Local Whittle estimation of long memory: standard versus bias-reducing techniques*

Javier García-Enríquez[†1] and Javier Hualde [‡2]

[1]Department of Econometrics and Statistics, University of the Basque Country UPV/EHU, Bilbao, Spain

[2]Department of Economics, Universidad Pública de Navarra, Pamplona, Spain

May 16, 2019

Abstract

Frequency domain semiparametric estimation of memory parameters belongs to the standard toolkit of applied time series researchers. These methods are based on a local approximation of the spectral density, which robustifies the estimation methods against misspecification, but induces a loss with respect to the parametric setting, where the spectral density is known up to a finite number of unknown parameters. In particular, standard semiparametric estimators have convergence rates no better than $T^{2/5}$, whereas the rate $T^{1/2}$ is achievable under parametric assumptions. Refinements of the local approximation have been developed by means of bias-reducing techniques, implying that rates arbitrarily close to the parametric one are achievable in the semiparametric setting. Two of these approaches to cover more general settings (including non-stationarity) are extended. A Monte Carlo experiment of finite sample performance is used to assess whether the asymptotic advantages of the bias-reducing methods materialize in better finite sample behaviour.

JEL Classification: C22, C32.

Keywords. Memory parameters; semiparametric estimation; standard versus bias-reducing techniques; fractionally integrated processes.

*The Supplementary Appendix contains the proofs of the propositions given in Sections 2 and 3 of the present paper.

†E-mail address: javier.garcia@ehu.es

‡Corresponding author. E-mail address: javier.hualde@unavarra.es

1 Introduction

Semiparametric estimation of memory parameters has become very popular in the time series literature from theoretical and empirical perspectives. These methods, unlike the rival parametric approach, do not impose any parametric assumption on the short run structure of the model. In particular, for a covariance stationary process with spectral density given by $f(\lambda)$, the semiparametric approach is motivated by the local approximation

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

where G is a positive finite constant, $-1/2 < \delta < 1/2$, and "$\sim$" denotes that the ratio between the left and right sides of the relation tends to one as $\lambda \to 0$. Under (1), f has a pole at $\lambda = 0$ when $\delta > 0$. In view of (1), frequency domain approaches to estimating δ are dominant and two main alternative methods are distinguished: the log periodogram approach, proposed by Geweke and Porter-Hudak (1983) and theoretically justified by Robinson (1995a), and the local Whittle approach, proposed by Künsch (1987) and analysed by Robinson (1995b). See Velasco (2006) for a complete review of these semiparametric procedures and several extensions.

These seminal contributions have been extended in various directions. First, noting that these estimation methods were initially designed to cover covariance stationary processes (with $\delta < 1/2$), Velasco (1999a,b) generalized those results to time series with possibly arbitrarily large memory by means of tapering. Other strategies to deal with the $\delta \geq 1/2$ case include the exact local Whittle approach of Shimotsu and Phillips (2005), which is based on fractional differences of the observed data (instead of the traditional approach of "whitening" the periodogram) and permits the domain of δ to have (at most) a 9/2 width. An alternative approach is the extended local Whittle estimator of Abadir, Distaso and Giraitis (2007), which is based on the so-called fully extended discrete Fourier transform and periodogram.

Second, a different strand of the literature has proposed variants of the standard semiparametric methods to deal with perturbed fractional processes, where the observable series is composed of a long memory process contaminated by an additive noise term. These include Perron and Qu (2010), Frederiksen, Nielsen and Nielsen (2012), McCloskey and Perron (2013) and Hou and Perron (2014), who extended earlier works by Deo and Hurvich (2001), Hurvich and Ray (2003), Arteche (2004, 2006) and Hurvich, Moulines and Soulier (2005).

Finally, a different type of extension focuses on obtaining estimators of δ with improved properties. This is motivated by the fact that the standard semiparametric estimators of memory parameters have convergence rates no better than $T^{2/5}$ (see Giraitis, Robinson and Samarov, 1997), where T denotes sample size throughout. This bound to the convergence rate reflects a natural loss with respect to the parametric setting, where the rate $T^{1/2}$ is achievable. It should also be noted that, depending on the smoothness of f around frequency zero, these semiparametric estimators can have a substantially slower rate than $T^{2/5}$. In order to improve the rate of convergence of these semiparametric estimators, different bias-reducing techniques have been proposed in the literature, which exploit the possible smoothness of f around frequency 0, and impose a richer structure to the approximation (1). This includes Andrews and Guggenberger (2003), Andrews and Sun (2004) (AS hereinafter), who extended the log periodogram and the local Whittle approaches, respectively, by means of local polynomials. In a similar fashion, Robinson and Henry (2003) (RH hereinafter) proposed a very general M-estimation procedure, nesting both the log periodogram and local Whittle approaches, employing higher-order kernels. Similar improvements can be achieved by a broadband approach (instead of a local one), although this requires global smoothing conditions of f outside frequency zero. This strategy was pursued by Moulines and Soulier (1999) and Hurvich and Brodsky (2001).

Importantly, these bias-reducing procedures might lead to memory parameter estimators enjoying convergence rates arbitrarily close to the parametric one. Given that no parametric assumptions are needed, this is a remarkable result with potentially important implications for applied work. Thus, in order to decide whether the use of these techniques should be encouraged in practice, a relevant question is to analyse if these theoretical asymptotic advantages translate to a better finite sample behaviour. This paper evaluates this issue by means of an extensive Monte Carlo experiment. Previous Monte Carlo analyses involving semiparametric memory estimation include Hauser (1997), with particular emphasis on testing, and the very complete study of Nielsen and Frederiksen (2005). This latter work includes some evidence of the behaviour of biased-reducing techniques, in particular results for the simplest version of the estimators in Andrews and Guggenberger (2003) and AS are given.

Building on these previous works, we here offer a richer evidence on the performance of the bias-reducing approaches. In particular we focus on the RH and AS proposals, which are inspired by the local Whittle principle. As it is well known, this approach leads to more efficient estimators than the log-periodogram alternative, and, in addition, unlike the broadband approaches, does not require global smoothness assumptions. Furthermore, providing further evidence about the performance of higher order kernels in the present setting appears to be particularly relevant because of this technique involves local averaging with negative weights, which could have a severe distortive effect. In fact, as will be seen, our study reflects some of the complications associated to the higher-order kernels. Formally, this materializes in the presence of a component (denoted V_q by RH) in the denominator of the asymptotic bias and variance of the RH estimators (see, e.g., equations (3.6), (3.7) in RH), which, depending on the order of the kernel (q), could take values very close to zero (see Table 1 in RH). This might have a dramatic effect on finite samples and our experiment sheds light on this potential problem, which appears to be extremely serious for $q = 3$. That is even the case for sample sizes which are much larger than those typically employed in many empirical analyses.

Incidentally, our paper also makes some theoretical contributions with independent interest. First, the properties of the bias-reducing techniques for memory estimation have been just derived for generic covariance stationary and invertible processes. Given the strong evidence of non-stationarity in many different contexts, this limits their practical application. Thus, it seems desirable to extend these methods to cover the very relevant non-stationary case as well. Although other possibilities could have been pursued, given that our focus is not mainly theoretical, we found that the simplest strategy to extend the bias-reducing estimators to the non-stationary setting is the use of tapering. This technique can nicely deal with arbitrarily large memory parameters and, in addition, it can remove time polynomials so these deterministic terms do not interfere in the estimation. Specifically, the tapered estimators are invariant to polynomial time trends if tapers with high enough order are employed. There is, however, a price to pay in the form of a variance increase due to the correlation of the tapered periodogram ordinates and likewise to the order of the taper.

A sensible alternative to tapering without incurring this variance increase is to develop bias-reducing techniques within the exact local Whittle framework of Shimotsu and Phillips (2005). This is a promising avenue for future research, but it should be noted that, even in this setting, our results appear to be useful: the exact local Whittle estimator assumes that the mean of the process is known and this is a serious drawback. This problem has been solved by Shimotsu (2010), who proposed an extension of the exact local Whittle approach to deal with the case of unknown mean and polynomial time trend. In particular, Shimotsu (2010) proposes a two-step local Whittle estimator which is efficient for any value of the integration order in the interval $(-1/2, 7/4)$. It should be noted, however, that this estimator also relies on tapering in an indirect way, because this two-step estimation method is based on a consistent (with a certain rate) first step estimator. This is precisely the tapered local Whittle estimator which we extend in the present paper. Thus, evaluating improved methods for first step estimation, as we do in the present paper, appears to be a useful exercise even from the point of view of exact local Whittle estimation.

As a second theoretical contribution, we deal with the issue of the rival definitions of non-stationary fractionally integrated processes, namely the so-called Type I and II (see, Robinson, 2005). In particular, we show that the properties of our biased-reducing tapered estimators based on either Type I or Type II processes enjoy identical first order asymptotic properties.

Our last theoretical contribution addresses with a difficulty associated to RH's proposal. Specifically, RH's estimator does not in general correspond to a global minimum of a suitable objective function, so consistency needs to be assumed. To circumvent this problem, we propose a two-step estimator, which updates Velasco's (1999a) local Whittle tapered estimator.

Finally, note that our theoretical results appear to be of special importance in fractional cointegration, a field which has recently attracted substantial attention from time series researchers. Estimators of the cointegrating relation with optimal asymptotic properties require, in general, the estimation of the memory parameters driving the long-run behaviour of the series included in the model subject of study. As justified by Hualde and Robinson (2006, 2010), these estimators should satisfy certain convergence requirements, and the bias-reducing techniques serve this purpose.

The theoretical content of the paper is concentrated in Sections 2 and 3, which cover the extensions of RH and AS, respectively. In Section 4 we present the finite sample results which compare the standard versus the bias-reducing techniques. Due to space restrictions, the proofs of the theoretical results are given in a Supplementary Appendix.

2 Tapered higher-order kernel local Whittle estimation of memory parameters

We introduce some notation before moving on to our proposed estimators. We emphasize, throughout, the use of fractional processes, which have been stressed in the literature to describe non-stationary (and indeed also stationary) processes. Here, denoting by u_t a covariance stationary weak dependent process (with finite and bounded away from zero spectral density at frequency zero), we say that a process v_t is Type I fractionally integrated of order δ if, setting $r = [\delta + 1/2]$, where $[\cdot]$ denotes integer part,

$$v_t = \Delta^{-r}\left\{\psi_t 1\left(t > 0\right)\right\}, \quad \psi_t = \Delta^{r-\delta} u_t = \sum_{j=0}^{\infty} \pi_j\left(\delta - r\right) u_{t-j}, \tag{2}$$

where $1(\cdot)$ is the indicator function, and we employ the difference operator $\Delta = 1 - L$, where L is the lag operator, and formally, for any real α, $\alpha \neq -1, -2, ..,$

$$(1-z)^{-\alpha} = \sum_{j=0}^{\infty} \pi_j(\alpha) z^j, \quad \pi_j(\alpha) = \frac{\Gamma(j+\alpha)}{\Gamma(\alpha)\Gamma(j+1)},$$

with Γ denoting the gamma function. Note that $\delta - r < 1/2$, so that ψ_t is well defined in mean square sense. Alternatively, the Type II fractionally integrated process has been defined as

$$\widetilde{v}_t = \Delta^{-\delta} \left\{ u_t 1 \left(t > 0 \right) \right) \} = \sum_{j=0}^{t-1} \pi_j \left(\delta \right) u_{t-j}. \tag{3}$$

When $r = 0$, v_t is covariance stationary, whereas $\widetilde{v}_t$ is non-stationary for any value of δ, although asymptotically stationary when $\delta < 1/2$. When $\delta \geq 1/2$ (so $r > 0$ in the Type I definition), v_t and $\widetilde{v}_t$ are purely non-stationary, and display in general different asymptotic properties. In fact, it can be shown that both processes, properly normalized, converge to different versions of the fractional Brownian motion (see Marinucci and Robinson, 1999, for further details about these types of processes).

Within this framework, the bias-reduced versions of the local Whittle estimators just cover v_t in case $r = 0$, and therefore, as mentioned before, we will extend these methods to cover the Type II definition and also situations where δ might be arbitrarily large. First, the RH method employs higher-order kernels to obtain "improved" estimators of memory parameters. Their M-estimation procedure is very general, so we focus on extending a particular case. As anticipated, this extension is mainly based on the use of tapering (see Velasco 1999a,b), which alleviates the problem of periodogram bias due to the leakage from zero frequency when the process is nonstationary, and on a two-step approach, which avoids assuming consistency. Before presenting our method, we introduce some regularity conditions.

ASSUMPTION 1. The process u_t, $t = 0, \pm 1, ...$, in (2), (3), has representation

$$u_t = b\left(L\right) \xi_t, \quad b\left(z\right) = \sum_{j=0}^{\infty} b_j z^j,$$

where

(i)

$$\left| b\left(z\right) \right| \neq 0, \ \ \left| z \right| \leq 1;$$

(ii) $b\left(e^{i\lambda}\right)$ is s-times differentiable in some neighbourhoods of zero with derivative of

order $s \geq 1$ in $Lip(\eta)$, $0 < \eta \leq 1$;

(iii) $E(\xi_t | F_{t-1}) = 0$, $E(\xi_t^2 | F_{t-1}) = 1$, $E(\xi_t^3 | F_{t-1}) = \mu_3$, $E(\xi_t^4 | F_{t-1}) = \mu_4$ almost surely, $t = 0, \pm 1, ...$, where μ_3, μ_4, are finite constants and F_t is the σ-field of events generated by ξ_s, $s \leq t$;

(iv) there exists a random variable ξ such that $E\xi^2 < \infty$, and for all $\varkappa > 0$ and some $K > 0$, $P(|\xi_t| > \varkappa) \leq KP(|\xi| > \varkappa)$.

Assumption 1 with $s \geq 1$ implies that Assumption 8 in Velasco (1999a) holds for ψ_t, with $\beta = \min\{s + \eta, 2\}$. As Velasco (1999a) acknowledges, he needs to use $\beta > 1$ in some of his theorems, as one cannot resort to the second moments of the tapered periodogram (see (4) below) as is done in the non-tapered case. Given that we used some of Velasco's (1999a) results in the proofs of the propositions below, there is the need to assume $s \geq 1$, $\eta > 0$.

Denoting the spectral density of u_t by $f_u(\lambda)$, the smoothness condition given in (ii) translates directly to $f_u(0)$. Defining $h(\lambda) = \left(2\sin(\lambda/2)\lambda^{-1}\right)^{-2\delta} f_u(\lambda)$, for any s for which Assumption 1 is satisfied, setting $q = [s/2]$, this condition implies that

$$h(\lambda) = f_u(0) + \sum_{i=0}^{q} \frac{h_i \lambda^{2i}}{(2i)!} + O\left(\lambda^{s+\eta}\right) \text{ as } \lambda \to 0,$$

where $h_0 = 0$ and for $i \geq 1$, h_i represents the $2i$-th derivative of $h(\lambda)$ at $\lambda = 0$. As established in RH, this result can be exploited by the use of a higher order kernel to reduce asymptotic bias when $q \geq 2$ (or equivalently $s \geq 4$), so we will concentrate on this case. Note that if $q = 1$, we are in the situation covered by Robinson (1995a,b) and Velasco (1999a,b), where the maximum rate of convergence achievable is $T^{2/5}$. For $s = 1$, following these references, our Assumption 1 permits the rate $T^{(1+\eta)/(3+2\eta)}$.

Defining a taper $\{g_t\}_{t=1}^T$ of order p as in Velasco (1999a,b), and a sequence ζ_t, the discrete Fourier transform and periodogram of the tapered sequence $g_t\zeta_t$ are

$$w_\zeta^p(\lambda) = \left(2\pi \sum_{t=1}^T g_t^2\right)^{-1/2} \sum_{t=1}^T g_t \zeta_t e^{it\lambda}, \; I_\zeta^p(\lambda) = \left|w_\zeta^p(\lambda)\right|^2. \tag{4}$$

For integer q, we introduce a real function $k_q(u)$, $q \geq 2$, $0 \leq u \leq 1$, satisfying

ASSUMPTION 2. $k_q(u)$, $0 \leq u \leq 1$ is a boundedly differentiable function such that $\int_0^1 k_q(u)\, du = 1$, and defining $U_{iq} = \int_0^1 (\log u + 1) u^{2i} k_q(u)\, du$, we have $U_{iq} = 0$, $0 \leq i \leq q-1$; $\quad U_{qq} \neq 0$.

RH described $k_q(u)$ as a higher-order kernel and proposed a particular characterization given by $k_q^*(u) = \sum_{j=0}^{q} \alpha_j u^{2j}$, for choices α_j given in (5.4)-(5.6) of RH. Following RH, for an integer m to be described subsequently such that m/p is integer, for suitable $q \geq 2$, $k_q(u)$, we define for an arbitrary sequence θ_t

$$S_\theta^p(c) = \frac{2m \sum' b_{q,j} \lambda_j^{2c} I_\theta^p(\lambda_j)}{\sum' k_{q,j} \lambda_j^{2c} I_\theta^p(\lambda_j)} \quad \text{and} \quad H_\theta^p(c) = \frac{4m\left(G_{2,\theta}^p(c)\, G_{0,\theta}^p(c) - \left(G_{1,\theta}^p(c)\right)^2\right)}{\left(G_{0,\theta}^p(c)\right)^2},$$

where $\lambda_j = 2\pi j/T$ are the Fourier frequencies, $\sum' = \sum_{j=p,2p,\ldots}^{m}$,

$$b_{q,j} = k_{q,j}\nu_{q,j},\ k_{q,j} = k_q(j/m),\ \nu_{q,j} = \log\lambda_j - \frac{\sum' k_{q,j}\log\lambda_j}{\sum' k_{q,j}}$$

and

$$G_{g,\theta}^p(c) = \frac{p}{m}\sum' k_{q,j}(\log\lambda_j)^g \lambda_j^{2c} I_\theta^p(\lambda_j),\ g = 0, 1, 2.$$

We now present our estimators of δ. Denote by $\overline{\delta}_G$, $\widetilde{\delta}_G$, the tapered local Whittle estimators based on processes v_t, $\widetilde{v}_t$, respectively, which optimize over the interval $\Theta = [\nabla_1, \nabla_2]$ the loss function of Velasco (1999a). Denoting by M the bandwidth employed in the computation of $\overline{\delta}_G$, $\widetilde{\delta}_G$ and assuming

$$\frac{1}{M} + \frac{M^{1+2\beta}(\log M)^2}{T^{2\beta}} \to 0, \text{ as } T \to \infty, \tag{5}$$

Velasco (1999a) established that $\overline{\delta}_G$ is $M^{1/2}$-consistent. We then define our estimators $\overline{\delta}_{RH}$, $\widetilde{\delta}_{RH}$ of δ based on v_t, $\widetilde{v}_t$, respectively, as

$$\overline{\delta}_{RH} = \overline{\delta}_G - \frac{S_v^p(\overline{\delta}_G)}{H_v^p(\overline{\delta}_G)} \quad \text{and} \quad \widetilde{\delta}_{RH} = \widetilde{\delta}_G - \frac{S_{\widetilde{v}}^p\left(\widetilde{\delta}_G\right)}{H_{\widetilde{v}}^p\left(\widetilde{\delta}_G\right)}.$$

Note that defining $q_\theta^p(c) = \frac{p}{m}\sum' b_{q,j}\left(I_\theta^p(\lambda_j)\lambda_j^{2c} - 1\right)$, $\overline{\delta}_{RH}$, $\widetilde{\delta}_{RH}$ would have identical first order asymptotic properties to the zeroes of $q_v^p(c)$, $q_{\widetilde{v}}^p(c)$, which are closest to $\overline{\delta}_G$, $\widetilde{\delta}_G$, respectively, which correspond to the qth-order kernel M-estimator proposed by RH for the choices $J = 1$, $g(\lambda) = \lambda$, $\psi(z) = \psi_1(z)$. Thus these estimators are higher-order kernel versions of the local Whittle estimators of Künsch (1987) and Robinson (1995b), with corresponding loss functions $Q_v^p(c)$, $Q_{\widetilde{v}}^p(c)$, where

$$Q_\theta^p(c) = m\left(\log G_\theta^p(c) - 2c\frac{\sum' k_{q,j}\log\lambda_j}{\sum' k_{q,j}}\right) \text{and} \quad G_\theta^p(c) = \frac{\sum' k_{q,j}\lambda_j^{2c} I_\theta^p(\lambda_j)}{\sum' k_{q,j}},$$

assuming the estimators do not fall on the boundary of Θ. Before presenting our results, we introduce an additional regularity condition related to the different bandwidths employed in the estimation, namely M and m.

ASSUMPTION 3. Let M and m be power roots of T such that (5) holds and, as $T \to \infty$,

$$\frac{m^{1/2}}{M} \to 0 \quad \text{and} \quad m = O\left(T^{4q/(4q+1)}\right). \tag{6}$$

The second condition in (6) (taken from RH) imposes the maximum rate at which the bandwidth m can grow. Defining

$$\Phi = \lim_{T\to\infty} \left(\sum\nolimits_{t=1}^{T} g_t^2\right)^{-2} \sum\nolimits_{k=0,p,2p,..}^{T-p} \left(\sum\nolimits_{t=1}^{T} g_t^2 \cos\left(t\lambda_k\right)\right)^2,$$
$$V_q = \int_0^1 \left(\log u + 1\right)^2 k_q\left(u\right) du \ \text{ and } \ W_q = \int_0^1 \left(\log u + 1\right)^2 k_q^2\left(u\right) du,$$

and denoting by δ_{RH}^* either $\overline{\delta}_{RH}$ or $\widetilde{\delta}_{RH}$, we establish the following result.

PROPOSITION 1. *Under Assumptions 1-3,* $\delta \in (\nabla_1, \nabla_2)$, $\nabla_1 > -1/2$, $p > \max\{1, \nabla_2, \delta + 1/2\}$, $q \geq 2$, then

$$m^{1/2}\left(\delta_{RH}^* - \delta\right) + \frac{(2\pi)^{2q} U_{qq} h_q}{2\,(2q)!\, f_u\left(0\right) V_q} \frac{m^{2q+1/2}}{T^{2q}} \to_d N\left(0, \frac{p\Phi W_q}{4V_q^2}\right). \tag{7}$$

Proposition 1 is justified in the Supplementary Appendix. One of the main implications of (7) is that letting m grow at rate $T^{4q/(4q+1)}$, the convergence rate of our estimators is $T^{2q/(4q+1)}$, which can be arbitrarily close to the parametric rate $T^{1/2}$ for q (and thus s) large enough. Note also that for the suggested choice of m the bias term in (7) has exact rate $O\left(1\right)$, while (6) prevents this bias from dominating.

3 Tapered local polynomial Whittle estimation of memory parameters

Similarly, we here propose an extension of the estimators in AS, obtaining similar achievements to those in our extension of RH. Thus, Proposition 2 below shows that similar results to those in AS apply to Type I or II fractional processes of arbitrarily large memory. Following AS, we consider the tapered local polynomial Whittle log-likelihood based on process ζ_t (for the particular choice of polynomial order $2q$)

$$Q^p_{\zeta,q}(c,G,\gamma) = \frac{p}{m}\sum_{j=p,2p,..}^{m}\left\{\log\left[G\lambda_j^{-2c}\exp\left(-\kappa_q(\lambda_j,\gamma)\right)\right] + \frac{I^p_\zeta(\lambda_j)}{G\lambda_j^{-2c}\exp\left(-\kappa_q(\lambda_j,\gamma)\right)}\right\},$$

where $\kappa_q(\lambda_j,\gamma) = \sum_{k=1}^{q}\gamma_k\lambda_j^{2k}$, $\gamma = (\gamma_1,...,\gamma_q)'$, and different tapered versions of the LPW-FOC estimators of (δ,ϕ), where ϕ is a $q\times 1$ vector with ith component given by

$$-\frac{1}{(2i)!}\frac{d^{2i}}{d\lambda^{2i}}\log h(\lambda)|_{\lambda=0}.$$

We denote these estimators by $\left(\overline{\delta}_{AS},\overline{\phi}'_{AS}\right)'$, $\left(\widetilde{\delta}_{AS},\widetilde{\phi}'_{AS}\right)'$, depending on whether $Q^p_{v,q}(c,G,\gamma)$ or $Q^p_{\widetilde{v},q}(c,G,\gamma)$ is used, respectively. Before presenting our results, we introduce a series of further regularity conditions.

ASSUMPTION 4

$$\frac{m^{2q+\frac{1}{2}}}{T^{2q}}\to\infty \text{ and } \frac{m^{s+\eta+\frac{1}{2}}}{T^{s+\eta}} = O(1),$$

as $T\to\infty$.

ASSUMPTION 5. ϕ belongs to the interior of a compact and convex set Ξ.

Note that Assumptions 1, 4 and 5, imply that Assumptions 1-5 in AS hold with their order of smoothness, given in our case by $s+\eta$, and with their polynomial order ($2r$ in their notation) given by $2q$. Note also that in AS' notation $\alpha(\lambda) = (1-e^{i\lambda})^{r-\delta}f_u^{1/2}(\lambda)$ in our framework, which satisfies part (c) of their Assumption 3. Denoting by δ^*_{AS} either $\overline{\delta}_{AS}$ or $\widetilde{\delta}_{AS}$, and by ϕ^*_{AS} either $\overline{\phi}_{AS}$ or $\widetilde{\phi}_{AS}$, respectively, we establish the following result.

PROPOSITION 2. *Under Assumptions 1, 4, 5, $\delta\in(\nabla_1,\nabla_2)$, $\nabla_1 > -1/2$, $p > \max\{1,\nabla_2,\delta+1/2\}$, we obtain*

$$\begin{pmatrix} m^{\frac{1}{2}}(\delta^*_{AS}-\delta) \\ m^{\frac{1}{2}}diag\left(\lambda_m^2,...,\lambda_m^{2q}\right)(\phi^*_{AS}-\phi)\end{pmatrix} - \Omega_q^{-1}\nu_T(q,s)\to_d N\left(0,p\Phi\Omega_q^{-1}\right),$$

where Ω_q, $\nu_T(q,s)$ are given in (4.4), (4.6) in AS.

Proposition 2 is justified in the Supplementary Appendix. This result is basically the same as in AS simply taking into account the effect of tapering given by $p\Phi$. Note that Assumption 4 holds for a bandwidth m^* which grows at rate $T^{2s/(1+2s)}$, which given our smoothness conditions, is not exactly the "optimal" bandwidth proposed by AS (say m', which grows at rate $T^{2(s+\eta)/(1+2(s+\eta))}$), but both are similar when s is large. We found it more informative to propose a bandwidth depending on the number of existing derivatives. Of course, the relevance of this choice is that should s be arbitrarily large, we obtain a rate of convergence $T^{s/(1+2s)}$ for our estimator of the order of integration based on v_t or $\widetilde{v}_t$, which could be arbitrarily close to the parametric rate $T^{1/2}$. Note also that for our choice of bandwidth m^*, the exact rate of $\nu_T(q,s)$ is $(m^*)^{s+\eta+1/2}\,T^{-(s+\eta)} = T^{-\eta/(1+2s)} = o(1)$ as $T \to \infty$. Finally, note that m^* is not allowed by (6) when s is odd (although it is when s is even), and in this case the rate $T^{s/(1+2s)}$ is larger than $T^{2q/(4q+1)}$ (although both approximate as s increases).

4 Finite sample performance

We perform a Monte Carlo experiment in order to understand the extent to which the bias-reducing techniques are worth employing in finite samples. The different estimators are applied to an observable process x_t, which is generated according to four different mechanisms (denoted as Models I, II, III and IV). For the first three models, $x_t = \widetilde{v}_t$ (see (3)), for three different error input processes u_t corresponding to

$$u_t = \zeta u_{t-1} + \varepsilon_t,$$

where ε_t is a Gaussian independent and identically distributed process with $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = 1$, and ζ takes three different values $(0, 0.5, 0.9)$, which correspond to Models I, II, III, respectively. The latter two models cover the autoregressive case, where a richer approximation of the spectral density might be exploited in estimation to improve results, so the bias-reducing techniques might offer advantages. For Model IV,

$$x_t = \widetilde{v}_t + w_t,$$

where $\widetilde{v}_t$ is generated as in Model I and w_t is a Gaussian independent and identically distributed process, independent of ε_t, such that $E(w_t) = 0$, $Var(w_t) = 1$. This model corresponds to a perturbed (or contaminated) fractionally integrated process. This setting might again be favourable to the bias-reducing methods.

The memory parameter δ is allowed to take two different values, namely $\delta = 0.3, 1.3$, where x_t is asymptotically stationary or purely nonstationary, respectively. We present results for various sample sizes $T = 256$, 1024, 3072, 9216. Some of these sample sizes are larger than the typical ones employed in applications but, noting that the potential advantages of the reducing-bias estimators are asymptotic, it is illustrative to evaluate the behaviour of the different estimation techniques for large T. The RH estimator was computed using the kernel $k_q^*(u)$ described in Section 2 and for both RH and AS three different bandwidth choices (m_1, m_2, m_3) were employed. In all cases $m_j = 2m_{j-1}$, $j = 2, 3$, where for $T = 256$, $m_1 = 10, 16, 23, 30$, for $q = 1, 2, 3, 4$, respectively; for $T = 1024$, $m_1 = 25, 50, 75, 100$, for $q = 1, 2, 3, 4$, respectively; for $T = 3072$, $m_1 = 55, 135, 215, 295$; finally, for $T = 9216$, $m_1 = 145, 390, 635, 880$. As Assumption 3 indicates, larger q implies that larger bandwidths can be used, and we accommodated this possibility in our experiment.

RH and AS were computed as described in Sections 2, 3, respectively, with the optimizing intervals being set as $[-.5, 2]$ in all cases. Throughout RH with $q = 1$ will refer to Velasco's (1999a) estimator (which, hereinafter, will be denoted as LW), noting also that LW takes the role of the first-step estimator when computing RH for $q > 1$. Whenever tapering is employed, we use the Zhurbenko taper with $p = 2$. Other possibilities like the Parzen taper could have been used (see Alekseev, 1996, for additional examples), but we believe that different taper choices would not alter our results significantly. We present results for Monte Carlo bias, standard deviation (SD) and coverage probabilities, all computed across 5000 replications.

Table 1 presents bias results for the untapered RH estimator in the stationary case. The first noticeable feature is the extremely unsatisfactory behaviour of RH when $q = 3$. This very poor performance also occurs in other settings which cover tapered estimation in stationary or nonstationary situations. In fact, our results indicate that the use of this particular type of higher order kernel should be heavily discouraged. As already mentioned, the reason for this disappointing behaviour of RH is the presence of V_q in (7), which takes values $V_q = 1, 0.4125, 0.0089, -0.2066$ for $q = 1, 2, 3, 4$, respectively. Thus for $q = 3$, even larger sample sizes than the ones employed in this experiment are needed for the asymptotic advantage to start being noticeable. Unless otherwise stated, our comments below will not refer to this specific estimator.

Table 1. Bias, $\delta = 0.3$, RH, Untapered estimators

M	$T \backslash q,m$	$1,m_1$	$1,m_2$	$1,m_3$	$2,m_1$	$2,m_2$	$2,m_3$	$3,m_1$	$3,m_2$	$3,m_3$	$4,m_1$	$4,m_2$	$4,m_3$
I	256	-.022	-.012	-.008	-.262	-.063	.044	-.043	-.104	-.186	-.034	.036	.064
	1024	-.009	-.003	-.003	.000	.034	.020	-.145	-.185	-.235	.061	.041	.012
	3072	-.001	.000	.000	.023	.008	.002	-.193	-.229	-.249	.014	.005	.000
	9216	-.002	-.001	.000	.001	.001	.001	-.229	-.225	-.224	-.002	.000	.001
II	256	-.007	.054	.160	-.258	.020	.206	.074	.238	.507	.039	.039	-.007
	1024	-.001	.020	.074	.002	.031	.112	-.034	.159	.588	.065	.010	.029
	3072	.000	.011	.041	.015	.016	.089	-.086	.179	.762	.005	-.002	.032
	9216	.003	.010	.033	.003	.016	.082	-.087	.301	.897	.000	.001	.034
III	256	.375	.575	.722	.145	.568	.787	.691	.801	.739	.305	.330	.608
	1024	.191	.387	.598	.268	.511	.716	.747	.980	1.06	.110	.331	.582
	3072	.113	.271	.492	.213	.439	.657	.944	1.21	1.22	.114	.315	.552
	9216	.086	.225	.442	.195	.407	.629	1.22	1.52	1.44	.112	.293	.524
IV	256	-.082	-.087	-.099	-.317	-.142	-.054	-.120	-.231	-.299	-.072	-.027	-.015
	1024	-.050	-.059	-.073	-.049	-.019	-.051	-.215	-.302	-.370	.045	.006	-.047
	3072	-.035	-.044	-.058	-.010	-.043	-.068	-.285	-.346	-.426	-.005	-.034	-.056
	9216	-.028	-.039	-.053	-.031	-.048	-.068	-.371	-.413	-.416	-.028	-.040	-.056

According to Table 1, increasing q only manages to equal (or slightly improve) the results for LW if T is very large for Models I and II (which represents the case of moderate short memory dependence). Model III is a very adverse case for memory estimation, because the short memory dependence is very strong. It is well known for this case that the memory estimation tends to overestimate the true memory parameter and this is indeed the case in our experiment, as bias is in general quite large, especially for the largest bandwidth. Interestingly LW provides better results than RH for larger sample sizes and worse for small T, which is counterintuitive. Finally, for Model IV, in general, RH shows better behaviour than LW, especially for $q = 4$ if T is not very large.

Table 2 presents bias results for the AS estimator. First, while we observe that increasing q clearly reduces the bias for Model I, this improvement does not occur for Model II, where bias either remains constant or worsens slightly. Here $q = 2$ appears to be the best choice. In the case of Model III, higher q leads to poorer results, and the results for Model IV are not very affected by q.

Table 2. Bias, $\delta = 0.3$, AS, Untapered estimators

M	$_{T}\backslash_{q,m}$	$1, m_1$	$1, m_2$	$1, m_3$	$2, m_1$	$2, m_2$	$2, m_3$	$3, m_1$	$3, m_2$	$3, m_3$	$4, m_1$	$4, m_2$	$4, m_3$
I	256	-.088	-.049	-.021	-.043	-.030	-.012	-.068	-.041	-.016	-.047	-.030	-.006
	1024	-.042	-.019	-.007	-.019	-.010	-.004	-.026	-.012	-.004	-.019	-.008	-.003
	3072	-.013	-.005	-.002	-.004	-.001	.000	-.006	-.002	.000	-.003	-.001	.000
	9216	-.004	-.002	.000	-.001	.000	.001	-.002	-.001	.000	-.001	.000	.002
II	256	-.087	-.043	.019	-.047	-.028	.025	-.073	-.038	.023	-.048	-.025	.056
	1024	-.038	-.017	.000	-.021	-.009	.011	-.026	-.010	.012	-.020	-.006	.019
	3072	-.012	-.006	.001	-.005	-.001	.009	-.007	-.001	.013	-.005	.000	.019
	9216	-.003	.000	.002	.000	.001	.008	.000	.001	.013	.000	.001	.019
III	256	.071	.315	.543	.109	.336	.558	.096	.340	.567	.128	.362	.616
	1024	.006	.141	.352	.050	.211	.436	.057	.237	.467	.075	.256	.492
	3072	.000	.075	.236	.040	.168	.382	.057	.210	.437	.074	.234	.468
	9216	.005	.055	.190	.036	.150	.360	.054	.196	.424	.070	.222	.458
IV	256	-.124	-.096	-.088	-.078	-.079	-.081	-.113	-.091	-.085	-.085	-.081	-.082
	1024	-.072	-.055	-.057	-.051	-.050	-.059	-.056	-.053	-.061	-.050	-.052	-.061
	3072	-.037	-.035	-.042	-.031	-.038	-.049	-.034	-.041	-.053	-.033	-.042	-.055
	9216	-.023	-.028	-.036	-.025	-.034	-.046	-.028	-.037	-.050	-.028	-.038	-.052

Overall, comparing the different estimation methods, we conclude that, in general, LW beats AS for Model I, especially for small T, and both estimates are better than RH (although for very large T the three methods give similar results). For Model II, except for very small T, AS clearly improves RH and, in most cases, also LW. In Model III, AS provides the smallest bias, with RH being better than LW if T is small and worse if T is large. Finally, the behaviour of AS is slightly better than LW in Model IV, and worse than RH for moderate T.

We also computed results for the tapered RH and AS estimators when $\delta = 0.3$, although for space reasons we do not report the results here (results are available upon request). Tapering is not necessary in this case, but in practice δ is unknown, so its use could be motivated by the (incorrect) belief that it falls within the non-stationary region. Our results indicate that, in general, tapering increases the bias: this is more evident for the AS, while Model II is the least affected scenario, especially when large bandwidths are used.

We complete our interpretation of the bias results for the stationary case with a brief description of the effect of the bandwidth choice m. First, the general pattern of behaviour differs depending on the model for LW or RH. For Model I, an increase of m reduces the bias, especially if T is large. On the contrary, for Models II-IV, typically, the opposite happens, although there are exceptions, especially for small T and tapered estimators. With respect to AS, bias reduces as m increases in Model I, although not in all cases if T is small (more noticeably with tapering). There is no clear pattern for Model II , while bias clearly worsens as m increases in Model III. Finally, for Model IV, bias increases as m increases for large T, with the opposite effect being evident for small T. If tapering is employed, this bias increase is just observed for large T and q.

Tables 3 and 4 present bias results for the non-stationary case $\delta = 1.3$ when tapering is used. Overall, a similar picture to the previously described is observed. Regarding RH, larger q leads to bias improvements for large T in Model I, with the asymptotic theory operating here. However, RH only manages to slightly beat LW for very large T and small m. Similarly, in Model II, an increase of q reduces the bias for T large enough. For example, if $q = 4$ and T is large, RH outperforms LW for any bandwidth. In Model III the order of the kernel just reduces the bias for small T. Thus LW appears to be preferable for large T, which, again, is counterintuitive. A similar result is observed for Model IV: bias does not decrease while increasing q. In fact, it often increases, so, here, LW is clearly better than RH.

Table 3. Bias, $\delta = 1.3$, RH, Tapered estimators

M	$T \backslash_{q,m}$	$1, m_1$	$1, m_2$	$1, m_3$	$2, m_1$	$2, m_2$	$2, m_3$	$3, m_1$	$3, m_2$	$3, m_3$	$4, m_1$	$4, m_2$	$4, m_3$
I	256	.047	.040	.012	-.499	-.270	-.159	-.398	-.249	-.410	-.228	-.221	-.175
	1024	.159	.023	.010	-.286	-.070	.014	-.349	-.373	-.460	-.284	.038	.054
	3072	.066	.014	.006	-.036	.018	.018	-.410	-.422	-.492	-.055	.033	.016
	9216	.026	.007	.004	.018	.009	.006	-.446	-.462	-.511	.014	.008	.005
II	256	.064	.120	.200	-.531	-.199	.087	-.379	.023	.078	-.211	-.109	-.071
	1024	.169	.051	.092	-.260	.004	.125	-.320	-.129	.160	-.253	.064	.050
	3072	.072	.027	.050	-.011	.024	.090	-.358	-.137	.293	-.026	.011	.036
	9216	.030	.018	.037	.014	.019	.081	-.388	-.048	.481	.012	.005	.034
III	256	.453	.599	.664	-.259	.297	.491	.074	.377	.237	.044	.276	.480
	1024	.390	.443	.614	.115	.498	.658	.203	.517	.443	-.187	.362	.574
	3072	.195	.297	.505	.251	.452	.651	.447	.642	.632	.080	.329	.558
	9216	.118	.236	.447	.207	.413	.630	.602	.682	.694	.119	.300	.527
IV	256	.037	.004	-.093	-.480	-.296	-.315	-.394	-.405	-.664	-.237	-.298	-.397
	1024	.156	.017	-.021	-.276	-.091	-.136	-.328	-.445	-.566	-.282	-.076	-.183
	3072	.066	.010	-.006	-.030	.031	.075	-.409	-.485	-.541	-.066	.027	-.041
	9216	.027	.006	-.006	.020	.021	.088	-.473	-.518	-.545	.016	.022	.068

Regarding AS, bias is reduced by increasing q in Model I, especially for large T. The improvement is not so clear for small T. In Model II, an increase of q does not have a clear effect on the bias, but it depends on the values of T and m. In Model III, there is a clear increase in the bias when increasing q, whereas the results remain fairly stable in Model IV. Finally, the bandwidth choice affects the bias in a similar way to that described for the stationary case.

Table 4. Bias, $\delta = 1.3$, AS, Tapered estimators

M	$_T\backslash_{q,m}$	$1, m_1$	$1, m_2$	$1, m_3$	$2, m_1$	$2, m_2$	$2, m_3$	$3, m_1$	$3, m_2$	$3, m_3$	$4, m_1$	$4, m_2$	$4, m_3$
I	256	-.201	-.013	.024	-.106	.031	.041	.700	.005	.031	-.097	.029	.039
	1024	.530	.024	.020	.048	.032	.023	.319	.023	.019	.040	.028	.021
	3072	.210	.021	.015	.112	.020	.013	.085	.015	.011	.078	.016	.011
	9216	.067	.011	.008	.014	.009	.006	.027	.007	.005	.010	.007	.005
II	256	-.171	-.002	.074	-.082	.037	.087	.700	.010	.078	-.068	.040	.117
	1024	.529	.027	.030	.043	.036	.039	.319	.027	.037	.038	.032	.043
	3072	.208	.020	.018	.110	.020	.022	.082	.016	.025	.076	.017	.029
	9216	.068	.014	.011	.017	.011	.014	.030	.009	.018	.014	.010	.021
III	256	-.020	.364	.573	.053	.395	.587	.700	.397	.591	.109	.430	.621
	1024	.560	.207	.407	.129	.280	.485	.410	.300	.509	.150	.323	.534
	3072	.227	.110	.263	.164	.200	.402	.158	.236	.453	.168	.260	.502
	9216	.076	.071	.201	.054	.162	.367	.088	.207	.429	.089	.232	.477
IV	256	-.185	.003	.027	-.094	.042	.049	.700	.018	.041	-.077	.041	.035
	1024	.533	.026	.024	.045	.035	.029	.323	.027	.027	.039	.031	.029
	3072	.211	.022	.263	.111	.022	.402	.085	.018	.453	.079	.020	.019
	9216	.063	.011	.009	.014	.010	.010	.027	.009	.012	.011	.009	.012

To summarise, as regards bias behaviour, in Model I, LW show the best behaviour, followed by AS. In Model II, in general, AS is the best method, followed by RH. AS is again the preferred one in Model III, with RH being better than LW if T is small and worse if T is large. Finally, LW improves AS, and both are better than RH in Model IV.

Tables 5, 6 present the SD for the stationary case. First, for RH, as expected, larger T and/or m reduce SD. Interestingly, the minimum SD is almost always achieved by $q = 1$ (the LW estimator), with the exception of Model III when T is large. When increasing q from 2 to 4, for Model I, SD remain similar; for II, SD are in general reduced, except for small T; for III, SD increase for small T, but decrease if T is large enough; finally, there is not a clear pattern for IV.

Table 5. SD, $\delta = 0.3$, RH, Untapered estimators

M	$_T\backslash_{q,m}$	$1, m_1$	$1, m_2$	$1, m_3$	$2, m_1$	$2, m_2$	$2, m_3$	$3, m_1$	$3, m_2$	$3, m_3$	$4, m_1$	$4, m_2$	$4, m_3$
I	256	.250	.152	.096	.494	.492	.423	.682	.685	.661	.723	.610	.417
	1024	.129	.085	.056	.443	.284	.144	.672	.665	.654	.440	.248	.129
	3072	.077	.052	.036	.211	.108	.060	.675	.664	.663	.178	.095	.062
	9216	.046	.031	.022	.074	.047	.033	.658	.662	.670	.072	.047	.034
II	256	.256	.152	.099	.479	.429	.250	.693	.710	.637	.758	.587	.327
	1024	.129	.083	.056	.386	.174	.070	.708	.710	.552	.386	.150	.073
	3072	.079	.052	.035	.177	.068	.038	.713	.681	.527	.137	.063	.037
	9216	.045	.030	.021	.068	.037	.021	.719	.632	.431	.061	.034	.020
III	256	.263	.159	.106	.468	.434	.374	.655	.634	.689	.817	.672	.428
	1024	.134	.089	.064	.221	.131	.096	.553	.427	.443	.218	.120	.099
	3072	.079	.056	.041	.078	.054	.044	.515	.351	.300	.076	.053	.042
	9216	.045	.033	.025	.040	.029	.023	.449	.238	.230	.040	.030	.024
IV	256	.249	.151	.097	.485	.501	.439	.669	.638	.638	.707	.609	.426
	1024	.131	.083	.055	.463	.315	.172	.649	.639	.595	.480	.289	.146
	3072	.079	.052	.036	.230	.118	.069	.631	.623	.554	.232	.115	.074
	9216	.045	.031	.022	.080	.052	.037	.575	.557	.583	.045	.056	.040

Table 6. SD, $\delta = 0.3$, AS, Untapered estimators

M	$_T\backslash_{q,m}$	$1, m_1$	$1, m_2$	$1, m_3$	$2, m_1$	$2, m_2$	$2, m_3$	$3, m_1$	$3, m_2$	$3, m_3$	$4, m_1$	$4, m_2$	$4, m_3$
I	256	.460	.277	.164	.462	.272	.162	.441	.262	.154	.433	.255	.137
	1024	.232	.141	.090	.196	.119	.077	.182	.112	.073	.176	.108	.069
	3072	.132	.085	.056	.099	.064	.044	.090	.058	.040	.084	.055	.036
	9216	.071	.048	.032	.052	.035	.025	.047	.032	.022	.045	.031	.019
II	256	.461	.276	.166	.460	.275	.164	.439	.263	.158	.433	.258	.139
	1024	.233	.139	.088	.193	.117	.076	.180	.110	.072	.171	.107	.070
	3072	.132	.085	.056	.099	.065	.043	.089	.058	.040	.084	.056	.039
	9216	.071	.048	.034	.052	.036	.025	.047	.033	.023	.045	.031	.023
III	256	.502	.287	.174	.496	.284	.172	.473	.271	.164	.460	.262	.151
	1024	.234	.144	.097	.197	.123	.085	.185	.117	.081	.176	.113	.079
	3072	.132	.085	.059	.099	.067	.049	.089	.061	.046	.085	.059	.044
	9216	.072	.049	.034	.053	.037	.029	.047	.035	.027	.046	.033	.026
IV	256	.451	.277	.165	.455	.271	.162	.433	.260	.156	.431	.255	.147
	1024	.236	.141	.091	.195	.120	.078	.182	.112	.073	.172	.109	.071
	3072	.132	.084	.056	.098	.065	.044	.089	.059	.040	.084	.056	.038
	9216	.071	.048	.033	.052	.036	.025	.047	.033	.023	.044	.031	.021

Focusing on AS, the general behaviour of the SD when m and/or T raise is similar: SD decrease except for some cases. Additionally, an increase of q also reduces the SD in all cases.

Comparing the three estimation methods we reach the following conclusions. In Model I, AS provides smaller SD than RH and, if T is sufficiently large, similar to those of LW. In II, the smaller SD are obtained with LW, followed by AS. In III, if T is large, RH has the smallest SD, while LW improves AS, especially if T small. Finally, in IV, AS beats RH, although LW provides results very similar to AS if T is large (and better if T is small).

Tables 7, 8 present corresponding results for the tapered estimates. As the theory predicted, the main phenomenon here is the increase in SD for all cases, with a similar picture being observed for the (unreported) stationary and non-stationary cases.

Table 7. SD, $\delta = 1.3$, RH, Tapered estimators

M	$T \backslash q,m$	$1,m_1$	$1,m_2$	$1,m_3$	$2,m_1$	$2,m_2$	$2,m_3$	$3,m_1$	$3,m_2$	$3,m_3$	$4,m_1$	$4,m_2$	$4,m_3$
I	256	.437	.262	.156	.910	.508	.498	.757	.650	.600	.841	.727	.621
	1024	.223	.136	.088	.580	.440	.332	.654	.651	.625	.679	.410	.277
	3072	.127	.081	.055	.371	.223	.115	.644	.648	.607	.475	.188	.133
	9216	.069	.047	.032	.148	.078	.054	.642	.654	.612	.148	.080	.064
II	256	.438	.255	.156	.908	.463	.444	.748	.640	.656	.855	.767	.575
	1024	.223	.134	.086	.572	.334	.136	.679	.700	.650	.717	.285	.138
	3072	.127	.081	.053	.320	.127	.061	.674	.697	.616	.403	.112	.058
	9216	.068	.046	.032	.137	.059	.033	.695	.679	.432	.118	.053	.031
III	256	.326	.151	.064	.783	.507	.513	.663	.620	.661	.830	.665	.409
	1024	.206	.132	.072	.434	.221	.105	.653	.511	.635	.755	.200	.111
	3072	.128	.084	.060	.143	.084	.048	.449	.308	.370	.191	.081	.063
	9216	.070	.048	.036	.063	.043	.034	.296	.203	.112	.061	.043	.035
IV	256	.446	.262	.160	.922	.530	.537	.756	.624	.549	.824	.697	.719
	1024	.222	.132	.087	.593	.490	.597	.643	.581	.438	.656	.548	.657
	3072	.126	.080	.054	.360	.277	.392	.626	.577	.411	.493	.300	.550
	9216	.069	.047	.032	.158	.095	.146	.646	.569	.420	.172	.137	.305

Table 8. SD, $\delta = 1.3$, AS, Tapered estimators

M	$_T\backslash_{q,m}$	$1, m_1$	$1, m_2$	$1, m_3$	$2, m_1$	$2, m_2$	$2, m_3$	$3, m_1$	$3, m_2$	$3, m_3$	$4, m_1$	$4, m_2$	$4, m_3$
I	256	.834	.496	.287	.801	.481	.282	.000	.467	.271	.785	.457	.234
	1024	.269	.241	.145	.341	.199	.122	.297	.186	.115	.304	.179	.110
	3072	.225	.136	.086	.162	.101	.067	.144	.091	.061	.135	.086	.058
	9216	.112	.072	.049	.079	.053	.037	.071	.048	.033	.066	.046	.031
II	256	.832	.487	.279	.802	.478	.277	.012	.462	.268	.773	.460	.229
	1024	.270	.236	.144	.339	.197	.123	.290	.184	.116	.297	.177	.111
	3072	.225	.136	.087	.163	.101	.066	.144	.091	.061	.135	.086	.058
	9216	.112	.073	.050	.079	.054	.037	.071	.049	.033	.066	.046	.032
III	256	.793	.401	.182	.766	.381	.171	.000	.370	.164	.736	.354	.102
	1024	.248	.234	.142	.335	.196	.122	.269	.181	.114	.296	.177	.107
	3072	.225	.135	.088	.162	.101	.071	.145	.092	.066	.133	.088	.081
	9216	.114	.075	.051	.082	.055	.041	.073	.051	.039	.069	.048	.054
IV	256	.823	.496	.290	.795	.490	.284	.007	.475	.271	.769	.458	.242
	1024	.265	.238	.144	.341	.198	.120	.292	.185	.114	.299	.178	.109
	3072	.223	.135	.088	.159	.100	.071	.144	.091	.066	.134	.087	.057
	9216	.114	.074	.050	.081	.054	.037	.073	.049	.034	.068	.047	.032

In Table 9 we present the coverage probabilities corresponding to the RH estimator for the non-stationary case (i.e. $\delta = 1.3$). As previously pointed out, RH estimator should be heavily discouraged when $q = 3$. Given the huge variance of the estimator for this case, the 95% confidence interval is so wide that coverage probabilities always equal to one. Focusing on the rest of kernel orders, Table 9 shows that, in general, an increase of T raises the coverage probabilities, placing them close to 0.95 value. This happens for all models, except for Model III when medium or large values of m are chosen. The effects of increasing m are not, however, so clear: whereas it is clearly harmful for model III, large and, especially, intermediate values of m lead to coverage probabilities closer to 0.95 value for the rest of the models. Finally, comparing the behavior of the LW and RH estimators, we observe that while the former outperforms the latter in models I and IV, using higher order kernels seems to be more appropriate in autoregressive cases (models II and III). The coverage probabilities for the stationary cases (with and without tapering) were also computed for RH estimates, giving similar results. We do not report these results here for the sake of space saving, but they are available upon request.

Table 9. Coverage probabilities, $\delta = 1.3$, RH, Tapered estimators

M	$T \backslash q,m$	$1,m_1$	$1,m_2$	$1,m_3$	$2,m_1$	$2,m_2$	$2,m_3$	$3,m_1$	$3,m_2$	$3,m_3$	$4,m_1$	$4,m_2$	$4,m_3$
I	256	.678	.772	.854	.578	.736	.720	1	1	1	.788	.646	.641
	1024	.676	.855	.890	.620	.756	.784	1	1	1	.600	.762	.778
	3072	.817	.902	.920	.763	.844	.866	1	1	1	.655	.836	.822
	9216	.887	.919	.933	.859	.906	.897	1	1	1	.840	.898	.848
II	256	.672	.724	.543	.561	.813	.771	1	1	1	.784	.588	.687
	1024	.680	.838	.707	.626	.848	.850	1	1	1	.566	.838	.932
	3072	.804	.887	.801	.767	.928	.832	1	1	1	.705	.946	.973
	9216	.885	.906	.750	.881	.946	.584	1	1	1	.890	.975	.955
III	256	.336	.065	.002	.755	.635	.113	1	1	1	.842	.480	.268
	1024	.291	.040	.000	.771	.135	.000	1	1	1	.523	.530	.003
	3072	.477	.028	.000	.661	.000	.000	1	1	1	.913	.075	.000
	9216	.500	.001	.000	.305	.000	.000	1	1	1	.819	.000	.000
IV	256	.674	.781	.789	.591	.690	.550	1	1	1	.791	.654	.584
	1024	.692	.867	.890	.609	.691	.471	1	1	1	.618	.696	.446
	3072	.816	.913	.920	.759	.791	.558	1	1	1	.642	.719	.307
	9216	.883	.923	.931	.845	.855	.556	1	1	1	.819	.734	.301

To summarise, the overall performance of the bias-reducing techniques is somewhat disappointing because, in many cases, these more sophisticated methods do not outperform the simpler LW alternative. More specifically, for Model I, LW is always the best method, both in stationary or non-stationary settings. For Model II, AS is better than LW for $\delta = 0.3, 1.3$, while RH beats LW in few cases (just if T is small and $\delta = 1.3$). For Model III, AS is again better than LW, and RH tends to outperform LW if T is small. Finally, for Model IV, AS tends to be better than LW when $\delta = 0.3$, especially if T is large. For the non-stationary case, RH beats LW just if T is small, although, overall, when $\delta = 1.3$ LW tends to be the best method. In any case, the performance of the higher order kernel estimator RH is particularly worrying, as its general behaviour is poor (and even extremely unsatisfactory for $q = 3$). As already mentioned the role of V_q in (7) gives an advantage to the choice $q = 1$ (i.e., to LW), which, at least for the sample sizes employed in the experiment, is not compensated by the faster convergence rates which higher q's permit. In fact, we provide additional evidence to illustrate the problem suffered by the higher order kernel approach. In Table 10 we report the number of replications (out of 5000) which falls on the boundary of the optimizing interval $[-.5, 2]$ for the $\delta = 1.3$ case (other cases present qualitatively identical results).

Table 10. Replications outside the optimizing interval, $\delta = 1.3$, RH, Tapered estimators

M	$T \backslash_{q,m}$	$1, m_1$	$1, m_2$	$1, m_3$	$2, m_1$	$2, m_2$	$2, m_3$	$3, m_1$	$3, m_2$	$3, m_3$	$4, m_1$	$4, m_2$	$4, m_3$
I	256	4	0	0	1508	432	512	1036	879	659	1624	1170	905
	1024	0	0	0	565	430	298	774	862	764	889	602	313
	3072	0	0	0	302	116	6	809	836	703	595	94	21
	9216	0	0	0	40	1	0	789	845	719	28	0	0
II	256	2	0	0	1424	367	565	1033	1258	1438	1721	1540	761
	1024	0	0	0	570	300	34	891	1271	1734	1119	296	11
	3072	0	0	0	240	19	0	930	1254	2239	426	4	0
	9216	0	0	0	31	0	0	983	1365	2705	5	0	0
III	256	1	0	0	928	1397	2967	1445	2859	2004	2160	2322	2133
	1024	0	0	0	549	1009	2928	1861	3418	3430	1342	461	1344
	3072	0	0	0	62	57	1646	2379	4267	4499	25	1	157
	9216	0	0	0	0	0	240	3473	4929	4953	0	0	0
IV	256	8	0	0	1584	470	507	1030	663	379	1534	903	1018
	1024	0	0	0	603	564	782	759	649	303	803	812	1121
	3072	0	0	0	309	229	512	746	635	291	613	315	936
	9216	0	0	0	47	4	60	831	609	306	50	24	419

Although results improve as T and m increase, the evidence is very worrying (especially for $q = 3$). It should be noted that this problem hardly affects LW and it does not affect AS at all, where none of the replications fell on the boundary of the optimizing set.

Acknowledgements

Research supported by the Spanish Ministry of Science and Innovation and ERDF grants ECO2015-64330-P and ECO2016- 76884-P

REFERENCES

Abadir, K.M., W. Distaso & L. Giraitis (2007) Nonstationarity-extended local Whittle estimation. *Journal of Econometrics* 141, 1353-1384.

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

Andrews, D.W.K. & P. Guggenberger (2003) A bias-reduced log-periodogram regression estimator for the long-memory parameter. *Econometrica* 71, 675-712.

Andrews, D.W.K. & Y. Sun (2004) Adaptive local polynomial Whittle estimation of long-range dependence. *Econometrica* 72, 569-614.

Arteche, J. (2004) Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models. *Journal of Econometrics* 119, 131-154.

Arteche, J. (2006). Semiparametric estimation in perturbed long memory series. *Computational Statistics & Data Analysis* 51, 2118-2141.

Deo, R.S. & C.M. Hurvich (2001) On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models. *Econometric Theory* 17, 686-710.

Frederiksen, P., F.S. Nielsen & M.O. Nielsen (2012) Local polynomial Whittle estimation of perturbed fractional processes. *Journal of Econometrics* 167, 426-447.

Hauser, M.A. (1997) Semiparametric and nonparametric testing for long memory: a Monte Carlo study. *Empirical Economics* 22, 247-271.

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

Giraitis, L., P.M. Robinson, & 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.

Hou, J. & P. Perron (2014) Modified local Whittle estimator for long memory processes in the presence of low frequency (and other) contaminations. *Journal of Econometrics* 182, 309-328.

Hualde, J. & P.M. Robinson (2006) Semiparametric estimation of fractional cointegration. STICERD Discussion Paper EM/2006/502.

Hualde, J. & P.M. Robinson (2010) Semiparametric inference in multivariate fractionally cointegrated systems. *Journal of Econometrics* 157, 492-511.

Hurvich, C.M. & J. Brodsky (2001) Broadband semiparametric estimation of the memory parameter of a long-memory time series using fractional exponential models. *Journal of Time Series Analysis* 22, 221-249.

Hurvich, C.M., E. Moulines & P. Soulier (2005) Estimating long memory in volatility. *Econometrica* 73, 1283-1328.

Hurvich, C.M. & B.K. Ray (2003) The local Whittle estimator of long-memory stochastic volatility. *Journal of Financial Economics* 1, 445-470.

Kim, C.S. & P.C.B. Phillips (1999) Modified log periodogram regression. Working Paper, Cowles Foundation, Yale University, New Haven.

Künsch, H.R. (1987) Statistical aspects of self-similar processes. In Y.A. Prohorov & V.V. Sazonov (eds.), *Proceedings First World Congress of the Bernoulli Society*, pp. 67-74. VNU Science Press.

Marinucci, D. & P.M. Robinson (1999) Alternative forms of fractional Brownian motion. *Journal of Statistical Planning and Inference* 80 111-122.

McCloskey, A. & P. Perron (2013) Memory parameter estimation in the presence of level shifts and deterministic trends. *Econometric Theory* 29, 1196-1237.

Moulines, E. & P. Soulier (1999) Broadband log-periodogram regression of time series with long-range dependence. *Annals of Statistics* 27, 1415-1439.

Nielsen, M.O. & P.H. Frederiksen (2005) Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews* 24, 405-443.

Perron, P. & Z. Qu (2010) Long-memory and level shifts in the volatility of stock market return indices. *Journal of Business and Economic Statistics* 28, 275-290.

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

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

Robinson, P.M. (2005) The distance between rival nonstationary fractional processes. *Journal of Econometrics* 128, 283-300.

Robinson, P.M. & M. Henry (2003) Higher-order kernel semiparametric M-estimation of long memory. *Journal of Econometrics* 114, 1-27.

Shimotsu, K. & P.C.B. Phillips (2005) Exact local Whittle estimation of fractional integration. *Annals of Statistics* 33, 1890-1933.

Shimotsu, K. (2010) Exact local Whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory* 26, 501-540.

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

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

Velasco, C. (2006) Semiparametric estimation of long-memory models. In K. Patterson & T.C. Mills (eds.), *Palgrave Handbook of Econometrics, Vol. 1. Econometric Theory*, pp. 353-395. Palgrave, MacMillan.

Supplementary Appendix

to

Local Whittle estimation of long memory: standard versus bias-reducing techniques

This supplements "Local Whittle estimation of long memory: standard versus bias-reducing techniques" by providing proofs of Propositions 1 and 2.

Proof of Proposition 1. First, we show the result for $\overline{\delta}_{RH}$. Clearly

$$m^{\frac{1}{2}}\left(\overline{\delta}_{RH}-\delta\right)=m^{\frac{1}{2}}\left(\overline{\delta}_{G}-\delta-\frac{S_v^p\left(\overline{\delta}_G\right)-S_v^p\left(\delta\right)}{H_v^p\left(\overline{\delta}_G\right)}\right)-m^{\frac{1}{2}}\frac{S_v^p\left(\delta\right)}{H_v^p\left(\overline{\delta}_G\right)},$$

so (7) holds on showing

$$m^{-\frac{1}{2}}S_v^p\left(\delta\right)-\frac{2\left(2\pi\right)^{2q}U_{qq}h_q}{(2q)!f_u\left(0\right)}\frac{m^{2q+1/2}}{T^{2q}}\rightarrow_d N\left(0,4p\Phi W_q\right), \tag{S.1}$$

$$m^{-1}H_v^p\left(\overline{\delta}_G\right)\rightarrow_p 4V_q>0 \quad \text{and} \tag{S.2}$$

$$m^{\frac{1}{2}}\left(\overline{\delta}_{G}-\delta-\frac{S_v^p\left(\overline{\delta}_G\right)-S_v^p\left(\delta\right)}{H_v^p\left(\overline{\delta}_G\right)}\right)=o_p\left(1\right). \tag{S.3}$$

First, we show (S.1). Now,

$$m^{-\frac{1}{2}}S_v^p\left(\delta\right)=2p^{\frac{1}{2}}\frac{A}{B}, \tag{S.4}$$

where

$$A=\left(\frac{p}{m}\right)^{\frac{1}{2}}\sum{}' b_{q,j}\left(\lambda_j^{2\delta}I_v^p\left(\lambda_j\right)-1\right) \quad \text{and} \quad B=\frac{p}{m}\sum{}' k_{q,j}\lambda_j^{2\delta}I_v^p\left(\lambda_j\right),$$

noting that $\sum' b_{q,j}=0$. Then, we could set $A=\sum_{i=1}^4 A_i$ where

$$\begin{aligned}
A_1&=2\pi\left(\frac{p}{m}\right)^{\frac{1}{2}}\sum{}' b_{q,j}\left\{E\left(I_\xi^p\left(\lambda_j\right)h\left(\lambda_j\right)\right)-E\left(f_u\left(0\right)I_\xi^p\left(\lambda_j\right)\right)\right\},\\
A_2&=2\pi f_u\left(0\right)\left(\frac{p}{m}\right)^{\frac{1}{2}}\sum{}' b_{q,j}\left\{I_\xi^p\left(\lambda_j\right)-E\left(I_\xi^p\left(\lambda_j\right)\right)\right\},\\
A_3&=2\pi\left(\frac{p}{m}\right)^{\frac{1}{2}}\sum{}' b_{q,j}\left\{I_\xi^p\left(\lambda_j\right)h\left(\lambda_j\right)-f_u\left(0\right)I_\xi^p\left(\lambda_j\right)\right.\\
&\quad\left.-E\left(I_\xi^p\left(\lambda_j\right)h\left(\lambda_j\right)-f_u\left(0\right)I_\xi^p\left(\lambda_j\right)\right)\right\} \quad \text{and}\\
A_4&=\left(\frac{p}{m}\right)^{\frac{1}{2}}\sum{}' b_{q,j}\left\{I_v^p\left(\lambda_j\right)\lambda_j^{2\delta}-2\pi h\left(\lambda_j\right)I_\xi^p\left(\lambda_j\right)\right\},
\end{aligned}$$

noting that $E\left(I_\xi^p\left(\lambda_j\right)\right)=1/\left(2\pi\right)$. We have

$$A_1=\left(\frac{p}{m}\right)^{\frac{1}{2}}\sum{}' b_{q,j}\left(h\left(\lambda_j\right)-f_u\left(0\right)\right)=\left(\frac{p}{m}\right)^{\frac{1}{2}}\sum{}' b_{q,j}\sum_{i=1}^{q}\frac{h_i\lambda_j^{2i}}{(2i)!}+O\left(m^{-\frac{1}{2}}\sum{}'\left|b_{q,j}\right|\lambda_j^{s+\eta}\right). \tag{S.5}$$

The first term on the right of (S.5) equals

$$\left(\frac{m}{p}\right)^{\frac{1}{2}}\left\{\sum_{i=1}^{q}\frac{h_i\lambda_m^{2i}}{(2i)!}U_{iq}+\sum_{i=1}^{q}\frac{h_i\lambda_m^{2i}}{(2i)!}\left[\frac{p}{m}\sum{}' b_{q,j}\left(\frac{j}{m}\right)^{2i}-\int_0^1\left(\log u+1\right)u^{2i}k_q\left(u\right)du\right]\right\}. \tag{S.6}$$

Then noting that

$$\sum_{i=1}^{q}\left(\frac{m}{T}\right)^{2i}\leq\frac{1}{1-\left(\frac{m}{T}\right)^2}\leq 0.75^{-1},$$

because $m/T\leq 0.5$, that by proceeding as in Lemma 5 of Velasco (1999a),

$$\frac{p}{m}\sum{}' b_{q,j}\left(\frac{j}{m}\right)^{2i}-\int_0^1\left(\log u+1\right)u^{2i}k_q\left(u\right)du=O\left(m^{-1}\log m\right),$$

and Assumption 2, then (S.6) is

$$\left(\frac{m}{p}\right)^{\frac{1}{2}}\frac{h_q\lambda_m^{2q}}{(2q)!}U_{qq}+O\left(\frac{\log m}{m^{\frac{1}{2}}}\right).$$

This implies that

$$A_1=\left(\frac{m}{p}\right)^{\frac{1}{2}}\frac{h_q\lambda_m^{2q}}{(2q)!}U_{qq}+O\left(\frac{\log m}{m^{\frac{1}{2}}}\right)+O\left(\frac{m^{s+\eta+\frac{1}{2}}\log T}{T^{s+\eta}}\right), \tag{S.7}$$

where by (6), the third term on the right of (S.7) is of smaller order than the first, while the second is $o\left(1\right)$.

For A_2, in view of the proof of Lemma 6 of Velasco (1999a), it is straightforward to show that

$$\left(\frac{p}{m}\right)^{\frac{1}{2}}\sum{}' b_{q,j}\left\{2\pi I_\xi^p\left(\lambda_j\right)-1\right\}\to_d N\left(0,W_q\Phi\right),$$

just noting that, like in (S.6),

$$\frac{p}{m}\sum{}' b_{q,j}^2=W_q+O\left(\frac{\left(\log m\right)^2}{m}\right).$$

Next, according to some of our previous arguments and (A23) in Velasco (1999a),

$Var\left(A_3\right)=o\left(1\right)$ and

$$A_4=\left(\frac{p}{m}\right)^{\frac{1}{2}}\sum{}' b_{q,j}h\left(\lambda_j\right)\left\{\frac{I_v^p\left(\lambda_j\right)}{h\left(\lambda_j\right)\lambda_j^{-2\delta}}-2\pi I_\xi^p\left(\lambda_j\right)\right\}$$
$$=O_p\left(m^{-1/2}\log^2 m+m^{\delta-p+1/2}\log^{3/2}m\right)=o_p\left(1\right),$$

by the condition we set on the tapering order p.

Expanding B in a similar way to A, we get $B=\sum_{i=0}^4 B_i$, where

$$B_0=\frac{2\pi f_u\left(0\right)p}{m}\sum{}' k_{q,j}E\left(I_\xi^p\left(\lambda_j\right)\right),$$
$$B_1=\frac{2\pi p}{m}\sum{}' k_{q,j}\left\{E\left(I_\xi^p\left(\lambda_j\right)h\left(\lambda_j\right)\right)-E\left(f_u\left(0\right)I_\xi^p\left(\lambda_j\right)\right)\right\},$$
$$B_2=\frac{2\pi f_u\left(0\right)p}{m}\sum{}' k_{q,j}\left\{I_\xi^p\left(\lambda_j\right)-E\left(I_\xi^p\left(\lambda_j\right)\right)\right\},$$
$$B_3=\frac{2\pi p}{m}\sum{}' k_{q,j}\left\{I_\xi^p\left(\lambda_j\right)h\left(\lambda_j\right)-f_u\left(0\right)I_\xi^p\left(\lambda_j\right)\right.$$
$$\left.-E\left(I_\xi^p\left(\lambda_j\right)h\left(\lambda_j\right)-f_u\left(0\right)I_\xi^p\left(\lambda_j\right)\right)\right\}\quad\text{and}$$
$$B_4=\frac{p}{m}\sum{}' k_{q,j}\left\{I_v^p\left(\lambda_j\right)\lambda_j^{2\delta}-2\pi h\left(\lambda_j\right)I_\xi^p\left(\lambda_j\right)\right\}.$$

In line with previous results

$$B_1=O\left(\left(\frac{m}{T}\right)^{2q}\right),\quad B_2=O_p\left(m^{-\frac{1}{2}}\right),\quad B_3=o_p\left(1\right)\quad\text{and}$$
$$B_4=O_p\left(m^{-1}\log m+m^{\delta-p}\log^{1/2}m\right),$$

whereas

$$B_0=\frac{f_u\left(0\right)p}{m}\sum{}' k_{q,j}=f_u\left(0\right)+O\left(m^{-1}\log m\right),$$

to complete the proof of (S.1).

Next, we show (S.2), which holds if

$$m^{-1}H_v^p\left(\delta\right)\to_p 4V_q>0\quad\text{and}\tag{S.8}$$

$$m^{-1}\left(H_v^p\left(\overline{\delta}_G\right)-H_v^p\left(\delta\right)\right)=o_p\left(1\right).\tag{S.9}$$

Regarding (S.8), clearly,

$$m^{-1}H_v^p\left(\delta\right)=\frac{4\left(F_{2,v}^p\left(\delta\right)F_{0,v}^p\left(\delta\right)-\left(F_{1,v}^p\left(\delta\right)\right)^2\right)}{\left(F_{0,v}^p\left(\delta\right)\right)^2},$$

where

$$F^p_{g,\theta}(c) = \frac{p}{m}\sum{}' k_{q,j}\left(\log\frac{j}{m}\right)^g \lambda_j^{2c} I^p_\theta(\lambda_j), \;\; g = 0,1,2,3.$$

By the same decomposition as that in the treatment of B in (S.4), it is easy to show that

$$F^p_{g,v}(\delta) \to_p f_u(0)\int_0^1 k_q(u)(\log u)^g\, du, \;\; g = 0,1,2,3, \tag{S.10}$$

so that (S.8) follows immediately by Assumption 2.

Next, (S.9) holds if

$$m^{-1}\left|\sum{}' k_{q,j}\left(\log\frac{j}{m}\right)^g \left(\lambda_j^{2\overline{\delta}_G} - \lambda_j^{2\delta}\right) I^p_v(\lambda_j)\right| = o_p(1), \tag{S.11}$$

for $g = 0,1,2$. By the mean value theorem, the term inside the modulus in (S.11) equals

$$2\left(\overline{\delta}_G - \delta\right)\sum{}' k_{q,j}\left(\log\frac{j}{m}\right)^g \log\lambda_j \lambda_j^{2\left(\overline{\delta}-\delta\right)}\lambda_j^{2\delta} I^p_v(\lambda_j), \tag{S.12}$$

where $\left|\overline{\delta} - \delta\right| \leq \left|\overline{\delta}_G - \delta\right|$. Then, noting that by Theorem 4 of Velasco (1999b), under our conditions $E\left|\lambda_j^{2\delta} I^p_v(\lambda_j)\right| \leq K$, (S.12) is bounded in probability by

$$K\left|\overline{\delta}_G - \delta\right|\sum{}'\left|\log\frac{j}{m}\right|^g \left|\log\lambda_j\right| \lambda_j^{-2\left|\overline{\delta}-\delta\right|} \leq KM^{-\frac{1}{2}}(\log T)^3 T^{M^{-\frac{1}{2}}}\sum{}' j^{-2M^{-\frac{1}{2}}}.$$

Then, noting that $T^{M^{-\frac{1}{2}}} = O(1)$, $\sum{}' j^{-2M^{-\frac{1}{2}}} = O(m)$, the left of (S.11) is $O_p\left((\log T)^3 M^{-1/2}\right) = o_p(1)$ by Assumption 3, to conclude the proof of (S.9).

Finally, we show (S.3). First, by the mean value theorem the left-hand side of (S.3) is

$$m^{\frac{1}{2}}\left(\overline{\delta}_G - \delta\right)\left(1 - \frac{H^p_v\left(\overline{\delta}\right)}{H^p_v\left(\overline{\delta}_G\right)}\right), \tag{S.13}$$

where $\left|\overline{\delta} - \delta\right| \leq \left|\overline{\delta}_G - \delta\right|$. Applying the mean value theorem again, (S.13) equals

$$m^{\frac{1}{2}}\left(\overline{\delta}_G - \delta\right)\left(\overline{\delta}_G - \overline{\delta}\right)\frac{J^p_v\left(\overline{\overline{\delta}}\right)}{H^p_v\left(\overline{\delta}_G\right)},$$

where $J^p_\theta(c) = dH^p_\theta(c)/dc$ and $\left|\overline{\overline{\delta}} - \overline{\delta}\right| \leq \left|\overline{\delta}_G - \overline{\delta}\right|$. After some tedious but straightforward manipulations, it can be shown that

$$J_\theta^p(c) = 8m\frac{F_{3,\theta}^p(c)\left(F_{0,\theta}^p(c)\right)^3 - 3F_{2,\theta}^p(c)\,F_{1,\theta}^p(c)\left(F_{0,\theta}^p(c)\right)^2 + 2\left(F_{1,\theta}^p(c)\right)^3 F_{0,\theta}^p(c)}{\left(F_{0,\theta}^p(c)\right)^4}.$$

Then, by (S.2), (S.10) and a very simple extension of (S.11) to cover the treatment of $F_{3,v}^p\left(\overline{\overline{\delta}}\right)$, it can be easily shown that $J_v^p\left(\overline{\overline{\delta}}\right)/H_v^p\left(\overline{\delta}_G\right) = O_p(1)$, which implies that the left of (S.3) is $O_p\left(m^{1/2}M^{-1}\right) = o_p(1)$ by Assumption 3, to complete the proof for $\overline{\delta}_{RH}$.

Regarding $\widetilde{\delta}_{RH}$, we first show that $\widetilde{\delta}_G$ is $M^{1/2}$-consistent. Following the proof strategy of Robinson (1995) and Velasco (1999a), we set $\Theta = \Theta_1 \cup \Theta_2$, with

$$\Theta_1 = \{c : \delta - 1/2 + \epsilon \le c \le \nabla_2\} \quad \text{and} \quad \Theta_2 = \{c : \nabla_1 \le c < \delta - 1/2 + \epsilon\},$$

for $\epsilon \in (0, 1/4)$ (taking Θ_2 to be empty in case $\nabla_1 \ge \delta - 1/2 + \epsilon$). The main steps of the proof consist of establishing

$$\sup_{c\in\Theta_1}\left|\frac{G_v^p(c) - G_{\widetilde{v}}^p(c)}{G^p(c)}\right| = o_p\left(\log^{-10} M\right), \tag{S.14}$$

where

$$G_\theta^p(c) = \frac{p}{M}\sum\nolimits'' \lambda_j^{2c} I_\theta^p(\lambda_j) \quad \text{and} \quad G^p(c) = f_u(0)\frac{p}{M}\sum\nolimits'' \lambda_j^{2(c-\delta)},$$

where throughout $\sum'' = \sum_{j=p,2p,..}^M$,

$$\Pr\left(\inf_{\Theta_2} S(c) \le 0\right) \to 0 \text{ as } T \to \infty, \tag{S.15}$$

where

$$S(c) = \log\frac{G_{\widetilde{v}}^p(c)}{G_{\widetilde{v}}^p(\delta)} - 2(c-\delta)\frac{p}{M}\sum\nolimits'' \log\lambda_j,$$

$$M^{-1}\sum\nolimits'' (\log j)^k \lambda_j^{2\delta}\left(I_v^p(\lambda_j) - I_{\widetilde{v}}^p(\lambda_j)\right) = o_p(1), \quad k = 0, 1, 2, \tag{S.16}$$

and

$$M^{-\frac{1}{2}}\sum\nolimits''\left(\log j - \frac{p}{M}\sum\nolimits'' \log k\right)\lambda_j^{2\delta}\left(I_v^p(\lambda_j) - I_{\widetilde{v}}^p(\lambda_j)\right) = o_p(1). \tag{S.17}$$

We first show (S.14). Now

$$\frac{G_v^p(c) - G_{\widetilde{v}}^p(c)}{G^p(c)} = \frac{\frac{p}{M}\sum'' \left(\frac{j}{M}\right)^{2(c-\delta)} \lambda_j^{2\delta}\left(I_v^p(\lambda_j) - I_{\widetilde{v}}^p(\lambda_j)\right)}{f_u(0)\frac{p}{M}\sum'' \left(\frac{j}{M}\right)^{2(c-\delta)}},$$

so that, by a similar reasoning to that in the proof of Theorem 5 of Velasco (1999a), the left side of (S.14) is bounded by

$$K \sup_{c\in\Theta_1} \frac{p}{M} \sum\nolimits'' \left(\frac{j}{M}\right)^{2(c-\delta)} \lambda_j^{2\delta} \left| I_v^p(\lambda_j) - I_{\widetilde{v}}^p(\lambda_j) \right| \leq K \frac{p}{M} \sum\nolimits'' \left(\frac{j}{M}\right)^{-1+2\epsilon} \lambda_j^{2\delta} \left| I_v^p(\lambda_j) - I_{\widetilde{v}}^p(\lambda_j) \right|$$
$$= O_p\left(M^{-1} \sum\nolimits'' \left(\frac{j}{M}\right)^{-1+2\epsilon} j^{\delta-r-1} \right) = O_p\left(M^{-2\epsilon}\right),$$

according to the Theorem of Robinson (2005), since $\delta - r < 1/2$ and $\epsilon < 1/4$, to justify (S.14). Next, we show (S.15). Setting $z = \exp\left(pM^{-1}\sum'' \log j\right)$, we have

$$\begin{aligned}\Pr\left(\inf_{\Theta_2} S(c) \leq 0\right) &= \Pr\left(\inf_{\Theta_2} \frac{p}{M} \sum\nolimits'' \left[\left(\frac{j}{z}\right)^{2(c-\delta)} - 1\right] \lambda_j^{2\delta} I_{\widetilde{v}}^p(\lambda_j) \leq 0\right) \\ &\leq \Pr\left(\frac{p}{M}\sum\nolimits'' \left[a_j - 1\right] \lambda_j^{2\delta} I_{\widetilde{v}}^p(\lambda_j) \leq 0\right), \end{aligned} \tag{S.18}$$

where

$$a_j = \begin{cases} \left(\frac{j}{z}\right)^{-1+2\epsilon}, \ 1 \leq j \leq z, \\ \left(\frac{j}{z}\right)^{2(\nabla_1-\delta)}, \ z < j \leq M. \end{cases}$$

Velasco (1999a) shows (see p.115) that $\Pr\left(\frac{p}{M}\sum'' \left[a_j - 1\right] \lambda_j^{2\delta} I_v^p(\lambda_j) \leq 0\right) = o(1)$, so that (S.18) is $o(1)$ by showing

$$\frac{p}{M}\sum\nolimits'' (a_j - 1) \lambda_j^{2\delta} \left(I_{\widetilde{v}}^p(\lambda_j) - I_v^p(\lambda_j)\right) = o_p(1). \tag{S.19}$$

According to the Theorem of Robinson (2005) the left side of (S.19) is bounded by

$$KM^{-1}\sum\nolimits_{j=p,2p,..}^{z} \left(\frac{j}{z}\right)^{-1+2\epsilon} j^{\delta-r-1} + KM^{-1}\sum\nolimits_{j=z+p,z+2p,..}^{M} \left(\frac{j}{z}\right)^{2(\nabla_1-\delta)} j^{\delta-r-1} + KM^{-1}\sum\nolimits'' j^{\delta-r-1},$$

which noting that z asymptotically equals M/e is $O\left(M^{-2\epsilon} + M^{-1+\delta-r} + M^{-1}\log M\right) = o(1)$ because $\delta - r < 1/2$. The proof of (S.16) is omitted as it almost identically follows arguments to that for (S.17). Finally, in line with previous arguments, the expectation of the absolute value of the left of (S.17) is bounded by

$$KM^{-\frac{1}{2}} \log M \sum\nolimits'' j^{\delta-r-1} = o(1),$$

to conclude the proof of $M^{1/2}$-consistency of $\widetilde{\delta}_G$.

Then, in line with previous arguments, the proof for $\widetilde{\delta}_{RH}$ holds, showing that

$$m^{-\frac{1}{2}}\left(S_v^p(\delta)-S_{\widetilde{v}}^p(\delta)\right)=o_p(1), \tag{S.20}$$

$$m^{-1}\left(H_v^p(\delta)-H_{\widetilde{v}}^p(\delta)\right)=o_p(1) \quad \text{and} \tag{S.21}$$

$$m^{-1}\left|\sum{}' k_{q,j}\left(\log\frac{j}{m}\right)^g\left(\lambda_j^{2\widehat{\delta}}-\lambda_j^{2\delta}\right)I_{\widetilde{v}}^p(\lambda_j)\right|=o_p(1), \tag{S.22}$$

for $g=0,1,2,3$, where $\widehat{\delta}-\delta=O_p\left(M^{-1/2}\right)$. First, (S.20) follows if

$$m^{-\frac{1}{2}}\sum{}' b_{q,j}\lambda_j^{2\delta}\left(I_v^p(\lambda_j)-I_{\widetilde{v}}^p(\lambda_j)\right)=o_p(1) \quad \text{and} \tag{S.23}$$

$$m^{-1}\sum{}' k_{q,j}\lambda_j^{2\delta}\left(I_v^p(\lambda_j)-I_{\widetilde{v}}^p(\lambda_j)\right)=o_p(1). \tag{S.24}$$

We just give the proof for (S.23), as that for (S.24) is significantly simpler. The expectation of the absolute value of the left side of (S.23) is bounded by

$$Km^{-\frac{1}{2}}\log T\sum{}'\left\{E\left(\lambda_j^{2\delta}\left|w_v^p(\lambda_j)-w_{\widetilde{v}}^p(\lambda_j)\right|^2\right)E\left(\lambda_j^{2\delta}\left|w_v^p(\lambda_j)+w_{\widetilde{v}}^p(\lambda_j)\right|^2\right)\right\}^{\frac{1}{2}},$$

which, by according to the Theorem of Robinson (2005) and results in Velasco (1999b) is

$$O\left(m^{-\frac{1}{2}}\log T\sum{}' j^{\delta-r-1}\right)=O\left(m^{-\frac{1}{2}}\log T\left(m^{\delta-r}1(\delta>r)+\log m1(\delta=r)+1(\delta<r)\right)\right),$$

which is $o(1)$ by (6), since $\delta-r<1/2$. Similarly, (S.21) holds because for $g=0,1,2$,

$$\begin{aligned}&m^{-1}\sum{}' k_{q,j}(\log\lambda_j)^g\lambda_j^{2\delta}\left(I_v^p(\lambda_j)-I_{\widetilde{v}}^p(\lambda_j)\right)\\&=O\left(m^{-1}\log^2 T\left(m^{\delta-r}1(\delta>r)+\log m1(\delta=r)+1(\delta<r)\right)\right)=o(1),\end{aligned}$$

by (6). Next, as in the discussion of (S.11), (S.22) holds if

$$m^{-1}\sum{}'\left|\log\frac{j}{m}\right|^g\left|\log\lambda_j\right|\lambda_j^{-2\left|\widehat{\delta}-\delta\right|}\lambda_j^{2\delta}\left|I_v^p(\lambda_j)-I_{\widetilde{v}}^p(\lambda_j)\right|=o_p(1), \tag{S.25}$$

which can be easily justified by previous arguments, to conclude the proof of the proposition.

Proof of Proposition 2. The proof for $\left(\overline{\delta}_{AS},\overline{\phi}_{AS}'\right)'$ follows from adapting Lemmae 2 and 3 in Andrews and Sun (2004) (AS hereinafter) to the tapered case. We initially show that the equivalent of AS' Lemma 2 holds for our tapered Hessian and score vectors (given by the tapered version of (4.2) in AS, noting that the summation now runs from

$p, 2p, \ldots$ to m). Defining

$$B_T^p = \left(\frac{m}{p}\right)^{\frac{1}{2}} diag\left(1, (\lambda_m)^2, \ldots, (\lambda_m)^{2q}\right), \; J_T^p = \sum_{j=p,2p,..}^{m} \widehat{X}_j \widehat{X}_j',$$

$$\widehat{X}_j = X_j - \frac{p}{m} \sum_{k=p,2p,..}^{m} X_k, \; X_j = \left(2\log j, (\lambda_j)^2, \ldots, (\lambda_j)^{2q}\right)' \quad \text{and}$$

$$\widetilde{X}_j = \left(2\log j, (j/m)^2, \ldots, (j/m)^{2q}\right)',$$

we first show the equivalent of Lemma 2 (a), which in our case is

$$\left(B_T^p\right)^{-1} J_T^p \left(B_T^p\right)^{-1} \to \Omega_q \text{ as } T \to \infty.$$

In view of Lemma 5 of Velasco (1999a), it suffices to show that for the different choices $a(j) = \log j, \log^2 j, j^k \log j/m^k, j^k/m^k$ for any finite $k \geq 2$,

$$\frac{1}{m} \int_0^m a(x)\, dx - \frac{p}{m} \sum_{j=p,2p,..}^{m} a(j) = o(1), \tag{S.26}$$

as $m \to \infty$. The left side of (S.26) can be written as m^{-1} times

$$-pa(p) + \int_0^p a(x)\, dx - \sum_{j=2p,3p,..}^{m} \int_{j-p}^{j} (a(j) - a(x))\, dx,$$

and (S.26) follows noting that

$$\begin{aligned} \sup_{j-p \leq x \leq j} |a(j) - a(x)| &\leq K j^{-1} \log j, \text{ for } a(j) = \log j, \; \log^2 j, \\ &\leq K j^{k-1} m^{-k} \log j, \text{ otherwise.} \end{aligned}$$

Next, the equivalent of AS' Lemma 2(b) holds if

$$\left\| \left(B_T^p\right)^{-1} \left(H_T^p(\delta, \phi(q)) - J_T^p\right) \left(B_T^p\right)^{-1} \right\| = o_p(1).$$

Here, apart from the obvious change of notation and changing $m^{-1} \sum_{j=1}^{m}$ by $pm^{-1} \sum_{j=p,2p,..}^{m}$, the only critical point refers to the order of

$$\sum_{j=p,2p,..}^{k} \left(\frac{I_v^p(\lambda_j)}{g_j} - 2\pi I_\xi^p(\lambda_j) \right), \tag{S.27}$$

where g_j is given in (A.10) in AS. Following arguments given in Velasco (1999a) (c.f. (A23)) and AS (c.f. (A.13)(i)) , it is straightforward to show that (S.27) is

$$O_p\left(k^{1-(s+\eta)/2}1\left(s+\eta<2\right)+\log k1\left(s+\eta\geq 2\right)+k^{\delta-p+1}\log^{\frac{1}{2}}k+k^{s+\eta+1}T^{-(s+\eta)}\right), \tag{S.28}$$

the rest of the proof following easily.

In line with previous arguments, the proof for our corresponding result to Lemma 2(c) and 2(d) follows straightforwardly, but the treatment for the corresponding proof of Lemma 2(e) needs to be discussed in more detail. Following AS' proof, concerning the corresponding result for $T_{1,T}$, the most delicate issue is to find an equivalent to the bound in (A.21) in AS. Clearly, the equivalent to (A.29) in AS in our setting is A_1+A_2, where

$$A_1=\sum_{j=p,2p,..}^{k}\left[\frac{I_v^p\left(\lambda_j\right)}{f_v\left(\lambda_j\right)}-2\pi I_\xi^p\left(\lambda_j\right)-E\left(\frac{I_v^p\left(\lambda_j\right)}{f_v\left(\lambda_j\right)}-2\pi I_\xi^p\left(\lambda_j\right)\right)\right]\frac{f_v\left(\lambda_j\right)}{g_j}\quad\text{and}$$

$$A_2=2\pi\sum_{j=p,2p,..}^{k}\left(I_\xi^p\left(\lambda_j\right)-E\left(I_\xi^p\left(\lambda_j\right)\right)\right)\left(\frac{f_v\left(\lambda_j\right)}{g_j}-1\right),$$

(c.f. (A.29) in AS), where $f_v\left(\lambda\right)$ is the pseudo-spectral density function of the process v_t. Now, following the bounds given in the proof of Theorem 5 of Velasco (1999a) and the proof of AS,

$$E\left(A_2^2\right)\leq K\left\{\sum_{j=p,2p,..}^{k}\lambda_j^{2(s+\eta)}+\sum_{i=2p,..}^{k}\sum_{j=p,2p,..}^{i-p}\left[(i-j)^{-2p}+(i+j)^{-2p}+T^{-1}\right]\lambda_i^{s+\eta}\lambda_j^{s+\eta}\right\}.$$

First,

$$\begin{aligned}\sum_{i=2p,..}^{k}\sum_{j=p,2p,..}^{i-p}(i-j)^{-2p}\lambda_i^{s+\eta}\lambda_j^{s+\eta}&=\left(\frac{2\pi}{T}\right)^{2(s+\eta)}\sum_{i=p,2p,..}^{k-p}i^{-2p}\sum_{j=p,2p,..}^{k-i}j^{s+\eta}\left(j+i\right)^{s+\eta}\\&\leq KT^{-2(s+\eta)}k^{2(s+\eta)+1}.\end{aligned}$$

Next,

$$\begin{aligned}\sum_{i=p,2p,..}^{k}\sum_{j=p,2p,..}^{i-p}(i+j)^{-2p}\lambda_i^{s+\eta}\lambda_j^{s+\eta}&\leq KT^{-2(s+\eta)}\sum_{i=p,2p,..}^{k}\sum_{j=p,2p,..}^{i-p}\frac{(ij)^{s+\eta}}{j^{2p}}\\&\leq KT^{-2(s+\eta)}\left\{k^{s+\eta+1}\log k+k^{2+2(s+\eta-p)}\right\},\end{aligned}$$

implying that

$$E\left(A_2^2\right) = O\left(T^{-2(s+\eta)}k^{2(s+\eta)+1}\right).$$

Next, as in (S.28),

$$E\left(A_1^2\right) = O\left(k^{2-(s+\eta)} + \log^2 k + k^{2(\delta-p+1)}\log k + k^{2(s+\eta+1)}T^{-2(s+\eta)}\right),$$

and following the proof in AS (pp. 601-602), it is straightforward to show that

$$T_{1,T} = O_p\left(m^{\frac{1-(s+\eta)}{2}} + m^{-\frac{1}{2}}\log m + m^{\delta-p+\frac{1}{2}}\log^{\frac{1}{2}} m + \left(\frac{m}{T}\right)^{s+\eta}\right) = o_p(1).$$

Moreover, noting that following Velasco (1999a, p. 113)

$$E\left(\frac{I_v^p(\lambda_j)}{f_v(\lambda_j)} - 1\right) = O\left(j^{-1} + j^{2(\delta-p)}\log j\right),$$

it is straightforward to show that

$$T_{2,T} = O\left(m^{-\frac{1}{2}}\log^2 m + m^{2(\delta-p)+\frac{1}{2}}\log^2 m\right) = o_p(1).$$

Next

$$T_{3,T} = \left(\frac{p}{m}\right)^{\frac{1}{2}}\sum_{j=p,2p,..}^{m}\left(2\pi I_\xi^p(\lambda_j) - 1\right)\zeta_j,$$

where

$$\zeta_j = \tau'\left(\widetilde{X}_j - \frac{p}{m}\sum_{k=p,2p,..}^{m}\widetilde{X}_k\right),$$

for any arbitrary $(q+1)\times 1$ vector $\tau \neq 0$, $\|\tau\| < \infty$. The result

$$T_{3,T} \rightarrow_d N(0, \Phi\Omega_q),$$

is straightforward in view of Lemma 6 in Velasco (1999a), just noting that in line with previous arguments

$$\frac{p}{m}\sum_{j=p,2p,..}^{m}\zeta_j^2 \rightarrow \tau'\Omega_q\tau \text{ as } T \rightarrow \infty,$$

and

$$|\zeta_j - \zeta_{j+p}| \leq \|\tau\|\left\|\widetilde{X}_j - \widetilde{X}_{j+p}\right\| \leq Kj^{-1}.$$

By the proof of AS, it is easy to show that

$$T_{4,T} = \left(\frac{p}{m}\right)^{\frac{1}{2}} \sum_{j=p,2p,..}^{m} \left(\frac{f_v(\lambda_j)}{g_j} - 1\right)\left(\widetilde{X}_j - \frac{p}{m}\sum_{k=p,2p,..}^{m}\widetilde{X}_k\right) = -\nu_T(q,s) + o(1),$$

to finish the proof of our equivalent to Lemma 2. Finally, the proof of the equivalent to Lemma 3 follows straightforwardly as in AS.

We next show the result for the estimator based on $\widetilde{v}_t$. First, taking into account that, as $T \to \infty$, uniformly in $j \in \{p, 2p, ..., m\}$

$$\sup_{\gamma \in \Xi} \exp\left(\left(\kappa_q(\lambda_j, \gamma)\right)\right) = 1 + o(1),$$

in view of the different steps in the proof of Lemma 2 of AS, the result follows on showing

$$m^{-\frac{1}{2}} \sum_{j=p,2p,..}^{m} \log j \left| I_{\widetilde{v}}^{p}(\lambda_j) - I_{v}^{p}(\lambda_j) \right| \lambda_j^{2\delta} = o_p(1), \tag{S.29}$$

which implies that the corresponding of Lemmae 2(b), 2(c) and 2(e) follow using the process $\widetilde{v}_t$ instead of v_t, and

$$\sup_{c:|c-\delta| \le \frac{K}{\log^5 m}} m^{-1} \sum_{j=p,2p,..}^{m} \left| I_{\widetilde{v}}^{p}(\lambda_j) - I_{v}^{p}(\lambda_j) \right| \lambda_j^{2c} \log^2 j = o_p\left(\log^{-2} m\right), \tag{S.30}$$

for any finite $K > 0$, which implies that the equivalent of Lemma 2(d) holds for $\widetilde{v}_t$. First, (S.29) follows by almost identical arguments to those in the proof of (S.23). Next, the left side of (S.30) is bounded by

$$K T^{\frac{2K}{\log^5 m}} m^{-1} \log^2 m \sum_{j=p,2p,..}^{m} \left| I_{\widetilde{v}}^{p}(\lambda_j) - I_{v}^{p}(\lambda_j) \right| \lambda_j^{2\delta}.$$

Noting that by Assumption 4, $m^{(1+1/4q)} T^{-1} \to \infty$ as $T \to \infty$, $T^{2K/\log^5 m}$ is dominated by $m^{2K(1+1/4q)/\log^5 m}$, and (S.30) follows by (S.29) just noting that

$$m^{\frac{2K(1+1/4q)}{\log^5 m}} = \exp\left(\log\left(m^{\frac{2K(1+1/4q)}{\log^5 m}}\right)\right) = \exp\left(\frac{2K(1+1/4q)}{\log^4 m}\right) \to 1 \text{ as } T \to \infty.$$

REFERENCES

Andrews, D.W.K. & Y. Sun (2004) Adaptive local polynomial Whittle estimation of long-range dependence. Econometrica 72, 569-614.

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

Robinson, P.M. (2005) The distance between rival nonstationary fractional processes. Journal of Econometrics 128, 283-300.

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

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