



ELSEVIER

Journal of Econometrics 114 (2003) 1–27

JOURNAL OF
Econometrics

www.elsevier.com/locate/econbase

Higher-order kernel semiparametric M-estimation of long memory

Peter M. Robinson^{a,*}, Marc Henry^b

^a*Department of Economics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK*

^b*Department of Economics, Columbia University, New York, NY 10027, USA*

Abstract

Econometric interest in the possibility of long memory has developed, as a flexible alternative to or compromise between the usual short memory or unit root prescriptions, for example in the context of modelling cointegrating or other relationships and in describing the dependence structure of nonlinear functions of financial returns. Semiparametric methods of estimating the memory parameter can avoid bias incurred by misspecification of the short memory component. We introduce a broad class of such semiparametric estimates that also covers pooling across frequencies. A leading “Box-Cox” sub-class, indexed by a single tuning parameter, interpolates between the popular local log-periodogram and local Whittle estimates, leading to a smooth interpolation of asymptotic variances. The bias of these two estimates also differs to higher order, and we also show how bias, and asymptotic mean-squared error, can be reduced, across the class of estimates studied, by means of a suitable version of higher-order kernels. We then calculate an optimal bandwidth (the number of low-frequency periodogram ordinates employed) which minimizes this mean-squared error. Finite sample performance is studied in a small Monte Carlo experiment, and an empirical application to intra-day foreign exchange returns is included. © 2002 Published by Elsevier Science B.V.

JEL classification: C22

Keywords: Long memory; Semiparametric methods; Higher-order kernel; M-estimation; Bias; Mean-squared error

1. Introduction

Semiparametric modelling has become popular in the investigation of possible long memory in economic and financial time series. A correct, fully parametric, model for

* Corresponding author. Tel.: +44-20-7955-7516; fax: +44-20-7831-1840.

E-mail address: p.m.robinson@lse.ac.uk (P.M. Robinson).

autocorrelations over all lags, or equivalently for the spectral density over all frequencies, computed from a series of length n , advantageously produces \sqrt{n} -consistent estimates of memory parameters. However, long memory is essentially a low-frequency phenomenon, and semiparametric modelling consequently entails a parameterization of the spectral density only near frequency zero, or equivalently of autocorrelations at only long lags. Hence, there is potential to avoid the misspecification incurred by incorrect parametric modelling, such as by a fractionally integrated autoregressive moving average (FARIMA) model, when either the autoregressive or moving average order is under-specified, or both are over-specified; such misspecification can produce inconsistent estimates not only of the autoregressive and moving average parameters, which essentially describe short memory behaviour, but also of the memory parameter, which measures any long memory. Although semiparametric estimates are less-than- \sqrt{n} -consistent, they do achieve consistency under broad conditions, without running such risk of misspecification, so their greater robustness might then be preferred, at least when n is large.

A semiparametric approach is consistent with standard econometric practice nowadays, in connection with autocorrelation-consistent variance estimation and unit root time series analysis. Both involve $I(0)$ series, by which we mean covariance stationary (or asymptotically covariance stationary) series with spectral density that is finite and positive at frequency zero, and thus exhibit neither long memory nor antipersistence. In the first case, a statistic which can be approximated by a partial sum of functions of $I(0)$ series has asymptotic variance which involves a quantity that itself can be interpreted as the spectral density of a certain $I(0)$ series at frequency zero (see e.g. Hannan, 1957). In the second case, due to the preponderance of low-frequency power in unit root series, correction for autocorrelation need entail estimation of the spectral density of the $I(0)$ input series at only zero frequency (see e.g. Phillips, 1987). In fact, criteria for semiparametric estimation of long memory have been used in testing the $I(0)$ hypothesis (see Lobato and Robinson, 1998), while autocorrelation-consistent variance estimates for the sample mean of a long memory series developed by Robinson (1994) involve semiparametric memory estimates.

We define the memory (or self-similarity) parameter H of a covariance stationary time series x_t , $t = 0, \pm 1, \dots$, by means of the semiparametric representation

$$f(\lambda) \sim C\lambda^{1-2H} \quad \text{as } \lambda \rightarrow 0+ \quad (1.1)$$

for the spectral density $f(\lambda)$ of x_t , defined by $\text{Cov}(x_0, x_j) = \int_{-\pi}^{\pi} f(\lambda) \cos j\lambda \, d\lambda$. In (1.1) “ \sim ” indicates that the ratio of left- and right-hand sides tends to one, and

$$0 < C < \infty, \quad (1.2)$$

$$0 < H < 1. \quad (1.3)$$

When $0 < H < 0.5$ (so $f(0) = 0$), there is said to be antipersistence or negative memory; when $H = 0.5$ (so $0 < f(0) < \infty$), there is said to be short memory; and when $0.5 < H < 1$ (so $f(0) = \infty$), there is said to be long memory. As discussed above, (1.1) is silent about the behaviour of $f(\lambda)$ away from frequency zero; though it must be integrable, nonnegative, and even on $[-\pi, \pi]$, it need not be smooth and can have

poles and/or zeroes. Indeed, Robinson (1995a, b) showed that two leading “semiparametric” estimates of H , based on an observed sequence x_t , $t = 1, \dots, n$, have desirable asymptotic properties in such a broad setting—these are the log-periodogram estimate which originated in Geweke and Porter-Hudak (1983), and the semiparametric Gaussian or local Whittle estimate which originated in Künsch (1987).

Both estimates depend on a bandwidth parameter m , being the number of low-frequency periodogram ordinates employed in the estimation, and both are $m^{1/2}$ -consistent, where m increases as n increases, but more slowly. Indeed their mean-squared errors are of order no smaller than $n^{-4/5}$ (Henry and Robinson, 1996; Hurvich et al., 1998), compared to the n^{-1} order of estimates based on a full parametric specification of $f(\lambda)$. This slow convergence is of concern, especially in macroeconomic series, where n is unlikely to be very large. In particular, semiparametric memory estimates are useful in fractional cointegration analysis, where observables can be non-stationary, or even stationary, fractional processes, and cointegrating errors can be $I(0)$ or have long memory, for example in connection with purchasing power parity analysis and volume-volatility comovements. Parametric modelling of dynamic structure is possible, but again there is danger of misspecification, and a semiparametric approach again fits in with ones used in standard cointegration analysis, which assumes that observables have unit roots and cointegrating errors are $I(0)$ (see, e.g. Phillips, 1991). Note also that in cointegration or other analysis of nonstationary series, where first differences may have long memory or be antipersistent, nonstationary memory parameters arise. Here, in contrast with the superconsistent rates of, say, autoregressive coefficient estimates, semiparametric memory estimates are still only \sqrt{m} -consistent (see Velasco, 2000) (just as parametric estimates are still only \sqrt{n} -consistent). Thus, a strong incentive for improving semiparametric memory estimates is identified.

Recently, Moulines and Soulier (1999), Hurvich and Brodsky (2001) have shown that a global version of log-periodogram estimation, in which robustness to high-frequency behaviour is achieved by a form of series expansion across all frequencies, rather than by omitting high frequencies, is capable of achieving a mean-squared error of order $\log n/n$. However, this depends on $f(\lambda)/\lambda^{1-2H}$ being globally analytic, across $[-\pi, \pi]$, which is true for FARIMA models but not, for example, for Gegenbauer models (see Gray et al., 1989), in which a spectral pole at some nonzero frequency is entertained, due to cyclic behaviour.

A striking feature of the asymptotic theory of Robinson (1995a, b) is that the regularity conditions correspond to those required in nonparametric spectral estimation for $I(0)$ series, due to the fact that $f(\lambda)\lambda^{2H-1}$ is itself the spectrum of an $I(0)$ series. These conditions are purely local, unlike those of Moulines and Soulier (1999), Hurvich and Brodsky (2001). Now improvements to nonparametric spectral estimates are long established (see e.g. Parzen, 1957; Cogburn and Davis, 1974), where in particular a lag window in a weighted sum of sample autocovariances is chosen to exploit a supposed degree of spectral smoothness (specifically the number of derivatives existing), in order to achieve a squared bias and a variance (and thence a mean-squared error) of order $n^{-\kappa}$, where $\kappa < 1$ but the index of smoothness κ can approach arbitrarily close to 1. In the frequency domain, this corresponds to periodogram averaging using a spectral window that behaves rather like the sort of higher-order kernel employed

in nonparametric probability density estimation. Because a frequency-domain approach seems much more elegant than a time-domain one in our setting, we employ an analogous approach in estimating H , and are again able to achieve a mean-squared error of order $n^{-\kappa}$, $\kappa < 1$, which can be nearly as good as that of [Moulines and Soulier \(1999\)](#) and [Hurvich and Brodsky \(2001\)](#), but imposes only local assumptions. The kernels that arise in our case differ from those in the spectral density estimation problem, and while both are perforce negative over some part of their support, nothing corresponding to the worrying consequences of a negative spectral density emerges in our estimation of H .

One feature of the semiparametric long memory literature is the wide variety of estimates that has emerged. A number of these has the disadvantage of a nonstandard limit distribution for some or all H , or, if asymptotically normal for some values of H , having asymptotic variance depending on H in a complicated way, while their convergence rate can also depend on H . This behaviour is perhaps associated with the relatively ad hoc character of such estimates, compared to the log-periodogram and Gaussian semiparametric estimates, which both employ a suitable “whitening” (like classical least squares and maximum likelihood estimates) and are asymptotically normal for all H , with rates of convergence and (simple) asymptotic variances that are free of H . Nevertheless, the dichotomy in the literature between the log-periodogram and semiparametric Gaussian estimate is still of some concern, as they tend to be treated individually, whereas they appear to be related and should be eligible for similar modification and, to some extent, be treatable simultaneously; the most obvious differences are that the former can be computed in closed form but is technically rather difficult to handle due to the logging of individual periodograms, whereas the second is only implicitly defined but more efficient. In the present paper we find that both are special cases of a class of M-estimates, so we can apply our higher-order kernel improvements to them both simultaneously (as well as to uncountably many other members of the class), thereby providing a unifying interpretation.

We focus on scalar stationary series satisfying (1.1), but it seems our higher-order kernel and M-estimation theory are extendable to multivariate series (cf. [Robinson, 1995a](#); [Lobato, 1999](#)), to nonstationary series (cf. [Velasco, 1999](#)), and to seasonal and cyclic long memory (cf. [Arteche, 2000](#)), while the methods pursued by [Moulines and Soulier \(1999\)](#), [Hurvich and Brodsky \(2001\)](#) might also be included in a more general M-estimate class.

While we provide some justification, with relevant conditions, for our results, we do not attempt anything approaching full, rigorous proofs, or a full set of sufficient conditions. Reasonably complete proofs would be extremely lengthy. This is in part due to the implicit nature of many of our estimates, which require an initial consistency proof; in fact, as in other applications of higher-order kernels, our estimate of H solves first-order conditions but does not in general correspond to a global minimum of an objective function, so that only existence of a consistent root can be established. In general, dealing with actual means and variances of our estimates is very difficult. We adopt an approach of a familiar type, assuming consistency, approximating our estimates by relatively simple statistics by neglecting terms of smaller order in probability, and then approximating the moments of these simple statistics. We do not even provide

full details of the latter step, because they are similar to ones in [Robinson \(1995a, b\)](#) and [Hurvich et al. \(2001\)](#), while a consistency proof would also follow in part that of [Robinson \(1995b\)](#); detailed proofs of this type thus no longer seem to have much novel interest in the semiparametric memory estimation context.

The following section introduces notation and our class of estimates. Aside from the M-estimation and higher-order kernel aspects discussed above, we consider two additional ones. One is the pooling across frequencies introduced by [Robinson \(1995a\)](#) to reduce the variance of log-periodogram estimates; there is interest in the effect of such pooling on other members of our class. We also allow for different implementations of the estimates corresponding to alternative asymptotically equivalent versions of (1.1), a leading one being

$$f(\lambda) \sim C \left(2 \sin \frac{\lambda}{2} \right)^{1-2H} \quad \text{as } \lambda \rightarrow 0 +. \tag{1.4}$$

For example, [Geweke and Porter-Hudak’s \(1983\)](#) original version of the log-periodogram estimate is based on (1.4), while [Robinson’s \(1995a\)](#) is based on (1.1), but both have the same first-order asymptotic properties (see [Robinson, 1995a](#)). However, the errors in the approximations in (1.1) and (1.4) can differ from each other, and this can affect a higher-order analysis.

Section 3 presents results for bias, variance, mean-squared error and optimal bandwidth. Their derivations are discussed in Section 4, and a special case is treated in Section 5. Monte Carlo simulations and an empirical application appear in, respectively, Sections 6 and 7.

2. Higher-order kernel M-estimates

For any integer $q \geq 1$, we introduce a user-chosen real-valued function $k_q(u)$, $0 \leq u \leq 1$, such that

$$\int_0^1 k_q(u) \, du = 1. \tag{2.1}$$

Also, defining

$$U_{iq} = \int_0^1 \{\log(u) + 1\} u^{2i} k_q(u) \, du, \quad 0 \leq i \leq q, \tag{2.2}$$

we suppose that $k_q(u)$ satisfies

$$U_{iq} = 0, \quad 0 \leq i \leq q - 1, \tag{2.3}$$

$$U_{qq} \neq 0. \tag{2.4}$$

We call $k_q(u)$ satisfying (2.1)–(2.4) a q th-order kernel. Perhaps the most analytically convenient class of $k_q(u)$ is

$$k_q^*(u) = \sum_{j=0}^q a_j u^{2j}, \quad 0 \leq u \leq 1 \tag{2.5}$$

for which, in Section 5, we derive the explicit formulae for a_j to satisfy (2.1)–(2.4).

Denote by $\psi(z)$ a user-chosen, real-valued monotonic function. We shall stress in particular the “Box-Cox” class

$$\psi_\alpha(z) = \frac{z^\alpha - 1}{\alpha}, \quad z > 0, \quad \alpha > 0, \tag{2.6}$$

$$\psi_0(z) = \log z, \quad z > 0. \tag{2.7}$$

Introduce the periodogram

$$I(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^n x_t e^{it\lambda} \right|^2, \tag{2.8}$$

based on a record of observations x_t , $1 \leq t \leq n$, and for brevity denote

$$I_j = I(\lambda_j), \quad \lambda_j = \frac{2\pi j}{n} \tag{2.9}$$

for integer j . To allow for pooling, for a fixed integer $J \geq 1$, introduce the J -averaged periodogram

$$F_{jj} = \frac{1}{J} \sum_{l=j-J+1}^j I_l.$$

Denote by m a user-chosen integer such that $1 \leq m < n/2$. Our estimates use the I_j for $1 \leq j \leq m$ only, or, more precisely, for $1 \leq j \leq [m/J]J$, $[\cdot]$ denoting integer part.

Finally denote by $g(\lambda)$ a user-chosen nonnegative function that is asymptotic to λ to the extent that

$$g(\lambda) = \lambda + G\lambda^3 + o(\lambda^3) \quad \text{as } \lambda \rightarrow 0+. \tag{2.10}$$

Two leading choices of $g(\lambda)$ are (cf. (1.1))

$$g^{(1)}(\lambda) = \lambda, \tag{2.11}$$

where $G = 0$, and (cf. (1.4))

$$g^{(2)}(\lambda) = 2 \sin \lambda/2, \tag{2.12}$$

where $G = -\frac{1}{24}$.

We now define a q th-order kernel M-estimate of H , $\hat{H}_{qm}(\psi, g, k, J)$, as a solution of the equation

$$\sum_j' k_{qj} v_{qj}(g) \psi(F_{jj} g_j^{2\hat{H}_{qm}(\psi, g, k, J)-1}) = 0, \tag{2.13}$$

where \sum_j' throughout denotes a sum over $j = J, 2J, \dots, [m/J]J$, and

$$v_{qj}(g) = \log g_j - \frac{\sum_j' k_{qj} \log g_j}{\sum_j' k_{qj}}, \tag{2.14}$$

$$g_j = g(\lambda_j), \quad k_{qj} = k_q \left(\frac{j}{m} \right). \tag{2.15}$$

The omission of frequency $\lambda_0=0$ (and $\lambda_n=2\pi$) from (2.13) permits x_t to have unknown mean, since I_j is invariant to location shift for $1 \leq j < n/2$.

To relate (2.13) to the existing literature on semiparametrically estimating H , we focus on the extreme choices of α in (2.6) and (2.7), where $\alpha = 0$ gives

$$\hat{H}_{qm}(\psi_0, g, k, J) = \frac{1}{2} \left\{ 1 - \frac{\sum_j' k_{qj} v_{qj}(g) \log F_{jj}}{\sum_j' k_{qj} v_{qj}(g) \log g_j} \right\} \tag{2.16}$$

and $\alpha = 1$ gives

$$\sum_j' k_{qj} v_{qj}(g) \{ F_{jj}^{2\hat{H}_{qm}(\psi_1, g, k, J) - 1} - 1 \} = 0. \tag{2.17}$$

Using formulae in Section 5 (2.3) for $i = 0, q = 1$ implies $a_1 = 0$ in (2.5), so that, applying also (2.1),

$$k_1^*(u) = a_0 = 1, \tag{2.18}$$

to explain why we took $q \geq 1$ at the start of the section. We thus find that $\hat{H}_{1m}(\psi_0, g^{(2)}, k^*, 1)$ is the log-periodogram estimate of Geweke and Porter-Hudak (1983), while $\hat{H}_{1m}(\psi_0, g^{(1)}, k^*, 1)$ is the log-periodogram estimate of Robinson (1995a), $\hat{H}_{1m}(\psi_0, g^{(1)}, k^*, J)$ is for $J > 1$ the pooled log-periodogram estimate of Robinson (1995a), whereas $\hat{H}_{1m}(\psi_1, g^{(1)}, k^*, 1)$ is the local Whittle estimate of Künsch (1987) and Robinson (1995b). (Strictly speaking, the latter estimate is defined as an extremum estimate, so if it falls on the boundary of the compact admissible parameter set, it need not solve the first-order conditions (2.17).) Choosing α between 0 and 1 interpolates between these estimates, and turns out to produce an intermediate asymptotic variance (when $J = 1$, between $\pi^2/24$ for $\alpha = 0$ and $1/4$ for $\alpha = 1$). On the other hand, for given ψ , a higher-order kernel, in particular taking $q \geq 2$ in (2.5), has the potential to reduce asymptotic bias, and thence asymptotic mean-squared error. The choice of g , among those satisfying (2.10), does not affect asymptotic variance but it can affect bias and thus mean-squared error, even possibly in terms of rate of convergence.

3. Asymptotic bias and mean-squared error

We refine (1.1) by assuming that

$$f(\lambda) = Cg^{1-2H}(\lambda)h(\lambda), \tag{3.1}$$

where (1.2), (1.3) and (2.10) hold, along with

$$h(\lambda) - 1 - \sum_{i=1}^q \frac{h_i \lambda^{2i}}{(2i)!} = O(\lambda^{2(q+1)}) \quad \text{as } \lambda \rightarrow 0+. \tag{3.2}$$

Note that h_i is the $2i$ th derivative of $h(\lambda)$ at $\lambda = 0$, so that (3.2) is tantamount to assuming that $f(\lambda)/g^{1-2H}(\lambda)$ is $2q$ -times continuously differentiable at $\lambda = 0$. In the central limit theory of Robinson (1995a, b) for standard log-periodogram and local Whittle estimates, q was allowed to be a continuous smoothness parameter. However, for pure FARIMA processes q can be arbitrarily large in (3.2), and whenever $q \geq 2$ we can exploit (3.2) by the q th-order kernel $k_q(u)$ to reduce the order of magnitude of the bias, and thence that of the mean-squared error.

For a given $f(\lambda)$, changing the choice of $g(\lambda)$ in (3.1) will affect the h_i , in a manner that depends also on H . For example, suppose we commence from the form

$$f(\lambda) = Cg^{(2)}(\lambda)^{1-2H}h^{(2)}(\lambda), \tag{3.3}$$

where $h^{(2)}(\lambda)$ is the spectrum of an autoregressive moving average (ARMA) process, so that x_t is a FARIMA. Then, we can instead write

$$f(\lambda) = Cg^{(1)}(\lambda)^{1-2H}h^{(1)}(\lambda), \tag{3.4}$$

where $h^{(1)}(\lambda) = \{g^{(2)}(\lambda)/g^{(1)}(\lambda)\}^{1-2H}h^{(2)}(\lambda)$. Indeed in case of a FARIMA(0, $H - 1/2$, 0) process, we have $h^{(2)}(\lambda) \equiv 1$, all of whose derivatives are zero, whereas $h^{(1)}(\lambda)$ has nonzero derivatives. For simplicity, we will later often suppress reference to g in our notation.

As acknowledged in Section 1, we do not list full regularity conditions but it is helpful to indicate what may constitute a large part of a set of sufficient conditions. We take x_t to be Gaussian. We take m to satisfy at least

$$m \rightarrow \infty, \quad m = O(n^{4q/(4q+1)}) \quad \text{as } n \rightarrow \infty.$$

For simplicity no trimming (omitting of contributions from the very lowest Fourier frequencies) was allowed for in (2.13). We can trim, and it may facilitate proofs for our general class, noting that Robinson (1995a) required trimming in his central limit theorem for $\hat{H}_{1m}(\psi_0, g^{(1)}, k^*, J)$. However, under somewhat different conditions, for the alternative log-periodogram estimate $\hat{H}_{1m}(\psi_0, g^{(2)}, k^*, 1)$, Hurvich et al. (1998) avoided trimming, while no trimming was required by Robinson (1995b) for the local Whittle estimate $\hat{H}_{1m}(\psi_1, g^{(1)}, k^*, 1)$. Some smoothness in $\psi(z)$ is necessary in order to demonstrate improvements as q increases, and for convenience we take it to be infinitely differentiable, such that, with $\psi^{(u)}(z)$ denoting u th derivative,

$$E|(CZ_J)^u \psi^{(u)}(CZ_J)| \leq K < \infty, \quad u \geq 1, \tag{3.5}$$

where K is a positive constant independent of u , and Z_J denotes a $\chi^2_{2J}/(2J)$ random variable. Notice that in case of (2.6) and (2.7), inequalities (3.5) are automatically satisfied—see (5.7). We take $k_q(u)$ to be boundedly differentiable except perhaps at finitely many points. Define, for $q \geq 1$,

$$V_q = \int_0^1 \{\log(u) + 1\}^2 k_q(u) du, \quad W_q = \int_0^1 \{\log(u) + 1\}^2 k_q(u)^2 du$$

and for $1 \leq i \leq q$,

$$Y_{iq} = \int_0^1 u^{2+2i} k_q(u) du - \int_0^1 u^2 k_q(u) du \int_0^1 u^{2i} k_q(u) du,$$

$$P_{ij}(\psi) = (2\pi)^{2i} \sum_{u=1}^i \frac{1}{u!} E((CZ_J)^u \psi^{(u)}(CZ_J)) \sum_{r_1+\dots+r_u=i} \left(\prod_{l=1}^u \frac{h_{r_l}}{(2r_l)!} \right),$$

$$Q_J(\psi) = JV(\psi(CZ_J)),$$

$$R_J(\psi) = 2E(CZ_J \psi^{(1)}(CZ_J)).$$

Bearing in mind that in general $\hat{H}_{qm}(\psi, g, k, J)$ is only implicitly defined, and remarks in Section 1, we consider asymptotic bias and asymptotic variance based on a formal expansion, first approximating $\hat{H}_{qm}(\psi, g, k, J) - H$ by $-A/B$, where

$$A = \frac{J}{m} \sum_j' k_{qj} v_{qj}(g) \psi(F_{jJ} g_j^{2H-1}),$$

$$B = \frac{2J}{m} \sum_j' k_{qj} v_{qj}(g) \log(g_j) F_{jJ} g_j^{2H-1} \psi^{(1)}(F_{jJ} g_j^{2H-1}).$$

Following arguments used in the discussion of A in the following section, we have

$$B = R_J(\psi) V_q + o_p(1).$$

We then take

$$AB = \text{Asybias}\{\hat{H}_{qm}(\psi, g, k, J)\} = -\frac{\text{Asy\textit{mea}}\{A\}}{R_J(\psi) V_q}, \tag{3.6}$$

$$AV = \text{Asyvar}\{\hat{H}_{qm}(\psi, g, k, J)\} = \frac{\text{Asy\textit{var}}\{A\}}{R_J^2(\psi) V_q^2}, \tag{3.7}$$

$$AM = \text{Asymse}\{\hat{H}_{qm}(\psi, g, k, J)\} = \frac{\text{Asy\textit{mse}}\{A\}}{R_J^2(\psi) V_q^2}, \tag{3.8}$$

where $\text{Asy\textit{mea}}\{A\}$ and $\text{Asy\textit{var}}\{A\}$ represent the leading terms in the mean and variance of an approximation to A that neglects terms of small order in probability, and $\text{Asy\textit{mse}}\{A\} = \text{Asy\textit{var}}\{A\} + (\text{Asy\textit{mea}}\{A\})^2$. We have

$$AB = -\frac{U_{qq} P_{qJ}(\psi) (m/n)^{2q} + (2\pi)^2 G Y_{1q} P_{1J}(\psi) (m/n)^4}{R_J(\psi) V_q}, \tag{3.9}$$

$$AV = \frac{Q_J(\psi) W_q}{\{R_J(\psi) V_q\}^2} \frac{1}{m} \tag{3.10}$$

$$AM = \left[Q_J(\psi) W_q / m + \left\{ U_{qq} P_{qJ}(\psi) \left(\frac{m}{n}\right)^{2q} + (2\pi)^2 G Y_{1q} P_{1J}(\psi) \left(\frac{m}{n}\right)^4 \right\}^2 \right] / \{R_J(\psi) V_q\}^2. \tag{3.11}$$

Then AM is minimized by $\hat{m} = m_q(\psi, q, k, J)$, given by

$$\hat{m} = \left[\frac{Q_J(\psi) W_q}{4q U_{qq}^2 P_{qJ}^2(\psi)} \right]^{1/(4q+1)} n^{4q/(4q+1)}, G = 0 \text{ or } q = 1, \tag{3.12}$$

$$\hat{m} = \frac{\{Q_J(\psi) W_q / 8\}^{1/9}}{\{U_{22} P_{2J}(\psi) + (2\pi)^2 G P_{1J}(\psi) Y_{12}\}^{2/9}} n^{8/9}, G \neq 0, q = 2, \tag{3.13}$$

$$\hat{m} = \left[\frac{Q_J(\psi)W_q}{32\pi^2 G^2 P_{1J}^2(\psi)Y_{1q}^2} \right]^{1/9} n^{8/9}, G \neq 0 \text{ and } q > 2. \tag{3.14}$$

The minimized MSE is given by

$$\frac{(4q + 1)[\{Q_J(\psi)W_q\}^{4q} U_{qq}^2 P_{qJ}^2(\psi)]^{1/(4q+1)}}{(4q)^{4q/(4q+1)} \{R_J(\psi)V_q\}^2 n^{4q/(4q+1)}}, G = 0 \text{ or } q = 1, \tag{3.15}$$

$$\frac{9\{Q_J(\psi)W_q\}^{8/9} \{U_{22}P_{2J}(\psi) + (2\pi)^2 GP_{1J}(\psi)Y_{12}\}^{2/9}}{8^{8/9} \{R_J(\psi)V_2\}^2 n^{8/9}}, G \neq 0 \text{ and } q = 2, \tag{3.16}$$

$$\frac{9[\{Q_J(\psi)W_q\}^8 \{(2\pi)^2 GP_{1J}(\psi)Y_{1q}\}^2]^{1/9}}{8^{8/9} \{R_J(\psi)V_q\}^2 n^{8/9}}, G \neq 0 \text{ and } q > 2. \tag{3.17}$$

The cases described in (3.13), (3.14), (3.16) and (3.17) are purely due to $G \neq 0$ in (2.10), that is, not taking $g(\lambda) = \lambda$. On the other hand, if we choose some other $g(\lambda)$, such as $2 \sin(\lambda/2)$ (2.12), the problem is avoided by replacing the factor u^{2i} by $g(u^{2i})$ in our specification of the U_{iq} in (2.2) (thereby, with (2.3), altering the precise specification of the $k_q(u)$) and considering an expansion for $h(\lambda)$ in powers of $g(\lambda)$ rather than λ in (3.2). We have stressed the specification (2.2) because of the consequent ability to determine $k_q(u)$ analytically, specifically the a_j in (2.5), and on the other hand we feel the implications of not thereby ‘matching’ $g(\lambda)$ with $k_q(u)$ are worth noting.

We mention connections with other work. Graf (1983) discussed M-estimation of parametric long memory models in the frequency domain, employing information from the entire Nyquist band, and with a heuristic treatment of asymptotic theory. On the other hand, Andrews and Guggenberger (2000) reduce the bias, and hence the mean-squared error, of the version of the log-periodogram estimate of Robinson (1995a). They do not employ higher order kernels, but instead incorporate suitable additional variables in the log-periodogram regression, to derive results that correspond to ours, providing rigorous proofs in their more specialized setting. A similar achievement is due to Andrews and Sun (2001), who propose a local polynomial modification to local Whittle estimation. This disadvantageously increases the dimension of the nonlinear optimization problem, though the additional parameter estimates are useful in optimal choice of m .

4. Partial details of derivations

We focus on the (semi)heuristic derivation of $Asym\{A\}$ and $Asyvar\{A\}$, and thence $Asymse\{A\}$, in (3.6)–(3.8). Write $A = A_1 + A_2 + A_3 + A_4 + A_5$, where

$$A_1 = \frac{J}{m} \sum_j' b_{qj} \{E\psi(Z_J f_j g_j^{2H-1}) - E\psi(CZ_J)\},$$

$$\begin{aligned}
 A_2 &= \frac{J}{m} \sum_j' b_{qj} \{ \psi(2\pi CF_{\varepsilon jJ}) - E\psi(2\pi CF_{\varepsilon jJ}) \}, \\
 A_3 &= \frac{J}{m} \sum_j' b_{qj} \{ \psi(2\pi F_{\varepsilon jJ} f_j g_j^{2H-1}) - \psi(2\pi CF_{\varepsilon jJ}) \\
 &\quad - E\psi(Z_J f_j g_j^{2H-1}) + E\psi(CZ_J) \}, \\
 A_4 &= \frac{J}{m} \sum_j' b_{qj} \left\{ \psi(F_{jJ} g_j^{2H-1}) - \psi \left(\frac{2\pi}{J} \sum_{l=j-J+1}^j I_{\varepsilon l} f_l g_j^{2H-1} \right) \right\}, \\
 A_5 &= \frac{J}{m} \sum_j' b_{qj} \left\{ \psi \left(\frac{2\pi}{J} \sum_{l=j-J+1}^j I_{\varepsilon l} f_l g_j^{2H-1} \right) - \psi(2\pi F_{\varepsilon jJ} f_j g_j^{2H-1}) \right\},
 \end{aligned}$$

where we write $f_j = f(\lambda_j)$, $b_{qj} = k_{qj} v_{qj}$, $v_{qj} = v_{qj}(g)$, define

$$I_{\varepsilon j} = \frac{1}{2\pi n} \left| \sum_{t=1}^n \varepsilon_t e^{it\lambda_j} \right|^2, \quad F_{\varepsilon jJ} = \frac{1}{J} \sum_{l=j-J+1}^j I_{\varepsilon l},$$

where the ε_t are NID(0, 1), and note that $E\psi(2\pi CF_{\varepsilon jJ}) = E\psi(CZ_J)$ and, from (2.17), $\sum_j' k_{qj} v_{qj} = 0$.

We later argue that A_4 and A_5 can be neglected, in the sense of being $o_p(m^{-1/2})$, so we consider $A_1 + A_2 + A_3$. By elementary inequalities,

$$\begin{aligned}
 E(A_1 + A_2 + A_3)^2 &= E(A_1 + A_2)^2 + 2E\{(A_1 + A_2)A_3\} + E(A_3^2) \\
 &= A_1^2 + V(A_2) + O(\{A_1^2 + V(A_2)\}^{1/2} \{V(A_3)\}^{1/2} + V(A_3)),
 \end{aligned}$$

because A_1 is nonstochastic and A_2 and A_3 have zero means.

Consider first A_1 . In the formal expansion

$$\psi(ab) = \psi(a) + \sum_{l=1}^{\infty} a^l \psi^{(l)}(a) \frac{(b-1)^l}{l!},$$

we have, taking $a = CZ_J$, $b = h(\lambda_j)$,

$$\begin{aligned}
 A_1 &= \frac{J}{m} \sum_j' b_{qj} \sum_{l=1}^{\infty} E((CZ_J)^l \psi^{(l)}(CZ_J)) \frac{(h(\lambda_j) - 1)^l}{l!} \\
 &= \frac{J}{m} \sum_j' b_{qj} \sum_{l=1}^{\infty} E((CZ_J)^l \psi^{(l)}(CZ_J)) \frac{1}{l!} \left(\sum_{i=1}^q \frac{h_i \lambda_j^{2i}}{(2i)!} \right)^l + o\left(\left(\frac{m}{n} \right)^{2q} \right)
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{J}{m} \sum_j' b_{qj} \sum_{i=1}^q \lambda_j^{2i} \sum_{u=1}^i \frac{1}{u!} \mathbb{E}((CZ_J)^u \psi^{(u)}(CZ_J)) \\
 &\quad \times \sum_{r_1+\dots+r_u=i} \left(\prod_{l=1}^u \frac{h_{r_l}}{(2r_l)!} \right) + o\left(\left(\frac{m}{n}\right)^{2q}\right).
 \end{aligned} \tag{4.1}$$

Now for $i = 1, \dots, q$,

$$\frac{J}{m} \sum_j' b_{qj} \left(\frac{j}{m}\right)^{2i} = U_{iq} + a_{1i} + a_{2i},$$

where

$$\begin{aligned}
 a_{1i} &= \frac{J}{m} \sum_j' k_{qj} \left\{ \log\{g_j/\lambda_j\} - \frac{\sum_j' k_{qj} \log\{g_j/\lambda_j\}}{\sum_j' k_{qj}} \right\} \left(\frac{j}{m}\right)^{2i}, \\
 a_{2i} &= \frac{J}{m} \sum_j' k_{qj} \left\{ \log\left(\frac{j}{m}\right) - \frac{\sum_j' k_{qj} \log(j/m)}{\sum_j' k_{qj}} \right\} \left(\frac{j}{m}\right)^{2i} - U_{iq}.
 \end{aligned}$$

In case $g(\lambda)$ is given by (2.11), so $G = 0$ in (2.10), we have $a_{1i} \equiv 0$. Otherwise, including under (2.12),

$$a_{1i} = \lambda_m^2 G Y_{iq} + o\left(\left(\frac{m}{n}\right)^2\right).$$

On the other hand, proceeding as in the proof of Lemma 2 of Robinson (1995b),

$$a_{2i} = O\left(\frac{\log m}{m}\right).$$

Thus (4.1) is

$$\begin{aligned}
 &\sum_{i=1}^q \left\{ U_{iq} + G Y_{iq} \lambda_m^2 + o\left(\left(\frac{m}{n}\right)^2\right) + O\left(\frac{\log m}{m}\right) \right\} \lambda_m^{2i} \\
 &\quad \times \sum_{u=1}^i \frac{1}{u!} \mathbb{E}((CZ_J)^u \psi^{(u)}(CZ_J)) \sum_{r_1+\dots+r_u=i} \left\{ \prod_{l=1}^u \frac{h_{r_l}}{(2r_l)!} \right\} + o\left(\left(\frac{m}{n}\right)^{2q}\right) \\
 &= U_{qq} P_{qJ}(\psi) \left(\frac{m}{n}\right)^{2q} + (2\pi)^2 G Y_{1q} P_{1J}(\psi) \left(\frac{m}{n}\right)^4 + o\left(\left(\frac{m}{n}\right)^{2q} + \left(\frac{m}{n}\right)^4\right) \\
 &\quad + O\left(\frac{m \log m}{n^2}\right).
 \end{aligned}$$

It follows that

$$A_1 = \begin{cases} U_{qq}P_{qJ}(\psi) \left(\frac{m}{n}\right)^{2q} + o\left(\left(\frac{m}{n}\right)^{2q}\right) + O\left(\frac{1}{n}\right), & G = 0 \text{ or } q = 1, \\ \{U_{22}P_{2J}(\psi) + (2\pi)^2GY_{12}P_{1J}(\psi)\} \left(\frac{m}{n}\right)^4 \\ + o\left(\left(\frac{m}{n}\right)^4\right) + O\left(\frac{1}{n}\right), & G \neq 0, q = 2, \\ (2\pi)^2GY_{1q}P_{1J}(\psi) \left(\frac{m}{n}\right)^4 + o\left(\left(\frac{m}{n}\right)^4\right) + O\left(\frac{1}{n}\right), & G \neq 0, q > 2. \end{cases}$$

Next, because the $2\pi F_{\epsilon jJ}$ are iid $\chi^2_{2J}/(2J)$ variates,

$$\begin{aligned} V(A_2) &= \frac{1}{m} Q_J(\psi) \frac{J}{m} \sum_j b_{qj}^2 \\ &= \frac{1}{m} Q_J(\psi) W_q + o\left(\frac{1}{m}\right) \quad \text{as } m \rightarrow \infty, \end{aligned}$$

after approximating the v_{qj} by replacing g_j by λ_j , if necessary, and approximating sums by integrals.

Next, in view of the proofs for A_1 and $V(A_2)$ it is readily seen that

$$V(A_3) = o\left(\frac{1}{m}\right).$$

We finally consider the probability orders of A_4 and A_5 . We have

$$A_4 = O_p\left(\left(\frac{\log m}{m}\right)^{2/3} + \frac{1}{m^{1/2}n^{1/4}}\right) = o_p\left(\frac{1}{m^{1/2}}\right).$$

In case $\alpha=1$ in (2.6), this follows directly from the proof of (4.8) of Robinson (1995b). The same method of proof is also instrumental when $\psi(z)$ is nonlinear, but in addition one has to consider a Taylor approximation to A_4 ; the “linear” term will dominate, but the I_j and $I_{\epsilon j}$ also appear in the denominator, and here the Gaussianity is useful, c.f. Hurvich et al. (1998). Finally, because

$$\begin{aligned} f_j - f_l &= g_j^{1-2H} \{h(\lambda_j) - h(\lambda_l)\} + h(\lambda_l)(g_j^{1-2H} - g_l^{1-2H}) \\ &= O\left(g_j^{1-2H} \left(\frac{j}{n^2} + \frac{1}{j}\right)\right), \quad j - J + 1 < l < j, \end{aligned}$$

we have

$$A_5 = O_p\left(\frac{\log m}{m} \sum_j \left(\frac{j}{n^2} + \frac{1}{j}\right)\right) = O_p\left(\frac{m \log m}{n^2} + \frac{(\log m)^2}{m}\right).$$

5. A special case

It is of interest to specialize the formulae of Section 3 in case of the higher-order kernel (2.5) with $q \geq 2$, and the Box-Cox family (2.6) and (2.7), focussing on the case (2.11) for g , so $G = 0$.

Noting first that, for $c > -1$,

$$\int_0^1 u^c du = \frac{1}{c+1}, \quad \int_0^1 \log(u)u^c du = -\left(\frac{1}{c+1}\right)^2, \tag{5.1}$$

we have under (2.5)

$$\sum_{j=0}^q \frac{a_j}{2j+1} = 1$$

from (2.1). Also,

$$\int_0^1 u^{2i} k_q^*(u) du = \sum_{j=0}^q \frac{a_j}{(2i+2j+1)},$$

$$\int_0^1 \log(u) u^{2i} k_q^*(u) du = -\sum_{j=0}^q \frac{a_j}{(2i+2j+1)^2},$$

so

$$U_{iq} = -\sum_{j=0}^q a_j \left\{ \frac{1}{(2i+2j+1)^2} - \frac{1}{2i+2j+1} \right\}$$

$$= \sum_{j=0}^q a_j \frac{2i+2j}{(2i+2j+1)^2}. \tag{5.2}$$

Thus, denoting by D the $(q+1) \times (q+1)$ matrix with (i, j) th element $2(i+j-2)/(2i+2j-3)^2$ for $i=1, \dots, q, j=1, \dots, q+1$ and $(q+1, j)$ th element $1/(2j-1), j=1, \dots, q+1$, and by d the $(q+1)$ vector all of whose elements are zero except for the $(q+1)$ st, which is unity, and writing $a_{(q)} = (a_0, \dots, a_q)'$, we have

$$a_{(q)} = -D^{-1}d. \tag{5.3}$$

For example, we deduce that

$$a_{(2)} = (0.45, 9.91, -13.77)', \tag{5.4}$$

$$a_{(3)} = (-0.15, 34.69, -106.23, 75.85)', \tag{5.5}$$

$$a_{(4)} = (-0.57, 64.71, -314.69, 484.75, -236.79)'. \tag{5.6}$$

For any q , we have from (5.2)

$$U_{qq} = U_{qq}^* = \sum_{j=0}^q a_j \frac{2q + 2j}{(2q + 2j + 1)^2}.$$

Further, because of (5.1) and also

$$\int_0^1 (\log u)^2 u^c \, du = \frac{2}{(c + 1)^3}, \quad c > -1,$$

we have

$$\begin{aligned} V_q &= V_q^* = \sum_{j=0}^q a_j \int_0^1 u^{2j} \{(\log u)^2 + 2 \log u + 1\} \, du \\ &= \sum_{j=0}^q a_j \left\{ \frac{2}{(2j + 1)^3} - \frac{2}{(2j + 1)^2} + \frac{1}{2j + 1} \right\} \\ &= \sum_{j=0}^q a_j \frac{4j^2 + 1}{(2j + 1)^3}, \end{aligned}$$

$$\begin{aligned} W_q &= W_q^* = \sum_{j=0}^q \sum_{l=0}^q a_j a_l \int_0^1 u^{2j+2l} \{(\log u)^2 + 2 \log u + 1\} \, du \\ &= \sum_{j=0}^q \sum_{l=0}^q a_j a_l \left\{ \frac{2}{(2j + 2l + 1)^3} - \frac{2}{(2j + 2l + 1)^2} + \frac{1}{2j + 2l + 1} \right\} \\ &= \sum_{j=0}^q \sum_{l=0}^q a_j a_l \frac{4(j + l)^2 + 1}{(2j + 2l + 1)^3}, \end{aligned}$$

$$Y_{1q} = \sum_{j=0}^q \frac{a_j}{2j + 5} - \left\{ \sum_{j=0}^q \frac{a_j}{2j + 3} \right\}^2.$$

Table 1 evaluates U_{qq}, V_q, W_q and Y_{1q} for $q = 1, 2, 3, 4$.

Table 1
 U_{qq}, V_q, W_q, Y_{1q}

q	U_{qq}	V_q	W_q	Y_{1q}
1	0.2222	1	1	0.0889
2	-0.0745	0.4125	0.9910	-0.0515
3	0.0143	0.0089	0.8554	-0.0018
4	-0.0005	-0.2066	0.5723	0.0011

Next, specializing $\psi(z)$ to $\psi_\alpha(z)$ given by (2.6) and (2.7), note that $\psi_\alpha(z)$ has u th derivative

$$\psi_\alpha^{(u)}(z) = \begin{cases} \frac{\Gamma(\alpha)z^{\alpha-u}}{\Gamma(\alpha-u+1)}, & u \geq 1, \quad 0 < \alpha < 1, \\ z^{-1}, & u = 1, \quad \alpha = 0, \\ (-1)^{u-1}\Gamma(u)z^{-u}, & u > 1, \quad \alpha = 0, \\ 1, & u = 1, \quad \alpha = 1, \\ 0, & u > 1, \quad \alpha = 1. \end{cases} \tag{5.7}$$

Thus, because $E(Z_J^\gamma) = \Gamma(\gamma+J)/\{\Gamma(J)J^\gamma\}$ (see e.g. Johnson and Kotz, 1970, p. 168), we deduce that

$$E((CZ_J)^u \psi^{(u)}(CZ_J)) = \begin{cases} \frac{C^\alpha \Gamma(\alpha) \Gamma(\alpha+J)}{\Gamma(\alpha-u+1) \Gamma(J) J^\alpha}, & u \geq 1, \quad 0 < \alpha < 1, \\ (-1)^{u+1} u!, & u \geq 1, \quad \alpha = 0, \\ C, & u = 1, \quad \alpha = 1, \\ 0, & u > 1, \quad \alpha = 1. \end{cases}$$

Thus,

$$R_J(\psi_\alpha) = \begin{cases} \frac{2C^\alpha \Gamma(\alpha+J)}{\Gamma(J) J^\alpha}, & 0 < \alpha < 1, \\ 2, & \alpha = 0, \\ 2C, & \alpha = 1. \end{cases}$$

Table 2 evaluates $R_J(\psi_\alpha)/(2C^\alpha) = \Gamma(\alpha+J)/\Gamma(J)J^\alpha$ for $\alpha = 0, 0.25, 0.5, 0.75, 1$ and $J = 1, 2, 3, 4$.

Table 2
 $R_J(\psi_\alpha)/(2C^\alpha)$

		α				
J	0	0.25	0.5	0.75	1	
1	1	0.9064	0.8862	0.9191	1	
2	1	0.9527	0.9400	0.9563	1	
3	1	0.9685	0.9594	0.9702	1	
4	1	0.9764	0.9693	0.9774	1	

Table 3
 $Q_J(\psi_\alpha)/C^2$

α					
J	0	0.25	0.5	0.75	1
1	1.6459	1.0346	0.8584	0.8616	1
2	1.2918	1.0328	0.9314	0.9259	1
3	1.1873	1.0253	0.9553	0.9496	1
4	1.1392	1.0202	0.9667	0.9618	1

From Johnson and Kotz (1970, pp. 168, 181), we deduce that

$$Q_J(\psi) = \begin{cases} \frac{C^{2\alpha}\{\Gamma(2\alpha + J)/\Gamma(J) - \Gamma(\alpha + J)^2/\Gamma(J)^2\}}{\alpha^2 J^{2\alpha-1}}, & 0 < \alpha < 1, \\ J \frac{d^2}{dz^2} \log \Gamma(z)|_{z=J}, & \alpha = 0, \\ C^2, & \alpha = 1. \end{cases}$$

Table 3 evaluates $Q_J(\psi_\alpha)/C^2$ for $\alpha = 0, 0.25, 0.5, 0.75, 1$ and $J = 1, 2, 3, 4$. Finally, we have

$$P_{iJ}(\psi_\alpha) = \frac{(2\pi)^{2i} C^\alpha \Gamma(\alpha) \Gamma(\alpha + J)}{\Gamma(J) J^\alpha} \sum_{u=1}^i \frac{1}{\Gamma(\alpha - u + 1) u!}$$

$$\times \sum_{r_1 + \dots + r_u = i} \left(\prod_{l=1}^u \frac{h_{r_l}}{(2r_l)!} \right), \quad 0 < \alpha < 1,$$

$$P_{iJ}(\psi_0) = (2\pi)^{2i} \sum_{u=1}^i (-1)^{u+1} \sum_{r_1 + \dots + r_u = i} \left(\prod_{l=1}^u \frac{h_{r_l}}{(2r_l)!} \right),$$

$$P_{iJ}(\psi_1) = (2\pi)^{2i} \frac{h_i}{(2i)!}.$$

Note that, for the Box-Cox class (2.6) and (2.7), formulae (3.9)–(3.17) are all invariant to C .

It is apparent that pooling does not affect the bias when $\alpha = 0, 1$, and out of these two cases it affects variance only when $\alpha = 0$. We give asymptotic bias, variance and optimal m for $\alpha = 0, 1$ with $J = 1$ but allowing for a general q th-order kernel and taking $G = 0$, i.e. with g given by (2.11). For $\alpha = 0$ (log-periodogram estimation),

$$\text{Asybias} = (2\pi)^{2q} \frac{U_{qq}}{2V_q} \sum_{u=1}^q (-1)^u \sum_{r_1 + \dots + r_u = q} \left(\prod_{l=1}^u \frac{h_{r_l}}{(2r_l)!} \right) \left(\frac{m}{n} \right)^{2q},$$

$$\text{Asyvar} = \frac{\pi^2}{24} \frac{W_q}{V_q^2} \frac{1}{m},$$

$$\hat{m} = \left[\pi^2 W_q / \left(24q(2\pi)^{4q} U_{qq}^2 \left\{ \sum_{u=1}^q (-1)^u \sum_{r_1+\dots+r_u=q} \left(\prod_{l=1}^u \frac{h_{r_l}}{(2r_l)!} \right) \right\}^2 \right) \right]^{1/(4q+1)} \times n^{4q/(4q+1)},$$

while for $\alpha = 1$ (local Whittle estimation),

$$\text{Asybias} = -(2\pi)^{2q} \frac{U_{qq} h_q}{2(2q)! V_q} \left(\frac{m}{n} \right)^{2q},$$

$$\text{Asyvar} = \frac{W_q}{4V_q^2} \frac{1}{m},$$

$$\hat{m} = [W_q \{(2q)!\}^2 / \{4q(2\pi)^{4q} U_{qq}^2 h_q^2\}]^{1/(4q+1)} n^{4q/(4q+1)}.$$

6. Finite sample properties

To examine further the implications for both choice of kernel order, and choice of α within the Box-Cox class of M-estimates, a Monte Carlo study was carried out. Three types of Gaussian model were employed:

1. FARIMA(0, $H - 0.5, 0$): $f(\lambda) \propto |\sin(\lambda/2)|^{1-2H}$;
2. FARIMA(1, $H - 0.5, 0$) with autoregressive (AR) coefficient 0.5: $f(\lambda) \propto |\sin(\lambda/2)|^{1-2H} (1.25 - \cos \lambda)^{-1}$;
3. Gegenbauer FARIMA with poles at frequencies 0 and $\pi/4$, of intensities $H/2 - 0.25$ and 0.25, respectively: $f(\lambda) \propto |\sin(\lambda/2)|^{1-2H} \times \{\sin(\lambda/2 + \pi/8) \sin(\lambda/2 - \pi/8)\}^{-0.5}$.

Model 1 is the simplest possible setting for comparisons across H , α and kernel order. In Model 2, the competing influence of an AR spectral peak near frequency zero is examined. The cyclic pole in Model 3 does not affect the asymptotic properties described in the previous section, but is liable to impact on finite sample behaviour to some degree (indeed disastrously if m is chosen large enough). For each of these models, three values of H were used:

$$H = 0.25, \quad H = 0.5, \quad H = 0.75,$$

corresponding, respectively, to moderate antipersistence, short memory, and moderate long memory. In each of the $3 \times 3 = 9$ cases, 1000 replications of series of length $n = 1000$ were generated, according to the approach used in Robinson and Henry (1999) for Models 1 and 2, and Gray et al. (1989) for Model 3.

For each series generated, $4 \times 3 \times 2 = 24$ estimates $\hat{H}_{qm}(\psi, g, k, 1)$ were computed, such that

$$\psi(z) = \psi_\alpha(z), \quad \alpha = 0, 0.5, 1,$$

$$k(u) = (2.5), \quad q = 1, 2, 3, 4 \text{ (see (2.18), (5.4)–(5.6))},$$

$$g(\lambda) = (2.11), (2.12).$$

Thus, besides versions of the familiar log-periodogram and local Whittle estimates ($\alpha = 0, 1$), we considered an intermediate, “square-root” estimate ($\alpha = 0.5$). The coefficients for the four kernels are given in (2.18) and (5.4)–(5.6); one expects, at least for large enough n , most scope for bias reduction when $g(\lambda) = (2.12)$ in case of Model 1, but not necessarily in Model 2. The bandwidth m is determined separately in each case, according to an automatic, data-dependent, optimal procedure in which the h_i are approximated by periodogram regression near zero frequency, extending the method proposed in Henry and Robinson (1996): their procedure covers the case $q = 1$ only, regressing on a constant and λ_j^2 , while for $q \geq 2$ we regress also on $\lambda_j^{2k}, k = 2, \dots, q$. The constraint $m \leq 499$ was imposed since we introduced m to satisfy $1 < m < n/2$.

Tables 4, 5 and 6 correspond to Models 1, 2 and 3, respectively, each table being split into two sub-tables according to the choice of g . Within each sub-table, for each combination of H , α and q , we give: (Monte Carlo) bias; two root-mean-squared errors (rmse), namely the Monte Carlo rmse followed by the theoretical asymptotic minimal rmse; two m 's, namely the automatic m averaged across Monte Carlo replications followed by the theoretical asymptotic optimal m . The theoretical asymptotic optimal m and theoretical asymptotic minimal rmse were computed using formulae (3.12)–(3.14) and (3.15)–(3.17), respectively, and the numerical results of Section 5, with the h_i being determined by expanding the appropriate $h(\lambda)$; these calculations were quite complicated.

The biases do not display very systematic behaviour, indeed since approximately optimal bandwidths were used, any improvements should appear in the Monte Carlo rmse only, and not in the biases. However, we discuss first the Monte Carlo m . These increase monotonically in kernel order q . The averaged m are smallest for $H = 0.75$ (long memory); in the FARIMA $(1, H - 0.5, 0)$, they are larger for $H = 0.5$ than for $H = 0.25$, but the converse is true for the other models. There is little apparent sensitivity to α except in Model 2 (where an autoregressive component is included) and m decreases in α in 17 out of 24 cases for kernel orders $q = 1, 2$, but decreases in α in 17 out of 24 cases for $q = 3, 4$. There is surprisingly little sensitivity of m to g even though, for FARIMA $(0, H - 0.5, 0)$, the h_i are all zero under (2.12) and all non-zero under (2.11).

We now consider the Monte Carlo rmse. For Model 1, rmse is always minimized by kernel order $q = 2$, while on the other hand, $q = 3$ performs worst in 16 out of 18 cases. For the FARIMA $(1, H - 0.5, 0)$, $q = 1$ is always best when $H = 0.25$ and 0.5 , and $q = 4$ is worst in 15 out of 18 cases. On the other hand, $q = 2$ is always most efficient when $H = 0.75$, and it is especially notable that $q = 1$ comes third in 2 cases and fourth in 2 cases, so that here the higher-order kernels with $q = 3$ and 4 also afford some improvement over the unweighted estimate. This support for higher-order kernels is underlined by the fact that the long memory case is often the most interesting in practice. Finally, for the Gegenbauer Model 3, all higher-order kernel procedures are very inefficient, presumably because of a failure of the automatic bandwidth selection procedure to respond to leakage from the pole at frequency $\pi/4$.

The rmse is noticeably sensitive to α , increasing with α in 92 out of 144 cases. However, interestingly, in a number of cases the new, intermediate estimate with $\alpha = 0.5$ performs best. Finally, there is no significant pattern of dependence of rmse's on H .

Table 4

FARIMA(0, $H - 1/2, 0$): For each combination of H , α and q we report Monte Carlo bias; Monte Carlo rmse followed by theoretical asymptotic minimal rmse; Monte Carlo automatic m followed by theoretical asymptotic optimal m

α	q	H								
		0.25 (antipersistence)			0.50 (short memory)			0.75 (long memory)		
		bias	rmse	m	bias	rmse	m	bias	rmse	m
$g(\lambda) = \lambda$										
0	1	-0.006	0.057, 0.035	217, 415	-0.006	0.053, 0	241, ∞	-0.007	0.063, 0.035	184, 415
	2	0.002	0.044, 0.054	293, 926	-0.002	0.044, 0	308, ∞	-0.010	0.046, 0.054	273, 926
	3	0.126	0.350, 1.95	403, 1258	0.016	0.304, 0	425, ∞	-0.070	0.284, 1.95	372, 1258
	4	-0.022	0.064, 0.058	478, 1760	-0.006	0.062, 0	493, ∞	0.004	0.056, 0.058	452, 1760
0.5	1	0.001	0.048, 0.032	209, 376	-0.004	0.044, 0	225, ∞	-0.004	0.055, 0.033	174, 376
	2	0.004	-0.038, 0.043	280, 1090	-0.001	0.034, 0	292, ∞	-0.004	0.039, 0.051	240, 787
	3	0.116	0.245, 1.66	417, 1356	0.019	0.171, 0	423, ∞	-0.009	0.099, 1.83	373, 1108
	4	-0.015	0.050, 0.049	486, 1871	0.001	0.050, 0	496, ∞	0.018	0.052, 0.053	459, 1585
1	1	0.002	0.051, 0.029	202, 376	-0.005	0.044, 0	227, ∞	-0.001	0.055, 0.029	170, 376
	2	0.005	0.038, 0.032	276, 1192	-0.001	0.035, 0	290, ∞	-0.002	0.037, 0.047	249, 732
	3	0.165	0.273, 1.39	404, 1514	0.086	0.179, 0	414, ∞	0.027	0.096, 1.68	377, 1031
	4	0.291	0.458, 0.041	499, 2099	0.048	0.160, 0	499, ∞	0.047	0.089, 0.048	400, 1486
$g(\lambda) = 2 \sin \lambda/2$										
0	1	-0.010	0.055, 0	217, ∞	0.009	0.052, 0	240, ∞	-0.002	0.058, 0	185, ∞
	2	-0.011	0.046, 0	301, ∞	-0.005	0.044, 0	308, ∞	0.002	0.044, 0	266, ∞
	3	0.056	0.314, 0	411, ∞	-0.014	0.314, 0	420, ∞	-0.056	0.269, 0	355, ∞
	4	-0.011	0.058, 0	479, ∞	-0.007	0.058, 0	492, ∞	0.005	0.052, 0	438, ∞
0.5	1	-0.007	0.049, 0	207, ∞	-0.003	0.044, 0	227, ∞	-0.001	0.054, 0	169, ∞
	2	-0.009	0.039, 0	284, ∞	-0.001	0.035, 0	290, ∞	0.004	0.039, 0	243, ∞
	3	0.118	0.251, 0	422, ∞	0.060	0.185, 0	425, ∞	-0.001	0.109, 0	362, ∞
	4	-0.007	0.049, 0	488, ∞	-0.001	0.048, 0	498, ∞	0.007	0.44, 0	451, ∞
1	1	-0.006	0.047, 0	204, ∞	-0.002	0.042, 0	229, ∞	-0.001	0.54, 0	161, ∞
	2	-0.008	0.038, 0	280, ∞	-0.001	0.035, 0	289, ∞	0.007	0.38, 0	239, ∞
	3	0.164	0.267, 0	408, ∞	0.111	0.209, 0	413, ∞	0.026	0.097, 0	366, ∞
	4	0.010	0.126, 0	499, ∞	0.001	0.049, 0	499, ∞	0.015	0.052, 0	499, ∞

With respect to theoretical rmse and m , there are a few cases in which they are close to their Monte Carlo counterparts, but mostly discrepancies are considerable. Indeed, in Table 4 with $g(\lambda) = \lambda$ for $H = 0.5$, and $g(\lambda) = 2 \sin(\lambda/2)$ for all H , we report $\text{rmse} = 0$, $m = \infty$ (so we did not impose the constraint $m \leq 499$, employed in the Monte Carlo); all h_i are zero, so the bias contribution to rmse is non-existent and rmse is minimized by $m = \infty$. These cases are degenerate, x_t being white noise for $H = 0.5$, while for $g(\lambda) = 2 \sin(\lambda/2)$ the semiparametric ‘approximation’ is actually the true model, at all frequencies. The theoretical rmses for $q = 3$ help to explain the Monte Carlo ones, indeed they are much worse. This is largely due to the numerically tiny V_3 (see Table 1); there is no reason why the factor W_q/V_q^2 in the asymptotic variance should be monotonic in q here, due to the $(\log u + 1)^2$ factor in the integrand of both V_q and W_q ,

Table 5

FARIMA(1, $H - 1/2, 0$): For each combination of H , α and q we report Monte Carlo bias; Monte Carlo rmse followed by theoretical asymptotic minimal rmse; Monte Carlo automatic m followed by theoretical asymptotic optimal m

α	q	H								
		0.25 (antipersistence)			0.50 (short memory)			0.75 (long memory)		
		bias	rmse	m	bias	rmse	m	bias	rmse	m
$g(\lambda) = \lambda$										
0	1	0.126	0.152, 0.088	170, 67	0.040	0.098, 0.088	83, 67	-0.010	0.135, 0.087	48, 67
	2	0.215	0.226, 0.154	231, 114	0.098	0.118, 0.154	132, 114	0.032	0.107, 0.154	77, 114
	3	0.138	0.287, 5.70	306, 148	0.146	0.167, 5.70	170, 148	0.009	0.192, 5.70	96, 148
	4	0.215	0.225, 0.170	406, 204	0.235	0.230, 0.170	230, 204	0.047	0.172, 0.170	127, 204
0.5	1	0.106	0.135, 0.081	142, 60	0.020	0.092, 0.081	62, 61	-0.040	0.176, 0.081	34, 61
	2	0.211	0.222, 0.145	217, 99	0.090	0.112, 0.145	122, 99	0.033	0.104, 0.145	70, 99
	3	0.084	0.153, 5.35	309, 130	0.162	0.168, 5.35	168, 130	0.006	0.160, 5.35	94, 130
	4	0.232	0.237, 0.158	423, 181	0.246	0.261, 0.158	229, 181	0.032	0.154, 0.158	126, 181
1	1	0.087	0.118, 0.072	120, 60	0.006	0.103, 0.072	51, 61	-0.057	0.208, 0.072	25, 61
	2	0.220	0.229, 0.133	214, 93	0.107	0.123, 0.133	132, 93	0.059	0.103, 0.132	85, 93
	3	0.104	0.176, 4.92	360, 121	0.163	0.171, 4.91	255, 121	0.180	0.185, 4.91	184, 121
	4	0.230	0.232, 0.145	499, 169	0.245	0.249, 0.145	478, 169	0.235	0.235, 0.145	378, 169
$g(\lambda) = 2 \sin \lambda/2$										
0	1	0.129	0.157, 0.088	174, 67	0.035	0.096, 0.088	82, 67	-0.021	0.144, 0.088	47, 67
	2	0.218	0.231, 0.137	233, 113	0.099	0.120, 0.137	131, 113	0.027	0.113, 0.137	75, 113
	3	0.030	0.176, 1.78	309, 1239	0.141	0.171, 1.78	169, 1239	0.001	0.187, 1.78	94, 1239
	4	0.213	0.225, 0.086	408, 657	0.234	0.251, 0.086	228, 657	0.033	0.176, 0.086	125, 657
0.5	1	0.100	0.128, 0.081	136, 61	0.014	0.096, 0.081	60, 61	-0.044	0.170, 0.081	33, 61
	2	0.208	0.221, 0.129	213, 99	0.085	0.107, 0.129	119, 99	0.033	0.102, 0.129	68, 99
	3	0.086	0.139, 1.61	304, 1173	0.161	0.175, 1.61	164, 1173	0.009	0.150, 1.61	91, 1173
	4	0.236	0.242, 0.077	416, 622	0.247	0.267, 0.077	223, 622	0.031	0.158, 0.077	122, 622
1	1	0.085	0.116, 0.072	119, 61	0.001	0.115, 0.072	48, 61	-0.059	0.211, 0.072	25, 61
	2	0.218	0.227, 0.118	212, 93	0.100	0.119, 0.118	128, 93	0.060	0.105, 0.118	84, 93
	3	0.075	0.104, 1.42	358, 1172	0.164	0.173, 1.42	249, 1172	0.184	0.189, 1.42	183, 1172
	4	0.228	0.232, 0.069	472, 622	0.245	0.250, 0.069	472, 622	0.236	0.237, 0.069	377, 622

absent in case of higher-order kernels in nonparametric spectral and probability density estimation. Note that our results also reflect our particular, polynomial, construction of $k_q(u)$. The fact that the theoretical rmse's increase over $q=1, 2, 4$ is also surprising, and is essentially due to the increase in W_q/V_q^2 overwhelming the decrease in $n^{-4q/(4q+1)}$ for our value of n ; for sufficiently large n this phenomenon would be reversed. The increase in theoretical optimal m with q is expected, but in Table 4 with $g(\lambda) = \lambda$ we have $m > n$ for $q \geq 2$ or 3. This is due in part to the small theoretical bias here, but also reflects the fact that our n is not large enough for the asymptotic formulae in (3.12)–(3.14) to be very reliable.

To analyse Monte Carlo bias, a different experiment was run with fixed bandwidths $m = 64, 128$ and 256, and focusing only on the FARIMA(0, 0.25, 0) case. The Monte

Table 6

Gegenbauer: For each combination of H , α and q we report Monte Carlo bias; Monte Carlo rmse followed by theoretical asymptotic minimal rmse; Monte Carlo automatic m followed by theoretical asymptotic optimal m

α	q	H								
		0.25 (antipersistence)			0.50 (short memory)			0.75 (long memory)		
		bias	rmse	m	bias	rmse	m	bias	rmse	m
$g(\lambda) = \lambda$										
0	1	0.009	0.057, 0.074	215, 95	-0.009	0.052, 0.074	230, 94	-0.008	0.060, 0.074	191, 93
	2	-0.001	0.044, 0.135	296, 148	-0.002	0.043, 0.135	305, 148	-0.007	0.046, 0.135	269, 148
	3	-0.130	0.363, 5.10	411, 185	-0.022	0.296, 5.10	421, 185	-0.068	0.276, 5.10	368, 185
	4	-0.025	0.066, 0.153	482, 249	-0.005	0.062, 0.113	490, 249	0.009	0.057, 0.153	451, 249
0.5	1	-0.003	0.050, 0.068	204, 86	-0.005	0.044, 0.068	228, 85	-0.001	0.052, 0.069	178, 84
	2	-0.001	0.038, 0.125	279, 133	-0.003	0.034, 0.125	296, 133	-0.002	0.037, 0.125	252, 133
	3	0.123	0.247, 4.70	416, 169	0.026	0.179, 4.70	427, 168	-0.001	0.098, 4.70	379, 168
	4	-0.019	0.051, 0.140	485, 230	-0.003	0.051, 0.141	498, 230	0.022	0.055, 0.141	465, 229
1	1	0.001	0.046, 0.060	200, 86	-0.004	0.044, 0.061	223, 85	-0.001	0.052, 0.161	163, 84
	2	0.002	0.036, 0.113	276, 128	-0.001	0.034, 0.113	288, 127	-0.002	0.037, 0.114	243, 127
	3	0.171	0.275, 4.26	405, 161	0.082	0.179, 4.26	413, 161	0.034	0.096, 4.27	372, 161
	4	0.254	0.431, 0.127	499, 220	0.053	0.165, 0.127	499, 220	0.041	0.085, 0.127	499, 219
$g(\lambda) = 2 \sin \lambda/2$										
0	1	-0.014	0.057, 0.074	219, 94	-0.007	0.050, 0.074	241, 94	-0.006	0.063, 0.074	181, 94
	2	-0.011	0.047, 0.118	302, 150	-0.004	0.043, 0.118	307, 150	0.001	0.047, 0.118	259, 150
	3	0.033	0.298, 1.62	412, 1498	-0.010	0.315, 1.62	421, 1498	-0.070	0.274, 1.62	352, 1498
	4	-0.015	0.060, 0.078	480, 795	-0.005	0.057, 0.078	491, 795	-0.002	0.055, 0.078	436, 795
0.5	1	-0.005	0.047, 0.068	201, 85	-0.007	0.044, 0.068	230, 85	-0.003	0.052, 0.068	171, 85
	2	-0.008	0.037, 0.110	278, 134	-0.004	0.035, 0.110	296, 134	0.001	0.037, 0.110	246, 134
	3	0.112	0.248, 1.46	412, 1417	0.047	0.187, 1.46	426, 1417	-0.013	0.112, 1.46	370, 1417
	4	-0.007	0.038, 0.071	485, 752	-0.003	0.047, 0.071	498, 752	0.005	0.043, 0.071	451, 752
1	1	-0.004	0.050, 0.061	201, 85	-0.002	0.044, 0.061	221, 85	0.001	0.059, 0.061	161, 85
	2	-0.001	0.039, 0.100	276, 128	-0.001	0.036, 0.100	289, 125	0.008	0.040, 0.100	240, 128
	3	0.185	0.292, 1.30	404, 1417	0.113	0.208, 1.30	412, 1417	0.024	0.098, 1.30	367, 1417
	4	0.018	0.141, 0.063	499, 752	0.003	0.051, 0.063	499, 752	0.013	0.052, 0.064	499, 752

Carlo biases and rmse (based still on 1000 replications) are reported in Table 7. In all cases, the best-performing bandwidth is undoubtedly $m = 256$, which shows that the automatic procedure for $q = 1$ tends to undersmooth.

As expected, the choices $q = 1$ and 2 are indistinguishable from the point of view of Monte Carlo rmse's when the bandwidth is the same. More surprisingly, Monte Carlo biases are also very similar. The sensitivity of bias to α is quite marked, on the other hand, increasing with α in 46 out of 48 cases, and this from slightly negative values when $\alpha = 0$ to slightly positive values when $\alpha = 1$ in 42 out of 48 cases. The rmse's decrease with α in 40 out of 48 cases, while of the eight remaining cases, the increase is 2% or less.

Table 7

FARIMA(0, $H - 1/2$, 0) and fixed bandwidths: For each combination of H , α and q and m we report Monte Carlo bias and Monte Carlo rmse

α	q	m		128		256	
		64		bias	rmse	bias	rmse
		bias	rmse	bias	rmse	bias	rmse
$g(\lambda) = \lambda$							
0	1	-0.007	0.111	-0.004	0.069	-0.004	0.045
	2	-0.006	0.110	-0.003	0.069	-0.005	0.045
	3	-0.016	0.123	-0.002	0.072	-0.001	0.057
	4	-0.010	0.114	-0.023	0.160	0.003	0.044
0.5	1	0.001	0.080	0.006	0.051	0.001	0.035
	2	0.002	0.080	0.006	0.051	0.001	0.035
	3	-0.006	0.085	0.010	0.057	0.008	0.048
	4	-0.001	0.081	0.013	0.131	0.002	0.035
1	1	0.008	0.079	0.008	0.052	0.002	0.034
	2	0.009	0.079	0.008	0.052	0.001	0.035
	3	-0.001	0.083	0.012	0.058	0.010	0.047
	4	0.004	0.079	0.028	0.131	0.002	0.034
$g(\lambda) = 2 \sin \lambda/2$							
0	1	-0.014	0.107	-0.005	0.069	0.006	0.049
	2	-0.013	0.107	-0.004	0.068	0.005	0.049
	3	0.024	0.119	0.003	0.071	0.008	0.063
	4	-0.017	0.111	-0.028	0.161	0.007	0.049
0.5	1	0.005	0.078	0.006	0.054	0.007	0.035
	2	0.006	0.078	0.007	0.054	0.007	0.036
	3	-0.003	0.084	0.012	0.061	0.009	0.048
	4	0.002	0.079	0.008	0.124	0.008	0.036
1	1	0.010	0.077	0.014	0.052	0.010	0.036
	2	0.011	0.077	0.014	0.051	0.010	0.037
	3	0.003	0.082	0.018	0.058	0.013	0.050
	4	0.007	0.078	0.031	0.133	0.010	0.037

Overall in this study, it is encouraging that the methods work better than the theoretical asymptotic minimal rmse and optimal m might suggest. We find evidence that a kernel of order $q = 2$ can improve estimation, but little that a larger order can exploit the local smoothness available in the models considered, at least for the sample size considered. The new estimate with $\alpha = 0.5$ emerges as a useful choice, intermediate between the log-periodogram and local Whittle estimates, and sometimes having better finite sample properties than either.

7. Application to intra-day foreign exchange volatility

The focus of this application is the estimation of long memory in nonlinear transformations of intra-day foreign exchange (FX) rates. We study three sets of FX returns, on the DEM/USD (Deutschmark/Dollar), JPY/USD (Yen/Dollar) and GBP/DEM (Sterling/Dollar), covering the period from the beginning of October 1992 to the beginning of March 1993 (26 weeks, week-ends excluded) and first investigated in the context of long memory estimation by Henry and Payne (1997). These return series are filtered transcriptions of the tick-by-tick quotation series appearing on the Reuters FFX page. Each quote encompasses a timestamp, bid and ask quotation pair, plus identifiers which allow one to determine the inputting bank and its location. In this study we ignore these identifications using the tick-by-tick data solely to construct a homogeneous time series in calendar time. The basic horizon over which we calculate returns is 10 min. This yields, for each currency, a time series with $n = 18,720$ observations (130 days with 144 observations each).

The choice of these series was motivated by the extensively documented daily seasonality of absolute, squared and log-squared returns, and other transformations. For an extensive description of this seasonal pattern and its market microstructural interpretations, see Payne (1996). This seasonal behaviour was modelled, albeit for different exchange rates, by a Gegenbauer model with possibly asymmetric poles at the daily frequency and its multiples, in Arteche (2000). It is therefore a series for which long-memory analysis may be best performed by means of purely local methods, as opposed to those in Hurvich and Brodsky (2001) and Moulines and Soulier (1999), which appeal more to global smoothness. For all three series, the strongest seasonal feature is a large peak in the estimated spectrum at the daily frequency (in this case $2\pi/144$), corresponding to periodogram harmonic 130, and at multiples of the daily frequency.

We estimated H from $x_t =$ absolute returns, squared returns, and square-rooted absolute returns, using the same α , k and g , and the same type of automatic m , as in the previous section. Of course, such x_t are certainly not Gaussian, though theoretical evidence from, for example, Robinson (1995b), Robinson and Henry (1999) and Velasco (2000) suggests that this may not affect asymptotic properties. More seriously, Deo and Hurvich (2001) have pointed out the inconsistency between stochastic volatility models and differentiability of $h(\lambda)$ in (3.1), where $f(\lambda)$ is the spectral density of squares or other instantaneous nonlinear functions of the data. Arguing rather informally, in case we observe $y_t = \varepsilon_t \sigma_t$ where $\{\varepsilon_t\}$ is a sequence of zero-mean independent variates, independent also of the stationary sequence $\{\sigma_t\}$, then the y_t are martingale differences and thus serially uncorrelated, and the spectral density of $x_t = |y_t|^\beta$, $\beta > 0$, is of form

$$f(\lambda) = \tilde{f}(\lambda) + \frac{\tau^2}{2\pi}, \quad (7.1)$$

where $\tilde{f}(\lambda)$ is the spectral density of $|\varepsilon_t \sigma_t|^\beta$ and $\tau^2 = V(|\varepsilon_t|^\beta) E|\sigma_t|^{2\beta}$. Then, if $\tilde{f}(\lambda) = C\lambda^{1-2H} \tilde{h}(\lambda)$, then no matter how smooth $\tilde{h}(\lambda)$ is, $f(\lambda)\lambda^{2H-1}$ satisfies a Lipschitz condition of order no greater than $2H - 1 < 1$. The same outcome occurs in more general

Table 8

Foreign exchange absolute returns: For each combination of α and q we report the estimate \hat{H} and automatic m

α	q	Deutsch Mark		Sterling		Yen	
		\hat{H}	m	\hat{H}	m	\hat{H}	m
$g(\lambda) = \lambda$							
0	1	0.82	492	0.73	495	0.81	487
	2	0.77	1228	0.73	1239	0.74	1210
	3	0.56	1609	0.63	1623	0.53	1586
	4	0.94	2135	0.81	2154	0.74	2105
0.5	1	0.87	902	0.84	890	0.82	891
	2	0.77	2796	0.76	2796	0.78	2812
	3	0.75	2396	0.76	2422	0.77	2390
	4	0.73	2887	0.72	2815	0.77	2826
1	1	0.93	1361	0.93	1273	0.89	1379
	2	0.88	2145	0.89	2065	0.86	2167
	3	0.84	3208	0.85	3160	0.84	3299
	4	0.81	7298	0.82	6480	0.83	6578
$g(\lambda) = 2 \sin \lambda/2$							
0	1	0.82	492	0.73	496	0.81	487
	2	0.77	1229	0.73	1241	0.74	1211
	3	0.57	1611	0.63	1625	0.56	1588
	4	0.94	2138	0.81	2156	0.74	2108
0.5	1	0.87	903	0.84	892	0.82	893
	2	0.77	2806	0.76	2806	0.79	2822
	3	0.75	2400	0.76	2412	0.77	2384
	4	0.73	2892	0.72	2834	0.78	2831
1	1	0.93	1364	0.93	1276	0.89	1382
	2	0.88	2159	0.89	2078	0.86	2181
	3	0.85	3213	0.85	3165	0.84	3304
	4	0.81	7308	0.82	6489	0.84	6587

‘signal-plus-noise’ models. In this context, there is no benefit to using higher-order kernels (though the issues of choice of M-estimate and pooling are still relevant). However, serial uncorrelatedness of $y_t = \varepsilon_t \sigma_t$ still holds if the independence properties of the ε_t are relaxed such that $E(\varepsilon_t | \varepsilon_{t-1}, \dots, \sigma_t, \sigma_{t-1}, \dots) = 0$, almost surely, indeed the y_t are still martingale differences, but this is insufficient for the representation (7.1), so the argument against the possibility of a smooth $h(\lambda)$ in (3.1) is no longer conclusive. More generally, though on the one hand many functions of mixing processes are mixing at the same rate, and on the other, taking a nonlinear instantaneous function of a Gaussian long memory process cannot increase memory, there seems to be no reason why there cannot exist y_t that has short memory autocorrelation, and is even a martingale difference sequence, while nonlinear functions $x_t = r(y_t)$ not only have long memory but have spectral density of form (3.1) with $h(\lambda)$ satisfying (3.2) with

$q \geq 1$. For example, if $x_t = y_t^2$ where y_t is serially uncorrelated, x_t has lag- j autocorrelation $2(Ey_t^2)^2 + \text{cum}(y_t, y_t, y_t, y_t)$ for $j = 0$, and $\text{cum}(y_t, y_t, y_{t+j}, y_{t+j})$ for $j \neq 0$, where considerable generality is possible in the fourth cumulant function. In fact, even in (7.1) it is, strictly speaking, possible for $\tilde{f}(\lambda)$ to include an additive component $-\tau^2/(2\pi)$, such that $f(\lambda)$ has sufficient smoothness to justify use of higher-order kernels.

As results for the three transformed data series are similar, we include only those for absolute returns, in Table 8. We notice first that varying q has no effect, and that the great similarities in results for the different exchange rates lend plausibility to the hypothesis of a common factor, which may be interpreted as resulting from a common information arrival process affecting volatility. All estimated values of H fall within the interval $(0.5, 1)$, and except when $\alpha = 0$ and $q = 3$, approximate 95% confidence intervals exclude $H = 1/2$. Thus, in general it appears that such a common factor might be well modelled by a stationary long memory component.

Looking at the results of Table 8 in more detail, we see that for $(\alpha, q) \neq (0, 3)$ all estimates of H lie in the interval $(0.73, 0.93)$. There is a clear increase in automatic bandwidths with q and α , which results in an increase in α of estimated H , while use of a higher-order kernel reduces the estimates in 25 out of 27 cases. Very similar results, not reported here, obtain with deseasonalized spectra (using a double window deseasonalization procedure described in Henry and Payne, 1997). This confirms both that the automatic bandwidth selection procedure is insensitive to low-leakage peaks in the spectrum (a rather undesirable feature), and that M-estimation of long memory (with or without higher-order kernels) is robust to the type of peaks present in the spectra of intra-day foreign exchange absolute returns.

Acknowledgements

We thank two referees for helpful comments which have led to improvements in presentation. We also thank Fabrizio Iacone for help in computing the theoretical bandwidths and theoretical mean-squared errors in Tables 4–6. The first author's research was supported by a Leverhulme Trust Personal Research Professorship and ESRC Grant R000238212. The second author's research was supported by an FRS Grant from IRES.

References

- Andrews, D.W.K., Guggenberger, P., 2000. A bias-reduced log-periodogram estimator for the long memory parameter. Preprint, Yale University.
- Andrews, D.W.K., Sun, Y., 2001. Local polynomial Whittle estimation of long-range dependence. Preprint, Yale University.
- Arteche, J., 2000. Gaussian semiparametric estimation in seasonal/cyclical long memory time series. Euskal Erriko Unibertsitatea, Mimeo.
- Cogburn, I., Davis, H.T., 1974. Periodic splines and spectral estimation. *Annals of Statistics* 2, 1108–1126.
- Deo, R.S., Hurvich, C.M., 2001. On the log-periodogram regression estimator of the memory parameter in long memory stochastic volatility models. *Econometric Theory* 17, 686–710.
- Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4, 221–238.

- Graf, H.-P., 1983. Long range correlations and estimation of the self-similarity parameters. Dissertation ETH, Zurich.
- Gray, H.L., Zhang, N.I., Woodward, W.A., 1989. On generalized fractional processes. *Journal of Time Series Analysis* 10, 233–257.
- Hannan, E.J., 1957. The variance of the mean of a stationary process. *Journal of the Royal Statistical Society Series B* 19, 282–285.
- Henry, M., Payne, R., 1997. An investigation of long range dependence in intra-day foreign exchange volatility. Discussion Paper 264, Financial Markets Group, London School of Economics.
- Henry, M., Robinson, P.M., 1996. Bandwidth choice in Gaussian semiparametric estimation of long range dependence. In: Robinson, P.M., Rosenblatt, M. (Eds.), *Memory of E.J. Hannan. Conference on Applied Probability and Time Series Analysis, Vol. II: Time Series Analysis*. Springer, New York, pp. 220–232.
- Hurvich, C.M., Brodsky, J., 2001. Broadband semiparametric estimation of the memory parameter of a long-memory time series using fractional exponential models. *Journal of Time Series Analysis* 22, 221–2496.
- Hurvich, C.M., Deo, R.S., Brodsky, J., 1998. The mean-squared error of Geweke and Porter-Hudak's estimation of the memory parameter of a long-memory time series. *Journal of Time Series Analysis* 19, 19–46.
- Johnson, N.L., Kotz, S., 1970. *Continuous Univariate Distributions*. Vol. 1. Wiley, New York.
- Künsch, H., 1987. Statistical aspects of self-similar processes. In: *Proceedings of the First World Congress of the Bernoulli Society*. VNU Science Press, pp. 67–74.
- Lobato, I., 1999. A semiparametric two-step estimator in a multivariate long memory model. *Journal of Econometrics* 90, 129–153.
- Lobato, I., Robinson, P.M., 1998. A nonparametric test for $I(0)$. *Review of Economic Studies* 65, 475–495.
- Moulines, E., Soulier, P., 1999. Broadband log-periodogram regression of time series with long range dependence. *Annals of Statistics* 27, 1415–1439.
- Parzen, E., 1957. On consistent estimates of the spectrum of a stationary time series. *Annals of Mathematical Statistics* 28, 329–348.
- Payne, R., 1996. Announcement effects and seasonality in the intra-day foreign exchange market. Discussion Paper 238, Financial Markets Group, London School of Economics.
- Phillips, P.C.B., 1987. Time series regression with a unit root. *Econometrica* 55, 277–301.
- Phillips, P.C.B., 1991. Spectral regression for cointegrated time series. In: Barnett, W.A., Powell, J., Tauchen, G. (Eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge University Press, Cambridge.
- Robinson, P.M., 1994. Semiparametric analysis of long-memory time series. *Annals of Statistics* 22, 515–539.
- 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., Henry, M., 1999. Long and short memory conditional heteroscedasticity in estimating the memory parameters of levels. *Econometric Theory* 15, 299–336.
- Velasco, C., 1999. Non-stationary log-periodogram regression. *Journal of Econometrics* 91, 525–571.
- Velasco, C., 2000. Non-Gaussian log-periodogram regression. *Econometric Theory* 16, 44–79.