = O_P\left(\eta + \frac{1}{m}\sum_1^m j^{d_o - p}(\log m)^{1/2}\right) + o_P(1) = o_P(1).$$

Thus as $n \to \infty$, $\sup_{\Theta_1} |A_p(d)| \to_P 0$ and, with (19) and (22), $\sup_{\Theta_1} |\widehat{G}_p(d)/G_p(d) - 1| \to_P 0$. In view of (18) it follows that (17) $\to 0$ as $n \to \infty$.

When $d_o \geq \frac{1}{2} + \nabla_1$ we have to consider the second probability on the right of (16). Set $q = q_m = \exp(pnn^{-1} \sum_j^m \log j)$ and $S_p(d) = \log\left[\widehat{D}_p(d)/\widehat{D}_p(d_o)\right]$, where

$$\widehat{D}_p(d) = \frac{p}{m}\sum_j^m \left(\frac{j}{q}\right)^{2(d-d_o)} j^{2d_o} I_j^T.$$

Because $1 \leq q \leq m$ and $\inf_{\Theta_2}(j/q)^{2(d-d_o)} \geq (j/q)^{2(\nabla - d_o)}$ for $1 \leq j \leq q$, while $\inf_{\Theta_2}(j/q)^{2(d-d_o)} \geq (j/q)^{2(\nabla_1 - d_o)}$ for $q < j \leq m$, it follows that

$$\inf_{\Theta_2} \widehat{D}_p(d) \geq \frac{p}{m}\sum_j^m a_j j^{2d_o} I_j^T,$$

where

$$a_j = \begin{cases} \left(\frac{j}{q}\right)^{2(\nabla - d_0)}, & 1 \leq j \leq q \\ \left(\frac{j}{q}\right)^{2(\nabla_1 - d_0)}, & q < j \leq m. \end{cases}$$

Thus

$$P\left\{\inf_{\Theta_2} S_p(d) \leq 0\right\} \leq P\left\{\frac{p}{m}\sum_j^m (a_j - 1) j^{2d_0} I_j \leq 0\right\}.$$

As $m \to \infty$, $q \sim \exp(\log m - 1) = m/e$ and

$$\sum_{1 \leq j \leq q} a_j \sim p^{-1} \cdot q^{2(d_o - \nabla)} \int_o^q x^{2(\nabla - d_o)} dx = \frac{q/p}{2(\nabla - d_o)+1} \sim \frac{(m/e)/p}{2(\nabla - d_o)+1}. \tag{29}$$

It follows that

$$\frac{p}{m}\sum_j^m (a_j - 1) \geq \frac{p}{m}\sum_{1 \leq j \leq q} a_j - 1 \sim \frac{1}{e[2(\nabla - d_o)+1]} - 1 \quad \text{as} \quad m \to 0.$$

Choose $\nabla < d_o - \frac{1}{2} + \frac{1}{4e}$, which we may do with no loss of generality. Then for all sufficiently large m, $(p/m)\sum_j^m (a_j - 1) \geq 1$ and thus (16) is bounded by

$$P\left\{\left|\frac{p}{m}\sum_j^m (a_j - 1)(\frac{I_j}{g_j} - 1)\right| \geq 1\right\}.$$

Now apply (25) again and first note from (26) and (27) that

$$\left|\frac{p}{m}\sum_{j}^{m}(a_j-1)(1-\frac{g_j}{f_j})\frac{I_j}{g_j}\right| = O_P\left(\frac{\eta}{m}\sum_{j}^{m}(a_j+1)\right) = O_P(\eta),$$

with (27) and

$$\sum_{q<j\leq m} a_j \sim p^{-1}\cdot q^{2(d_o-\nabla_1)}\int_q^m x^{2(\nabla_1-d_o)}dx = O(m)$$

and

$$\sum_{j}^{m} a_j^2 = O\left(m^{4(d_o-\nabla)}+m\log m\right).$$

Observe that after equation (29) in Robinson (1995b) we need to choose in fact $\nabla < d_o - \frac{1}{2}+\frac{1}{4e}$, and not $\nabla < d_o - \frac{1}{2}+\frac{e}{4}$, without loss of generality. Due to this modification, we have to proceed in a different way to bound the next expression,

$$\left|\frac{p}{m}\sum_{1}^{m}\frac{(a_j-1)}{f_j}\left(I_j-|(1-e^{i\lambda_j})^{-s}\alpha_j|^2 I_{\epsilon j}\right)\right| \tag{30}$$

$$= O_P\left(\frac{1}{m}\sum_{1}^{m}(a_j+1)\left[j^{-1/2}+j^{d_o-p}(\log m)^{1/2}\right]\right)$$

$$= O_P\left(\frac{1}{m}\sum_{1}^{m}a_j[j^{-1/2}+j^{d_o-p}](\log m)^{1/2}+m^{-1/2}+\frac{1}{m}\sum_{1}^{m}j^{d_o-p}(\log m)^{1/2}\right), \tag{31}$$

Next, since $q\sim m/(ep)$,

$$\sum_{1}^{q}a_j j^{d_o-p} = q^{2(d_o-\nabla)}\sum_{1}^{q}j^{2(\nabla-d_o)+d_o-p} = O(m^{d_o-p+1}),$$

if $2(\nabla-d_o)+d_o-p>0$, and $O(m^{2(d_0-\nabla)}\log m)$ if $2(\nabla-d_o)+d_o-p\leq 0$. Also

$$\sum_{1}^{q}a_j j^{-1/2} = q^{2(d_o-\nabla)}\sum_{1}^{q}j^{2(\nabla-d_o)-1/2} = O(m^{1/2}),$$

if $2(\nabla-d_o)-1/2>0$, and $O(m^{2(d_0-\nabla)}\log m)$ if $2(\nabla-d_o)-1/2\leq 0$.

Then, using $\sum_q^m a_j = O(m)$ and $\sup_{j>q} j^{d_o-p} = O(q^{d_o-p}) = O(m^{d_o-p})$, we obtain that (31) is

$$O_P\left(m^{-1/2}+m^{-1}[m^{d_o-p+1}+m^{1/2}+m^{2(d_0-\nabla)}](\log m)^{3/2}\right) = o_P(1),$$

with $d_0<p$ and $d_0-\frac{1}{2}<\nabla$.

Finally, using Theorem 4 and proceeding as before,

$$\begin{aligned}
&\mathrm{Var}\left[\frac{p}{m}\sum_{j}^{m}(a_j-1)(2\pi I_{\epsilon j}^T-1)\right]\\
&= \frac{p^2}{m^2}\sum_{j}^{m}(a_j-1)^2\mathrm{Var}\left[2\pi I_{\epsilon j}^T\right]+\frac{p^2}{m^2}\sum_{j\neq}^{m}\sum_{k}^{m}(a_j-1)(a_k-1)\mathrm{Cov}\left[2\pi I_{\epsilon j}^T, 2\pi I_{\epsilon k}^T\right]\\
&= O\left(m^{-2}\sum_{j}^{m}(a_j-1)^2\sum_{k\neq j}^{m}(|j-k|^{-2p}+|j+k|^{-2p})\right)\\
&= O\left(m^{-2}\left[m+\sum_{j}^{m}a_j^2\right]\right) = O\left(m^{-2}\left[m\log m+m^{4(d_o-\nabla)}\right]\right)\\
&= O\left(m^{-1}\log m+m^{2[2(d_o-\nabla)-1]}\right) = o(1),
\end{aligned}$$

and the proof is completed. •

Proof of Theorem 6. We can adapt all the steps in the proof of Theorem 2 in Robinson (1995b) to the situation for $p > 1$ as we have done in the proof of Theorem 5. This accounts basically to the redefinition of the sums to frequencies $\lambda_p, \lambda_{2p}, \ldots, \lambda_m$ only.

The main step here is to bound in probability the quantity (cf. equation (4.7) in Robinson (1995b)), for $0 < \delta < \frac{1}{2}$,

$$A\sum_{r=1}^{m}\left(\frac{r}{m}\right)^{1-2\delta}\frac{1}{r^2}\left|\sum_{1}^{r}\left(\frac{I_j}{g_j}-1\right)\right| + B\frac{1}{m}\left|\sum_{1}^{m}\left(\frac{I_j}{g_j}-1\right)\right| \tag{32}$$

to be $o_P((\log m)^{-6})$, where A and B are two finite constants depending on p and ∇_2 (see equation (24) above).

Now, using the same procedure as in the proof of Theorem 5 (cf. Robinson's equation (3.17) and the following lines), using $\beta > 1$, with $r \leq m$,

$$\sum_{p}^{r}\left(\frac{I_j}{g_j} - 2\pi I_{\epsilon j}\right) = O_P\left(r^{1-\beta/2} + \log r + r^{d_o-p+1}(\log r)^{1/2} + r^{\beta+1}n^{-\beta}\right) \tag{33}$$

where the term $\log r$ shows up when $\beta = 2$ (or when $d_o - p = -1$), and the term $r^{\beta+1}n^{-\beta}$ is exactly the same as in his expression (4.8), see also the equation after (4.25). Note that in this case we have followed a much direct approach than Robinson's (1995b) proof, using a stronger assumption on the smoothness of the function f, namely $\beta > 1$. This is in part for convenience and in part because the correlation between adjacent tapered periodogram ordinates invalidates the approach using second moments of the periodogram as in Robinson's page 1648, and used above when $p = 1$ and $d_o < \frac{3}{4}$. A similar approach was used in Velasco (1997a) to analyze the log-periodogram ordinate for non-Gaussian stationary observations. Note that this procedure is only valid if we use a tapered periodogram with $p \geq 2$ but not otherwise: we use the lower bias of tapering, avoiding the increment of correlation.

Now, the bound in probability at the end of page 1643 in the reference is now, using (32) and (33) as $n \to \infty$

$$O_P\left(\left[m^{-1/2}\log m + m^{d_o-p}(\log m)^{3/2} + m^{\beta}n^{-\beta}\right](\log m)^2\right) = o_P(1).$$

From there we can reach the same limit as in expression (4.10) and the equivalent to expression (4.11) in Robinson's paper is now

$$\left\{2(m/p)^{-1/2}\sum_{p}^{m}\nu_j(2\pi I_{\epsilon j} - 1)\right.$$
$$\left. + O_P\left(\left[m^{(1-\beta)/2}(\log m)^2 + m^{d_o-p+1/2}(\log m)^2 + m^{\beta+1/2}n^{-\beta}\right]\log m\right)\right\}(1 + o_P(1)),$$

where $\nu_j = \log j - (p/m)\sum_p^m \log j$ satisfies $\sum_p^m \nu_j = 0$, which is, from the assumptions of the Theorem,

$$\left\{2(m/p)^{-1/2}\sum_{p}^{m}\nu_j(2\pi I_{\epsilon j} - 1) + o_P(1)\right\}(1 + o_P(1)).$$

Using Lemma 6 we can obtain the asymptotic distribution of $(m/p)^{-1/2}\sum_p^m \nu_j(2\pi I_{\epsilon j} - 1)$ and the Theorem is proved. •

12 Appendix: Technical Lemmas

Lemma 1 *Under the Assumptions of Theorem 3, $d_o \in [\frac{1}{2}, 1)$,*

$$\begin{aligned}
&\sum_{1}^{r}\left(\frac{I_j}{g_j} - 2\pi I_{\epsilon j}\right) \\
&= O_P\Big(r^{1/(5-4d_o)}(\log r)^{2/(5-4d_o)} + r^{\beta+1}n^{-\beta} \\
&\quad + r^{2d_o-1}\log r + n^{-1/2}r^{(1+d_o)/2}(\log n)^{5/4} + n^{-1/4}r^{d_o}(\log r)^{1/2}\Big)
\end{aligned}$$

Proof. We only consider the case $d_o \geq \frac{1}{2}$, since the stationary situation follows as in the proof of Theorem 2 in Robinson (1995b), with stronger results. Choosing an integer $1 < \ell < r$, for $d_o \in [\frac{1}{2}, 1)$ from (11) and $E[2\pi I_{\epsilon j}] = 1$,

$$E\left|\sum_{1}^{\ell}\left(\frac{I_j}{g_j} - 2\pi I_{\epsilon j}\right)\right| = O(\ell),$$

and also from (11) and Assumption 2,

$$E\left|\sum_{\ell+1}^{r}\left(\frac{I_j}{g_j} - \frac{I_j}{f_j}\right)\right| \leq C\sum_{\ell+1}^{r}\left|1 - \frac{g_j}{f_j}\right| = O\left(\frac{r^{\beta+1}}{n^{\beta}}\right).$$

Next, we consider

$$E\left[\left\{\sum_{\ell+1}^{r}\left(\frac{I_j}{f_j} - 2\pi I_{\epsilon j}\right)\right\}^2\right] = (2\pi)^2(a+b),$$

with the same definitions as in p. 1648 of Robinson (1995b). Further if we split the terms $a = a_1 + a_2$ and $b = b_1 + b_2$ corresponding to second and fourth cumulants, we find that when $d_o \in [\frac{1}{2}, 1)$, with Theorem 1

$$a_1 = O\left(\sum_{\ell+1}^{r} j^{2(d_o-1)}\log j\right) = O(r^{2d_o-1}(\log r)^2)$$

and

$$\begin{aligned}
b_1 &= O\left(\sum_{j=\ell+1}^{r}\sum_{k>j}^{r}\left\{(jk)^{2(d_o-1)}(\log k)^2 + (j^{2(d_o-1)}\log k)^2\right\}\right) \\
&= O\left((\log r)^2\sum_{k=\ell+2}^{r}\sum_{j=\ell+1}^{k-1} j^{4(d_o-1)}\right) \\
&= O(r\,\ell^{4d_o-3}(\log r)^2),
\end{aligned}$$

since we will only use $r = O(n)$ at most. Choosing $\ell \sim r^{1/(5-4d_o)}(\log r)^{2/(5-4d_o)}$ this gives the first term of this order in the lemma, since $(a_1)^{1/2}$ is of smaller order of magnitude

When $d_o \geq \frac{1}{2}$ we obtain the same expressions for a_2 and b_2 as in Robinson (1995b), and substituting $\alpha(\lambda)$ by $(1-e^{i\lambda})^{-1}\alpha(\lambda)$ and α_j by $(1-e^{i\lambda_j})^{-1}\alpha_j$, and defining here

$$P_j = \int_{-\pi}^{\pi}\left|\frac{\alpha(\lambda)(1-e^{i\lambda_j})}{(1-e^{i\lambda})\alpha_j} - 1\right|^2 K(\lambda-\lambda_j)d\lambda$$

where $K(\lambda) = (2\pi n)^{-1}(\sin n\lambda/2)\sin\lambda/2$ is Fejér kernel, the same bounds hold here. However the bound for the second type of summand considered by Robinson in b_2, $O(P_jP_k^{1/2})$, is improved in Lemma 2 to

$O(n^{-1}P_k^{1/2}(\log n)^2)$. This allows the consideration of values of the parameter $d_o < \frac{3}{4}$, which otherwise would be restricted to $d_o < \frac{2}{3}$. For a_2 we can still use the bound given by Robinson.

Then applying Lemma 3, with Lemma 2,

$$\begin{aligned}
a_2 &= O\left(\sum_{j=1}^{r}\left\{\frac{(\log j)^2}{j^{4(1-d_o)}}+\frac{(\log j)^{3/2}}{j^{3(1-d_o)}}+\frac{n^{-1/2}\log j}{j^{2(1-d_o)}}\right\}\right)\\
&= O(r^{3d_o-2}(\log r)^{3/2}+n^{-1/2}r^{2d_o-1}\log j+(\log r)^3)\\
b_2 &= O\left(\sum_{j=1}^{r}\sum_{k>j}^{r}\left\{\frac{(\log r)^2}{(jk)^{2(1-d_o)}}+\frac{(\log k)^{1/2}(\log n)^2}{nk^{1-d_o}}+\frac{n^{-1/2}\log r}{(jk)^{1-d_o}}\right\}\right)\\
&= O\Big(r^{2(2d_o-1)}(\log r)^2+n^{-1}r^{1+d_o}(\log n)^{5/2}+n^{-1/2}r^{2d_o}\log r+(\log r)^4\Big)
\end{aligned}$$

and the lemma follows. •

Lemma 2 *Under Assumptions 1 and 5, $d_o \in [\frac{1}{2},1)$,*

$$\begin{aligned}
&\frac{1}{(2\pi n)^2}\int_{\Pi^3}\left\{\frac{\alpha(\lambda+\mu+\zeta)(1-e^{i\lambda_j})}{(1-e^{i(\lambda+\mu+\zeta)})\alpha_j}-1\right\}\left\{\frac{\alpha(-\mu)(1-e^{-i\lambda_j})}{(1-e^{-i\mu})\bar{\alpha}_j}-1\right\}\left\{\frac{\alpha(-\zeta)(1-e^{-i\lambda_k})}{(1-e^{-i\zeta})\bar{\alpha}_k}-1\right\}\\
&\quad\times E_{jk}(\lambda,\mu,\zeta)d\lambda d\mu d\zeta \qquad (34)\\
&= O\Big(n^{-1}k^{d_o-1}(\log n)^2(\log k)^{1/2}\Big),
\end{aligned}$$

where

$$E_{jk}(\lambda,\mu,\zeta)=D(\lambda_j-\lambda-\nu-\zeta)D(\lambda_k+\lambda)D(\mu-\lambda_j)D(\zeta-\lambda_k),$$

and $D(\lambda)=\sum_t e^{i\lambda t}$ is Dirichlet kernel.

Proof. Making a change of variable and using the periodicity of D, (34) is

$$\begin{aligned}
&\frac{1}{(2\pi n)^2}\int_{\Pi^3}\left\{\frac{\alpha(\omega)(1-e^{i\lambda_j})}{(1-e^{i\omega})\alpha_j}-1\right\}\left\{\frac{\alpha(-\mu)(1-e^{-i\lambda_j})}{(1-e^{-i\mu})\bar{\alpha}_j}-1\right\}\left\{\frac{\alpha(-\zeta)(1-e^{-i\lambda_k})}{(1-e^{-i\zeta})\bar{\alpha}_k}-1\right\}\\
&\times D(\lambda_j-\omega)D(\lambda_k+\omega-\mu-\zeta)D(\mu-\lambda_j)D(\zeta-\lambda_k)d\omega d\mu d\zeta,
\end{aligned}$$

and this is less in absolute value than

$$\frac{1}{2\pi n}P_k^{1/2}\int_{-\pi}^{\pi}\left|\frac{\alpha(\omega)(1-e^{i\lambda_j})}{(1-e^{i\omega})\alpha_j}-1\right||D(\lambda_j-\omega)|d\omega\int_{-\pi}^{\pi}\left|\frac{\alpha(-\mu)(1-e^{-i\lambda_j})}{(1-e^{-i\mu})\bar{\alpha}_j}-1\right||D(\mu-\lambda_k)|d\mu.$$

Now using the bound for P_k in Lemma 3 and

$$\int_{-\pi}^{\pi}\left|\frac{\alpha(\omega)(1-e^{i\lambda_j})}{(1-e^{i\omega})\alpha_j}-1\right|D(\lambda_j-\omega)d\omega=O(\log n)\,, \qquad (35)$$

the lemma follows. To prove (35) we consider now

$$\int_{-\pi}^{\pi}\left|\frac{\alpha(\omega)(1-e^{i\lambda_j})}{(1-e^{i\omega})\alpha_j}-1\right||D(\lambda_j-\omega)|d\omega\le\left|\frac{\alpha(\lambda_j)}{1-e^{i\lambda_j}}\right|^{-1}\int_{-\pi}^{\pi}\left|\frac{\alpha(\lambda_j-\omega)}{1-e^{i(\lambda_j-\omega)}}-\frac{\alpha(\lambda_j)}{1-e^{i\lambda_j}}\right||D(\omega)|d\omega$$

and the following intervals of integration,

$$\begin{aligned}
\left|\int_{-\lambda_j/2}^{\lambda_j/2}\right| &\le \left|\frac{\alpha(\lambda_j)}{1-e^{i\lambda_j}}\right|^{-1}\sup_{-\lambda_j/2\le\omega\le\lambda_j/2}\left|\frac{d}{d\omega}\frac{\alpha(\lambda_j-\omega)}{1-e^{i(\lambda_j-\omega)}}\right|\int_{-\lambda_j/2}^{\lambda_j/2}|\omega||D(\omega)|d\omega\\
&= O\Big(\lambda_j^d\lambda_j^{-d_o-1}\lambda_j\Big)=O(1).
\end{aligned}$$

Next

$$\begin{aligned}
\left|\int_{\lambda_j/2}^{3\lambda_j/2}\right| &\leq \left|\frac{\alpha(\lambda_j)}{1-e^{i\lambda_j}}\right|^{-1}\int_{\lambda_j/2}^{3\lambda_j/2}\left\{\left|\frac{\alpha(\lambda_j-\omega)}{1-e^{i(\lambda_j-\omega)}}\right|-\left|\frac{\alpha(\lambda_j)}{1-e^{i\lambda_j}}\right|\right\}|D(\omega)|d\omega \\
&= O\left(\sup_{\lambda_j/2\leq\omega\leq 3\lambda_j/2}|D(\omega)|\left[\lambda_j^d\int_{\lambda_j/2}^{3\lambda_j/2}\left|\frac{\alpha(\lambda_j-\omega)}{1-e^{i(\lambda_j-\omega)}}\right|d\omega+\int_{\lambda_j/2}^{3\lambda_j/2}d\omega\right]\right) \\
&= O\left(\lambda_j^{-1}\left[\lambda_j^{d_o}\int_0^{\lambda_j}\omega^{-d_o}d\omega+\lambda_j\right]\right)=O(1),
\end{aligned}$$

since $d_o<1$ (note that $|\alpha(\lambda)(1-e^{i\lambda})^{-1}|=\{2\pi f(\lambda)\}^{1/2}$ is integrable because $d_o<1$.) Then, choosing $\epsilon>0$, fixed, as small as we want, such that Assumption 1 holds for $|\lambda|<\epsilon$, as in the proof of Theorem 2 of Robinson (1995a),

$$\left|\int_{-\epsilon}^{-\lambda_j/2}\right|\leq\left|\frac{\alpha(\lambda_j)}{1-e^{i\lambda_j}}\right|^{-1}\sup_{-\epsilon\leq\omega\leq-\lambda_j/2}\left\{\left|\frac{\alpha(\lambda_j-\omega)}{1-e^{i(\lambda_j-\omega)}}\right|-\left|\frac{\alpha(\lambda_j)}{1-e^{i\lambda_j}}\right|\right\}\int_{-\pi}^{\pi}|D(\omega)|d\omega=O(\log n),$$

$$\left|\int_{3\lambda_j/2}^{\epsilon}\right|\leq\left|\frac{\alpha(\lambda_j)}{1-e^{i\lambda_j}}\right|^{-1}\sup_{3\lambda_j/2\leq\omega\leq\epsilon}\left\{\left|\frac{\alpha(\lambda_j-\omega)}{1-e^{i(\lambda_j-\omega)}}\right|-\left|\frac{\alpha(\lambda_j)}{1-e^{i\lambda_j}}\right|\right\}\int_{-\pi}^{\pi}|D(\omega)|d\omega=O(\log n),$$

and the same bound holds for the remaining intervals of integration. •

Lemma 3 *Under Assumptions 1 and 5, $d_o\in[\frac{1}{2},1)$*

$$P_j=\int_{-\pi}^{\pi}\left|\frac{\alpha(\lambda)(1-e^{i\lambda_j})}{(1-e^{i\lambda})\alpha_j}-1\right|^2K(\lambda-\lambda_j)d\lambda=O(j^{2(d_o-1)}\log j)$$

Proof. This is Lemma 3 of Robinson (1995b) generalized to cover the non-stationary situation $d_o\in[\frac{1}{2},1)$ and follows considering the same intervals of integration, where for the interval $[-\lambda_j/2,\lambda_j/2]$ we can adapt the proofs of Theorem 1 or Theorem 6 ($p=1$) in Velasco (1997a), since $f(\lambda)$ is not integrable at the origin, to obtain a $O(j^{2(d_o-1)}\log j)$ contribution. •

Lemma 4 *For $p=1,2,\ldots,\varepsilon\in(0,1]$ and $C\in(\varepsilon,\infty)$, as $m\to\infty$,*

$$\sup_{\varepsilon\leq\gamma\leq C}\left|\frac{\gamma\cdot p}{m}\sum_{j=p,2p,\ldots}^{m}\left(\frac{j}{m}\right)^{\gamma-1}-1\right|=O\left(\frac{1}{m^{\varepsilon}}\right).$$

Proof. As in Lemma 1 of Robinson (1995b), $\int_0^a x^{\gamma-1}dx=a^{\gamma}/\gamma$ for $\gamma>0$,

$$\begin{aligned}
\left|\frac{\gamma\cdot p}{m}\sum_{j=p,2p,\ldots}^{m}\left(\frac{j}{m}\right)^{\gamma-1}-1\right| &\leq \gamma\int_0^{p/m}\left\{\left(\frac{p}{m}\right)^{\gamma-1}-x^{\gamma-1}\right\}dx \\
&\quad+\gamma\sum_{j=2p,3p,\ldots}^{m}\int_{(j-p)/m}^{j/m}\left\{\left(\frac{j}{m}\right)^{\gamma-1}-x^{\gamma-1}\right\}dx \\
&\leq \frac{\gamma}{(m/p)^{\gamma}}+\frac{1}{(m/p)^{\gamma}}+\frac{\gamma|\gamma-1|}{(m/p)^2}\sum_{j=p}^{m}\left(\frac{j}{m}\right)^{\gamma-2},
\end{aligned}$$

by the mean-value theorem. The last term is $O(\gamma^2m^{-1})$ for $\gamma>1$, zero for $\gamma=1$ and $O(m^{-\gamma})$ for $0<\gamma<1$. •

Lemma 5 *For all $m \geq 2p$, $p = 2, 3, \ldots$,*

$$\left| \frac{p}{m} \sum_{j=p,2p,\ldots}^{m} \log j - \log m + 1 \right| \leq \frac{2p \log p - p + 1 + \log(m-p)}{m}.$$

Proof. Because $\int_0^m \log x dx = m(\log m - 1)$, the left-hand side of (5.1) is

$$\begin{aligned}
&\left| \frac{p \log p}{m} + \frac{1}{m} \int_0^p \log x dx - \frac{1}{m} \sum_{j=2p,3p,\ldots}^{m} \int_{j-p}^{j} \log\left(\frac{j}{x}\right) dx \right| \\
&\leq \frac{p(2\log p - 1)}{m} + \frac{1}{m} \sum_{j=p,2p,\ldots}^{m-p} \frac{1}{j} \\
&\leq \frac{p(2\log p - 1)}{m} + \frac{1 + \log(m-p)}{m} \\
&= \frac{2p \log p - p + 1 + \log(m-p)}{m}. \qquad \bullet
\end{aligned}$$

Lemma 6 *If the sequence $\{h_j\}$ is a data taper of order p as defined previously, and the random variables $\{\epsilon_j\}$ satisfy Assumption 6, with $\nu_j = \log j - (p/m) \sum_{j=p}^{m} \log j$,*

$$Z_n = (m/p)^{-1/2} \sum_{p}^{m} \nu_j (2\pi I_\epsilon^T(\lambda_j) - 1) \to_D N(0, \Phi)$$

where Φ is given in (10).

Proof. We will follow Robinson (1995b, pp. 1644-1647), adapting his non-tapered proof to the tapered case. We have that $Z_n = 2\sum_{t=1}^{n} z_t$ and

$$\begin{aligned}
z_t &= h_t \epsilon_t \sum_{s=1}^{t-1} h_s \epsilon_s c_{t-s}, \\
c_s &= 2\left(\sum_{r=1}^{n} h_r^2\right)^{-1} \left(\frac{m}{p}\right)^{-1/2} \nu_j \cos s\lambda_j,
\end{aligned}$$

remembering that $\sum_{r=1}^{n} h_r^2 \sim b \cdot n$. Now the z_t form a zero-mean martingale difference array, and from a standard CLT we can deduce that $\sum z_t$ tends to a $N(0, \Phi)$ random variable in distribution if

$$\sum_{t=1}^{n} E[z_t^2 | F_{t-1}] - \Phi \quad \to_p \quad 0, \tag{36}$$

$$\sum_{t=1}^{n} E[z_t^2 I(|z_t| > \rho)] \quad \to \quad 0 \quad \text{for all} \quad \rho > 0. \tag{37}$$

Now the left hand side of (36) is

$$\left\{ \sum_{t=1}^{n} h_t^2 \sum_{s=1}^{t-1} h_s^2 \epsilon_s^2 c_{t-s}^2 - \Phi \right\} + \sum_{t=1}^{n} h_t^2 \sum_{s=1}^{t-1} \sum_{r \neq s}^{t-1} h_s \epsilon_s h_r \epsilon_r c_{t-s} c_{t-r}. \tag{38}$$

The term in braces is

$$\left\{ \sum_{t=1}^{n-1} h_t^2 (\epsilon_t^2 - 1) \sum_{s=1}^{n-t} h_{s+t}^2 c_s^2 \right\} + \left\{ \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 c_s^2 - \Phi \right\}. \tag{39}$$

Now

$$\begin{aligned}
\sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 c_s^2 &= \frac{4p}{m\left(\sum_r h_r^2\right)^2} \sum_{j=p}^{m}\sum_{k=p}^{m} \nu_j\nu_k \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \cos s\lambda_j \cos s\lambda_k \\
&= \frac{4p}{m\left(\sum_r h_r^2\right)^2} \sum_{j=p}^{m} \nu_j^2 \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \cos^2 s\lambda_j \\
&\quad + \frac{2p}{m\left(\sum_r h_r^2\right)^2} \sum_{j=p}^{m}\sum_{k\neq j}^{m} \nu_j\nu_k \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \left[\cos s(\lambda_j-\lambda_k) + \cos s(\lambda_j+\lambda_k)\right].
\end{aligned}$$

Next, using part (A) of Lemma 7, for n large enough,

$$\frac{4p}{m}\left(\sum_r h_r^2\right)^{-2} \sum_{j=p}^{m} \nu_j^2 \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \cos^2 s\lambda_j = \frac{p}{m}\sum_{j=p}^{m} \nu_j^2 + O\left(m^{-1}(\log m)^2 + n^{-1}\right), \tag{40}$$

and using part (B) of Lemma 7,

$$\begin{aligned}
&\frac{2p}{m}\left(\sum_r h_r^2\right)^{-2} \sum_{j=p}^{m}\sum_{k\neq j}^{m} \nu_j\nu_k \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \left[\cos s(\lambda_j-\lambda_k) + \cos s(\lambda_j+\lambda_k)\right] \\
&= \frac{p}{m}\left(\sum_r h_r^2\right)^{-2} \sum_{j}^{m}\sum_{k\neq j}^{m} \nu_j\nu_k \left\{\left[\sum_1^n h_t^2 \cos t(\lambda_j-\lambda_k)\right]^2 + \left[\sum_1^n h_t^2 \cos t(\lambda_j+\lambda_k)\right]^2\right\} \qquad (41) \\
&+O(mn^{-1}(\log m)^2).
\end{aligned}$$

Noting that, $1 \le j \le n/2$,

$$\left(\sum_1^n h_t^2\right)^{-1} \sum_1^n h_t^2 \cos t\lambda_j = O(j^{-p}), \tag{42}$$

(see, e.g. Lemmas 1 and 2 in Velasco (1997a)) so

$$\begin{aligned}
\frac{p}{m}\left(\sum_1^n h_t^2\right)^{-2} \sum_{j=p}^{m}\sum_{k\neq j}^{m} \nu_j\nu_k \left[\sum_1^n h_t^2 \cos t(\lambda_j+\lambda_k)\right]^2 &= O\left(m^{-1}\sum_{j=p}^{m}\sum_{k=p}^{m} \nu_j\nu_k (j+k)^{-2p}\right) \\
&= O\left(m^{-1}(\log m)^2 \sum_{j=p}^{m}\sum_{k=p}^{m} j^{-p}k^{-p}\right) \\
&= O\left(m^{-1}(\log m)^2\right),
\end{aligned}$$

the second term in the brackets of (41) can be neglected. For the other term in (41) we can write, including simultaneously the first component of the right hand side of (40), $0 \le \eta(n) \le m$,

$$\begin{aligned}
&\frac{p}{m}\left(\sum_1^n h_t^2\right)^{-2} \sum_{j=p}^{m}\sum_{k=p}^{m} \nu_j\nu_k \left[\sum_1^n h_t^2 \cos t(\lambda_j-\lambda_k)\right]^2 \\
&= \frac{p}{m}\left(\sum_1^n h_t^2\right)^{-2} \sum_{j=p}^{m} \sum_{k:|j-k|\le\eta} \nu_j\nu_k \left[\sum_1^n h_t^2 \cos t(\lambda_j-\lambda_k)\right]^2 \\
&\quad + \frac{p}{m}\left(\sum_1^n h_t^2\right)^{-2} \sum_{j=p}^{m} \sum_{k:|j-k|>\eta} \nu_j\nu_k \left[\sum_1^n h_t^2 \cos t(\lambda_j-\lambda_k)\right]^2 \\
&= \frac{p}{m}\left(\sum_1^n h_t^2\right)^{-2} \sum_{j=p}^{m} \nu_j^2 \sum_{k:|j-k|\le\eta} \left[\sum_1^n h_t^2 \cos t(\lambda_j-\lambda_k)\right]^2
\end{aligned}$$

$$+O\left(m^{-1}\left(\sum_1^n h_t^2\right)^{-2}\sum_{j=p}^m\sum_{k:|j-k|\le\eta}|\nu_j|\sup_{|j-k|\le\eta}|\nu_j-\nu_k|\left[\sum_1^n h_t^2\cos t(\lambda_j-\lambda_k)\right]^2\right)$$
$$+O\left(m^{-1}(\log m)^2\sum_{j=p}^m\sum_{k:|j-k|>\eta}|j-k|^{-2p}\right)$$

and this is

$$\frac{p}{m}\sum_{j=p}^m\nu_j^2\left(\sum_1^n h_t^2\right)^{-2}\sum_{k=0,p,2p,\ldots}^{n-p}\left(\sum_1^n h_t^2\cos t\lambda_k\right)^2$$
$$+O\left(m^{-1}\sum_{j=p}^m\nu_j^2\sum_{k>\eta}^n k^{-2p}\right)+O\left(m^{-1}\log m\sum_{j=p}^m\frac{\eta}{j}\sum_{k:|j-k|\le\eta}|j-k|^{-2p}\right)+O\left(\eta^{1-2p}(\log m)^2\right),$$

which using

$$\frac{p}{m}\sum_p^m\nu_j^2=1+O\left(\frac{(\log m)^2}{m}\right),$$

is

$$\left(\sum_1^n h_t^2\right)^{-2}\sum_{k=0,p,2p,\ldots}^{n-p}\left(\sum_1^n h_t^2\cos t\lambda_k\right)^2+O\left(\eta^{1-2p}\right)+O\left(\frac{\eta}{m}(\log m)^2\right)+O\left(\eta^{1-2p}(\log m)^2\right)+o(1)$$
$$=\quad\Phi+o(1),$$

the errors being $o(1)$ on choosing, e.g., $\eta\sim m^{1/2}$, and $\Phi<\infty$ exists due to (42). Then the second term in (39) is $o(1)$ as $n\to\infty$. The first component of (39) has zero mean and variance

$$O\left(\sum_{t=1}^{n-1}h_t^2\left\{\sum_{s=1}^{n-t}h_{s+t}^2c_s^2\right\}^2\right).$$

Now using the same bounds for $|c_s|$ as in Robinson (1995b) and noting that $\sup_t|h_t|\le 1$, we obtain that this is $O((\log n)^4/n)$, so (39) is $o_P(1)$. The second component of (38) has zero mean and variance

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

because the weights $\{h_t\}$ are symmetric around $\lfloor n/2\rfloor$. As in Robinson's paper, the first term on the right is $O((\log m)^4/n)$, and the second has absolute value bounded by

$$4\sum_{t=3}^n\sum_{u=2}^{t-1}\left(\sum_s^{u-1}c_{t-r}^2\sum_{r\ne s}^{u-1}c_{u-r}^2\right)\le 4\left(\sum_1^n c_t^2\right)\left(\sum_{t=3}^n\sum_{u=2}^{t-1}\sum_{r=t-u+1}^{t-1}c_r^2\right),$$

since $\sup_t|h_t|\le 1$, and using the same arguments in that reference this is $O((\log m)^4/m^{1/3})$ and thus we have verified (36). To prove (37) we also check the sufficient condition

$$\sum_1^n E[z^4]\to 0\qquad\text{as }\ n\to\infty.$$

The left-hand side of this equals

$$\begin{aligned}
\mu_4 \sum_2^n E\left[\left(\sum_1^{t-1} h_s \epsilon_s c_{t-s}\right)^4\right] &\leq C \sum_2^n E\left[\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}\sum_{q=1}^{t-1}\sum_{p=1}^{t-1} h_s h_r h_q h_p \epsilon_s \epsilon_r \epsilon_q \epsilon_p c_{t-s} c_{r-s} c_{q-s} c_{p-s}\right] \\
&\leq C \sum_1^n \left(\sum_1^n c_{t-s}^4\right) + C \sum_1^n \sum_1^{t-1} \sum_1^{t-1} c_{t-s}^2 c_{t-r}^2 = O\left(\frac{(\log m)^4}{n}\right)
\end{aligned}$$

using the bound for h_t and the given reference, completing the proof. •

Lemma 7 *If the sequence $\{h_j\}$ is a data tapers of order p as defined previously, $0 < |j| < n/2$,*

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

and, $0 < |j| < n$,

$$\text{(B)} \quad \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \cos s\lambda_j = \frac{1}{2}\left(\sum_{1}^{n} h_t^2 \cos t\lambda_j\right)^2 + O(n).$$

Proof. *Proof of (B).* We have

$$\begin{aligned}
\sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \cos s\lambda_j &= \sum_{t=1}^{n-1} h_t^2 \sum_{s=1-t}^{0} h_{s+t}^2 \cos s\lambda_j + O(n) \\
&= \frac{1}{2}\sum_{t=1}^{n-1} h_t^2 \sum_{s=1-t}^{n-t} h_{s+t}^2 \cos s\lambda_j + O(n) \\
&= \frac{1}{2}\sum_{t=1}^{n} h_t^2 \cos t\lambda_j \sum_{s=1}^{n} h_s^2 \cos s\lambda_j + O(n).
\end{aligned}$$

The first two lines follow by symmetry, because $h_t = h_{n-t}$ and $\psi_t = \phi_{n-t}$, where $\psi_t = \sum_{s=1}^{n-t} h_{s+t}^2 \cos s\lambda_j$ and $\phi_t = \sum_{s=1-t}^{0} h_{s+t}^2 \cos s\lambda_j$, the error terms are due to end effects, and the last step follows because

$$\begin{aligned}
\sum_{t=1}^{n-1} h_t^2 \sum_{s=1-t}^{n-t} h_{s+t}^2 \cos s\lambda_j &= \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n} h_s^2 \cos(s-t)\lambda_j \\
&= \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n} h_s^2 \left(\cos s\lambda_j \cos t\lambda_j + \sin s\lambda_j \sin t\lambda_j\right) \\
&= \sum_{t=1}^{n} h_t^2 \sum_{s=1}^{n} h_s^2 \cos s\lambda_j \cos t\lambda_j + O(n),
\end{aligned}$$

since the sine terms cancel out by symmetry again.

Proof of (A). Again, by symmetry, changing variable in the sum index, using trigonometric identities and the proof of property (B),

$$\begin{aligned}
\sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \cos^2 s\lambda_j &= \frac{1}{2}\sum_{t=1}^{n-1} h_t^2 \sum_{s=1-t}^{n-t} h_{s+t}^2 \cos^2 s\lambda_j + O(n) \\
&= \frac{1}{2}\sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n} h_s^2 \cos^2(s-t)\lambda_j + O(n) \\
&= \frac{1}{4}\sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n} h_s^2 \left\{1 - \cos 2(s-t)\lambda_j\right\} + O(n)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{4}\sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n} h_s^2 - \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n} h_s^2 \cos(s-t)\lambda_{2j} + O(n) \\
&= \frac{1}{4}\left(\sum_{t=1}^{n} h_t^2\right)^2 + O\left(n + \left[\sum_{s=1}^{n} h_s^2 \cos s\lambda_{2j}\right]^2\right),
\end{aligned}
$$

and the lemma follows on using (42). •

Acknowledgements

I am grateful to P.M. Robinson, an Associate Editor and a referee for helpful discussions and suggestions. I also wish to thank J. Arteche and A. Pérez for valuable comments. Part of this research was carried out while the author was at the Departments of Statistics, London School of Economics and University of Oxford. Research Supported by ESRC grant 235892.

References

[1] AGIAKLOGLOU, C., NEWBOLD, P. and WOHAR, M. (1993) Bias in an estimator of the fractional difference parameter. *J. Time Ser. Anal.* 14, 235-246.

[2] ALEKSEEV, V.G. (1996) Jackson- and Jackson-Vallée Poussin-type kernels and their probability applications. *Theory Probab. Appl.* 41, 137-143.

[3] BAILLIE, R.T., BOLLERSLEV, T. AND MIKKELSEN, H.O. (1996) Fractionally integrated generalized autoregressive conditional heteroskedasticidity. *J. Economet.* 74, 3-30.

[4] BLOOMFIELD, P. (1991) Time series methods. In *Statistical Theory and Modeling*, edited by D.V. Hinkley, N. Reid and E.J. Snell. London: Chapman and Hall, pp. 152-176.

[5] BOLLERSLEV, T. AND MIKKELSEN, H.O. (1996) Modeling and pricing long memory in stock market volatility. *J. Economet.* 73, 151-184.

[6] DAHLHAUS, R. (1985) Asymptotic Normality of Spectral Estimates. *J. Multivariate Anal.* 16, 412-431.

[7] _____ (1988) Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *Ann. Stat.* 16, 808-841.

[8] FOX, R. AND TAQQU, M.S. (1986) Large-sample properties of parameter estimates for strongly dependent stationary Gaussian times series. *Ann. Stat.* 14, 517-532.

[9] GEWEKE, J. AND PORTER-HUDAK, S. (1983) The estimation and application of long memory time series models. *J. Time Ser. Analysis* 4, 221-238.

[10] GIRAITIS, L. AND SURGAILIS, D. (1990) A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate. *Probab. Theory Relat. Fields* 86, 87-104.

[11] GIRAITIS, L., ROBINSON, P.M. and SAMAROV, A. (1995) Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long range dependence. *J. Time Ser. Anal.* 18,49-60.

[12] HASSLER, U. (1992) Unit root test: the autoregressive approach in comparison with the periodogram regression. *Statistical Papers* 34, 67-82.

[13] HURVICH, C.M. AND RAY, B.K. (1995) Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *J. Time Ser. Anal.* 16, 17-42.

[14] JENSEN, M.J. (1995) Ordinary Least Squares Estimate of the Fractional Differencing Parameter Using Wavelets as Derived from Smoothing Kernels. Southern Illinois University, Carbondale.

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

[16] LOBATO, I. and ROBINSON, P.M. (1996) Averaged periodogram estimation of long memory. *J. Economet.* 73, 303-324.

[17] McCOY, E.J. and WALDEN, A.T. (1996) Wavelet Analysis and Synthesis of Stationary Long-Memory Processes. *J. of Comp. and Graphical Stats.* 5, 26-56.

[18] ROBINSON, P.M. (1986) On the errors-in-variables problem for time series. *J. Multivariate Anal.* 19, 240-250.

[19] _____ (1991) Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *J. Economet.* 47, 67-84.

[20] _____ (1994a) Semiparametric analysis of long-memory time series. *Ann. Stat.* 22, 515-539.

[21] _____ (1994b) Rates of convergence and optimal spectral bandwidth for long range dependence. *Probability Theory and Related Fields* 99, 443-473.

[22] _____ (1995a) Log-periodogram regression of time series with long range dependence. *Ann. Stat.* 23, 1048-1072.

[23] _____ (1995b) Gaussian semiparametric estimation of long range dependence. *Ann. Stat.* 23, 1630-1661.

[24] VELASCO, C. (1997a) Non-Stationary log-periodogram regression. Preprint.

[25] _____ (1997b) Non-Gaussian log-periodogram regression. Preprint.

[26] ZHURBENKO, I.G. (1979) On the efficiency of estimates of a spectral density. *Scandinavian J. Statistics* 6, 49-56.

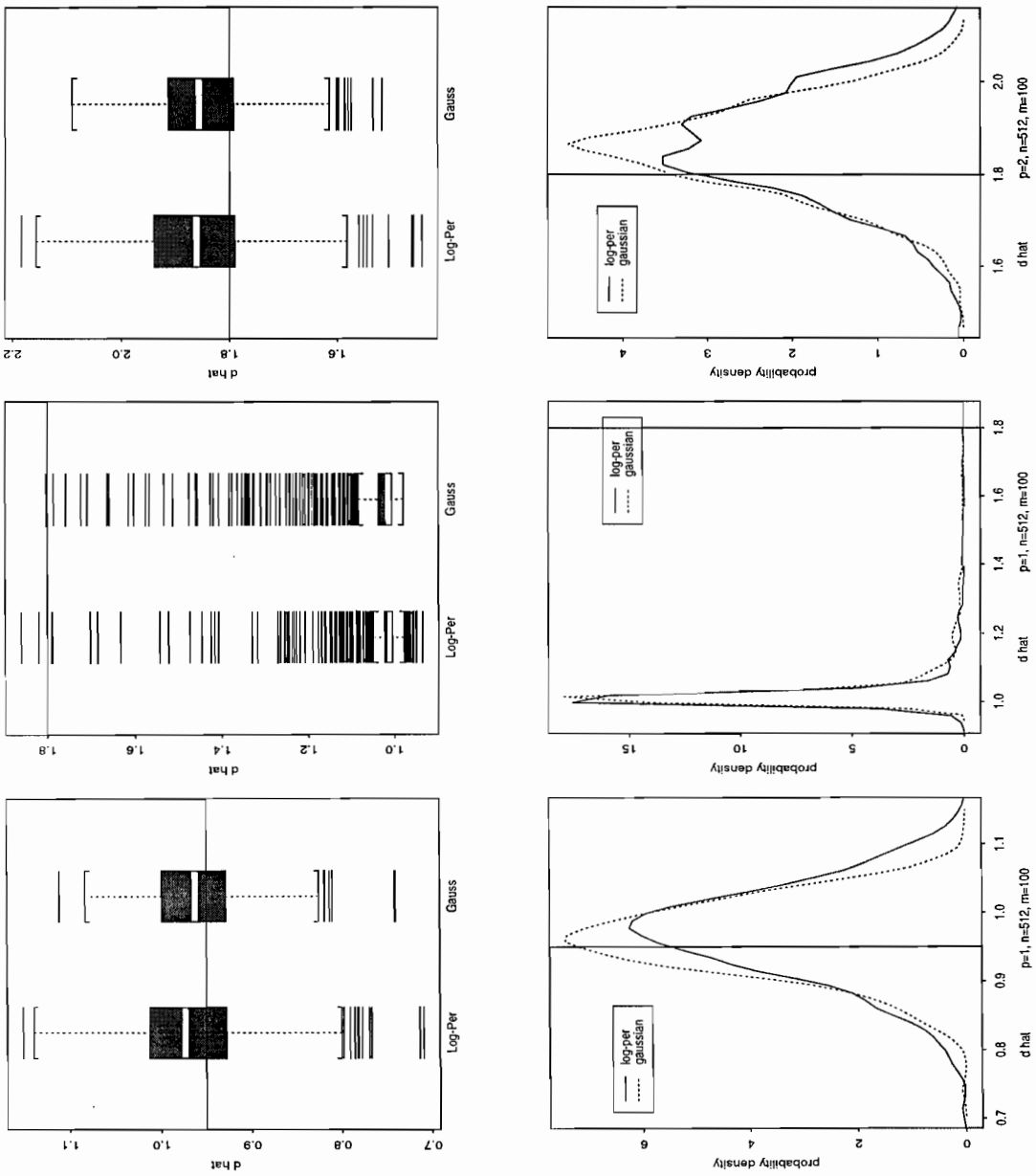

Figure 2: Gaussian semiparametric and Log-periodogram estimates, $n = 512$, $m = 100$, Gaussian ARFIMA$(0, d, 0)$, 1000 Replications

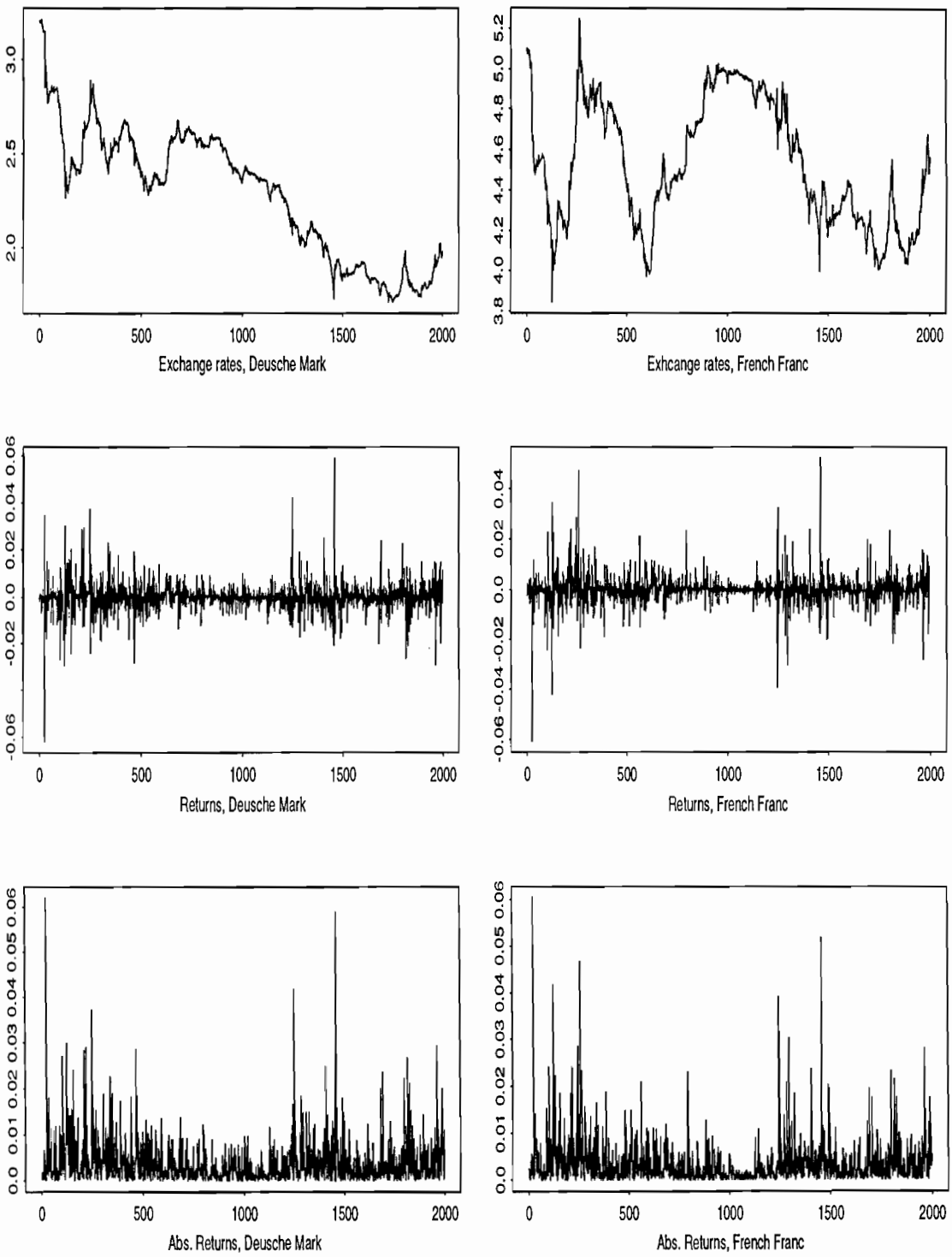

Figure 3: Exchange rates, returns and absolute returns for the Deusche Mark and French Franc against US dollar, November 1973 - January 1981

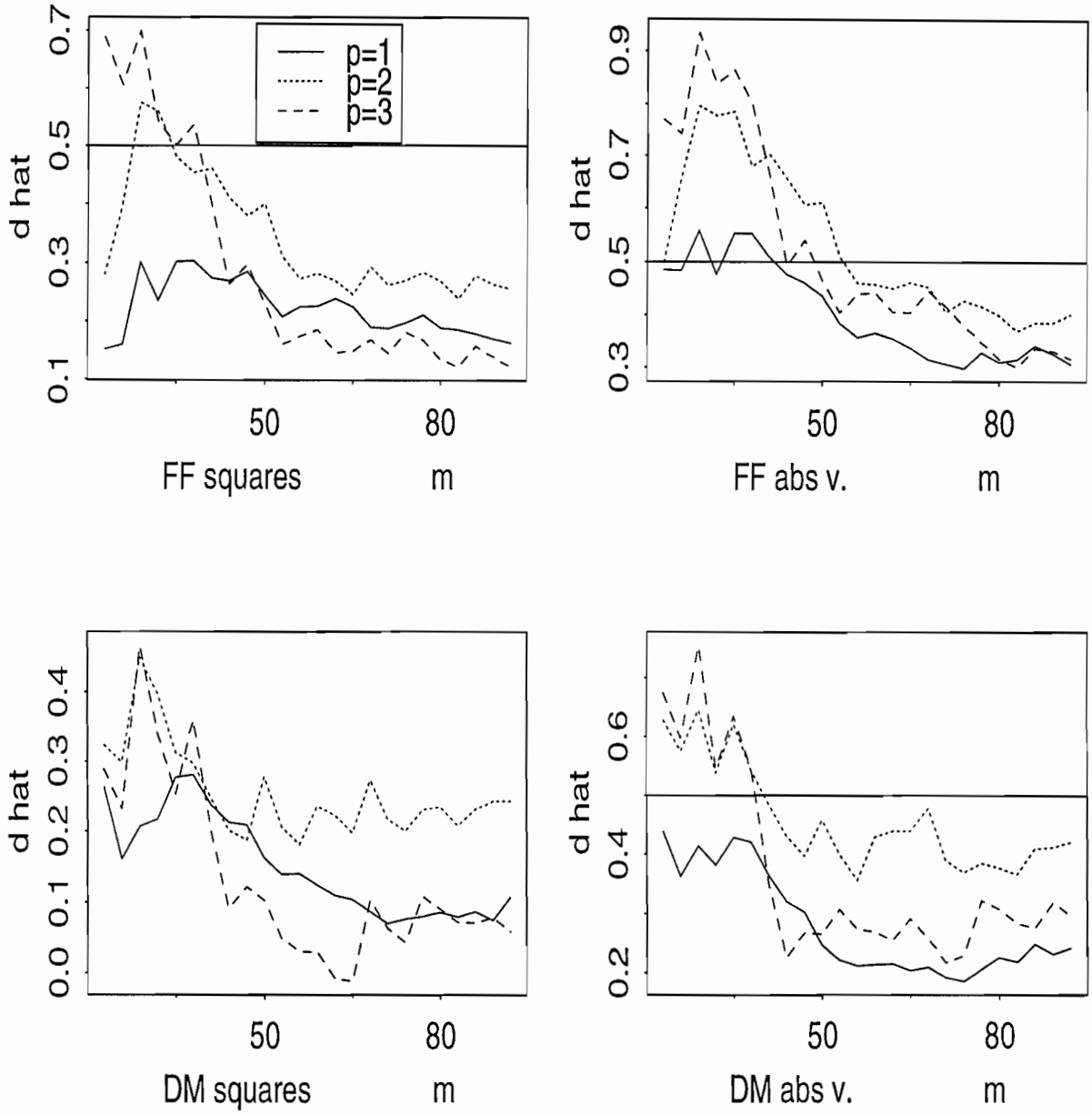

Figure 4: Gaussian semiparametric estimates of the persistence for French Franc and Deusche Mark