Working Paper 98-19
Statistics and Econometrics Series 12
February 1998

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

NON-STATIONARY LOG-PERIODOGRAM REGRESSION.

Carlos Velasco*

Abstract

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

Keywords:
Non-stationary time series, log-periodogram regression, 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. I am grateful to P.M. Robinson for helpful discussions and suggestions. I also wish to thank P. Zaffaroni, C. Michelacci, F.J. Hidalgo, L. Giraitis and L.A. Gil Alaña for valuable comments. The first version of this paper was written while the author was at the London School of Economics and Political Science. Financial support from the Fundación Ramón Areces (Spain) and the Economic and Social Research Council (ESRC) grant n. R000235892 is gratefully acknowledged. Research funded by the Spanish Dirección General de Enseñanza Superior, Ref. n. PB95-0292.

Non-Stationary Log-Periodogram Regression

Carlos Velasco *†

Department of Statistics and Econometrics

Universidad Carlos III de Madrid

February 2, 1998

Abstract

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

Key words: Non-stationary time series; log-periodogram regression; semiparametric inference; tapering.

JEL classification: C11; C22

*Mailing address: Carlos Velasco, Department of Statistics and Econometrics, Universidad Carlos III de Madrid. C. Madrid 126, 28903 Getafe (Madrid), Spain. Tel. +34 1 6249887. Fax. +34 1 6249849. E-mail: cavelas@est-econ.uc3m.es

†I am grateful to P.M. Robinson for helpful discussions and suggestions. I also wish to thank P. Zaffaroni, C. Michelacci, F.J. Hidalgo, L. Giraitis and L.A. Gil Alaña for valuable comments. The first version of this paper was written while the author was at the London School of Economics & Political Science. Financial support from the Fundación Ramón Areces (Spain) and the Economic and Social Research Council (ESRC) grant no. R000235892 is gratefully acknowledged.

case. We find that in the Gaussian case the log-periodogram estimate is asymptotically normal for $d < \frac{3}{4}$ and still consistent for $d < 1$. Here we are trying to approximate a different function than in the stationary situation, explaining the discrepancy with respect to the estimates which use previously differentiated observations. When we taper the periodogram with the cosine window, as suggested by Hurvich and Ray (1995), we adapt Velasco (1997) results to show that the estimate is asymptotically normal even for $d < \frac{3}{2}$.

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

Finally, we analyse empirically the performance of the estimates for finite sample sizes. We show how to base a choice of the degree of tapering, identifying when it produces biased estimates for all possible choices of a bandwidth parameter. Then we illustrate the theory with the application of the log-periodogram estimate with different degrees of tapering and bandwidth choices to two macroeconomic time series.

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

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

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

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

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

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

$$\left|\frac{d}{d\lambda} f(\lambda)\right| = O(\lambda^{-1-2d}) \quad \text{as } \lambda \to 0^+. \tag{4}$$

These assumptions could have been formulated in terms of the functions f^* and/or f, since we are precisely interested in the implications they have on the function f, (2) to (4). 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_ϵ has.

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

$$\begin{aligned} w(\lambda_j) &= \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t \exp(i\lambda_j t) \\ &= \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \sum_{k=1}^{t} \epsilon_k \exp(i\lambda_j t), \end{aligned} \tag{5}$$

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

Defining the periodogram of X_t as

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

and for $J = 1, 2, \ldots$, fixed, (assuming $(m - \ell)/J$ integer),

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

the estimate considered in Robinson (1995) for stationary and invertible time series is

$$\widehat{d} = \left(\sum_k \Lambda_k^2\right)^{-1} \left(\sum_k \Lambda_k Y_k^{(J)}\right).$$

(a) $E[v(\lambda_j)\overline{v}(\lambda_j)] = 1 + O\left(\delta_{j,j} + [j/n]^{\alpha}\right)$,

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

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

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

Proof. See Appendix A.

This result is valid only for $d < 1$ and makes sense with Hurvich and Ray (1995) observation that the bias of the periodogram decreases as j grows only for $d < 1$ (but otherwise increases).

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

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

as a convolution of f and the Fejér kernel

$$K(\lambda) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} e^{i\lambda t}\right|^2 = \frac{1}{2\pi n}\frac{\sin^2[n\lambda/2]}{\sin^2[\lambda/2]},$$

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

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

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

$$\frac{1}{m-\ell} + \frac{\log m}{\ell^{2(1-d)}} + \frac{(\log n)^2}{m} + \frac{m}{n} \to 0 \qquad \text{as } n \to \infty, \tag{6}$$

$\widehat{d}$ is consistent for d.

Proof. From Theorem 1, for $d < 1$ and frequencies λ_j, $j = \ell, \ldots, m$, with ℓ increasing slowly with n, from (6), the normalized discrete Fourier transforms of X_t have exactly the same first two moments structure as in the stationary and invertible case ($-\frac{1}{2} < d < \frac{1}{2}$). Then, given the Gaussianity assumption for ϵ_t, the Fourier transforms are also Gaussian distributed because they are a linear combination of Gaussian variables from (5). Then, following Remark 8 of Robinson (1992), the estimate of $d < 1$ will be consistent with condition (6). •

We observe that the trimming has to be more important (i.e. ℓ increasing faster) as d approaches 1. For values $d \geq 1$ the periodogram is not unbiased for the function f as j increases, and therefore $\widehat{d}$

4 Cosine bell tapered periodogram

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

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

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

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

$$w^T(\lambda_j) = \frac{1}{\sqrt{6}}\left[-w(\lambda_{j-1}) + 2w(\lambda_j) - w(\lambda_{j+1})\right]. \tag{8}$$

Then, the spectral kernel for the tapered periodogram, corresponding to Fejér kernel $K(\lambda)$ for the periodogram is

$$\begin{aligned} K^T(\lambda_j - \lambda) = \frac{1}{2\pi \sum h_t^2}\left|D^T(\lambda_j - \lambda)\right|^2 &= \frac{1}{2\pi \sum h_t^2}\left|\sum_{t=1}^{n} h_t \exp\{it(\lambda_j - \lambda)\}\right|^2 \\ &= \frac{1}{2\pi \sum h_t^2} \sin^2[n(\lambda_j - \lambda)/2]\, H_j^2(\lambda). \end{aligned}$$

where $D^T(\lambda)$ is the equivalent of the Dirichlet kernel $D(\lambda)$ in the non tapered case, from (8) equal to

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

with

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

Then $K^T(\lambda)$ is even, positive, integrates to one and satisfies (see, e.g., Bloomfield (1976) or Hannan (1970, p. 265)):

- $\sup_{\lambda,n} |K^T(\lambda)| = O(n)$.
- $\sup_{\lambda,n} |K^T(\lambda)| = O(n^{-5}|\lambda|^{-6})$.

These properties derive from the fact that $\sup_{\lambda,n} |D^T(\lambda)| = O(\min\{n, n^{-2}|\lambda|^{-3}\})$. From this property of K^T, the tapered periodogram have improved asymptotic properties with respect to the usual periodogram, since the tails of the kernel $K^T(\lambda)$ decrease much faster with the frequency and with the sample size that the tails of Fejér kernel K. Therefore, we will be able to reduce the bias of the periodogram on the tails, even for frequencies close to a singularity and non-integrable functions, if they are smooth enough.

Theorem 5 *Under the assumptions of Theorem 4, ϵ_t Gaussian and*

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

we obtain

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

Proof. First, we observe that for $d \in [\frac{1}{2}, \frac{3}{2})$ the uniform bound for the bias errors in the covariance matrix of the tapered discrete Fourier transforms at the frequencies considered in the definition of d^T, is

$$o\left(m^{-1/2}\right),$$

using the last condition in (9). Hence, under (9) the asymptotic uncorrelatedness and then independence of w^T are enough to make valid all the asymptotic results of Robinson (1995) for $\widehat{d}^T$ and non-stationary processes with memory parameter $d \in [\frac{1}{2}, \frac{3}{2})$. The conditions on the bandwidths are now slightly milder, since we do not have the term in $\log m$ thanks to tapering. •

This result is in line with Hurvich and Ray (1995) empirical findings for $\widehat{d}^T$ and $d \geq 1$. In this case the choice of bandwidth and trimming numbers does not depend on the value of d, even when it is arbitrary close to $\frac{3}{2}$. Also it tells us that, for any value of d, although tapering might reduce the bias of the periodogram and therefore of the estimate of d, it always will increase the variance by a factor of 3, due to the modification in the definition of $\widehat{d}^T$ with respect to $\widehat{d}$. We conjecture that this modification could be avoided, using all the Fourier frequencies, resulting in an increment of the variance of the estimate due to the autocorrelation between adjacent Fourier transforms, which follow approximately an MA(2) process. However in this case Robinson's (1995) results can not be applied directly since they are based on the asymptotic independence of those transforms.

5 General non-stationary processes

In this section we propose a general model for non-stationary time series $d \geq \frac{1}{2}$ and show how to extend the previous ideas to the estimation of the memory parameter d when we use appropriate data tapers.

We say that the observed sequence X_t, $t = 1, \ldots, n$, has memory parameter $d > -\frac{1}{2}$ if $\Delta^s X_t = \epsilon_t^{(s)}$, $s = \lfloor d + \frac{1}{2} \rfloor$, is stationary with mean μ, possibly different from zero, and spectral density $f_{\epsilon^{(s)}}(\lambda)$ behaving as $\lambda^{-2(d-s)}$ around the origin. In Section 2 we have considered the case $s = 1$.

Denote for $r = 1, 2, \ldots, s$,

$$\Delta^r X_t = \epsilon_t^{(r)},$$

and the function

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

1 Introduction

Statistical inference for stationary long range dependent time series is often based on semiparametric estimates that avoid parameterization of the short run behaviour. One of most popular semiparametric estimates in the frequency domain is the log-periodogram regression, proposed initially by Geweke and Porter-Hudak (1983). Robinson (1995) showed the consistency and asymptotic normality of a version of that estimate for stationary and invertible Gaussian vector time series. He assumed that the spectral density $f(\lambda)$ of the observed stationary sequence satisfies for one constant $0 < G < \infty$,

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

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

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

Robinson (1995) advocated an initial differentiation (integration) of the observed time series when non-stationarity (non-invertibility) is suspected, to obtain a value of d in the stationary and invertible interval $(-\frac{1}{2}, \frac{1}{2})$ and then perform the periodogram regression on the transformed series, adjusting the estimate with the number of differences (integrations) taken. However, the simulation work of Hassler (1993) and Hurvich and Ray (1995) suggests that, at least for values $d \in [\frac{1}{2}, 1)$, the estimation procedure using the original series can be consistent, although it will not coincide in general with the pre-differenced estimate.

Using Hurvich and Ray's definitions we extend Robinson (1995) results to cover the non-stationary

Condition (12) holds if for this sequence of frequencies λ_j $D^T(\lambda_j)$ has s-th derivative equal to zero. The Dirichlet kernel is zero for all Fourier frequencies λ_j, $0 < j < n$, but its derivative is not zero. The same holds for the cosine bell taper.

Summarising, to estimate the parameter d for a general non-stationary process X_t as defined above, we need sequences of data tapers $\{h_t\}$ with the following requirements, in terms of their Fourier transform $D^T(\lambda_j)$, for some sequence of Fourier frequencies λ_j,

1. $D^T(\lambda_j)$ need to have zeroes at these frequencies of order at least s (to make the expectation of the tapered periodogram finite and remove the influence from the past).

2. $D^T(\lambda_j)$ need to have all derivatives up to order s equal to zero at these frequencies (to remove time trends).

3. The tails of $D^T(\lambda)$ have to converge uniformly to zero with n and λ as fast as possible (to reduce the bias of the periodogram for f).

In previous analysis of tapering properties only condition 3. has been required (see, for example, Condition C1 in Robinson (1986) or Dahlhaus (1988) assumptions), but to deal with a general form of non-stationarity, conditions 1. and 2. are essential. However, in Robinson (1986, p. 246) is noted that not only $D(\lambda)$, but also its derivatives should be small away from the origin if we want to control the trending behaviour due to non-random smooth functions in t (e.g. polynomial). Luckily, several data tapers that have good convergence properties of type 3. also satisfy 1. and 2. for some $s > 1$ and some Fourier frequencies. Note that the cosine bell taper improve 3. with respect to Fejér kernel, but as this kernel, satisfy 1. only for $s = 1$, but not 2., so we assumed $\mu = 0$ (e.g. they do not work even for random walks with drift). Before defining a general class of data tapers, we consider two examples.

For sample size $n = 4N$, N integer, the weights given by the Parzen window

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

satisfy (12) 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) use 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

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

2 Assumptions and definitions

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

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

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

$$X_t = \sum_{k=1}^{t} \epsilon_k + R, \qquad R = \sum_{k=-\infty}^{0} \epsilon_k,$$

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

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

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

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

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

$$f_\epsilon(\lambda) = G \cdot \lambda^{-2(d-1)} + O(\lambda^{-2(d-1)+\alpha}) \qquad \textit{as } \lambda \to 0^+.$$

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

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

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

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

$$f_\epsilon(\lambda) = G \cdot \lambda^{-2(d-1)} + G\, E_\alpha \cdot \lambda^{-2(d-1)+\alpha} + o(\lambda^{-2(d-1)+\alpha}) \qquad \textit{as } \lambda \to 0^+.$$

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

$$\frac{f(\lambda)}{g(\lambda)} = 1 + E_\alpha \cdot \lambda^\alpha + o(\lambda^\alpha) \qquad \text{as } \lambda \to 0^+. \tag{3}$$

at frequencies λ_{jp}, $0 < j < N$, condition (12) is satisfied.

Note that these last two conditions are due to the presence of the function $\sin^p[n\lambda/2p]$ in the numerator of D_p^T, and that the presence in the denominator of $\sin^p[\lambda/2]$ will allow a relatively easy of treatment of the asymptotic moments of the Fourier transform in terms of the function $f(\lambda)$. Note also that these conditions apply directly to the previous definition of the Fourier transform ($p = 1$), to the triangular-Barlett window ($p = 2$) and to Parzen's weights ($p = 4$), but for the Zhurbenko-Kolmogorov's ones (for the same p) we need a rescaling of the weights given before. The cosine bell taper belongs to this class with $p = 1$, but has some improved convergence properties of type 3., corresponding to tapers of order $p = 3$.

We now analyse the covariance matrix of the (normalized) tapered Fourier transform with tapers of order p. We obtain that the periodogram is unbiased for any $d < p$. The main problem here are the covariance terms: tapering destroys the orthogonality of the sine and cosine functions and the solution we employed for the cosine taper is no longer valid. Therefore, we are led to consider frequencies which are moving closer somewhat slower than n^{-1}.

Theorem 6 *Under Assumptions 2 and 3 [$d > -\frac{1}{2}$, $0 < \alpha \leq 2$] for $f_{\epsilon^{(s)}}$, a data taper of order $p = 2, 3, \ldots$, with $p \geq s+1$ [or just $p > d$ if $\mu = 0$], for any increasing sequences of positive integers $k = k(n)$ and $j = j(n)$, $k < j$, and $\eta = \eta(n)$, $\eta < \inf(j-k)$, such that $j/n \to 0$,*

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

and

$$\frac{\log n}{\eta} \to 0 \qquad as \ n \to \infty, \tag{13}$$

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

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

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

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

Proof. See Appendix A. •

Condition (13) is not strictly necessary, but simplifies the bounds obtained in the proof. In Theorem 8 we will have to assume something stronger about η and $\min k$ to obtain a consistent estimate of d.

Given the results of the previous theorem, we have to adapt consequently the definition of $\widehat{d}$, taking $J = 1$ for simplicity,

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

Here $\Lambda_k = z_k - \overline{z}$, $\overline{z} = \{J/(m-\ell)\} \sum_k z_k$ and $z_k = -2\log \lambda_k$. The number m is an integer smaller than n and ℓ is a user-chosen trimming number. In the asymptotics both numbers tend to infinity with the sample size n, but more slowly.

The main idea to show that Robinson (1995) results go through in the non-stationary case ($d \geq \frac{1}{2}$) is to analyse the asymptotic behaviour of the discrete Fourier transform of X_t for frequencies λ_j, $\ell < j \leq m$. We will show that under some assumptions this behaviour is equivalent to the stationary case. Therefore, assuming Gaussianity for the ϵ_k's, we could repeat the steps in Robinson (1995) to obtain the consistency and asymptotic distribution of the log periodogram estimate of the parameter d for non-stationary processes. This is possible, because the proof of Theorem 3 in Robinson (1995) only uses the error in the estimation of the covariance matrix of the discrete Fourier transforms at low frequencies and the Gaussianity of the discrete Fourier transform of X_t (implied by (5)).

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

3 Non-tapered periodogram

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

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

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

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

definition of $d^{(p)}$ is again $o((m\eta)^{-1/2})$, using the first two and the last conditions in (15) and $p \geq s+1$, so $p > d + \frac{1}{2}$. Hence, under (15) the asymptotic uncorrelatedness and Gaussianity of w_p^T is enough to make valid all the asymptotic results of Robinson (1995) for $\widehat{d}_p^T$ and non-stationary processes with $d \geq \frac{1}{2}$. •

6 Non-invertible processes

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

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

In this section we analyse if tapering with higher order ($p > 1$) tapers may be fruitful to estimate the memory d of non-invertible time series satisfying the semiparametric model (1). We shall obtain, using the techniques of Theorems 6 and 8, that with p big enough (for d fixed), $\widehat{d}_p^T$ is consistent and asymptotic normal for any $d \leq -\frac{1}{2}$. The main intuition is the following. With $d < 0$ the process is stationary, so there are no problems with the definition of the spectral density $f(\lambda)$ or with its integrability. Here, given the required normalization for the moments of the discrete Fourier transform ($f(\lambda_j) = O(\lambda_j^{-2d}) = o(1)$ for $d < 0$ and $j/n \to 0$), the issue is how to avoid leakage from high frequencies (i.e. outside a neigbourhood of the origin, where we do not assume anything for f apart from integrability) to the zero frequency, where the spectral density f has a zero of order $-2d > 0$. This problem can be controlled easily by the fast uniform convergence of the tails of $K_p^T(\lambda)$ with n and λ when p is chosen suitably. We could consider all Fourier frequencies, but this will not improve in principle the estimation procedure, given the high correlation for adjacent periodogram ordinates when tapering.

For the covariance matrix of the tapered periodogram we have

Theorem 9 *Under Assumptions 1 and 3 [$0 < \alpha \leq 2$] for $f(\lambda)$, $d \leq -\frac{1}{2}$, we can chose a data taper of order $p = 2, 3, \ldots$, such that for any increasing sequences of positive integers $j = j(n)$ and $k = k(n)$, $j > k$, $\inf(j-k) > \eta$, with $j/n \to 0$, $\eta = \eta(n) \to \infty$,*

$$\frac{\log n}{\eta} \to 0 \qquad \text{as} \quad n \to \infty,$$

can not be consistent. The asymptotic normality of $\widehat{d}$ needs stronger assumptions on the trimming and bandwidth numbers to control the bias and can only be obtained for $d < \frac{3}{4}$:

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

$$\frac{m^{1/2}\log m}{\ell^{2(1-d)}} + \frac{\ell(\log n)^2}{m} + \frac{m^{1+1/2\alpha}}{n} \to 0 \qquad as \;\; n \to \infty, \tag{7}$$

we obtain

$$m^{1/2}(\widehat{d} - d) \to_d N(0, \frac{J}{4}\psi'(J)),$$

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

Proof. Further to the comments in the proof of the previous theorem, from equation (5.17) in Robinson (1995), we need the error terms in the covariance matrix to be $o(m^{-1/2})$ and that ℓ is tending to infinity slower than m (only possible for $d < \frac{3}{4}$). From Theorem 1, (7) is sufficient for that. •

Note that this result for $d < \frac{3}{4}$ is exactly the same as in the stationary case, and that the asymptotic distribution does not depend on any unknown parameter. However when d is very close to the boundary $\frac{3}{4}$ the choice of the numbers ℓ and m is very limited by the first condition in (7), and will depend on the true value of d. The limitations in the asymptotics are due to the extra bias in the estimation of the elements of the covariance matrix of the discrete Fourier transform because of the behaviour of f when $d \geq \frac{1}{2}$. Basically, the periodogram is asymptotically unbiased at λ_j as j increases only when $d < 1$, and the order magnitude of the bias depends on the value of d, unlike in the stationary case. Furthermore, the bounds for the biases of the covariance matrix of the Fourier transforms are not sufficient for the asymptotic normality for $d \geq \frac{3}{4}$.

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

As both functions, $K(\lambda)$ and $K^T(\lambda)$, integrate to one, there has to be a trade off between the behaviour of the kernels at the origin and at the tails, i.e., the tails of K^T are less thicker than those of K, but the central lobe is much wider. This is the reason why we only can consider tapered periodogram ordinates or discrete Fourier transforms that are at least three basic frequencies $\lambda_1 = 2\pi/n$ away.

Furthermore, the order of the zero of K^T at λ_j, $j = 1, 2, \ldots, n-1$, given by the function $\sin^2[n\lambda/2]$, is of the same order, 2, as in the case of Fejér kernel, so we cannot consider functions f with $d \geq \frac{3}{2}$, as the expectation of the periodogram will always diverge.

Define as before the normalized tapered Fourier transform $v^T(\lambda) = w^T(\lambda)/(G^{1/2}\lambda^{-d})$.

Theorem 4 *Under Assumptions 2 and 3 [$0 < \alpha \leq 2$], $d \in [\frac{1}{2}, \frac{3}{2})$ for any increasing sequences of positive integers $j = j(n)$ and $k = k(n)$, $j > k + 2$, such that $j/n \to 0$ and*

$$\gamma_{j,k} \equiv (jk)^{d-3} \log j \to 0 \qquad \text{as } n \to \infty,$$

(a) $E[v^T(\lambda_j)\overline{v^T}(\lambda_j)] = 1 + O\left(\min\{j^{-\alpha}, j^{-1}\} + [j/n]^{\alpha} + \gamma_{jj}\right)$,

(b) $E[v^T(\lambda_j)v^T(\lambda_j)] = O\left(j^{-4} + \gamma_{jj}\right)$,

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

(d) $E[v^T(\lambda_j)v^T(\lambda_k)] = O\left(k^{-1} + \gamma_{jk}\right)$.

Proof. See Appendix B. The proof of this theorem results much easier after the one for Theorem 6 in Appendix A below. •

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

This result confirms Hurvich and Ray (1995) observation that the tapered periodogram is unbiased, even for values of d close to $\frac{3}{2}$.

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

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

with

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

and using Theorem 4, we can obtain, similarly to Theorem 3, and $n = 90$ on the first row of Figure 1. We can observe that the bigger the order p, the smoother is the transition in the extremes of the taper weights in the observed interval $1, \ldots, n$. The spectral windows $K_p^T(\lambda)$ in the second row exhibit zeroes at different frequencies and central lobes with width increasing with p. In the third row we have the same simulated ARFIMA$(0, 1.45, 0)$ tapered series for all p considered. For any $p > 1$ the tapered series is hardly comparable with the original, $p = 1$, the shape of the tapering scheme is dominating. Finally, the last row of pictures corresponds to the periodograms of the different tapered series in log-log coordinates, all of them being approximately linear, at least for the lower frequencies, though with different slopes as a consequence of the properties of the tapered periodogram for each p.

For $d \in [-1, 1.5]$ Hurvich and Ray (1995) provide an extensive simulation exercise for the log-periodogram estimate of d using the raw and the tapered (cosine bell) periodograms, confirming the results of our own simulations. Just to report the performance of $\widehat{d_p^T}$ for big p and d, we calculated 1000 replications of the log-periodogram estimate with $p = 8$ for Gaussian ARFIMA$(0, 4.45, 0)$ and different bandwidths numbers m/p, covering all the reasonable range of values for $n = 512$. We use no trimming and only frequencies $\lambda_8, \lambda_{16}, \ldots$, so, for example for $m/p = 60$ and $\eta = 1$ we are using about 7 frequencies in the regression. The series were simulated with the S-Plus function `arima.fracdiff.sim` with $d = .45$ and then integrated four times.

The results of the simulation exercise are summarised in Table I and in the box-plots of Figure 2 for $\eta = 1, 2$. Here m or me represents the value j for the maximum Fourier frequency λ_j employed, $m\eta p$ in the notation of Theorem 8. For $\eta = 1$ we can observe that for small m the estimates have positive bias, which could be in part due to the use of no trimming and to the use of a too small number of frequencies. For big m, close to $n/2$, the bias is negative, and the variance is always decreasing with m, as we could expect, although of relatively big magnitude in all cases. Part of the high variability of the estimates can be due to the presence of correlation between different periodogram ordinates when tapering, so we decided to try with $\eta = 2$, but keeping in the simulations the same maximum frequency used (and therefore using half of *observations* in each regression). Now, the bias is reduced for small values of m, but the variance increases, as the reduction of the number of periodogram ordinates seems to compensate in excess the lower correlation between them.

is defined as before in terms of the spectral density of the stationary sequence $\epsilon_t^{(s)}$.

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

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

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

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

$$
\begin{aligned}
w^T(\lambda_j) &= \frac{1}{\sqrt{2\pi \sum h_t^2}} \sum_{t=1}^{n} w_t X_t \exp(i\lambda_j t) \\
&= \frac{1}{\sqrt{2\pi \sum h_t^2}} \sum_{t=1}^{n} w_t \left(\sum_{r=1}^{s} R^{(r)} p^{(r)}(t) + \mu p_\mu(t) \right) \exp(i\lambda_j t) \qquad (10) \\
&\quad + \frac{1}{\sqrt{2\pi \sum h_t^2}} \sum_{t=1}^{n} w_t \sum_{j_1}^{t} \sum_{j_2}^{j_1} \cdots \sum_{j_s}^{j_{s-1}} \epsilon_{j_s}^{(*)} \exp(i\lambda_j t) \qquad (11)
\end{aligned}
$$

We think of the term (10) as a nuisance term which comprises the information in $\{X_t\}_1^n$ from the past. To make inferences about d we need to eliminate this dependence on the past or on the initial conditions as we did when $s = 1$, making this expression equal to zero, at least for certain frequencies λ_j, using certain orthogonality properties of the weights h_t, i.e.

$$
\sum_{t=1}^{n} h_t \left(1 + t + t^2 + \cdots + t^s\right) \exp(i\lambda_j t) = 0. \qquad (12)
$$

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 $\mu = 0$, so we only need to eliminate the influence from the polynomial $p^{(1)}(t) = 1$ of order 0 (a constant with respect to t) and both the raw and cosine bell-tapered Fourier transforms satisfy that condition (but not any of higher order).

The definition of X_t in terms of the s-th integration of a stationary process, allows the inclusion of not only s unit roots but also of deterministic time trends up to order s. Then, if condition (12) holds, these trends are removed in the calculation of $w^T(\lambda_j)$ without need of estimate them by any means.

an indication that the memory of the series is bigger that the value of p used, so we need to use higher order tapers (or differentiate).

The last situation considered is an example of the misleading that the presence of deterministic trends may cause on the estimates for different p. We took the series with memory parameter $d = 2.45$ and added to it a cubic trend. Of the estimates considered, only $\widehat{d_4^T}$ is resistant to that modification, as it is clear from the estimation results. Here $\widehat{d_3^T}$ estimates almost always 3, although for this series only $d = 2.45$: it takes wrongly the cubic trend as an indication of more memory than what actually is.

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

In Figure 4 we repeat the same exercise as before, but now differencing the original series ($d = .45$) two, three and four times. In each case only the procedures with $p > |d| - 1$ give consistent estimates, taking into account that no deterministic trends are present. It can be observed that in all cases the leakage from high frequencies when m is big leads to positive biases.

We now illustrate all these points with two data sets. They are taken from Engle and Granger (1987) and correspond to the logs of the US Consumer Price Index and production worker wages in manufacturing over the 50, 60 and 70's. These are monthly observations and we have $n = 360$ observations. We have calculated as before all $\widehat{d_p^T}$ from $mp = 12$ to $n/4 \approx 88$, with steps of 4, so for $p = 4$ and $mp = 12$ we only use 3 points to carry out the regression. In Figure 5 we plot the logarithm of the tapered periodogram with $p = 2$ and 3, since in previous studies it has been sustained that these series have two unit roots. We can see how for the wages series there is a significant seasonal (monthly) component that will condition all the analyses that use frequencies above $\frac{2}{12}\pi$ (i.e. λ_j with $j \geq 30$). Therefore, we will expect an important negative bias for all estimates of d we report when $m \geq 30$, i.e. for all but the first 5 estimates $\widehat{d_p^T}$ for each p. In the case of the prices series the seasonal component is much less clear, but in any case, the use of large m will probably cause serious downwards bias.

For the wages series we give the results of the analysis in Figure 6. In the first row we plot the original and transformed series after trend removal. We estimate, by OLS, linear and quadratic trends successively. Then, in the bottom row graphics, we plot the result of the estimation of d in each case using Zhurbenko tapers, $p = 1, 2, 3, 4$. We have chosen not to plot the confidence intervals, to avoid too complicated graphics, but they can be obtained easily from the theorems above. From the analysis with the original series, focusing on the first estimates (for small m), we are already in conditions of saying that d is about, perhaps larger than, 2, using the results with $p = 2, 3, 4$, although the estimates are very dependent on the m chosen. The estimate $p = 1$ has a "suspicious" behaviour, taking values around 1 for all m, indicating that it is not able to deal with the memory present in the series. When we

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)). Its exact value can be calculated easily for each n as well as the set of weights $\{h_t\}$. Therefore, this class of taper weights for $p = 1, 2, \ldots$, fixed in the asymptotics, and $n = pN$ satisfies conditions 1. and 2. with $s \leq p-1$ at frequencies λ_{jp}, $0 < j < N$. Zhurbenko considered both p and N increasing with the sample size n. We find more natural to fix p and then define $N = n/p$, regarding p as the *order* of the taper, which will indicate the maximum value of d we can estimate.

When $p = 4$, these weights are very close to Parzen's ones and both have the same asymptotic properties. Kolmogorov weights correspond to the p-th convolution of the uniform density, so for $p = 1$ we obtain the raw Fourier transform and with $p = 2$ we are using Barlett's or triangular window. The properties that these kernels share derive for the function $\left(\sin^2[n\lambda/2p] \,/\, \sin^2[\lambda/2]\right)^p$ that appears in their spectral kernels $K^P(\lambda)$ and $K^Z(\lambda)$.

We will only consider tapers symmetric around $n/2$, with $\max h_t = 1$. We say then that a sequence of data tapers $\{h_t\}_1^n$ is of order p if the following two conditions are satisfied:

- For $N = n/p$ (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 a 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

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

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

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

with

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

in such a way that for $\eta = 1, 2, \ldots$ we are still using about m *observations* in the regression [ignoring the trimming], so the variance of $\widehat{d}_p^T$ can be of order m^{-1} if $\eta > 1$. Now using Theorem 6, we obtain

Theorem 7 *Under the assumptions of Theorem 6, $p \geq s+1$, $p > 1$ [or just $p > d$ if $\mu = 0$], $\epsilon_t^{(s)}$ Gaussian, $\eta = 1$ and*

$$\frac{1}{m-\ell} + \frac{\log m}{\ell^{2(p-d)}} + \frac{(\log n)^2}{m} + \frac{m}{n} \to 0 \qquad as\ \ n \to \infty, \tag{14}$$

we obtain $\widehat{d}_p^T \to_P d$.

Proof. This follows from Theorem 6 in the same way as Theorem 2 with the consideration of the first two moments of $\widehat{d}_p^T$, identifying in the covariance terms with periodogram ordinates at close frequencies and far apart. •

For the asymptotic distribution we need in the definition of $\widehat{d}_p^T$ that η increases with n to obtain approximate independence of the tapered ordinates used in the estimate. In this way, when we use the same number of periodogram values, the variance of the estimate is decreasing with η.

Theorem 8 *Under the assumptions of Theorem 6, $p \geq s+1$, $p > 1$ [or just $p > d$ if $\mu = 0$], $\epsilon_t^{(s)}$ Gaussian and*

$$\frac{m^{1/(2p-1)}}{\eta} + \frac{m^{p/(2p-1)}}{\ell} + \frac{\ell(\log n)^2}{m\eta} + \frac{(m\eta)^{1+1/2\alpha}}{n} \to 0 \qquad as\ \ n \to \infty, \tag{15}$$

we obtain

$$m^{1/2}\left(\widehat{d}_p^T - d\right) \to_d \mathcal{N}(0, \frac{\pi^2}{24}).$$

Note that the lower growth rate required for ℓ is significatively larger than for η. We observe that the improved convergence properties of tapering are used to keep the bias under control. However, for any d fixed, to increase p will not reduce significatively the bias in the covariance matrix of the Fourier transform, due mainly to the covariance terms, unless we increment at the same time α (i.e., the smoothness of f near the origin). For example with $\alpha = 2$ and $p > d+1$, we need

$$\frac{m^{p/2(2p-1)}}{\ell} \to 0 \qquad \text{as}\ \ n \to \infty,$$

where the exponent of m is tending to $\frac{1}{4}$ as p increases, so the trimming required is not specially big, because the bias is reduced.

Proof of Theorem 8. First, we observe that for any d the uniform bound for the bias errors in the covariance matrix of the tapered discrete Fourier transforms at the frequencies considered in the eter d, including non-stationary situations ($d \geq \frac{1}{2}$) with deterministic trends, and non-invertible ($d \leq -\frac{1}{2}$) time series.

- We have described what are the ultimate reasons why certain tapering schemes are resistant to particular non-stationary behaviours, but not to all. As Robinson (1986, p. 242) and Zhurbenko (1979) remark, the benefits of tapering only show up for certain data windows but not by tapering the data with any general smooth function.

- The results of this paper can be applied directly to obtain the asymptotic properties of nonparametric *spectral* estimates of functions f (of discrete average type) for fixed (Fourier) frequencies away from the origin, showing why traditional spectral nonparametric methods work in non-stationary situations for which they were not designed in first instance, justifying the conjecture of Robinson (1986, p. 246). This also confirms the observation of Granger (1966) about the shape of the spectral density of possibly non-stationary economic time series estimated from the original data.

- The bounds for the moments of the discrete tapered Fourier transform for non-stationary processes obtained in this paper are only valid when evaluated at some particular Fourier frequencies λ_{jp}, $0 < j < N$ since it is only there where the spectral kernel of the Fourier transform (Fejér kernel, if not tapered) has special properties. Thus, they do not extend for any continuously smoothed estimate of f or tapered autocovariances, and only to non-stationarity in other frequencies different from zero if they coincide with a suitable Fourier frequency.

- We have shown how to apply these theoretical findings to the analysis of real data, gaining great insight on the underlying structure of the observed time series without a priori assumptions.

- It is very likely that the results of this paper about the asymptotic properties of the tapered periodogram can be be adapted to carry out statistical inference for other semiparametric and parametric models of non-stationary (and non-invertible) observations without explicit specification of the degree of non-stationarity (non-invertibility).

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

and with p big enough ($p > |d| + \frac{1}{2}$) *such that*

$$\gamma_k = n^a k^b \to 0 \qquad as \quad n \to \infty,$$

$a = \frac{4d^2}{2(p-d)-1}$, $b = 2d - \frac{(2p-1)^2}{2(p-d)-1}$,

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

(b) $E[v_p^T(\lambda_{jp})v_p^T(\lambda_{jp})] = O\left(j^{-1} + \gamma_j\right)$,

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

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

Proof. See Appendix C. •

The exact choice of p to obtain a bound $O(j^{-1})$, say, can be made explicit, but it will depend on d and on the asymptotic relationship between j, k and n, so that $\gamma_k \to 0$. Hence, with the definition of $\widehat{d_p^T}$ as in the previous section, using exactly the same arguments as for Theorems 3, 5 or 8,

Theorem 10 *Under the assumptions of Theorem 9, X_t Gaussian and p big enough such that*

$$\frac{m^{1/(2p-1)}}{\eta} + \frac{m^{1/2}}{\ell} + \frac{\ell(\log n)^2}{m\,\eta} + \frac{(m\,\eta)^{1+1/2\alpha}}{n} + \frac{n^{8d^2} m^{2(p-d)-1}}{\ell^c} \to 0 \qquad as \;\; n \to \infty, \tag{16}$$

$c = -4d[2(p-d)-1] + 2(2p-1)^2$, *we obtain*

$$m^{1/2}\left(\widehat{d_p^T} - d\right) \to_d N(0, \frac{\pi^2}{24}).$$

The last condition in (16) corresponds to $\gamma_\ell = o(m^{-1/2})$. The exponents of all the quantities are positive, the one for ℓ growing very fast with p. For example, with $d = -2$ and $p = 4$ this is implied by $nm^{1/3}\ell^{-5.6} \to 0$, so if $m \sim n^{4/5}$ a choice of $\ell \sim n^{1/4}$ is sufficient. When $d = -2$ and $p = 3$ the condition is implied by $nm^{.29}\ell^{-3.82} \to 0$, so if $m \sim n^{4/5}$ again, a choice of $\ell \sim n^{.31}$ would suffice. These conditions are in the same line with the ones required by, e.g. the log-periodogram regression estimate for stationary and invertible processes.

Finally we note that this theorem is valid as is stated for the cosine-bell taper when we fix $p = 3$, since in the proofs we only use the uniform bounds for the tails of the kernel $D^T(\lambda)$ and not the properties of this kernel at any particular Fourier frequency.

7 Empirical results

In this section we describe briefly the practical implementation of the previous estimates of the memory parameter d, with both simulated and real non-stationary data. We will concentrate on Zhurbenko-Kolmogorov tapers with different values of p. We have plotted these data tapers for $p = 1, 2, 3$ and 4 and

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

Then, the theorem follows from Robinson's (1995) Theorem 2 proof, where he considered the stationary and invertible case $d \in (-\frac{1}{2}, \frac{1}{2})$. For the interval around the origin, $[-\lambda_j/2, \lambda_j/2]$, where $f(\lambda)$ is no longer integrable when $d \geq \frac{1}{2}$, use the proof of Theorem 6 below with $p = 1$, $\eta = 1$ and $d \in [\frac{1}{2}, 1)$, using the exact orthogonality of the sine and cosine components in the discrete Fourier transforms. •

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

Lemma 1 *For a data taper of order $p > 1$, $j = j(n)$, such that $1/j + j/n \to 0$,*

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

Proof. By symmetry we only need to consider $\lambda > 0$. Then

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

and the bound follows using the integrability in $[-\pi,\pi]$ of $D_p^T(\lambda)$, $p \geq 2$ for all n. •

Lemma 2 *For a data taper of order $p > 1$, $j = j(n)$, $k = k(n)$, $k < j$, such that $1/k + j/n \to 0$, and $\inf j - k > \eta$, $\eta = \eta(n)$, $1/\eta + \eta/n \to 0$,*

$$n^{-1}\int_{-\pi}^{\pi} |D_p^T(\lambda_{jp}-\lambda)D_p^T(\lambda-\lambda_{kp})|d\lambda = O(\eta^{-p})$$

Proof. Considering the intervals of integration

$$\int_{-\pi}^{\lambda_{(k+j)p/2}} + \int_{\lambda_{(k+j)p/2}}^{\pi}$$

and that $(\lambda_{jp}-\lambda_{kp})^{-1} = O(n\eta^{-1})$, we have, for example,

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

and the bound follows as before using the integrability of D_p^T, $p > 1$. •

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

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

Table I. Log-Periodogram estimate, Gaussian ARFIMA(0.4.45, 0), $n = 512$						
	$\eta = 1$			$\eta = 2$		
m	bias	s.d.	MSE	bias	s.d.	MSE
60	0.22235	0.39519	0.20559	-0.03244	0.82988	0.68975
80	0.10558	0.28835	0.09428	-0.03496	0.48934	0.24068
100	0.05336	0.24762	0.06416	-0.09801	0.44337	0.20618
130	-0.04428	0.21347	0.04753	-0.15503	0.34308	0.14174
160	-0.13664	0.18276	0.05207	-0.26321	0.29573	0.15673
190	-0.21491	0.15990	0.07175	-0.33477	0.27429	0.18731
230	-0.37179	0.15251	0.16148	-0.50896	0.23226	0.31298

The conclusions that we can draw of this and other related simulations we performed, and which will guide further analysis and comments are

- Except for very large samples sizes, there seems not to be special advantage in taking $\eta > 1$.
- We would expect positive bias for *small* m, and negative bias for *big* m.

Of course, this would be conditioned by the presence of other significant features in the dynamics of the process, like seasonal and cyclical components which may dominate the shape of $f(\lambda)$ at certain frequencies. It is important to note that model (1) is approximately valid for ARFIMA(0,d,0) processes for all frequencies, so to increase m may reduce sometimes the bias, but this will not be the case for more general models.

Given the general class of estimates $\widehat{d_p^T}$ defined by the Zhurbenko-Kolmogorov weights it is interesting to study their different properties depending on the value of d and on the presence or not of deterministic trends in the observed time series. In Figure 3 we show the typical behaviour of $\widehat{d_p^T}$ for the same time series when integrated and/or added trends. The starting series is as before Gaussian ARFIMA(0,0.45,0), $n = 1024$, and we integrate it once, twice and three times and also when integrated twice, we added a cubic trend to it. Then we obtained the values of $\widehat{d_p^T}$, $p = 1, \ldots, 4$, for a range of values of mp from about 25 to $n/2$, with increments of $4 = \max p$.

When $d = 1.45$ the estimate $\widehat{d_1^T} = \widehat{d}$ does not work, as expected (if fact this is the usual log-periodogram estimate, valid only for $d < 1$, following Theorem 2). For $p > 2$ the results are much better and we can regard the estimates as consistent, the best results obtained here for $p = 2, 3$. When $d = 2.45$ and $d = 3.45$ we can see that only with $p = 3, 4$ and $p = 4$, respectively, we capture the true features of the data, the estimates with $p < s + 1$ ($s = 2, 3$) converging invariantly to the value of p. This behaviour has been observed in all simulations and could be regarded, among other problems, as

increase the value of m, all estimates produce much smaller estimates to a different degree depending on how robust they are to leakage. One interesting point for this series, is that $\widehat{d_2^T}$ starts to have an equivalent behaviour to $\widehat{d_p^T}$ with $p = 3, 4$ when we have removed a quadratic trend, but not with only a linear one. This might indicate that for this series it is likely the presence of a quadratic deterministic trend.

For the prices series we give the analysis in Figure 8. Here the method is the same as with the wages. Again, from the original series, and looking at the estimates with small m, the memory of the series seems to be between 1.8 and 2.2. The estimate $\widehat{d_1^T}$ has problems to estimate the memory, since $d > 1$. When we take into account estimates with much bigger values of m, all estimates but $\widehat{d_1^T}$ give quite different answers than those using only low frequencies as with the wages series. Given that $\widehat{d_2^T}$ gives the same answers as the estimates $p = 3, 4$ with the original series and also with a linear trend removed, and that when a quadratic trend is removed $\widehat{d_2^T}$ seems to start to estimate less memory than before, we confirm that the quadratic trend is not appropriate here.

To finish the analysis we study what happens when we differentiate the observed time series enough number of times. In Figure 7 we give this analysis when we differentiate the wages series from one to three times. The estimates of the memory of the original series we report are $\widehat{d_p^T}$ for the transformed series plus the number of differences taken. These results confirm the previous remarks, specially the comments about the choice of m. With one difference, all estimates $p > 1$ give similar answers, in the line with the previous analysis. Now $\widehat{d_2^T}$ seems not to have problems, since in the case that a quadratic trend were present in the original series, it deals appropriately with the remaining linear trend in the "differenced" one. When we differentiate twice, so $\widehat{d_1^T}$ is consistent assuming that the original series had memory around 2 and a quadratic trend, it estimates d above 1.6 for the relevant choices of m, confirming the previous detrending analysis. When we differentiate three times, the estimate with $p = 1$ is no longer able to deal with the strong non-invertibility, in the sense that it cannot avoid the leakage from higher frequencies towards the zero at $f(0)$.

An equivalent analysis for the prices series is given in Figure 9 below. Here, if we differentiate once, so $\widehat{d_1^T}$ is consistent assuming that the original series had memory slightly less than 2 and a linear trend, it gives estimates around 1.8, agreeing with the previous analysis. When we differentiate twice or three times, $\widehat{d_2^T}$ starts, in fact, to estimate slightly less memory than before, indicanting an excess of differentiation and that the deterministic quadratic factor is probably not present in the series.

8 Conclusions

We summarise the main findings of this paper in the following points:

- We have given a unified asymptotic theory for the log-periodogram estimate of the memory paramwhich making a change of variable is equal to

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

and this is not greater than a constant times

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

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

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

and

$$\begin{aligned}A&\le\frac{2^{2p}}{n^{2p}}\int_{-\pi jp}^{\pi jp}\frac{\sin^{2p}[\lambda/2p]}{[(2\pi jp-\lambda)/2n]^{2p}}\left|\frac{\lambda}{2\pi j}\right|^{-2d}d\lambda\\&\le j^{2d}2^{4p}(2\pi)^{2d}\int_{-\pi jp}^{\pi jp}\frac{\sin^{2p}[\lambda/2p]}{(2\pi jp-\lambda)^2}|\lambda|^{-2d}d\lambda.\end{aligned}$$

Now, using that $2(p-d)>-1$ from $p\ge s+1$ we see that

$$\int_{-\pi jp}^{\pi jp}\sin^{2p}[\lambda/2p]\,|\lambda|^{-2d}d\lambda=O(\log j),\tag{18}$$

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

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

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

$$A=O\left(j^{2(d-p)}\log j\right).$$

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

$$\begin{aligned}\left|\int_{\lambda_{jp}/2}^{3\lambda_{jp}/2}[f(\lambda)-f(\lambda_{jp})]\,K_p^T(\lambda_{jp}-\lambda)d\lambda\right|&=\left|\int_{-\lambda_{jp}/2}^{\lambda_{jp}/2}[f(\lambda_{jp}-\lambda)-f(\lambda_{jp})]\,K_p^T(\lambda)d\lambda\right|\\&=\left|\int_{-\lambda_{jp}/2}^{\lambda_{jp}/2}[\lambda\cdot f'(\lambda_{jp})+O(\lambda_{jp}^{-\alpha-2d}|\lambda|^{\alpha})]\,K_p^T(\lambda)d\lambda\right|\\&=O\left(\lambda_{jp}^{-\alpha-2d}\int_{-\lambda_{jp}/2}^{\lambda_{jp}/2}|\lambda|^{\alpha}K_p^T(\lambda)d\lambda\right),\end{aligned}$$

since K_p^T is even and we are integrating in a symmetric interval around 0. Now, with $\alpha\in(1,2]$,

$$\begin{aligned}\int_{-\lambda_{jp}/2}^{\lambda_{jp}/2}|\lambda|^{\alpha}K_p^T(\lambda)&=2\left\{\int_0^{n^{-1}}+\int_{n^{-1}}^{\lambda_{jp}/2}\right\}\lambda^{\alpha}K_p^T(\lambda)d\lambda\\&=O\left(n\int_0^{n^{-1}}\lambda^{\alpha}d\lambda+n^{1-1p}\int_{n^{-1}}^{\lambda_{jp}/2}\lambda^{\alpha-2p}d\lambda\right)\\&=O(n^{-\alpha}).\end{aligned}$$

9 Appendix A: Proof of Theorems 1 and 6

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

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

Now

$$\begin{aligned} \sum_{t_1=1}^{n}\sum_{k_1=1}^{t_1}\exp\{i\lambda_j t_1\}\exp\{-ik_1\lambda\} &= \sum_{t_1=1}^{n}\exp\{i\lambda_j t_1\}\exp\left\{-i\frac{\lambda(t_1+1)}{2}\right\}\frac{\sin t_1\lambda/2}{\sin\lambda/2} \\ &= \sum_{t_1=1}^{n}\exp\{-i\lambda/2 - it_1(\lambda/2-\lambda_j)\}\frac{\sin t_1\lambda/2}{\sin\lambda/2} \\ &= \frac{\exp\{-i\lambda/2\}}{2i\sin\lambda/2}\sum_{t_1=1}^{n}[\exp\{-it_1\lambda_j\} - \exp\{it_1(\lambda_j-\lambda)\}] \\ &= -\frac{\exp\{-i\lambda/2\}}{2i\sin\lambda/2}\exp\left\{i(\lambda_j-\lambda)\frac{n+1}{2}\right\}\frac{\sin n(\lambda_j-\lambda)/2}{\sin(\lambda_j-\lambda)/2}. \end{aligned}$$

Repeating the same arguments for the sums in $\exp\{-i\lambda_j t_2\}\exp\{ik_2\lambda\}$, we get

$$\begin{aligned} E[I(\lambda_j)] &= \frac{1}{2\pi n}\int_{-\pi}^{\pi}\left[\frac{\sin n(\lambda_j-\lambda)/2}{\sin(\lambda_j-\lambda)/2}\right]^2\frac{f_\epsilon(\lambda)}{4\sin^2\lambda/2}d\lambda \\ &= \int_{-\pi}^{\pi}K(\lambda_j-\lambda)f(\lambda)d\lambda. \end{aligned}$$

For the other moments of the discrete Fourier transform,

$$\begin{aligned} E[w^2(\lambda_j)] &= \frac{1}{2\pi n}\sum_{t_1}\sum_{k_1}\sum_{t_2}\sum_{k_2}\exp\{i\lambda_j(t_1+t_2)\}\gamma_\epsilon(k_1-k_2) \\ &= \frac{1}{2\pi n}\int_{-\pi}^{\pi}\sum_{t_1}\sum_{k_1}\sum_{t_2}\sum_{k_2}\exp\{i\lambda_j(t_1+t_2)\}\exp\{-i\lambda(k_1-k_2)\}f_\epsilon(\lambda)d\lambda \\ &= \frac{1}{2\pi n}\int_{-\pi}^{\pi}\frac{\sin n(\lambda_j-\lambda)/2}{\sin(\lambda_j-\lambda)/2}\frac{\sin n(\lambda+\lambda_j)/2}{\sin(\lambda+\lambda_j)/2}\exp\{i\lambda_j(n+1)\}\frac{f_\epsilon(\lambda)}{4\sin^2\lambda/2}d\lambda \\ &= \frac{1}{2\pi n}\int_{-\pi}^{\pi}f(\lambda)D_n(\lambda_j-\lambda)D_n(\lambda+\lambda_j)d\lambda, \end{aligned}$$

where $D_n(\lambda) = \sum_{t=1}^{n}e^{it\lambda}$ is Dirichlet kernel. Finally

$$\begin{aligned} E[w(\lambda_j)\bar{w}(\lambda_k)] &= \frac{1}{2\pi n}\sum_{t_1}\sum_{\ell_1}\sum_{t_2}\sum_{\ell_2}\exp\{i\lambda_j t_1 - i\lambda_k t_2)\}\gamma_\epsilon(\ell_1-\ell_2) \\ &= \frac{1}{2\pi n}\int_{-\pi}^{\pi}\sum_{t_1}\sum_{k_1}\sum_{t_2}\sum_{k_2}\exp\{i\lambda_j t_1 - i\lambda_k t_2)\}\exp\{-i\lambda(\ell_1-\ell_2)\}f_\epsilon(\lambda)d\lambda \\ &= \frac{1}{2\pi n}\int_{-\pi}^{\pi}\frac{\sin n(\lambda_j-\lambda)/2}{\sin(\lambda_j-\lambda)/2}\frac{\sin n(\lambda-\lambda_k)/2}{\sin(\lambda-\lambda_k)/2}\exp\left\{i(\lambda_j-\lambda_k)\frac{n+1}{2}\right\}\frac{f_\epsilon(\lambda)}{4\sin^2\lambda/2}d\lambda \\ &= \frac{1}{2\pi n}\int_{-\pi}^{\pi}f(\lambda)D_n(\lambda_j-\lambda)D_n(\lambda-\lambda_k)d\lambda, \end{aligned}$$

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

$$\left|\int_{-\lambda_{jp}/2}^{\lambda_{jp}/2}\right| = O\left(\max_{-\lambda_{jp}/2\le\lambda\le\lambda_{jp}/2}\left|D_p^T(\lambda_{jp}-\lambda)D_p^T(\lambda_{jp}+\lambda)\right|\int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} f(\lambda_{jp})d\lambda\right)$$
$$+O\left(\int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} D_p^T(\lambda_{jp}-\lambda)D_p^T(\lambda_{jp}+\lambda)f(\lambda)d\lambda\right) \quad (20)$$

were the first term on the right hand side is

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

The second term (20) is also $O(f(\lambda_{jp})\cdot j^{1-2p})$ when f is integrable [see the bound for (17)], and when not, making a change of variable similar as before, normalizing by $1/f(\lambda_{jp})$ and substituting $2\sin[\lambda/2]$ by λ, the contribution of the integral in the interval $[-\lambda_{jp}/2,\lambda_{jp}/2]$ is of the same order of magnitude as

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

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

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

as $1/j+j/n\to 0$. Therefore the bound for part (b) follows.

Let study now the covariance term, $k<j$, with $j-k>\eta$,

$$E[w_p^T(\lambda_{jp})\overline{w_p^T(\lambda_{kp})}]$$
$$= \frac{1}{2\pi b n^{2p-1}}\int_{-\pi}^{\pi}\frac{a(\lambda_{jp}-\lambda)a(\lambda-\lambda_{kp})}{(2\sin[\lambda/2])^{2s}}$$
$$\times\left(\frac{\sin[n(\lambda_{jp}-\lambda)/2p]}{\sin[\lambda/2]\sin[(\lambda_{jp}-\lambda)/2]}\frac{\sin[n(\lambda_{kp}-\lambda)/2p]}{\sin[\lambda/2]\sin[(\lambda_{kp}-\lambda)/2]}\right)^p f_\epsilon(\lambda)d\lambda$$
$$= \frac{1}{2\pi b n^{2p-1}}\int_{-\pi}^{\pi}a(\lambda_{jp}-\lambda)a(\lambda-\lambda_{kp})\left(\frac{\sin[n(\lambda_{jp}-\lambda)/2p]}{\sin[(\lambda_{jp}-\lambda)/2]}\frac{\sin[n(\lambda_{kp}-\lambda)/2p]}{\sin[(\lambda_{kp}-\lambda)/2]}\right)^p f(\lambda)d\lambda$$
$$= \frac{1}{2\pi\sum h_t^2}\int_{-\pi}^{\pi}D_p^T(\lambda_{jp}-\lambda)D_p^T(\lambda-\lambda_{kp})f(\lambda)d\lambda.$$

This covariance can be expanded, with error $O(\eta^{-p})$ from Lemma 2, due to the loss of orthogonality as

$$\frac{1}{2\pi\sum h_t^2}\left[\int_{(\lambda_{kp}+\lambda_{jp})/2}^{2\lambda_{jp}}\{f(\lambda)-f(\lambda_{jp})\}D_p^T(\lambda_{jp}-\lambda)D_p^T(\lambda-\lambda_{kp})d\lambda\right. \quad (21)$$
$$+\int_{\lambda_{kp}/2}^{(\lambda_{kp}+\lambda_{jp})/2}\{f(\lambda)-f(\lambda_{kp})\}D_p^T(\lambda_{jp}-\lambda)D_p^T(\lambda-\lambda_{kp})d\lambda \quad (22)$$
$$-\int_{\lambda_{kp}/2}^{(\lambda_{kp}+\lambda_{jp})/2}\{f(\lambda_{jp})-f(\lambda_{kp})\}D_p^T(\lambda_{jp}-\lambda)D_p^T(\lambda-\lambda_{kp})d\lambda \quad (23)$$
$$\left.+\left\{\int_{2\lambda_{jp}}^{\pi}+\int_{-\pi}^{\lambda_{kp}/2}\right\}\{f(\lambda)-f(\lambda_{jp})\}D_p^T(\lambda_{jp}-\lambda)D_p^T(\lambda-\lambda_{kp})d\lambda\right]. \quad (24)$$

similarly to the stationary case, like if $f(\lambda)$ were the pseudo-spectrum of X_t and using the corresponding spectral kernel for tapering.

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

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

We consider the same intervals of integration to analyse the bias in

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

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

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

using the properties of $K_p^T(\lambda)$ and the integrability of f outside the origin. Next,

$$\begin{aligned}
\left|\int_{-\epsilon}^{-\lambda_{jp}/2}\right| &\leq f(\lambda_{jp}) \int_{-\epsilon}^{-\lambda j/2} |K_p^T(\lambda_{jp} - \lambda)| \, d\lambda + \int_{-\epsilon}^{-\lambda j/2} f(\lambda) \, |K_p^T(\lambda_{jp} - \lambda)| \, d\lambda \\
&= O\left(f(\lambda_{jp}) \cdot n^{1-2p} \int_{\lambda_{jp}/2}^{\infty} \lambda^{-2p} d\lambda + n^{1-2p} \int_{\lambda_{jp}/2}^{\infty} \lambda^{-2d-2p} d\lambda \right) \\
&= O\left(f(\lambda_{jp}) \cdot j^{1-2p} \right).
\end{aligned}$$

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

$$\begin{aligned}
\int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} &\leq \left[\max_{-\lambda_{jp}/2 \leq \lambda \leq \lambda_{jp}/2} K_p^T(\lambda_{jp} - \lambda) \right] \int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} f(\lambda_{jp}) d\lambda \\
&\quad + \int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} K_p^T(\lambda_{jp} - \lambda) f(\lambda) d\lambda
\end{aligned} \tag{17}$$

Now the first term on the right hand side is

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

If $d \in (-\frac{1}{2}, \frac{1}{2})$ the other contribution, (17), of the interval $[-\lambda_{jp}/2, \lambda_{jp}/2]$ is

$$O\left(\sup_{-\lambda_{jp}/2 \leq \lambda \leq \lambda_{jp}/2} K_p^T(\lambda_{jp} - \lambda) \int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} f(\lambda) d\lambda \right) = O(n^{1-2p} \lambda_{jp}^{-2p} \lambda_{jp}^{1-2d}) = O(f(\lambda_{jp}) \cdot j^{1-2p}).$$

When $f(\lambda)$ is not integrable, $d \geq \frac{1}{2}$, to bound (17) we normalize by $1/f(\lambda_{jp})$ and in the definition of f in terms of f^*, we substitute $2\sin[\lambda/2]$ by λ, since the terms $O(|\lambda|^3)$ will cause negligible error in that interval. Then this contribution is of order

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

Therefore

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

When $\alpha \leq 1$ using similar methods, the bound is seem to be $O(f(\lambda_{jp})\cdot j^{-1})$ with the Mean Value Theorem. The proof of (a) is now complete.

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

$$\begin{aligned} E[w(\lambda_{jp})^2] &= \frac{1}{2\pi b n^{2p-1}}\int_{-\pi}^{\pi}\frac{a(\lambda-\lambda_{jp})a(\lambda_{jp}+\lambda)}{(2\sin[\lambda/2])^{2s}} \\ &\quad\times\left(\frac{\sin[n(\lambda_{jp}-\lambda)/2p]}{\sin[(\lambda_{jp}-\lambda)/2]}\right)^p\left(\frac{\sin[n(\lambda_{jp}+\lambda)/2p]}{\sin[(\lambda_{jp}+\lambda)/2]}\right)^p f_\epsilon(\lambda)d\lambda \\ &= \frac{1}{2\pi b n^{2p-1}}\int_{-\pi}^{\pi} a(\lambda-\lambda_{jp})a(\lambda_{jp}+\lambda)\frac{\sin[n(\lambda_{jp}-\lambda)/2p]}{\sin[(\lambda_{jp}-\lambda)/2]}\frac{\sin[n(\lambda_{jp}+\lambda)/2p]}{\sin[(\lambda_{jp}+\lambda)/2]}f(\lambda)d\lambda. \\ &= \frac{1}{2\pi\sum h_t^2}\int_{-\pi}^{\pi} D_p^T(\lambda_{jp}-\lambda)D_p^T(\lambda_{jp}+\lambda)f(\lambda)d\lambda \end{aligned}$$

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

$$\frac{1}{2\pi\sum h_t^2}\int_{-\pi}^{\pi}[f(\lambda)-f(\lambda_{jp})]\,D_p^T(\lambda_{jp}-\lambda)D_p^T(\lambda_{jp}+\lambda)d\lambda + O(f(\lambda_{jp})j^{-p}) \tag{19}$$

where the last term follows form the approximate orthogonality for frequencies that are moving apart (Lemma 1). Now, we can study the integral in (19) splitting the range of integration in the following intervals,

$$\left|\int_{-\pi}^{-\epsilon}+\int_{\epsilon}^{\pi}\right| = O\left(\frac{1}{n^{2p-1}\epsilon^6}\int_{\epsilon}^{\pi}[f(\lambda)+f(\lambda_{jp})]\,d\lambda\right) = O([1+f(\lambda_{jp})]n^{1-2p}) = O(f(\lambda_{jp})j^{-p}),$$

$$\begin{aligned} \left|\int_{-\epsilon}^{-2\lambda_{jp}}+\int_{2\lambda_{jp}}^{\epsilon}\right| &= O\left(f(\lambda_{jp})n^{-1}\int_{2\lambda_{jp}}^{\pi}\left|D_p^T(\lambda_{jp}-\lambda)D_p^T(\lambda_{jp}+\lambda)\right|d\lambda\right) \\ &\quad+O\left(n^{-1}\int_{2\lambda_{jp}}^{\epsilon}f(\lambda)\left|D_p^T(\lambda_{jp}-\lambda)D_p^T(\lambda_{jp}+\lambda)\right|d\lambda\right) \\ &= O\left(f(\lambda_{jp})\cdot n^{1-2p}\int_{2\lambda j}^{\pi}\lambda^{-2p}d\lambda + n^{1-2p}\int_{\lambda j}^{\infty}\lambda^{-2d-2p}d\lambda\right) = O\left(f(\lambda_{jp})\cdot j^{1-2p}\right). \end{aligned}$$

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

$$\begin{aligned} \left|\int_{-2\lambda_{jp}}^{-\lambda_{jp}/2}+\int_{\lambda_{jp}/2}^{2\lambda_{jp}}\right| &= O\left(n^{-1}\sup_{\lambda_{jp}/2\leq\lambda\leq 2\lambda_{jp}}|f'(\lambda)D_p^T(\lambda_{jp}+\lambda)|\int_{\lambda_{jp}/2}^{2\lambda_{jp}}|\lambda_{jp}-\lambda||D_p^T(\lambda_{jp}-\lambda)|d\lambda\right) \\ &= O\left(n^{-1}f(\lambda_{jp})\lambda_{jp}^{-1}n^{1-p}\lambda_{jp}^{-p}\int_0^{2\lambda_{jp}}\lambda|D_p^T(\lambda)|d\lambda\right) = O(f(\lambda_{jp})j^{-1-p}\log n), \end{aligned}$$

because

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

and

$$\begin{aligned}\left|\int_{-\epsilon}^{-\lambda_{jp}}+\int_{2\lambda_{jp}}^{\epsilon}\right| &= O\left(f(\lambda_{jp})n^{-1}\left\{\int_{-\epsilon}^{-\lambda_{jp}}+\int_{2\lambda_{jp}}^{\pi}\right\}\left|D_p^T(\lambda_{jp}-\lambda)D_p^T(\lambda-\lambda_{kp})\right|d\lambda\right)\\ &\quad +O\left(n^{-1}\left\{\int_{-\epsilon}^{-\lambda_{jp}}+\int_{2\lambda_{jp}}^{\epsilon}\right\}f(\lambda)\left|D_p^T(\lambda_{jp}-\lambda)D_p^T(\lambda-\lambda_{kp})\right|d\lambda\right)\\ &= O\left(f(\lambda_{jp})\cdot n^{1-2p}\int_{\lambda_{jp}}^{\infty}\lambda^{-2p}d\lambda+n^{1-2p}\int_{\lambda_{jp}}^{\infty}\lambda^{-2d-2p}d\lambda\right)\\ &= O\left(f(\lambda_{jp})\cdot j^{1-2p}\right)=O\left(f_{jk}\cdot k^{1-2p}\right).\end{aligned}$$

Next,

$$\begin{aligned}\left|\int_{-\lambda_{jp}}^{-\lambda_{kp}/2}\right| &= O\left(n^{-1}\sup_{-\lambda_{jp}\le\lambda\le-\lambda_{kp}/2}f(\lambda)|D_p^T(\lambda_{jp}-\lambda)||D_p^T(\lambda-\lambda_{kp})|\int_{-\lambda_{jp}}^{-\lambda_{kp}/2}d\lambda\right)\\ &= O\left(n^{-1}[f(\lambda_{kp})+f(\lambda_{jp})]n^{2-2p}\lambda_{kp}^{-p}\lambda_{jp}^{1-p}\right)\\ &= O([f(\lambda_{kp})+f(\lambda_{jp})]\cdot k^{-p}j^{1-p})=O(f_{jk}\,k^{-p}),\end{aligned}$$

because $p>d$. Finally,

$$\begin{aligned}\left|\int_{-\lambda_{kp}/2}^{\lambda_{kp}/2}\right| &= O\left(n^{-1}\max_{-\lambda_{kp}/2\le\lambda\le\lambda_{kp}/2}\left|D_p^T(\lambda-\lambda_{kp})D_p^T(\lambda_{jp}-\lambda)\right|\int_{-\lambda_{kp}/2}^{\lambda_{kp}/2}f(\lambda_{jp})d\lambda\right)\\ &\quad +O\left(n^{-1}\int_{-\lambda_{kp}/2}^{\lambda_{kp}/2}\left|D_p^T(\lambda-\lambda_{kp})D_p^T(\lambda_{jp}-\lambda)\right|f(\lambda)d\lambda\right)\end{aligned}\tag{25}$$

The first term is

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

and the second term (25) is $O(f_{jk}\,k^{1-2p})$ if f is integrable, and otherwise, making a change of variable similar as before, the contribution from this interval, after normalization by f_{jk} and with the obvious notation, is of the same order as

$$\begin{aligned}&\frac{1}{2\pi bn^{2p-1}}\int_{-\lambda_{kp}/2}^{\lambda_{kp}/2}\frac{|a||a|f^*(\lambda)}{[f^*(\lambda_{jp})f^*(\lambda_{kp})]^{1/2}}\left|\frac{\sin[n(\lambda_{jp}-\lambda)/2p]}{\sin[(\lambda_{jp}-\lambda)/2]}\frac{\sin[n(\lambda_{kp}-\lambda)/2p]}{\sin[(\lambda_{kp}-\lambda)/2]}\right|^p\lambda_{jp}^d\lambda_{kp}^d|\lambda|^{-2d}d\lambda\\ =&\ \frac{1}{2\pi bn^{2p}}\cdot\int_{-\pi kp}^{\pi kp}\frac{|a||a|f^*(\lambda/n)}{[f^*(\lambda_{jp})f^*(\lambda_{kp})]^{1/2}}\left|\frac{\sin^2[\lambda/2p]}{\sin[(2\pi jp-\lambda)/2n]\sin[(2\pi kp-\lambda)/2n]}\right|^p\\ &\times\left(\frac{2\pi kp}{n}\right)^d\left(\frac{2\pi jp}{n}\right)^d\left|\frac{\lambda}{n}\right|^{-2d}d\lambda,\end{aligned}$$

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

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

as $1/j+j/n\to 0$, and

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

Define $f_{jk} = (\lambda_k \lambda_j)^{-d}$. Now (21) can be bounded by

$$
\begin{aligned}
& O\left(n^{-1} \sup_{(\lambda_{kp}+\lambda_{jp})/2 \le \lambda \le 2\lambda_{jp}} |f'(\lambda) D_p^T(\lambda - \lambda_{kp})| \int_{(\lambda_{kp}+\lambda_{jp})/2}^{2\lambda_{jp}} |\lambda_{jp} - \lambda||D_p^T(\lambda_{jp} - \lambda)| d\lambda\right) \\
= \; & O\left(n^{-1} \cdot f(\lambda_{jp}) \lambda_{jp}^{-1} \cdot n^{1-p} \lambda_{(j-k)p/2}^{-p} \int_0^{\lambda_{jp}} \lambda |D_p^T(\lambda)| d\lambda\right) \\
= \; & O(f(\lambda_{jp}) \cdot j^{-1} \eta^{-p} \log n) = O(f_{jk}\, k^{-1}\, \eta^{-p} \log n),
\end{aligned}
$$

because $\int_0^{2\lambda_{jp}} \lambda |D_p^T(\lambda)| d\lambda = O(n^{-1} \log n)$ for $p \ge 2$.

Next, for $k \ge j/2$, (22) is

$$
\begin{aligned}
& O\left(n^{-1} \sup_{\lambda_{kp}/2 \le \lambda \le (\lambda_{kp}+\lambda_{jp})/2} |f'(\lambda) D_p^T(\lambda_{jp} - \lambda)| \int_{\lambda_{kp}/2}^{(\lambda_{kp}+\lambda_{jp})/2} |\lambda - \lambda_{kp}||D_p^T(\lambda - \lambda_{kp})| d\lambda\right) \\
= \; & O\left(n^{-1} \cdot f(\lambda_{kp}) \lambda_{kp}^{-1} \cdot n \eta^{-p} \int_0^{\lambda_{jp}} \lambda |D_p^T(\lambda)| d\lambda\right) \\
= \; & O(f(\lambda_{kp}) \cdot k^{-1} \eta^{-p} \log n) = O(f_{jk}\, k^{-1} \eta^{-p} \log n),
\end{aligned}
$$

since $f(\lambda_{kp}) = O(f_{jk})$ if $k \ge j/2$, and when $k < j/2$, (22) is

$$
\begin{aligned}
& O\left(n^{-1} \sup_{\lambda_{kp}/2 \le \lambda \le (\lambda_{kp}+\lambda_{jp})/2} |f(\lambda) + f(\lambda_{kp})||D_p^T(\lambda_{jp} - \lambda)| \int_{\lambda_{kp}/2}^{(\lambda_{kp}+\lambda_{jp})/2} |D_p^T(\lambda - \lambda_{kp})| d\lambda\right) \\
= \; & O\left(n^{-1} \cdot [f(\lambda_{kp}) + f(\lambda_{jp})] \cdot n^{1-p} \lambda_{(j-k)p/2}^{-p} \int_0^{\lambda_{jp}} |D_p^T(\lambda)| d\lambda\right) \\
= \; & O([f(\lambda_{kp}) + f(\lambda_{jp})] \cdot j^{-p}) = O(f_{jk}\, k^{-p}),
\end{aligned}
$$

using that $j - k > j/2$, $p > d$, and $\int_0^{\lambda_{jp}} |D_p^T(\lambda)| d\lambda$ for $p \ge 2$.

For $k \ge j/2$, (23) is

$$
\begin{aligned}
& O\left(n^{-1}(\lambda_{jp} - \lambda_{kp}) \sup_{\lambda_{kp} \le \lambda \le \lambda_{jp}} |f'(\lambda)||D_p^T(\lambda_{jp} - \lambda)| \int_{\lambda_{kp}/2}^{(\lambda_{kp}+\lambda_{jp})/2} |D_p^T(\lambda - \lambda_{kp})| d\lambda\right) \\
= \; & O\left(n^{-1} \eta \cdot f(\lambda_{kp}) \cdot \lambda_{kp}^{-1} \eta^{-p} \int_0^{\lambda_{jp}} |D_p^T(\lambda)| d\lambda\right) \\
= \; & O(f(\lambda_{kp}) \cdot k^{-1} \eta^{1-p}) = O(f_{jk}\, k^{-1} \eta^{1-p}),
\end{aligned}
$$

and when $k < j/2$

$$
\begin{aligned}
& O\left(n^{-1} \sup_{\lambda_{kp}/2 \le \lambda \le (\lambda_{kp}+\lambda_{jp})/2} |f(\lambda) + f(\lambda_{kp})||D_p^T(\lambda_{jp} - \lambda)| \int_{\lambda_{kp}/2}^{(\lambda_{kp}+\lambda_{jp})/2} |D_p^T(\lambda - \lambda_{kp})| d\lambda\right) \\
= \; & O\left(n^{-1} \cdot [f(\lambda_{kp}) + f(\lambda_{jp})] \cdot n^{1-p} \lambda_{j-k}^{-p} \int_0^{\lambda_{jp}} |D_p^T(\lambda)| d\lambda\right) \\
= \; & O([f(\lambda_{kp}) + f(\lambda_{jp})] \cdot j^{-p}) = O(f_{jk}\, k^{-p}),
\end{aligned}
$$

with $j - k > j/2$ and $p > d$ again.

Then, for n and ϵ chosen as before, for the following intervals in (24) we have

$$
\left| \int_{-\pi}^{-\epsilon} + \int_{\epsilon}^{\pi} \right| = O\left(\frac{1}{n^{1-2p}\epsilon^6} \int_{\epsilon}^{\pi} [f(\lambda) + f(\lambda_{jp})]\, d\lambda\right) = O([1 + f(\lambda_{jp})] \cdot n^{1-2p}) = O(f_{jk}\, k^{1-p}),
$$

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

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

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

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

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

$$\begin{aligned} E[w^T(\lambda_j)^2] &= \frac{1}{8\pi\sum h_t^2}\int_{-\pi}^{\pi}\exp\left(\frac{-i(n+1)(\lambda_j-\lambda)}{2}-\frac{i(n+1)(\lambda_j+\lambda)}{2}\right) \\ &\quad\times\frac{\sin[n(\lambda_j-\lambda)/2]}{\sin[\lambda/2]}\frac{\sin[n(\lambda_j+\lambda)/2]}{\sin[\lambda/2]}H_j(\lambda)H_j(-\lambda)f_\epsilon(\lambda)d\lambda \\ &= \frac{1}{2\pi\sum h_t^2}\int_{-\pi}^{\pi}\exp\left(-i(n+1)\lambda_j\right) \\ &\quad\times\sin[n(\lambda_j-\lambda)/2]\sin[n(\lambda_j+\lambda)/2]H_j(\lambda)H_j(-\lambda)f(\lambda)d\lambda. \end{aligned}$$

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

$$|H_j(\lambda/n)H_j(-\lambda/n)| = O(n^2 j^{-6}),$$

and the same bound for that integral as for B. The contribution from the other intervals is from Velasco (1996) $O(j^{-4})$, so the results follows.

The obvious modifications apply for the cross terms at the frequencies λ_j and λ_k and the function H_k, exactly as in the non-tapered case, using now parts (c) and (d) of Theorem 2 in Velasco (1996). •

11 Appendix C: Proof of Theorem 9

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

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

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

$$\left|\int_{-\pi}^{-\epsilon}+\int_{\epsilon}^{\pi}\right| \leq \max_{|\lambda|\geq\epsilon}|K_p^T(\lambda_{jp}-\lambda)|\int_{-\varepsilon}^{\pi}|f(\lambda)-f(\lambda_{jp})|d\lambda = O(n^{1-2p}),$$

Therefore, the bound is $O\left((jk)^{d-p}\log j\right)$, following part (c) of the theorem.

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

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

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

10 Appendix B: Proof of Theorem 4

The proof of this theorem follows the lines of the previous one with $p = 3$, though the cosine bell is not of order 3 (but 1), except for the integration of the convolutions around the origin of f. Alternatively, we can use the proof of Theorem 2 in Velasco (1996) for the tapered Fourier transform using the cosine bell for stationary processes, taking special care of those intervals.

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

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

Again we only need to concentrate in the interval $[-\lambda_j/2, \lambda_j/2]$. All the other intervals contribute $O(\min\{j^{-1}, j^{-\alpha}\})$ and the term in $[j/n]^\alpha$ is the bias term due to the normalization. [In Velasco (1996) only the case $\alpha \geq 1$ is considered, but the extension for any $0 < \alpha < 1$ is straightforward].

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

$$\begin{aligned} B &\equiv \frac{|\sin[\lambda_j/2]|^{2d}}{n2\pi\sum h_t^2}\int_{-\pi j}^{\pi j}\frac{f^*(\lambda/n)}{f^*(\lambda_j)}\sin^2[\lambda/2]\,|\sin[\lambda/2n]|^{-2d}H_j^2(\lambda/n)d\lambda \\ &= O\left(n^{-2}j^{2d}\int_{-\pi j}^{\pi j}\sin^2[\lambda/2]\,|\lambda|^{-2d}H_j^2(\lambda/n)d\lambda\right). \end{aligned}$$

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

$$\sin a = a - \frac{1}{6}a^3 + \frac{1}{5!}a^5 - \frac{1}{7}a^7 + O(a^9),$$

so

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

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

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

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

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

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

$$\begin{aligned}
\left| \int_{-\delta_n}^{-\lambda_{jp}/2} \right| &\leq \left[\max_{\lambda_{jp}/2 \leq \lambda \leq \delta_n} f(\lambda) + f(\lambda_{jp}) \right] \int_{-\delta_n}^{-\lambda j/2} K_p^T(\lambda_{jp} - \lambda) d\lambda \\
&= O\left(\delta_n^{-2d} n^{1-2p} \int_{\lambda_{jp}/2}^{\pi} \lambda^{-2p} d\lambda \right) = O\left(\delta_n^{-2d} \cdot j^{1-2p}\right) \\
&= O\left(f(\lambda_{jp}) \cdot (j/n)^{2d} \delta_n^{-2d} \cdot j^{1-2p}\right).
\end{aligned}$$

and,

$$\begin{aligned}
\left| \int_{-\epsilon}^{-\delta_n} \right| &\leq \left[\max_{\delta_n \leq \lambda \leq \epsilon} f(\lambda) + f(\lambda_{jp}) \right] \int_{-\pi}^{-\delta_n} K_p^T(\lambda_{jp} - \lambda) d\lambda \\
&= O(n^{1-2p} \cdot \delta_n^{1-2p}) \;=\; O\left(f(\lambda_{jp}) \cdot (j/n)^{2d} \cdot n^{1-2p} \delta_n^{1-2p}\right).
\end{aligned}$$

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

$$\begin{aligned}
\left| \int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} \right| &\leq \left[\max_{-\lambda_{jp}/2 \leq \lambda \leq \lambda_{jp}/2} K_p^T(\lambda_{jp} - \lambda) \right] \int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} [f(\lambda_{jp}) + f(\lambda)] d\lambda \\
&= O(f(\lambda_{jp}) \cdot j^{1-2p}) \;=\; O(f(\lambda_{jp}) \cdot j^{-1}).
\end{aligned}$$

Finally, using Assumption 3 and the mean value theorem, for $|\lambda^*| \leq \lambda_{jp}/2$

$$\begin{aligned}
\left| \int_{\lambda_{jp}/2}^{3\lambda_{jp}/2} [f(\lambda) - f(\lambda_{jp})] K_p^T(\lambda_{jp} - \lambda) d\lambda \right| &\leq \int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} |\lambda \cdot f'(\lambda^*)| K_p^T(\lambda) d\lambda \\
&= O\left(\lambda_{jp}^{-1-2d} \int_{-\lambda_{jp}/2}^{\lambda_{jp}/2} |\lambda| K_p^T(\lambda) d\lambda \right) \\
&= O(f(\lambda_{jp}) \cdot j^{-1}).
\end{aligned}$$

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

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

so the bounds are

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

For $d < 0$, $p > 1$, the exponent of n is positive, but for d fixed, can be made arbitrarily small with p big enough, and the exponent of j is negative, and can be made as big as we want in absolute value, increasing p as necessary.

For the covariance terms the reasoning is exactly the same as, for example, in Theorem 6, using Lemmas 1 and 2, and considering the intervals with the optimal sequence δ_n to control the leakage from high frequencies. •

References

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

[2] Bloomfield, P., 1976, Fourier Analysis of Time Series: An Introduction (Wiley, New York).

[3] Bloomfield, P., 1991, Time series methods, in: D.V. Hinkley, N. Reid and E.J. Snell, eds., Statistical Theory and Modeling (Chapman and Hall, London) 152-176.

[4] Brillinger, D.R., 1975, Time Series. Data Analysis and Theory (Holt, Rinehart and Winston, New York).

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

[6] Durlauf, S. and P.C.B. Phillips, 1988, Trends versus random walks in time series analysis, Econometrica 56, 1333-1354.

[7] Engle, R.F. and C.W.J. Granger, 1983, Co-integration and error correction: representation, estimation and testing, Econometrica 55, 251-276.

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

[9] Giraitis, L., P.M. Robinson and A. Samarov, 1997, Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long range dependence, Journal of Time Series Analysis 18, 49-60.

[10] Granger, C.W.J., 1996, The typical spectral shape of an economic variable, Econometrica 34, 150-161.

[11] Hannan, E.J., 1970, Multiple Time Series (Wiley, New York).

[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 K.I. Beltrao, 1993, Asymptotics for the low-frequency ordinates of the periodogram of a long-memory time series, Journal of Time Series Analysis 14, 455-472.

[14] Hurvich, C.M. and B.K. Ray, 1995, Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes, Journal of Time Series Analysis 16, 17-42.

[15] Percival, D.B. and A.T. Walden, 1993, Spectral Analysis for Physical Applications (Cambridge University Press, Cambridge).

[16] Robinson, P.M., 1986, On the errors-in-variables problem for time series, Journal of Multivariate Analysis 19, 240-250.

[17] Robinson, P.M., 1992, Log-periodogram regression of time series with long range dependence, Preprint London School of Economics.

[18] Robinson, P.M., 1994a, Rates of convergence and optimal spectral bandwidth for long range dependence, Probability Theory and Related Fields 99, 443-473.

[19] Robinson, P.M., 1994b, Efficient tests of nonstationary hypotheses, Journal of the American Statistical Association 89, 1420-1437.

[20] Robinson, P.M., 1995, Log-periodogram regression of time series with long range dependence, Annals of Statistics 23, 1048-1072.

[21] Velasco, C., 1997, Non-Gaussian log-periodogram regression. Preprint.

[22] Zhurbenko, I.G., 1979, On the efficiency of estimates of a spectral density, Scandinavian Journal of Statistics 6, 49-56.

[23] Zhurbenko, I.G., 1980, On the efficiency of spectral estimates density estimators of a stationary process I, Theory of Probability and its Applications 6, 466-480.

[24] Zhurbenko, I.G., 1982, On the efficiency of spectral estimates density estimators of a stationary process II, Theory of Probability and its Applications 8, 409-419.

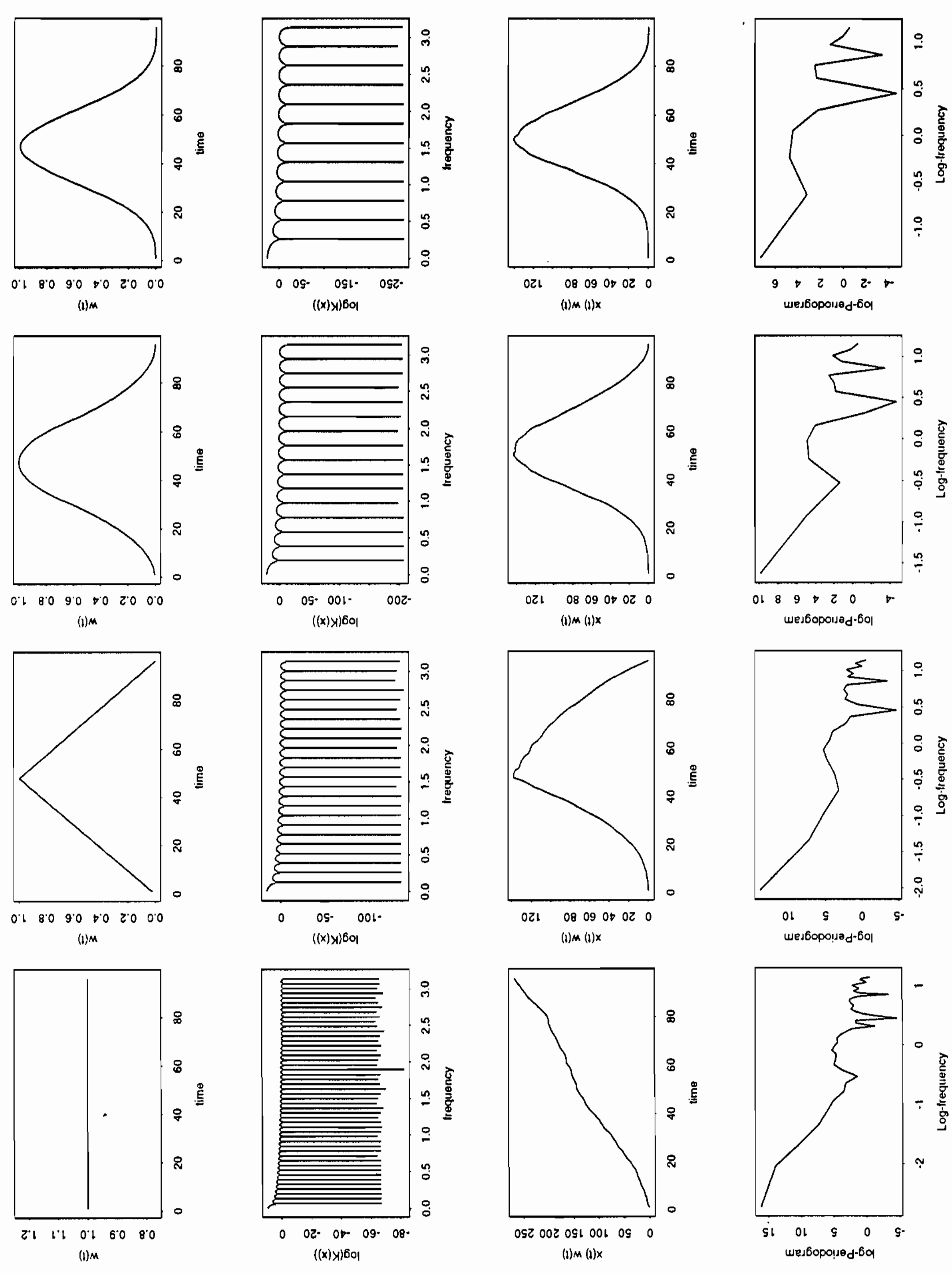

Figure 1: Zhurbenko-Kolmogorov tapers, $p = 1, 2, 3$ and 4

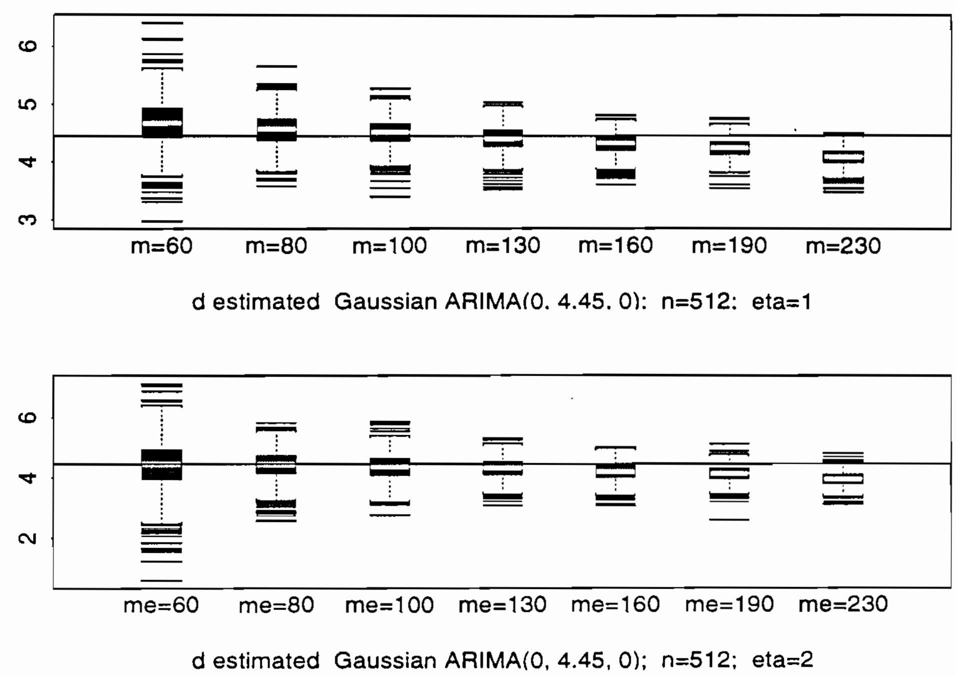

Figure 2: Monte Carlo $\widehat{d}_8^T$, Gaussian ARFIMA(0,4.45,0), n=512, 1000 replications

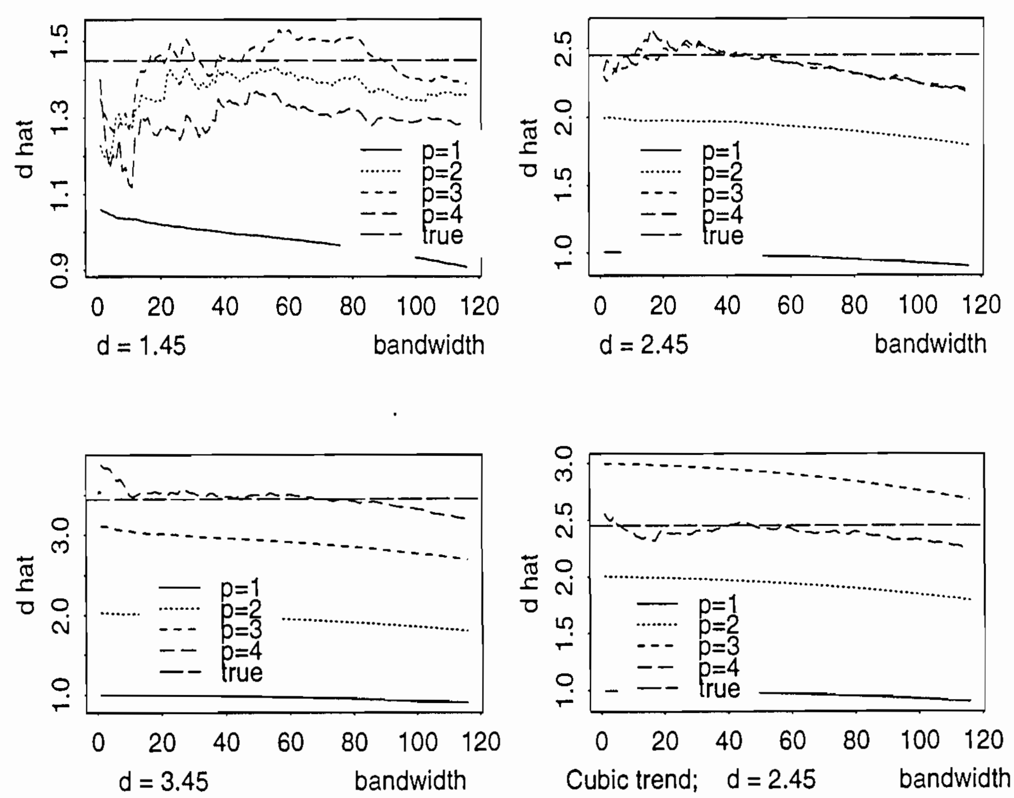

Figure 3: Non-stationary and trending analysis. Gaussian ARFIMA(0,d,0), n=1024

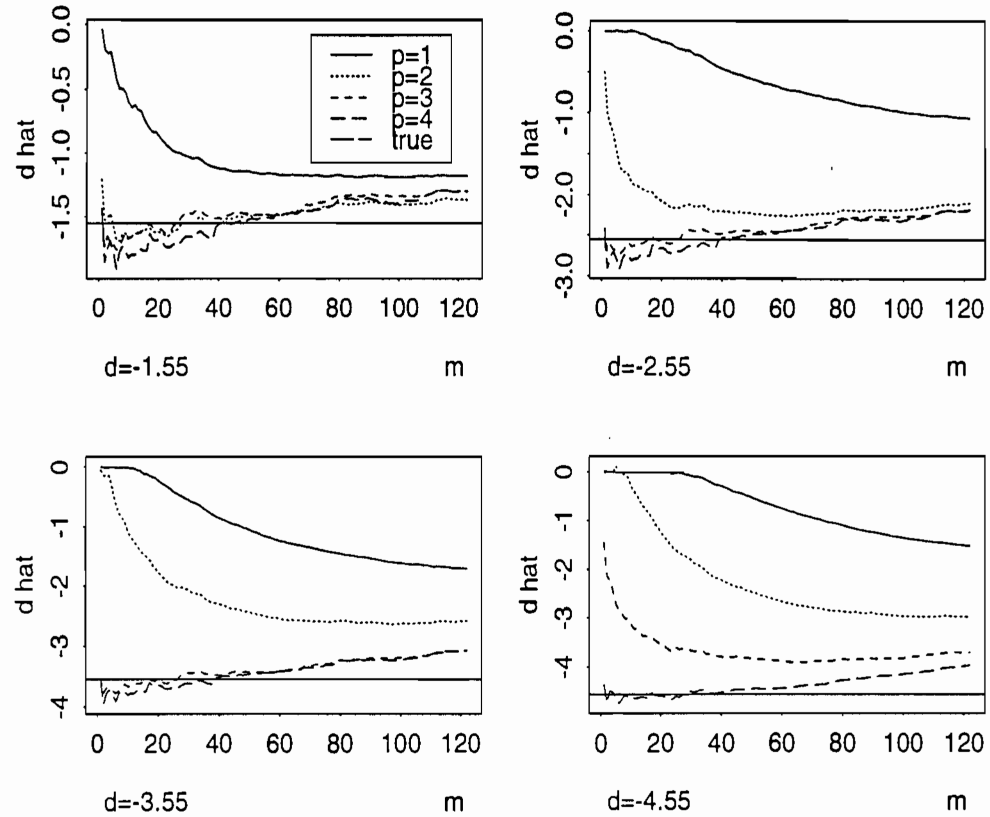

Figure 4: Differencing analysis. Gaussian ARFIMA(0,d,0), n=1024

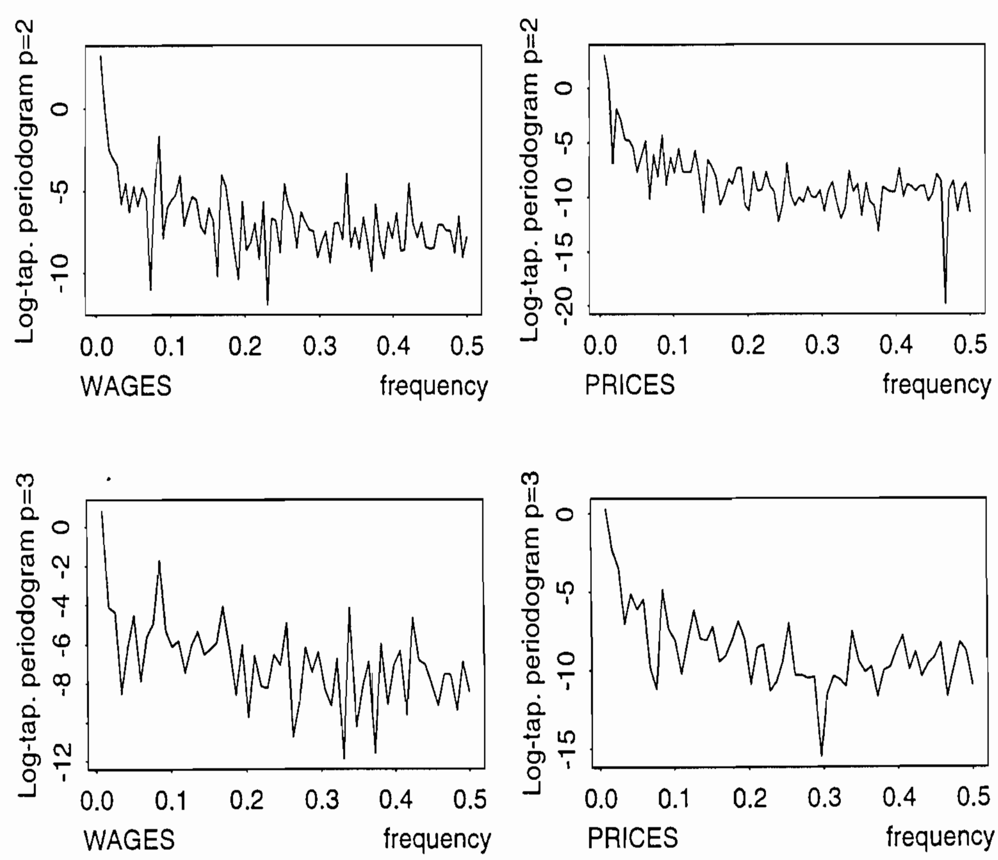

Figure 5: Logarithm of the tapered periodogram, $p = 2, 3$, for Wages and Prices

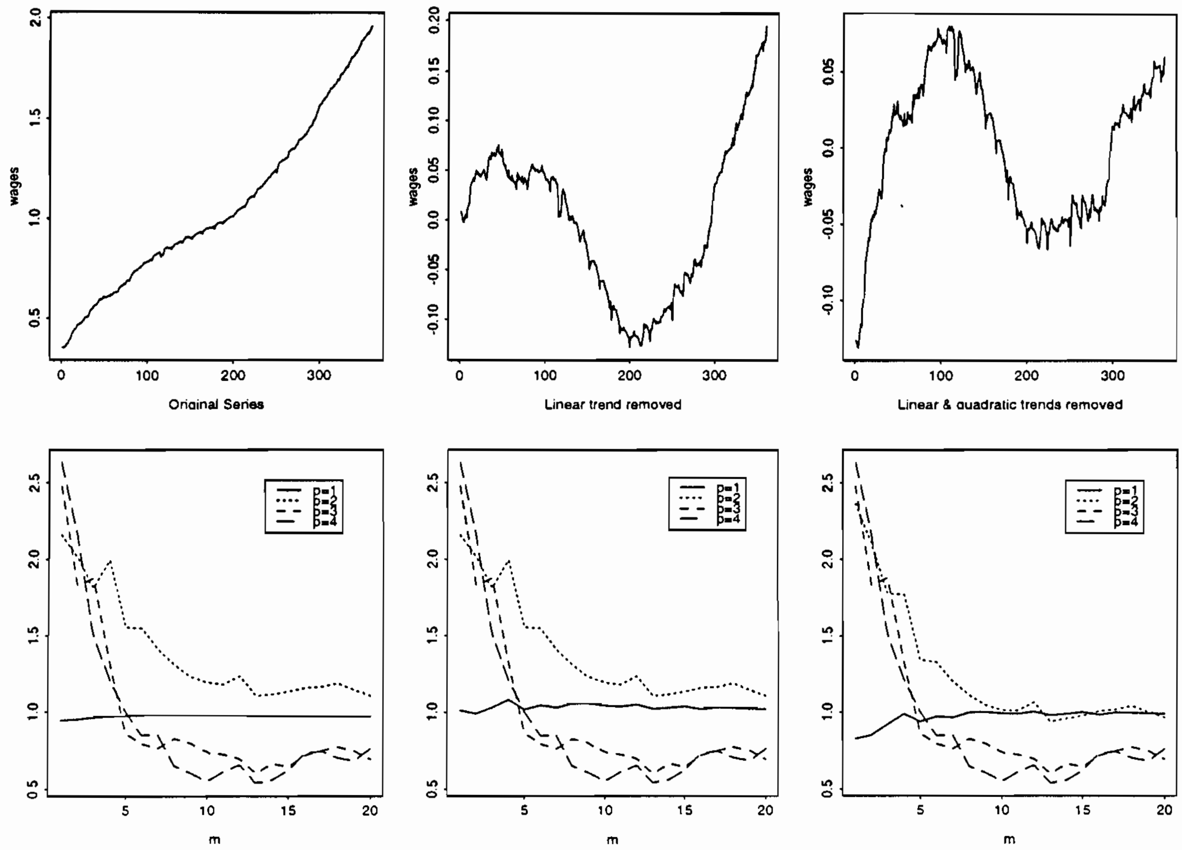

Figure 6: Wages detrending analysis

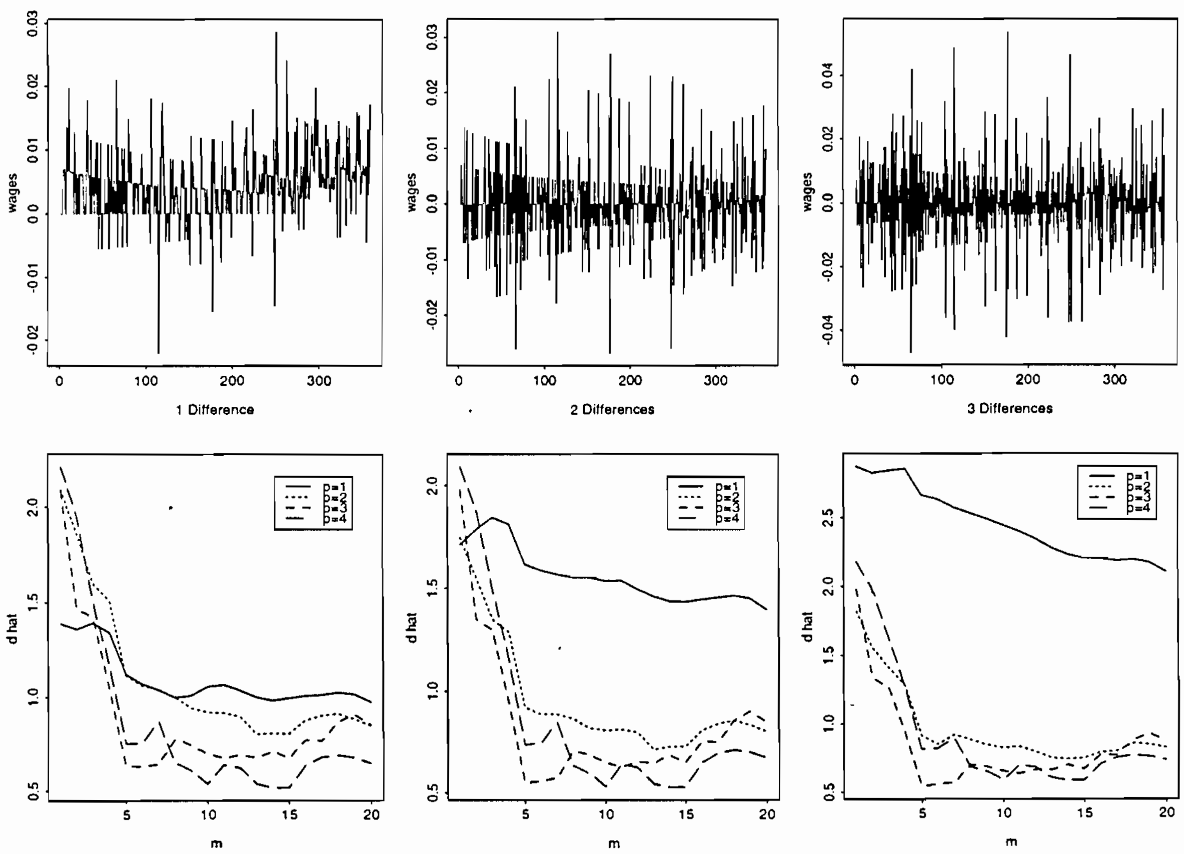

Figure 7: Wages differentiation analysis

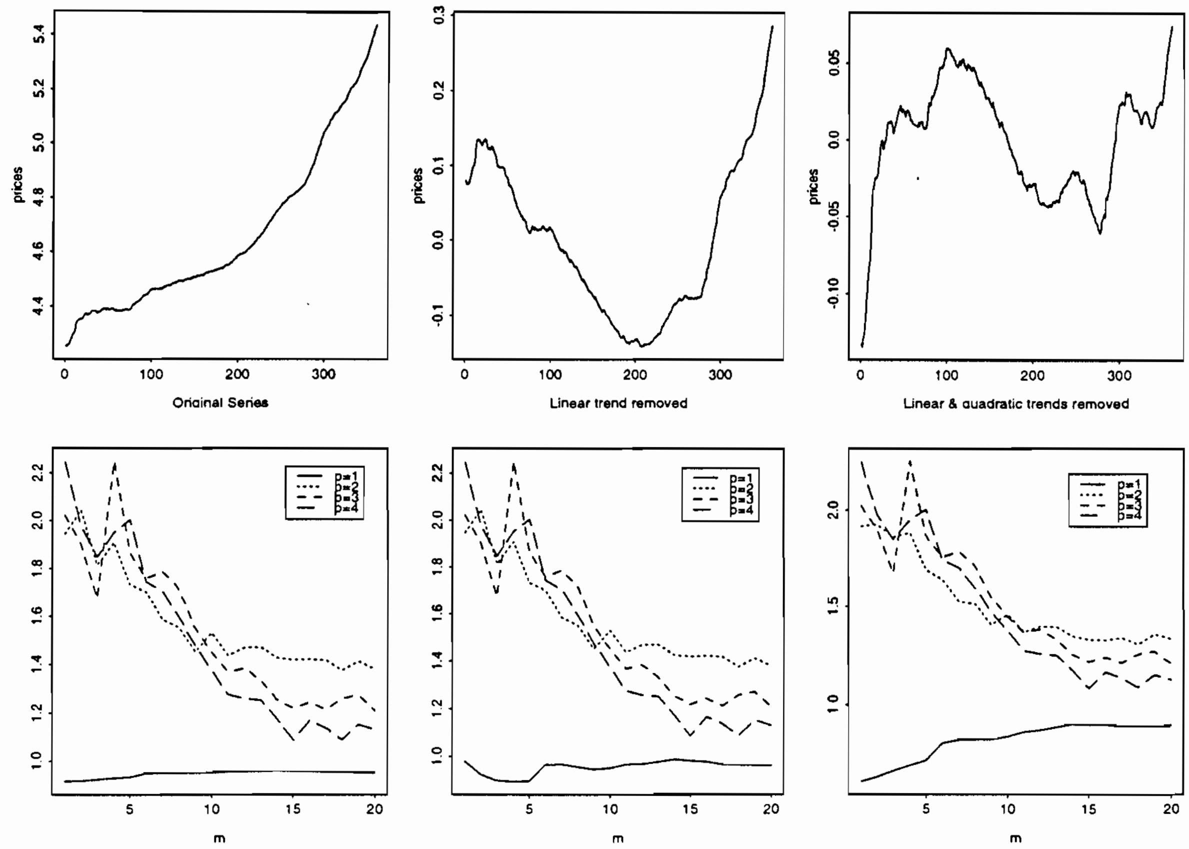

Figure 8: Prices detrending analysis

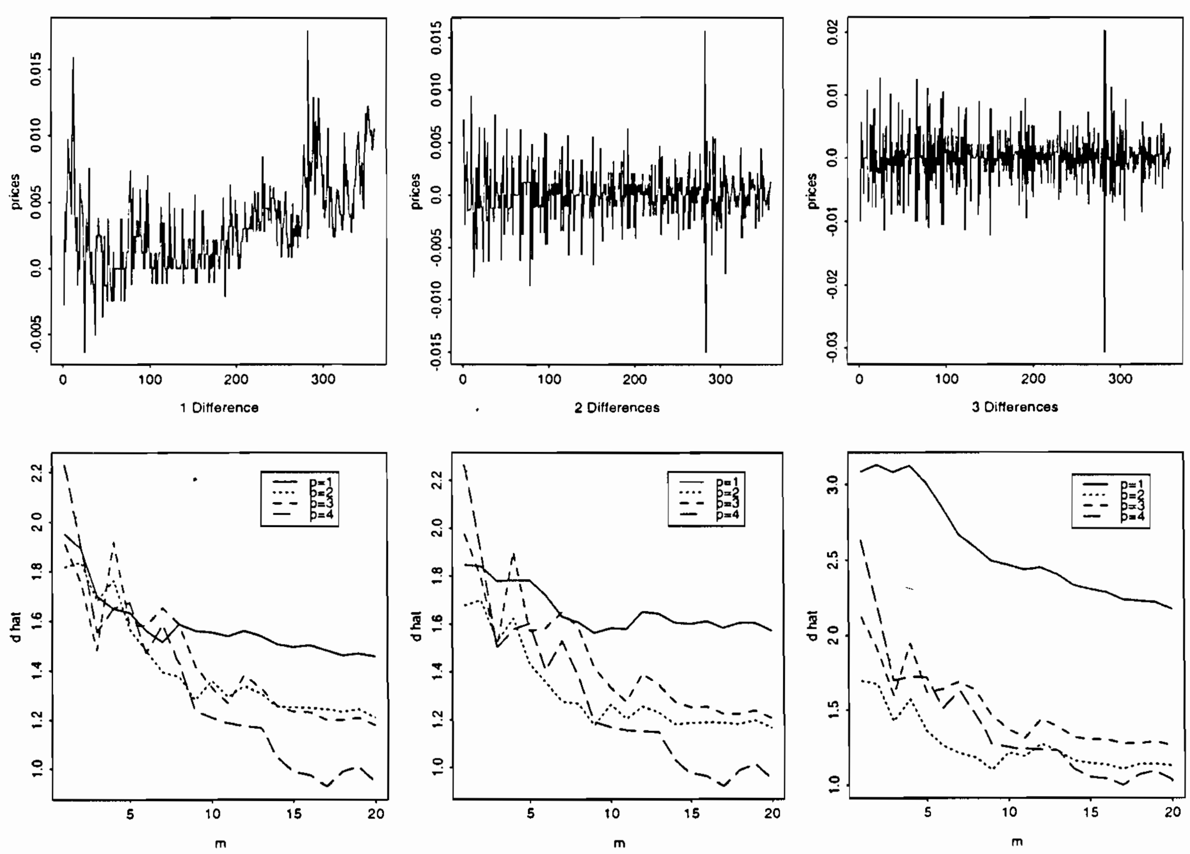

Figure 9: Prices differentiation analysis

WORKING PAPERS 1998

Economics Series

98-05 (01) Alfonso Alba
"The effect of unemployment compesation on the re-employment probability in Spain"

98-06 (02) Michael Jerison
"Nonrepresentative representative consumers"

98-10 (03) Laurence Kranich
"Equality and individual responsibility"

98-11 (04) Laurence Kranich
"Altruism and the political economy of income taxation"

98-13 (05) Jaume García and Sergi Jiménez
"Claim, offer and information in wage bargaining"

Economic History and Institutions Series

98-03 (01) Pedro Fraile and Alvaro Escribano
"The spanish 1898 disaster: the drifts towards national protectionism"

98-04 (02) Juan Carmona and James Simpson
"The "rebassa morta" in catalan viticulture: the rise and decline of a long term sharecropping contract, 1670s-1920s"

Statistics and Econometrics Series

98-01 (01) Jesús Gonzalo and Jean-Yves Pitarakis
"Specification via model selection in vector error correction models"

98-02 (02) Tom Engsted, Jesús Gonzalo and Niels Haldrup
"Testing for multicointegration"

98-07 (03) Felipe Aparicio and Alvaro Escribano
"A nonlinear framework for the characterization of integrated and cointegrated series using induced-order statistics"

98-08 (04) Felipe Aparicio and Alvaro Escribano
"Cointegration testing using first differences of ranges: an exploratory analysis"

98-09 (05) Felipe Aparicio
"Testign linearity with information-theoretic statistics and the bootstrap"

98-12 (06) Uwe Hassler and Francesc Marmol
"Fractional cointegrating regressions in the presence of linear time trends"