SEMIPARAMETRIC ESTIMATION OF LONG-MEMORY MODELS

Carlos Velasco*
Department of Economics
Universidad Carlos III de Madrid
28903 Getafe, Spain

April 5, 2005

Abstract

This article revises semiparametric methods of inference on different aspects of long memory time series. The main focus is on estimation of the memory parameter of linear models, analyzing bandwidth choice, bias reduction techniques and robustness properties of different estimates, with some emphasis on nonstationarity and trending behaviors. These techniques extend naturally to multivariate series, where the important issues are the estimation of the long run relationship and testing of fractional cointegration. Specific techniques for the estimation of the degree of persistence of volatility for nonlinear time series are also considered.

Keywords and Phrases: Long range dependence, log-periodogram regression, local Whittle estimation, nonstationarity, fractional differencing, cointegration, bandwidth choice, stochastic volatility.

*The author is grateful to Kerry Patterson, Josu Arteche and an Associate editor for their useful comments and suggestions on a previous version of the paper. Research supported by the Spanish Ministerio de Educación y Ciencia reference number SEJ2004-04583/ECON. Corresponding Address: Carlos Velasco, Departamento de Economía, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe, Spain. Carlos.Velasco@uc3m.es. Fax: +34 916249875

1 Introduction

The concepts of long memory and long range dependence describe the property that many time series models possess, despite being stationary, higher persistence than that predicted by usual short run linear models, such as ARMA processes. The same type of persistence, with slow decay in the autocorrelation function, has also been observed in many economic series, such as the increments of trending data, measures of volatility, and errors in long run equilibrium relationships; see Henry and Zaffaroni (2003) for a review of applications of long memory time series in economics.

Although several long memory parametric models can be found in the literature, such as Fractionally Integrated ARMA (ARFIMA) models (Hosking (1981), Granger and Joyeaux (1980)) or fractional Gaussian noise (e.g. Sinai (1976)), there has long been an interest in modelling separately the long and short run features of time series. Since parametric models and the weak limit of partial sums of a large class of long memory processes describe the degree of persistency by means of a memory parameter, usually denoted as d in the econometrics literature, much attention has been paid to stating alternative, semiparametric definitions of long range dependent behavior and, based on them, corresponding estimates of d that avoid the specification of short memory properties.

If X_t is a covariance stationary sequence, long memory is described in the time domain by means of the asymptotic relation

$$\gamma_X(j) = \mathrm{Cov}\left(X_t, X_{t+j}\right) \sim c_X j^{2d-1}, \quad \text{as } j \to \infty, \tag{1}$$

$|c_X| > 0$, where $a \sim b$ means that the limit of a/b is 1. The constant c_X can be replaced by a slowly varying function at infinity to achieve greater generality. Equation (1) states that the autocovariance function $\gamma_X(j)$ decays to zero as a power function of the lag j, where the decay rate is determined by the long memory parameter d, so that when $d > 0$ we have:

$$\sum_{j=-\infty}^{\infty} \gamma_X(j) = \infty, \tag{2}$$

and it is required that $d < 0.5$ for covariance stationarity.

Alternatively, long range dependence is reflected in the spectral density $f_X(\lambda)$ of X_t, defined by

$$\gamma_X(j) = \int_{-\pi}^{\pi} f_X(\lambda) \exp(ij\lambda) d\lambda, \quad j = 0, \pm 1, \ldots,$$

through its behavior at low frequencies,

$$f_X(\lambda) \sim G_X |\lambda|^{-2d} \quad \text{as } \lambda \to 0, \tag{3}$$

for some finite constant $G_X > 0$. Therefore, the spectral density has a pole at zero frequency when $d > 0$, agreeing with (2) and reflecting the increasing contribution of low frequency components to the variance decomposition of X_t. Negative values of d can be allowed, although they are not likely to occur in practice unless some differencing has first been applied to X_t. In this case (3) indicates that there is no contribution from the zero frequency to the variance of X_t, as would happen after first differencing a stationary time series, and such a property is termed negative memory or 'antipersistency'. However as long as $d > -0.5$, the series remains invertible. When $d = 0$, $f_X(0)$ is bounded and positive, and we will say that the series is weakly dependent. Note that (1) does not specify the behavior of γ_X for short lags nor does (3) gives the properties of f_X for cyclical, seasonal or short run frequencies.

Long memory behaviour is reflected also in the fact that the sample mean converges to the true expectation of X_t at the rate $T^{d-\frac{1}{2}}$, slower than the usual root-T rate for uncorrelated and weakly dependent sequences, where T is the sample size (Adenstedt (1974)). Similarly, the asymptotic properties of other basic statistics, such as partial sums (Mandelbrot and Van Ness (1968)), autocovariances (Hosking (1996)) or least squares (LS) regression coefficients (Yajima, (1988, 1991)), depend primarily on the value of d.

The long memory concept can also cross the stationarity border $d = \frac{1}{2}$ and it can be useful to characterize the long run behavior of nonstationary time series by means of the class of integrated $I(d)$ processes. This class nests the unit root $I(1)$ processes as well as the $I(0)$ weakly dependent processes. Here the concept of integration refers to the application of fractional difference/integration filters, defined by the formal binomial expansion of $(1-L)^d$ in terms of the lag operator L, such that for any real $d \neq 1, 2, \ldots$

$$(1-L)^d = \sum_{j=0}^{\infty} \psi_j(d) L^j, \quad \psi_j(d) = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)},\ j = 0, 1, \ldots, \tag{4}$$

where $\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx$ is the gamma function and $\Gamma(0)/\Gamma(0) = 1$. Thus $\psi_0(d) = 1$ and $\psi_j(d) = \psi_{j-1}(d)(j-d-1)/j$, $j \geq 1$, and, using Stirling's formula, the coefficients $\psi_j(d)$ behave as $\Gamma(-d)^{-1} j^{-d-1}$ for $j \to \infty$. When d is a positive integer, only the first $d+1$ terms are nonzero and we obtain the usual definition of the d-th difference operator. Then X_t is $I(d)$, i.e. integrated of order d, if $(1-L)^d X_t$ is weakly dependent. Note that the transfer function associated with the fractional filter $(1-L)^d$ is

$$\left|1-e^{i\lambda}\right|^{2d} = (2\sin|\lambda/2|)^{2d} \sim |\lambda|^{2d} \quad \text{as} \quad \lambda \to 0, \tag{5}$$

giving a simple intuition of the effect of fractional differencing in the frequency domain by means of annihilating the contribution at zero frequency.

Under this framework, the concept of long memory and fractional integration are key to the modelling of long run relationships among nonstationary trending time series. As proposed by Granger (1981), the series are (fractionally) cointegrated if a linear combination has reduced memory compared with the original series, reflecting a long run equilibrium (at least when the linear combination is stationary). When the memory levels are no longer an a priori assumption, as under the CI$(1,0)$ paradigm stressed since Engle and Granger (1987) with $I(1)$ levels and $I(0)$ errors, the inference problems complicate because of the unknown degree of cointegration.

We will first focus in Section 2 on semiparametric methods of estimating d in the frequency domain. These are the most used in practice and many extensions, including studies of validity and subsequent refinements, have appeared. We also provide a guide to the choice of the range of frequencies where the relationship (3) holds approximately for a particular problem with a given sample size, trying to balance bias and variability. In this vein we present several proposals for bias reduction, borrowing ideas from nonparametric statistics, as well as methods that consider all frequencies but are semiparametric in essence. In Section 3 we consider the extension of the previous methods to nonstationary fractionally integrated series and discuss the possibility of long memory at other nonzero frequencies, such as cyclical and seasonal ones. Section 4 describes applications of semiparametric methods to the analysis of economic series, stressing those to cointegrated multivariate nonstationary time series and to white noise series with persistence in their volatility.

2 Memory estimation

Each of the different asymptotic characterizations of long memory can lead to alternative estimates of the memory parameter where population quantities are replaced by sample equivalents. The rate of convergence of partial sums was early exploited by the rescaled range, or R/S, analysis introduced by Hurst (1951) and Mandelbrot and Wallis (1968). Time domain estimates proposed by Robinson (1994c) where analyzed by Hall, Koul and Turlach (1997), while Geweke and Porter-Hudak (1983), GPH henceforth, propose to use frequency domain estimates.

Frequency domain semiparametric methods exploit the asymptotic relationship (3) as a valid (semiparametric) model for the spectral density at low frequencies, in particular, the first m Fourier frequencies, $\lambda_j = 2\pi j/T$, $j = 1, \ldots, m$, where

$$\frac{1}{m} + \frac{m}{T} \to 0 \text{ as } T \to \infty, \tag{6}$$

so that m is increasing with the sample size T in the asymptotics, but a slower rate. These local methods are also termed narrow band estimates, because only a degenerating band of the spectrum around $\lambda = 0$ is modelled, basically in terms of the long memory parameter d. The idea behind all of the estimates in the frequency domain is to compare the spectral density f_X with its sample counterpart, the periodogram, across this range of frequencies and to find the value of d that best suits the data by alternative criteria. Define the discrete Fourier transform (DFT) of X_t, for a sample of T observations, $t = 1, \ldots, T$, as

$$w_X(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} X_t \exp(i\lambda t),$$

and the periodogram of X_t as

$$I_X(\lambda) = |w_X(\lambda)|^2 .$$

Note that $w_X(\lambda_j)$, $0 < j < T$, is invariant to shifts in mean, rendering periodogram based methods independent of mean estimation by dropping the zero frequency.

2.1 Log-periodogram estimation

GPH observed that by taking logs of both sides of (3) we obtain

$$\log f_X(\lambda_j) \sim \log G_X - 2d \log \lambda_j, \quad j = 1, \ldots, m$$

and on substituting $\log f_X(\lambda_j)$ by $\log I_X(\lambda_j)$, we obtain the linear regression model on the log-periodogram,

$$\log I_X(\lambda_j) = \alpha + d z_j + u_j, \quad j = 1, \ldots, m \tag{7}$$

with regressor $z_j = -2\log\lambda_j$ and $\alpha = \log G_X - \eta$, $\eta = 0.5772\ldots$ being Euler's constant. The error term $u_j = \log I_X(\lambda_j) / G_X \lambda_j^{-2d} + \eta$ is expected to be asymptotically homoskedastic with zero mean, since, at least for weakly dependent Gaussian time series, each $\log I_X(\lambda_j) / f_X(\lambda_j)$ is approximately an independent and identically distributed (iid) $\log \chi_2^2/2$ random variate, with expectation $-\eta$. Based on this fact, GPH proposed running an ordinary LS (OLS) regression to estimate d in (7). Robinson (1995a) justified such a procedure for multivariate Gaussian time series with possibly different memory parameters in the interval $(-0.5, 0.5)$ by trimming the first ℓ Fourier frequencies, since for fixed j, $I_X(\lambda_j)$ is asymptotically biased for $f_X(\lambda_j)$ when $d \neq 0$. Later, Hurvich, Deo and

Brodsky (1998) showed that such trimming is not necessary for the log-periodogram (LP) estimate to have nice asymptotic properties. It is also possible to replace the regressor z_j in the LP regression by some asymptotically equivalent sequence, such as $-2\log(2\sin\lambda_j/2)$, which arises naturally if X_t is fractionally integrated, cf. (5), as proposed by GPH.

Robinson (1995a) proposed pooling a finite number of adjacent periodogram ordinates to improve efficiency. For $K = 1, 2, \ldots$, fixed, (assuming m/K is integer), define

$$Y_{X,j}^{(K)} = \log\left(\sum_{k=1}^{K} I_X(\lambda_{j+k-K})\right), \qquad j = K,\ 2K,\ \ldots,\ m,$$

so the (pooled) LP estimate considered in Robinson (1995a) for a stationary and invertible time series is

$$\widehat{d}_m^{LP} = \left(\sum_j \Lambda_j^2\right)^{-1}\left(\sum_j \Lambda_j Y_{X,j}^{(K)}\right).$$

In this section all summations in j run for $j = K,\ 2K,\ \ldots,\ m$ and $\Lambda_j = z_j - \overline{z}_m$, $\overline{z}_m = (K/m)^{-1}\sum_j z_j$. Obviously for $K = 1$, $\widehat{d}_m^{LP}$ is the OLS coefficient in (7). Shimotsu and Phillips (2002) considered the case where K is allowed to grow in the asymptotics with T.

The asymptotic distribution of $\widehat{d}_m^{LP}$ is given by

$$2m^{1/2}\left(\widehat{d}_m^{LP} - d\right) \xrightarrow{d} \mathcal{N}\left(0, K\dot{\psi}(K)\right) \tag{8}$$

where $\psi(z) = (\mathrm{d}/\mathrm{d}z)\log\Gamma(z)$ is the digamma function, and the upper dot denotes first derivative. Under (6), semiparametric estimates with root-m convergence as in (8) are infinitely inefficient compared to usual parametric estimates which are standardized by $T^{1/2}$, but, by contrast, are more robust to misspecification. Note that the variance of the log of a $\chi^2_{2K}/2$ random variable (which is the weak limit of the centered $Y_{X,j}^{(K)}$) is equal to $\dot{\psi}(K)$ and $\sum_j \Lambda_j^2 \sim 4m/K$ as $m \to \infty$. For $K = 1$ we find that $K\dot{\psi}(K) = \pi^2/6$ and using that $\dot{\psi}(K+1) = \dot{\psi}(K) - K^{-2}$ it can be shown that $K\dot{\psi}(K)$ decreases with K, so choosing K large increases the (asymptotic) efficiency. In regular cases for which (3) is a good approximation, including ARFIMA processes, m can be chosen to just satisfy

$$\frac{\log^2 T}{m} + \frac{m^5}{T^4} \to 0 \tag{9}$$

as $T \to \infty$ (see Robinson (1995a, Assumption 6) and Hurvich et al. (1998, Theorem 2)). A consistent estimate of G_X can be obtained as $\widehat{G}_m^{LP} = \exp\left(\widehat{\alpha}_m - \psi(K)\right)$, where $\widehat{\alpha}_m$ is the OLS intercept and noting that the expectation of a $\log(\chi^2_{2K}/2)$ variate is $\psi(K)$.

A multivariate $N \times 1$ time series $\mathbf{X}_t$, with possibly different memory parameters, can be considered if we assume that (3) holds for the spectral density of each of the components of $\mathbf{X}_t$, that is, the diagonal elements of the spectral density matrix $\mathbf{f_X}(\lambda)$, defined implicitly by

$$\mathbf{\Gamma_X}(j) = \mathrm{Cov}(\mathbf{X}_t, \mathbf{X}_{t+j}) = \int_{-\pi}^{\pi} \mathbf{f_X}(\lambda)e^{ij\lambda}d\lambda,$$

satisfy $f_{nn}(\lambda) \sim G_{nn}|\lambda|^{-2d_n}$, $n = 1, \ldots, N$, as $\lambda \to \infty$. Defining the coherence between the r-th and s-th components of $\mathbf{X}_t$ by

$$R_{rs}(\lambda) = \frac{f_{rs}(\lambda)}{\left(f_{rr}(\lambda) f_{ss}(\lambda)\right)^{1/2}}, \quad r, s, = 1, \ldots, N,$$

Robinson (1995a) justified the LP regression when the coherence matrix at zero frequency is non-singular, so the long run variance matrix of $\mathbf{X}_t$ is full rank, and the elements of $\mathbf{f}_{\mathbf{X}}(\lambda)$ satisfy $(\mathrm{d}/\mathrm{d}\lambda)\log f_{rs}(\lambda) = O(|\lambda|^{-1})$ as $\lambda \to \infty$.

The simultaneous estimation of $\mathbf{d} = (d_1, \ldots d_N)'$ and of the intercept coefficients $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)'$ is obtained by means of a multivariate regression with dependent vector $\mathbf{Y}_j^{(K)} = \left(Y_{1,j}^{(K)}, \ldots, Y_{N,j}^{(K)}\right)'$,

$$\begin{pmatrix} \widehat{\boldsymbol{\alpha}}_m \\ \widehat{\mathbf{d}}_m^{LP} \end{pmatrix} = \operatorname{vec}\left[\sum_j \mathbf{Y}_j^{(K)}\mathbf{Z}_j \left(\sum_j \mathbf{Z}_j\mathbf{Z}_j'\right)^{-1}\right]$$

where $\mathbf{Z}_j = (1, z_j)'$. Then $\widehat{G}_{n,m}^{LP} = \exp\left(\widehat{\alpha}_{n,m} - \psi(K)\right), n = 1, \ldots, N$, and the asymptotic properties of the LP regression coefficients are described by

$$\begin{bmatrix} \frac{m^{1/2}}{\log T}\left(\widehat{\boldsymbol{\alpha}}_m - \boldsymbol{\alpha}\right) \\ 2m^{1/2}\left(\widehat{\mathbf{d}}_m^{LP} - \mathbf{d}\right) \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0,\ K\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \otimes \Omega^{(K)}\right),$$

where, paralleling standard OLS theory, $\Omega^{(K)}$ can be estimated consistently by

$$\tilde{\Omega}_m^{(K)} = \frac{K}{m}\sum_j \tilde{\mathbf{u}}_j\tilde{\mathbf{u}}_j',$$

using the OLS residual vector $\tilde{\mathbf{u}}_j$. The factor $K/(4m)$ in the approximate variance of $\widehat{\mathbf{d}}_m^{LP}$, $\Omega^{(K)}K/(4m)$, can be replaced in finite samples by $(\sum_j \Lambda_j^2)^{-1}$ to match the standard computation of standard errors in linear regression. Velasco (2000) and Hurvich, Moulines and Soulier (2002) show the robustness of these results to some non-Gaussian linear processes.

Following Robinson (1995a), we can define the Wald-type tests of the hypothesis that $H_0 : \mathbf{P}\mathbf{d} = \boldsymbol{\varrho}$, where $\mathbf{P}$ is a given $q \times N$ matrix and $\boldsymbol{\varrho}$ is a $q \times 1$ vector, which rejects the null if

$$W_m = \frac{4m}{K}\left(\mathbf{P}\widehat{\mathbf{d}}_m^{LP} - \boldsymbol{\varrho}\right)'\left(\mathbf{P}\tilde{\Omega}_m^{(K)}\mathbf{P}'\right)^{-1}\left(\mathbf{P}\widehat{\mathbf{d}}_m^{LP} - \boldsymbol{\varrho}\right)$$

is significantly large compared to the χ_q^2 distribution. If some restrictions on $\mathbf{d}$ are assumed, we can obtain more efficient estimates by using this information. Thus, if it is known that $d_1 = \cdots = d_N = d$, so that $\mathbf{d} = d\mathbf{1}_N$, $\mathbf{1}_N$ being the $N \times 1$ vector of ones, the following generalized LS (GLS) type of estimate is proposed,

$$\hat{d}_m^{GLS} = -\frac{\sum_j \Lambda_j \mathbf{1}_N'\tilde{\Omega}_m^{(K)-1}\mathbf{Y}_j^{(K)}}{2\mathbf{1}_N'\tilde{\Omega}_m^{(K)-1}\mathbf{1}_N\sum_j \Lambda_j^2},$$

whose asymptotic variance can be consistently estimated by $(K/4)\left(\mathbf{1}_N'\tilde{\Omega}_m^{(K)-1}\mathbf{1}_N\right)^{-1}$.

The idea of the LP regression has been extended to variance decompositions other that the frequency domain one given by the periodogram. The tapered periodogram provides a first possibility, which is analyzed in Section 3.1 in the context of memory estimation of nonstationary processes. A second proposal is related to wavelet analysis in the context of self-similar processes, which are characterized by a scale invariant property and whose increments display long range dependence (Taqqu (2003)). The sample wavelet coefficients are the counterpart of the DFT giving a decomposition of the variance of X_t at different scales. The wavelet coefficients present some fundamental characteristics similar to those of the DFT, i.e. they reproduce in the wavelet domain the power laws defining the scale invariance of self-similar and long memory processes, being weakly correlated (see, e.g. Twefik and Kim (1992) and Bardet, Lang, Moulines and Soulier (2000)). This has been exploited in the design of estimates similar to the LP regression by Jensen (1999) and Bardet et al. (2000) among others. Furthermore, the wavelet coefficients under appropriate choices of the mother wavelet function can satisfy some higher order properties (zero moments) which guarantee robustness to deterministic trends, similar to those that can be obtained through tapering (see equation (30) below). These properties, together with computationally efficient multiresolution algorithms, make wavelets amenable for the analysis of long range dependent series with possible trending or nonstationary behaviors.

2.2 Local Whittle estimation

Based on a proposal of Künsch (1987), Robinson (1995b) studied the Gaussian semiparametric estimate of d based on the minimization of a local Whittle frequency domain (minus) log-likelihood,

$$\mathcal{L}_m(d,G) = \frac{1}{m}\sum_{j=1}^{m}\left\{\log G\lambda_j^{-2d} + \frac{I_X(\lambda_j)}{G\lambda_j^{-2d}}\right\},$$

using the semiparametric model (3), which is valid for such frequencies. Given the interval of admissible estimates of d by $D = [\nabla_1, \nabla_2]$, where ∇_1 and ∇_2 are numbers such that $-\frac{1}{2} < \nabla_1 < \nabla_2 < \frac{1}{2}$, the local Whittle (LW) estimates are defined by

$$(\widehat{d}_m^{LW}, \widehat{G}_m^{LW}) = \arg\min_{d\in D, 0<G<\infty} \mathcal{L}_m(d,G).$$

Concentrating out $\widehat{G}_m^{LW}$ we obtain that

$$\widehat{d}_m^{LW} = \arg\min_{d\in D} R_m(d),$$

where

$$R_m(d) = \log \widehat{G}_m^{LW}(d) - 2d\frac{1}{m}\sum_{j=1}^{m}\log\lambda_j, \qquad \widehat{G}_m^{LW}(d) = \frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2d} I_X(\lambda_j). \tag{10}$$

For linear time series with homoskedastic martingale difference innovations, with spectral density satisfying the same regularity conditions as for LP estimation, Robinson (1994b) found that $\widehat{d}_m^{LW}$ is consistent and its asymptotic normal distribution is

$$2m^{1/2}\left(\widehat{d}_m^{LW} - d\right) \xrightarrow{d} \mathcal{N}(0,1), \tag{11}$$

so its asymptotic variance is free of nuisance parameters and smaller than that of the LP estimate. Therefore, when using the same number of frequencies, LW estimation is more efficient that the LP regression. The bandwidth m has to satisfy

$$\frac{1}{m} + \frac{m^5\log^2 m}{T^4} \to 0 \tag{12}$$

if the approximation (3) has error $O\left(|\lambda|^{2-2d}\right)$ as $\lambda \to 0$.

Following Lobato and Robinson (1998) and Lobato (1999), we can propose a joint estimate of the memory parameters of a vector $\mathbf{X}_t$ based on the semiparametric multivariate model for the spectral density matrix $\mathbf{f}_{\mathbf{X}}(\lambda)$, such as

$$\mathbf{f}_{\mathbf{X}}(\lambda) \sim \mathbf{\Lambda}(\mathbf{d})\,\mathbf{\Xi}_{\mathbf{X}}\mathbf{\Lambda}(\mathbf{d}) \quad \text{as } \lambda \to +0, \tag{13}$$

where $\mathbf{\Xi_X}$ is a positive definite (complex) hermitian matrix and $\mathbf{\Lambda}(\mathbf{d}) = \mathrm{diag}\left(\lambda^{-d_1}, \ldots, \lambda^{-d_N}\right)$. For fractional models with long run variance matrix $\mathbf{G_X}$ we have that $\mathbf{\Xi_X} = \mathbf{\Xi_X}(\mathbf{d}) = \Phi(\mathbf{d})\,\mathbf{G_X}\Phi^*(\mathbf{d})$, where $\mathbf{G_X} = \{g_{ab}\}$ is a real positive definite matrix, $\mathbf{\Phi}(\mathbf{d}) = \mathrm{diag}\left(e^{i\pi d_1/2}, \ldots, e^{i\pi d_N/2}\right)$, and $*$ means simultaneous transposition and complex conjugation. Therefore (13) ignores some information about how the real and imaginary parts of $\mathbf{\Xi_X}(\mathbf{d})$ relate in terms of $\mathbf{d}$.

The local Whittle likelihood, as a function of the memory parameters $\mathbf{d}$ and the scale matrix $\mathbf{\Xi}$, is given by

$$\mathcal{L}_m(\mathbf{d}, \mathbf{\Xi}) = \frac{1}{m}\sum_{j=1}^{m}\left\{\log\det\left[\mathbf{\Lambda}_j(\mathbf{d})\,\mathbf{\Xi}\mathbf{\Lambda}_j(\mathbf{d})\right] + tr\left[(\mathbf{\Lambda}_j(\mathbf{d})\,\mathbf{\Xi}\mathbf{\Lambda}_j(\mathbf{d}))^{-1}\mathbf{I_X}(\lambda_j)\right]\right\}, \tag{14}$$

where $\mathbf{\Lambda}_j(\mathbf{d}) = \mathrm{diag}\{\lambda_j^{-d_1}, \ldots, \lambda_j^{-d_N}\}$ and $\mathbf{I_X}(\lambda_j) = \mathbf{w_X}(\lambda_j)\mathbf{w_X}(\lambda_j)^*$ is the periodogram matrix of $\mathbf{X}_t$. Since

$$\mathcal{L}_m(\mathbf{d}, \mathbf{\Xi}) = \frac{1}{m}\sum_{j=1}^{m}\left\{2\log\det\left[\mathbf{\Lambda}_j(\mathbf{d})\right] + \log\det\left[\mathbf{\Xi}\right] + tr\left[\mathbf{\Xi}^{-1}\mathbf{\Lambda}_j^{-1}(\mathbf{d})\,\mathbf{I_X}(\lambda_j)\mathbf{\Lambda}_j^{-1}(\mathbf{d})\right]\right\}$$

we find that

$$\frac{\partial}{\partial\mathbf{\Xi}}\mathcal{L}_m(\mathbf{d}, \mathbf{G}) = \frac{1}{m}\sum_{j=1}^{m}\left\{\mathbf{\Xi}^{-1} - \mathbf{\Xi}^{-1}\left[\mathbf{\Lambda}_j^{-1}(\mathbf{d})\,\mathbf{I_X}(\lambda_j)\mathbf{\Lambda}_j^{-1}(\mathbf{d})\right]\mathbf{\Xi}^{-1}\right\},$$

and on setting

$$\widehat{\mathbf{\Xi}}_m(\mathbf{d}) = \frac{1}{m}\sum_{j=1}^{m}\mathbf{\Lambda}_j^{-1}(\mathbf{d})\,\mathbf{I_X}(\lambda_j)\mathbf{\Lambda}_j^{-1}(\mathbf{d}), \tag{15}$$

we obtain from (14) the following concentrated objective function (Lobato (1999))

$$\Upsilon_m(\mathbf{d}) = -\frac{2}{m}\sum_{i=1}^{N} d_i \sum_{j=1}^{m}\log(\lambda_j) + \log\det\left[\widehat{\mathbf{\Xi}}_m(\mathbf{d})\right],$$

because $\log\det\left[\mathbf{\Lambda}_j(\mathbf{d})\right] = -\log(\lambda_j)\sum_{i=1}^{N} d_i$.

The estimation procedure proposed by Lobato (1999) is a two step estimator based on this objective function. The first step is to compute the univariate local Whittle estimate for every series (denote that vector by $\widehat{\mathbf{d}}_m^{(1)}$) and the second step is obtained through the following expression

$$\widehat{\mathbf{d}}_m^{(2)} = \widehat{\mathbf{d}}_m^{(1)} - \left(\left.\frac{\partial^2\Upsilon_m(\mathbf{d})}{\partial\mathbf{d}\partial\mathbf{d}'}\right|_{\widehat{\mathbf{d}}_m^{(1)}}\right)^{-1}\left(\left.\frac{\partial\Upsilon_m(\mathbf{d})}{\partial\mathbf{d}}\right|_{\widehat{\mathbf{d}}_m^{(1)}}\right). \tag{16}$$

Further iterations could be considered, e.g. $\widehat{\mathbf{d}}_m^{(s)}$, $s = 2, 3, \ldots$, having the same first order efficiency. An estimator of the long run variance $\mathbf{G_X}$ can be constructed as

$$\widehat{\mathbf{G}}_{\mathbf{X},\mathbf{m}} = \widehat{\mathbf{G}}_{\mathbf{X},\mathbf{m}}\left(\widehat{\mathbf{d}}_m^{(2)}\right) = \mathrm{Re}\left\{\widehat{\mathbf{\Phi}}_m^*\widehat{\mathbf{\Xi}}_{\mathbf{X},m}\widehat{\mathbf{\Phi}}_m\right\} \tag{17}$$

where $\widehat{\mathbf{\Xi}}_{\mathbf{X},m} = \widehat{\mathbf{\Xi}}_m(\widehat{\mathbf{d}}_m^{(2)})$, $\widehat{\mathbf{\Phi}}_m = \mathbf{\Phi}\left(\widehat{\mathbf{d}}_m^{(2)}\right)$ and Re stands for real part.

Extending Robinson's (1995b) analysis, Lobato (1999) showed that

$$m^{1/2}\left(\widehat{\mathbf{d}}_m^{(2)} - \mathbf{d}\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{E}^{-1}\right),$$

where $\mathbf{E} = 2\left(I_N + \mathbf{\Xi_X} \circ \mathbf{\Xi_X^{*-1}}\right)$ and $\circ$ denotes the element by element Hadamard matrix product. He assumed that $\mathbf{X}_t$ is a linear process given by

$$\mathbf{X}_t = \mu + \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}$$

where ε_t is a martingale difference sequence with constant first four conditional moments, and the transfer function $A(\lambda) = \sum_{j=0}^{\infty} A_j e^{ij\lambda}$ is differentiable around $\lambda = 0$. The asymptotic variance can be estimated by using the previous estimate of $\mathbf{\Xi_X}$, obtaining $\hat{\mathbf{E}}_m = 2\left(I_N + \widehat{\mathbf{\Xi}}_{\mathbf{X},m} \circ \widehat{\mathbf{\Xi}}^{*-1}_{\mathbf{X},m}\right)$.

In order to achieve more efficient estimation of the vector $\mathbf{d}$, we could consider explicitly that $\mathbf{\Xi_X}$ is a function of $\mathbf{d}$, $\mathbf{\Xi_X}(\mathbf{d}) = \Phi(\mathbf{d})\,\mathbf{G_X}\Phi(\mathbf{d})^*$, see Shimotsu (2003). Furthermore, if we want to impose a rate for the semiparametric approximation (13) valid for fractional time series, we could consider

$$\mathbf{f_X}(\lambda) = \bar{\mathbf{\Lambda}}(\mathbf{d})\,\mathbf{G_X}\bar{\mathbf{\Lambda}}^*(\mathbf{d})\left(1 + O\left(\lambda^2\right)\right) \quad \text{as } \lambda \to +0, \tag{18}$$

where now $\bar{\mathbf{\Lambda}}(\mathbf{d}) = \operatorname{diag}\left(\lambda^{-d_1} e^{i(\pi-\lambda)d_1/2}, \ldots, \lambda^{-d_N} e^{i(\pi-\lambda)d_N/2}\right)$, so we can obtain $\bar{\mathbf{\Lambda}}(\mathbf{d})\,\mathbf{G_X}\bar{\mathbf{\Lambda}}^*(\mathbf{d}) \sim \mathbf{\Lambda}(\mathbf{d})\,\mathbf{\Xi_X}(\mathbf{d})\,\mathbf{\Lambda}(\mathbf{d})$ as $\lambda \to +0$.

As with LP estimates, efficient improvements are possible if a valid restriction on the vector $\mathbf{d}$ is used. If $d_1 = \cdots = d_N = d$, then we can set

$$\tilde{d}_m^{LW} = \arg\min_{d \in D} \tilde{\Upsilon}_m(d) \tag{19}$$

where

$$\tilde{\Upsilon}_m(d) = -\frac{2N}{m} d \sum_{j=1}^{m} \log(\lambda_j) + \log\det\left[\widehat{\mathbf{G}}_{\mathbf{X},m}(d\mathbf{1}_N)\right],$$

and now $\widehat{\mathbf{G}}_{\mathbf{X},m}(d\mathbf{1}_N) = \operatorname{Re}(\widehat{\mathbf{\Xi}}_{\mathbf{X},m})$, given the restriction of a unique d. The asymptotic variance of $\tilde{d}_m^{LW}$, $1/4N$, reflects the extra information used.

Wald tests are easily implemented as with LP estimates, but using the objective function $\Upsilon_m(d)$ we can also use the Lagrange Multiplier and Likelihood Ratio principles. An LM-type test of $d_1 = \cdots = d_N = 0$ was proposed by Lobato and Robinson (1998). The LM test for $H_0 : \mathbf{Pd} = \boldsymbol{\varrho}$ uses the statistic

$$LM_m = m \frac{\partial \Upsilon_m(\tilde{\mathbf{d}}_m^{LW})}{\partial \mathbf{d}'} \left[\mathbf{P}\hat{\mathbf{E}}_m \mathbf{P}'\right] \frac{\partial \Upsilon_m(\tilde{\mathbf{d}}_m^{LW})}{\partial \mathbf{d}},$$

compared to a χ_q^2 distribution, where $\tilde{\mathbf{d}}_m^{LW}$ minimizes $\Upsilon_m(\mathbf{d})$ subject to $\mathbf{Pd} = \boldsymbol{\varrho}$ and $\hat{\mathbf{E}}_m$ can be computed also under this restriction. For the test of $d_1 = \cdots = d_N = d_0$, the LM statistic reduces to

$$LM_m = \frac{m}{4N}\left[\frac{\partial \Upsilon_m^*(d_0)}{\partial d}\right]^2$$

with $q = 1$.

2.3 Averaged periodogram estimation

Many alternative semiparametric estimates have been proposed, both in the time and frequency domain. We now describe briefly the proposal of Robinson (1994a), the averaged periodogram estimate of d,

$$\hat{d}_{m,q}^{AP} = \frac{1}{2} - \frac{\log\left\{F_{X,T}(q\lambda_m)/F_{X,T}(\lambda_m)\right\}}{2\log q},$$

where $q \in (0,1)$ is a user chosen tuning parameter and $F_{X,T}$ is the averaged periodogram (AP),

$$F_{X,T}(\lambda) = \frac{2\pi}{T} \sum_{j=1}^{[T\lambda/2\pi]} I_X(\lambda_j).$$

The AP will be important in the discussion of narrow band estimates of long run relationships.

Note that $\hat{d}_{m,q}^{AP} \in (-\infty, 0.5]$, so it cannot estimate nonstationary values of d. Robinson (1994a) showed that $F_{X,T}(\lambda_m)$ is a consistent estimate of $F_X(\lambda_m) = \int_0^{\lambda_m} f_X(z)dz$ if $m^{-1} + mT^{-1} \to 0$ with the sample size when X_t is a linear process with martingale innovations. Thus it is easy to show that, under (3), $\hat{d}_{m,q}^{AP}$ is also consistent for d.

Lobato and Robinson (1996) analyzed the asymptotic distribution of the AP estimate. This is only normal for $d < 0.25$, for which $f_X(\lambda)$ is square integrable around $\lambda = 0$. In particular, if $d \in (0, 0.25)$,

$$m^{1/2}\left(d_{m,q}^{AP} - d\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\left(1 + q^{-1} - 2q^{-2d}\right)}{\log^2 q} \frac{(0.5-d)^2}{1-4d}\right).$$

For $d \in (0.25, 0.5)$ the asymptotic distribution of $d_{m,q}^{AP}$ is a functional of a Rosenblatt variate. The asymptotic variance of $d_{m,q}^{AP}$ when $d < 0.25$ depends on q and d, and by its minimization Lobato and Robinson (1996) find that for each d there is an optimal value of q. Lobato (1997) extends some of these results to a multivariate time series framework and Robinson and Marinucci (2000) to nonstationary vectors, see Section 3.

2.4 Bias reduction and bandwidth choice

The most important issue when applying any semiparametric memory estimate is the decision on the number of Fourier frequencies m to be used. For these frequencies we regard the model (3) as approximately valid, but increasing m leads to a reduction of the variance of estimates at the cost of an increment in bias due to the consideration of too high frequencies where the semiparametric model is not appropriate. We concentrate in this section on univariate and no pooled ($K = 1$) estimates.

Under the assumption that

$$f_X(\lambda) = |2\sin(\lambda/2)|^{-2d} f^*(\lambda) \tag{20}$$

where $f^*(\lambda)$ is nonnegative, even, integrable, twice continuously differentiable and positive at $\lambda = 0$, Hurvich et al. (1998) obtained the Mean Square Error (MSE) of the LP estimate and derived the expression for the MSE-optimal bandwidth,

$$m_{LP}^{opt} = T^{4/5}\left[\frac{27}{128\pi^2}\left(\frac{f^*(0)}{\ddot{f}^*(0)}\right)^2\right]^{1/5},$$

assuming for the second derivative $\ddot{f}^*$ of f^* that $\ddot{f}^*(0) \neq 0$. This expression gives the MSE-optimal rate for m, $T^{4/5}$, but depends on the short run dynamics of X_t described by $\ddot{f}^*$. Based on this formula, Hurvich and Deo (1999) devised a plug-in estimate of the optimal constant in m_{LP}^{opt} by means of an augmented LP regression,

$$\log I_X(\lambda_j) = \alpha + dw_j + \rho\frac{\lambda_j^2}{2} + u_j, \quad j = 1, \ldots, m_\rho, \tag{21}$$

where now $w_j = -2\log(2\sin(\lambda_j/2))$. Noting that $\dot{f}^*(0) = 0$ by the evenness of $f^*(\lambda)$, the OLS estimate of ρ in the regression (21) is consistent for $b_2 = \ddot{f}^*(0)/f^*(0)$, since we can write

$$\log f^*(\lambda) = \log f^*(0) + b_2 \frac{\lambda^2}{2} + O(\lambda^3)$$

if $f^*(\lambda)$ is smooth enough. The initial choice of the auxiliary bandwidth m_ρ is given by $m_\rho = AT^a$ for some $a > 3/4$ and some positive constant A. Note that when m is proportional to $T^{4/5}$, $m = BT^{4/5}$ say, the bandwidth conditions for the asymptotic normality (8), e.g. (9), are no longer valid, so asymptotic inference has to be adapted to take into account the asymptotic bias. Thus, introducing a bias correction, it is obtained in this case that

$$2m^{1/2}\left(\widehat{d}_m^{LP} - d\right) - \frac{4}{9}\pi^2 b_2 B^{4/5} \xrightarrow{d} \mathcal{N}\left(0, \frac{\pi^2}{6}\right),$$

leading to bias corrected versions of $\widehat{d}_m^{LP}$ when using MSE-optimal bandwidths.

Andrews and Guggenberger (2003) generalize the previous idea to obtain LP estimates with reduced bias in augmented regressions. To that end, it is assumed that

$$\log f^*(\lambda) = \log f^*(0) + \sum_{k=1}^{r} b_{2k} \frac{\lambda^{2k}}{(2k)!} + o(\lambda^{2r}) \quad \text{as } \lambda \to 0+, \tag{22}$$

where

$$b_k = \left(\frac{\mathrm{d}}{\mathrm{d}\lambda}\right)^k \log f^*(\lambda)\Big|_{\lambda=0},$$

and the polynomial LP (PLP) estimate $\widehat{d}_{r,m}^{PLP}$ of order r is given by the corresponding OLS coefficient in the linear regression

$$\log I_X(\lambda_j) = \alpha + dz_j + \sum_{k=1}^{r} \rho_k \lambda_j^{2k} + u_j. \tag{23}$$

Andrews and Guggenberger (2003) show that the OLS estimate of this regression satisfies

$$2m^{1/2}\left(\widehat{d}_{r,m}^{PLP} - d\right) - \upsilon_T(r) \xrightarrow{d} \mathcal{N}\left(0, c_r \frac{\pi^2}{6}\right) \tag{24}$$

if $m = O\left(T^{2\phi/(2\phi+1)}\right)$, $\phi = 2 + 2r$, where $\upsilon_T(r)$ is the asymptotic bias, $c_0 = 1$ and $c_r = \left(1 - \mu_r'\Gamma_r^{-1}\mu_r\right)^{-1}$, $r \geq 1$, with

$$\begin{aligned}
\mu_{r,k} &= \frac{2k}{(2k+1)^2}, \quad k = 1, \dots, r \\
\Gamma_{r,ik} &= \frac{4ik}{(2k+2i+1)(2i+1)(2k+1)}, \quad i, k = 1, \dots, r.
\end{aligned}$$

Assuming enough smoothness of $\log f^*$ in (22) so that the error term is $O(\lambda^{2r+2})$, the asymptotic bias $\upsilon_T(r)$ is given by

$$\upsilon_T(r) = m^{(5/2)+2r} T^{-(2+2r)} b_{2+2r} \tau_r,$$

where

$$\begin{aligned}
\tau_r &= \frac{\kappa_r c_r}{2}\left(1 - \mu_r'\Gamma_r^{-1}\xi_r\right), \\
\kappa_r &= \frac{(2\pi)^{2+2r}(2+2r)}{(3+2r)!(3+2r)}
\end{aligned}$$

and $\xi_r = \left(\xi_{r,1}, \ldots, \xi_{r,r}\right)'$ with

$$\xi_{r,k} = \frac{2k(3+2r)}{(2r+2k+3)(2k+1)}, \quad k = 1, \ldots, r.$$

The bias term $v_T(r)$ disappears in the asymptotic distribution (24) when a slower bandwidth $m = o\left(T^{2\phi/(2\phi+1)}\right)$ is used instead of the MSE-optimal one. This analysis also allows calculation of the asymptotic MSE of the PLP estimate, generalizing the expression for the optimal bandwidth,

$$m_{PLP}^{opt} = T^{(4+4r)/(5+4r)} \left[\frac{\pi^2 c_r}{24(4+4r)\tau_r^2 b_{2+2r}^2} \right]^{1/(5+4r)}.$$

The unknown b_{2+2r} can be estimated by means of an augmented regression similar to (21).

Similar studies have been conducted for other semiparametric estimates. Thus, for spectral densities satisfying

$$f(\lambda) = G\lambda^{-2d}\left(1 + E_\gamma \lambda^\gamma + o(\lambda^\gamma)\right), \quad \text{as } \lambda \to 0, \tag{25}$$

for some $\gamma \in (0, 2]$, $E_\gamma \neq 0$, Henry and Robinson (1996) approximate the MSE of the LW estimate, for which the optimal bandwidth is given by

$$m_{LW}^{opt} = T^{2\gamma/(1+2\gamma)} \left[\frac{(1+\gamma)^4}{2\gamma^3 E_\gamma^2 (2\pi)^{2\gamma}} \right]^{1/(1+2\gamma)}, \tag{26}$$

and propose an iterative method to estimate the unknown constant E_γ in m_{LW}^{opt}.

For spectral densities satisfying (20), Andrews and Sun (2004) investigate the MSE, optimal bandwidth and asymptotic properties of a generalization of the LW estimate similar to the PLP. They consider the local polynomial Whittle likelihood

$$\mathcal{L}_{r,m}(d, G, \boldsymbol{\theta}) = \frac{1}{m} \sum_{j=1}^{m} \left\{ \log\left[G\lambda_j^{-2d} \exp\left(-p_r(\lambda_j; \boldsymbol{\theta})\right) \right] + \frac{I_X(\lambda_j)}{G\lambda_j^{-2d} \exp\left(-p_r(\lambda_j; \boldsymbol{\theta})\right)} \right\}, \tag{27}$$

for $r \geq 0$, where

$$p_r(\lambda_j; \boldsymbol{\theta}) = \sum_{k=1}^{r} \theta_k \lambda_j^{2k}, \quad \boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'.$$

This likelihood includes higher order terms from expanding $\log f^*(\lambda)$ around $\lambda = 0$ up to order $2r$, as in (22). Concentrating out G we obtain that

$$\left(\widehat{d}_{r,m}^{PLW}, \boldsymbol{\theta}_{r,m}^{PLW} \right) = \arg \min_{d, \boldsymbol{\theta} \in D} R_{r,m}(d, \boldsymbol{\theta}),$$

where

$$\begin{aligned} R_{r,m}(d, \boldsymbol{\theta}) &= \log \widehat{G}_{r,m}^{PLW}(d, \boldsymbol{\theta}) - \frac{1}{m} \sum_{j=1}^{m} p_r(\lambda_j; \boldsymbol{\theta}) - 2d \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j + 1, \\ \widehat{G}_{r,m}^{PLW}(d, \boldsymbol{\theta}) &= \frac{1}{m} \sum_{j=1}^{m} \lambda_j^{2d} \exp\left(-p_r(\lambda_j; \boldsymbol{\theta})\right) I_X(\lambda_j). \end{aligned}$$

Andrews and Sun (2004) show the consistency and asymptotic normality of $\left(\widehat{d}_{r,m}^{PLW}, \boldsymbol{\theta}_{r,m}^{PLW} \right)$ when $m^{4r+1}/T^{4r} \to \infty$ and $m^{2\phi+1}/T^{2\phi} = O(1)$, $\phi = 2 + 2r$, under similar regularity assumptions to Robinson (1995b) together with (22). In particular,

$$2m^{1/2} \left(\widehat{d}_{r,m}^{PLW} - d \right) - v_T(r) \xrightarrow{d} \mathcal{N}\left(0, c_r\right).$$

Therefore, the PLW estimate has the same asymptotic bias as the PLP estimate, but retains its efficiency even if we consider r correcting terms. Similarly, it is possible to generalize the expression for the optimal bandwidth of the LW estimate,

$$m_{PLW}^{opt} = T^{(4+4r)/(5+4r)} \left[\frac{c_r}{16(1+r)\tau_r^2 b_{2+2r}^2} \right]^{1/(5+4r)}.$$

Note that for $r = 0$ and $\gamma = 2$ this gives the same expression (26) for m_{LW}^{opt} given by Henry and Robinson (1996), noting that $E_2 = b_2/2$, which is the only unknown in the optimal bandwidth.

Robinson and Henry (2003) provide a different approach to the problem of bias reduction. They propose M-estimates based on higher order kernels that are able to nest different classes of semiparametric estimates of d, such as versions of the LP and LW estimates. This class of M-estimates is based on a kernel function $k_q(u)$ with the property that

$$\int_0^1 k_q(u)du = 1,$$

and setting $U_{iq} = \int_0^1 (1 + \log u)\, u^{2i} k_q(u)du$, it is required that $U_{iq} = 0$, $i = 1, \ldots, q-1$, and $U_{qq} \neq 0$ for some $q \geq 1$. Other important ingredients are a real valued monotonic function ψ, which is particularized to the Box-Cox transformation, $\psi_\alpha(z) = (z^\alpha - 1)/\alpha$ for $\alpha > 0$, and $\psi_0(z) = \log z$, and a function $g(\lambda)$ which is asymptotically equivalent to λ, in the sense that

$$g(\lambda) = \lambda + G\lambda^3 + o(\lambda^3) \quad \text{as} \quad \lambda \to 0+,$$

such as $g(\lambda) = 2\sin(\lambda/2)$. Then the q-order kernel M-estimate of d, $\hat{d}_m^{Mq}$, is given as a solution of the equation

$$\sum_j k_q\left(\frac{j}{m}\right) v_{qj}(g)\psi_\alpha\left(\bar{I}_X^{(K)}(\lambda_j)g(\lambda_j)^{2\hat{d}_m^{Mq}}\right) = 0$$

where

$$v_{qj}(g) = \log g(\lambda_j) - \frac{\sum_j k_q\left(\frac{j}{m}\right)\log g(\lambda_j)}{\sum_j k_q\left(\frac{j}{m}\right)}$$

and

$$\bar{I}_X^{(K)}(\lambda_j) = \sum_{k=1}^{K} I_X(\lambda_{j+k-K}), \quad j = K, 2K, \ldots, m.$$

Thus, when $K = 1, q = 1$ and $\alpha = 0$ we obtain the LP estimate, whereas with $\alpha = 1$ we get the LW estimate and with values of $\alpha \in (0, 1)$ we interpolate between both methods. The asymptotic variance also varies with α between those found for $\alpha = 0, 1$. On the other hand, for a particular ψ, choosing a higher order kernel with $q \geq 2$ allows room for potential bias and MSE reduction, under the assumption that $f^*(\lambda) = f(\lambda)g(\lambda)^{2d}$ is $2q + 1$ times differentiable, cf. (22). Finally, the choice of g does not affect the asymptotic variance, but may have important effects on bias given this previous assumption.

The reviewed bias reduction and estimation of optimal bandwidth techniques rely on a priori assumptions on the smoothness of the spectral density, given by the value of the parameter γ or the number of derivatives of $\log f^*$. In practice, it is not easy to obtain information about such restrictions, so adaptive techniques have been developed. Giraitis, Robinson and Samarov (1997) showed that the LP estimate is rate optimal among a class of semiparametric estimates for processes with spectral density of γ degree of smoothness, cf. (25), whereas Giraitis, Robinson and Samarov

(2000) propose an adaptive estimate to the unknown degree of smoothness based on a modified LP regression, whose asymptotic risk is larger than the optimal risk only by a logarithmic factor. Related results have been obtained for the polynomial LP estimate by Andrews and Guggenberger (2003) and for the LW estimate by Andrews and Sun (2004). Alternatively, Hurvich and Beltrao (1994) propose a cross-validation method to estimate the integrated local MSE of any estimate of d around zero frequency and base a bandwidth choice on the minimization of such an estimate. For the AP estimate, Robinson (1994a) provides expressions for the MSE and the optimal bandwidth, which Delgado and Robinson (1996a) estimate by means of a plug-in iterative method, and Delgado and Robinson (1996b) find optimal kernels for the averaging.

2.5 Global methods: FEXP and FAR estimates

The fractional exponential (FEXP) estimate, proposed by Robinson (1994c), consists of a LP regression similar to (23), but expanding $\log f^*$ on a cosine basis, so that the coefficients $\theta_k = \int_{-\pi}^{\pi} \log f^*(\lambda) \cos k\lambda d\lambda$ define the cepstrum of f^*. The FEXP estimate $\hat{d}_r^{FEXP}$ of d is given by the corresponding coefficient in the OLS estimation of

$$Y_{X,j}^{(K)} = dz_j + \sum_{k=0}^{r} \theta_k \cos k\lambda_j + u_j, \quad j = K, 2K, \ldots, m,$$

allowing for periodogram pooling, $K \geq 1$. If we let $r \to \infty$ with T, then we can approximate nonparametrically the whole of f^*, so these methods are called global, in contrast with local methods, such as the LP or LW estimates. The asymptotic properties of the FEXP estimate have been analyzed by Moulines and Soulier (1999) for Gaussian series and by Hurvich, Moulines and Soulier (2002) for non-Gaussian series, see also Hurvich and Brodsky (2001). The analysis relies on the smoothness of f^*, so that the θ_k are square summable, and on a related restriction on r,

$$\frac{1}{r} + \frac{r \log^5 T}{T} + \left(\frac{T}{r}\right)^{1/2} \sum_{k=r}^{\infty} |\theta_k| \ \to 0.$$

Then

$$\left(\frac{T}{r}\right)^{1/2} \left(\hat{d}_r^{FEXP} - d\right) \xrightarrow{d} \mathcal{N}\left(0, K\Omega^{(K)}\right),$$

showing that the convergence rate of $\hat{d}_r^{FEXP}$ can be very close to the parametric rate of $T^{1/2}$ if r can be chosen very small so that it approximates f^* with fidelity. Iouditsky, Moulines and Soulier (2002) investigate an adaptive FEXP estimate, extending results of Hurvich (2001), who had proposed a local version of Mallow's C_L criterion to select a FEXP model by minimizing the asymptotic MSE.

Another global estimate in a similar spirit is the Fractional AutoRegressive (FAR) estimate, which is based on fitting an ARFIMA$(r, d, 0)$ model with r increasing with sample size T. Bhansali and Kokoszka (2001) have showed the consistency of this estimate when based on a full-band Whittle estimate.

3 Extensions

We consider in this section two natural extensions of the semiparametric model (3). The first relaxes the assumption of stationarity, $d < 0.5$, so that it is possible to check the robustness of the previous methods to the trending nonstationary of fractionally integrated series for large d. In this case special modifications of semiparametric memory estimates might be necessary to robustify inference against possible nonstationarity of unknown degree. The second extension considers the possibility of persistence at frequencies different from zero, which poses new problems and requires some extra care when applying semiparametric methods.

3.1 Nonstationary long memory

There are alternative ways of defining possibly nonstationary trending processes with persistence characterized by a long memory parameter d which can take values larger than 0.5, nesting in this way $I(1)$ unit root processes. Following Hurvich and Ray (1995), we can say that the non-stationary process $\{X_t\}$ has memory parameter $d \in \left[\frac{1}{2}, \frac{3}{2}\right)$ if the zero mean covariance stationary process $\Delta X_t = (1-L)\, X_t$ has spectral density

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

where $f^*(\lambda)$ is as in (20). Then, we can write, for any $t \geq 1$,

$$X_t = X_0 + \sum_{k=1}^{t} \nu_k, \tag{28}$$

where $\nu_t = \Delta X_t$ and X_0 is a random variable not depending on time t. We also need now to define a generalized spectral density function, which should be equal to the usual spectral density function for stationary X_t, but without restrictions on the value of d when not. From (5), the natural option is to extend the definition of f_X to

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

when $d \geq 0.5$, so that $f_X(\lambda)$ satisfies (3) for some $d < 1.5$, irrespective of X_t being stationary or not. Note that for nonstationary X_t ($d \geq 0.5$), f_X is not integrable in $[-\pi, \pi]$ and is not a proper spectral density. We do not assume that f^* is the spectral density of a stationary and invertible ARMA process as would be the case if ν_t followed a fractional ARIMA model. For example, f^* may have (integrable) poles or zeros at frequencies beyond the origin.

When estimating the memory of nonstationary series, the above definition of nonstationarity based on the increments leads to the so called 'differencing and adding back' method. This consists of taking first differences when it is known that $d \in (0.5, 1.5)$, estimating the memory of the increments by $\hat{d}_m^{\Delta X}$, say, and then setting $\hat{d}_m = \hat{d}_m^{\Delta X} + 1$. Similarly, series with higher degree of nonstationarity, $d \geq 1.5$, can be defined in terms of successive partial sums, and the actual memory estimated with successive differencing to guarantee that the true d is in $(-0.5, 0.5)$. However, such a method requires some a priori knowledge on the degree of nonstationarity of the observed series, which in many cases is difficult to obtain, such as when we suspect that $d \approx 0.5$. A first approach to this problem is the analysis of the previous semiparametric methods, designed for covariance stationary series, under this more general nonstationary framework without assumptions on whether $d < 0.5$ or $d \geq 0.5$. This study is based on some robustness properties of the periodogram.

For short memory processes, the periodogram is an inconsistent but asymptotically unbiased estimate of f_X at continuity points of the spectral density and is approximately independent across frequencies λ_j. Robinson (1995a) extended such results for stationary long range dependent series.

Interestingly, the normalized periodogram $I_X(\lambda_j)/f_X(\lambda_j)$ still has a limit expectation equal to one (and the DFTs at different frequencies are asymptotically uncorrelated) for non-stationary integrated time series at Fourier frequencies moving slowly away from the origin (Hurvich and Ray (1995), Velasco (1999b)). Note also that the DFT is invariant to X_0 at nonzero Fourier frequencies. In fact, it is possible to show the consistency of the LP estimate when $d < 1$, while the asymptotic distribution remains the same, cf. (8), but only when $d < 0.75$ (Velasco (1999b)). Velasco (1999a) obtained related results for the LW estimate, which is also asymptotically normal with an asymptotic variance of 0.25 when $d < 0.75$.

3.2 Tapering

The limitations of the applicability of usual semiparametric inference for large d when there is no a priori assumption on the degree of nonstationarity is due to the periodogram bias caused by the leakage from the nonstationary zero frequency. To alleviate this problem, the traditional remedy in time series analysis is tapering. Define the tapered DFT of X_t, for $t = 1, \ldots, T$ and a taper sequence $\{h_t\}_{t=1}^T$, as

$$w_X^{(h)}(\lambda) = \left(2\pi \sum_{t=1}^T h_t^2\right)^{-1/2} \sum_{t=1}^T h_t X_t \exp(i\lambda t),$$

and the tapered periodogram as $I_X^{(h)}(\lambda) = \left|w_X^{(h)}(\lambda)\right|^2$. The usual DFT has $h_t \equiv 1$. Typically, h_t downweights the observations at both extremes of the sequence, leaving largely unchanged the central part of the data. The improved bias properties of the tapered periodogram also have an immediate counterpart in terms of the DFT. Thus, if h_t is differentiable and vanishes at the boundaries, we obtain by summation by parts that

$$w_X^{(h)}(\lambda) \approx \frac{e^{i\lambda}}{1 - e^{i\lambda}} \left[w_{\Delta X}^{(h)}(\lambda) + \frac{w_X^{(\dot{h})}(\lambda)}{T} \right] \tag{29}$$

for $\lambda \neq 0$, explaining why a sufficiently smooth taper can reproduce the usual properties of the DFT with difference-stationary series. Furthermore, if for some positive integer p, the tapered DFT of integer powers of time t satisfies

$$w_{t^\ell}^{(h)}(\lambda_{jp}) = 0, \qquad \ell = 0, 1, \ldots, p-1, \tag{30}$$

then the taper scheme h_t is able to remove polynomial trends in the observed sequence when concentrating on the restricted set of frequencies λ_{jp}, $jp \neq 0$. This property generalizes the shift invariant property of the usual DFT and helps to define a class of tapers of order p. Lobato and Velasco (2000) provide an application of this property to avoid the effect of nonlinear trends in the traded volume of stocks when estimating its persistence.

There are several alternative tapering schemes having desirable properties to control leakage from remote frequencies. Following Velasco (1999a,b), we may consider a general class of so called tapers of type-I and orders $p = 1, 2, \ldots$, denoted as $\{h_t^{(1,p)}\}$, whose non-scaled DFT satisfies

$$\sum_{t=1}^T h_t^{(1,p)} e^{it\lambda} = \frac{a(\lambda)}{T^{p-1}} \left(\frac{\sin[T\lambda/2p]}{\sin[\lambda/2]} \right)^p, \tag{31}$$

where $a(\lambda)$ is a complex function whose modulus is positive and bounded. Some examples of tapers which satisfy (31) are the triangular Bartlett window ($p = 2$), the Parzen window ($p = 4$) or

Zhurbenko's (1979) class for integer p. Zhurbenko's tapers are obtained by increasingly smooth convolutions of the uniform density, and when $p=1$ give the nontapered DFT weights, $h_t \equiv 1$; when $p=3$ they are similar to the full cosine bell $h_t = (1-\cos\lambda_t)/2$; while for $p=4$ they are very close to Parzen's, given when $T=4N$ by

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

Type-I tapers provide interesting insights into the behavior of the periodogram of time series with spectral densities displaying peaks or troughs, but have the undesirable property of introducing some extra dependence among adjacent periodogram ordinates. This leads to some restrictions in the design and inference of frequency domain memory estimates. The use of a restricted set of Fourier frequencies, such as in (30), to guarantee orthogonality generally leads to an efficiency loss (Velasco (1999b)). To reduce the size of such sets of omitted frequencies, Hurvich, Moulines and Soulier (2002) and Hurvich and Chen (2000) propose alternative type-II complex data tapers,

$$h_t^{(2,p)} = h_{t,T}^{(2,p)} = (1-\exp(i\lambda_t))^{p-1}, \quad p=1,2\ldots, \tag{32}$$

so the tapered periodogram and DFT are obtained by

$$I_X^{(2,p)}(\lambda) = |w_X^{(2,p)}(\lambda)|^2 = \left(2\pi\sum_{t=1}^{T}|h_t^{(2,p)}|^2\right)^{-1}\left|\sum_{t=1}^{T}h_t^{(2,p)}X_t e^{it\lambda}\right|^2.$$

It can be shown that $\sum_{t=1}^{T}|h_t^{(2,p)}|^2 = Ta_p$, where $a_p = \binom{2(p-1)}{p-1}$. Here the order p is equivalent to $p-1$ as set by Hurvich et al. (2002), but is equivalent to the order p of Velasco (1999a,b) or Hurvich and Chen (2000), so both tapers of order $p=1$ give the usual DFT and periodogram. However, for higher order tapers, tapered DFTs at Fourier frequencies are correlated, though type-II tapers are not asymptotically correlated as $T\to\infty$ if $|j-k|\ge p$.

This correlation between tapered periodogram ordinates can be taken into account in different ways when the LP regression is designed. One alternative is to use only asymptotically uncorrelated periodograms. For type-II tapers, such an approach would imply neglecting $p-1$ frequencies of every p in the LP regression. To alleviate the efficiency loss incurred following this policy, Hurvich et al. (2002) use a pooling of periodogram ordinates as proposed by Robinson (1995a). However, as the correlation dies out very fast in $|j-k|$ for both types of tapered DFT, we can consider the use of all frequencies in the LP regression as in the nontapered case, that is, use all

$$\bar{I}_X^{(v,p)}(\lambda_j) = \sum_{k=1}^{K} I_X^{(v,p)}(\lambda_{j+k-K}), \qquad j=K,\ 2K,\ \ldots,\ m,$$

and let that the correlation among adjacent $\log\bar{I}_X^{(v,p)}$ appear in the asymptotic variance of the LP estimates. For type-II tapers the correlation affects at most a fixed number of adjacent periodograms, but for type-I tapers all periodograms display correlation.

Robinson (1995a), for $p=1$ and all K, and Hurvich et al. (2002), for $p>1$ and large K, give explicit expressions for the expectation and variance of the pooled LP, $\log\bar{I}_X^{(2,p)}(\lambda_j)$, which can be used to estimate the asymptotic variance of the LP regression memory estimate. Alternatively, we can use a consistent estimate of the asymptotic variance based on the LP residuals, which takes into account the LP correlation across Fourier frequencies,

$$\sigma_{K,v,p}^2(k) = \lim_{T\to\infty}\operatorname{Cov}\left[\log\bar{I}_X^{(v,p)}(\lambda_j), \log\bar{I}_X^{(v,p)}(\lambda_{j+k})\right], \quad k=0,\pm K,\pm 2K,\ldots.$$

Note that in the nontapered case, $\sigma^2_{K,v,1}(k) = 0$ for $k \neq 0$, $v = 1, 2$. This correlation appears in the asymptotic variance of the LP estimates, i.e. under standard conditions,

$$2m^{1/2}\left(\hat{d}_K^{(v,p)} - d\right) \xrightarrow{d} \mathcal{N}\left(0,\ K\Omega_K^{(v,p)}\right)$$

where

$$\Omega_K^{(v,p)} = \lim_{T\to\infty} \frac{4m}{K}\left(\sum_j \Lambda_j^2\right)^{-2} \sum_j \sum_k \Lambda_j \Lambda_k \sigma^2_{K,v,p}(j-k).$$

A feasible estimate of $\Omega_K^{(v,p)}$, proposed by Arteche and Velasco (2004) along the lines of Robinson (1995a), is

$$\hat{\Omega}_K^{(v,p)} = \frac{4m}{K}\left(\sum_j \Lambda_j^2\right)^{-2} \sum_j \sum_{|k|\leq \ell} \Lambda_j \Lambda_{j+k} \hat{\sigma}^2_{K,v,p}(k),$$

where ℓ is a fixed integer such that $\ell \geq K\left[1 + (p-1)/K\right]$ when $v = 2$, and $\hat{\sigma}^2_{K,v,p}$ are the sample residual autocovariances

$$\hat{\sigma}^2_{K,v,p}(k) = \frac{K}{m}\sum_j \hat{u}^{(v,p)}_{m,j}\hat{u}^{(v,p)}_{m,j+|k|}, \quad k = 0, \pm K, \pm 2K, \ldots,$$

based on the observed residuals $\hat{u}^{(v,p)}_{m,j}$ of the LP regression. Arteche and Velasco (2004) show the consistency of such estimate for type-II tapers in a related context. For type-I tapers, $v = 1$ and the lag number ℓ should be chosen to increase with T such that $\ell^{-1} + \ell m^{-1} \to 0$ as $T \to \infty$, to account asymptotically for the correlation among all the tapered periodograms, as in usual HAC asymptotic variance estimation.

Asymptotics of tapered LW estimates are considerably simpler than those of LP estimates and, using all Fourier frequencies, $j = 1, 2, \ldots, m$, inference can be conducted according to

$$2m^{1/2}(\tilde{d}_{LW}^{(v,p)} - d) \xrightarrow{d} \mathcal{N}\left(0, \Phi^{(v,p)}\right)$$

where

$$\Phi^{(v,p)} = \lim_{T\to\infty} T\left(\sum_{t=1}^{T}\left|h_t^{(v,p)}\right|^2\right)^{-2}\sum_{t=1}^{T}\left|h_t^{(v,p)}\right|^4 \tag{33}$$

is a well known tapering inflation factor, $\Phi^{(v,p)} \geq 1$, see Velasco (1999a) for more details.

These asymptotic results on tapered semiparametric memory estimates go through for nonstationary series if enough tapering is applied, i.e. if p is large enough compared to d. In particular, for series with stationary increments, $d < 1.5$, any of the previous tapering schemes with $p > 1$ provide consistent and asymptotically normal LP and LW estimates, where the asymptotic variances are not affected by the possible nonstationarity, only by the tapering employed.

3.3 Alternative nonstationary fractional processes

There are other ways to define nonstationary long memory or fractionally nonstationary processes. Thus, it is possible to consider (e.g. Robinson and Marinucci (2001), Phillips (1999)) processes ζ_t of memory α generated by a truncated fractional filtering as

$$\zeta_t = (1-L)^{-\alpha}\left\{\eta_t 1_{t>0}(t)\right\} = \sum_{j=0}^{t-1}\psi_j(-\alpha)\eta_{t-j}, \qquad t = 1, 2, \ldots, \tag{34}$$

where $1_A(\cdot)$ is the indicator function of the set A, so all the past weakly dependent stationary innovations η_t, $t \leq 0$, are ignored. Truncation in the definition of ζ_t is necessary because the coefficients $\psi_j(-\alpha)$ are not square-summable for $\alpha \geq \frac{1}{2}$. This convention makes essential the date of the start of the observations. However, this framework can easily be generalized by allowing a warming up period where the inflow of information can begin before we actually observe the process. The filtered process ζ_t, though with finite variance for fixed t, is non-stationary for any value of $\alpha \neq 0$. However if $\alpha < 0.5$, it converges in mean square as $t \to \infty$ to the covariance stationary X_t obtained by

$$X_t = (1-L)^{-\alpha}\eta_t = \sum_{j=0}^{\infty} \psi_j(-\alpha)\,\eta_{t-j}, \quad \alpha < 0.5, \tag{35}$$

cf. (4), for the same sequence of innovations η_j, $j = 1, \ldots, t$. As $\Gamma(\alpha)\psi_j(-\alpha) \sim j^{\alpha-1}$ as $j \to \infty$, when $\alpha \geq 0.5$ the variance of ζ_t grows without limit with t and ζ_t is nonstationary long-range dependent in the sense of Heyde and Yang (1997). The long-range properties of the processes (34) and (35) are described by the memory parameter α, and under regularity conditions and appropriately normalized, such processes converge to different versions of fractional Brownian motion with parameter $\alpha > 0.5$ respectively (see Marinucci and Robinson (2000) for a discussion). This reflects the fact that alternative definitions of nonstationary fractional processes differ in the treatment of initial conditions, which are transmitted through a long range dependent process ν_t in (28), while the stationary dynamics depend on the weakly dependent process η_t in (34).

Sufficient conditions for valid large sample LP inference on α for Gaussian processes defined by (34) are investigated in Velasco (2004) using local conditions on the spectral density of η_t. Several extensions of model (34) are considered, such as series with negative memory ($\alpha < 0$), which are relevant for statistical inference on fractionally differenced data; processes with filters initialized at a remote point in the past; and fractional differencing and integration of stationary long memory time series with η_t satisfying (3) with $0 < |d| < 0.5$ (see Marinucci and Robinson (2001)). Robinson (2004) considered bounds for the difference between the DFT of both types of nonstationary processes, useful to investigate the asymptotic behavior of a large class of estimates linear in the periodogram. The consistency of the LW estimate for asymptotically stationary processes given by (34), $|\alpha| < 0.5$, is studied in Marmol and Velasco (2004) for linear η_t. Also Shimotsu and Phillips (2004) have studied the behavior of the LW estimate for series generated by (34) for the nonstationary and unit root cases, showing similar results to when the series is given by a partial sum process, cf. (28).

However, the knowledge that ζ_t is given by (34) can be used directly in the estimation of α, either through numerical properties of the DFT (similar to (29) but taking into account end effects) or by using directly the time-domain truncated fractional differencing structure of ζ_t. The first route is followed in Phillips (1999) and Kim and Phillips (2000) for the LP estimates and in Shimotsu and Phillips (2000) for the LW estimates. The second option is pursued in Shimotsu and Phillips (2005a), where the following *exact LW* log-likelihood is analyzed,

$$\mathcal{L}_m^E(a,G) = \frac{1}{m}\sum_{j=1}^{m}\left\{\log G\lambda_j^{-2a} + \frac{I_{\Delta^a\zeta}(\lambda_j)}{G}\right\},$$

where $I_{\Delta^a\zeta}$ denotes the periodogram of the series

$$\Delta^a\zeta_t = \sum_{j=0}^{t-1}\psi_j(a)\zeta_{t-j}, \quad t = 1, 2, \ldots, T.$$

The normalization of the periodogram by λ_j^{2d} used in $\mathcal{L}_m$ is replaced in $\mathcal{L}_m^E$ by the fractional differencing of the original data, allowing in principle any value of α to be considered. The ELW estimates are defined by minimization of $\mathcal{L}_m^E(a, G)$ and, as usual, concentrating out G we obtain

$$\widehat{\alpha}_m^{ELW} = \arg\min_{a \in D} R_m^E(a),$$

where

$$R_m^E(a) = \log \widehat{G}_m^{ELW}(a) - 2a\frac{1}{m}\sum_{j=1}^{m}\log\lambda_j, \qquad \widehat{G}_m^{ELW}(a) = \frac{1}{m}\sum_{j=1}^{m} I_{\Delta^a\zeta}(\lambda_j).$$

Under conditions slightly more restrictive than those of Robinson (1995b), Shimotsu and Phillips (2005a) found that $\widehat{\alpha}_m^{ELW}$ is consistent and asymptotically normal with the usual 1/4 asymptotic variance when

$$\frac{1}{m} + \frac{m^{1+2\gamma}\log^2 m}{T^{2\gamma}} + \frac{\log T}{m^{\epsilon}} \to 0, \qquad as \ \ T \to \infty,$$

for some $\epsilon > 0$, where the parameter γ is equivalent to that given in (25) but for the spectral density of η_t in (34). The interest in this procedure, which is somewhat more cumbersome than that of the usual LW, is based on the fact that nonstationary values of α can be included in D, with the only restriction being that $\nabla_2 - \nabla_1 < 9/2$, which requires limited prior information on the value of α, avoiding in this way the efficiency loss of tapering. The relationship between these variants and the traditional version of the LW estimator are discussed by Shimotsu and Phillips (2005b).

3.4 Cyclical and seasonal long memory

It is possible to conceive of stochastic processes X_t that show strong persistence at some frequency $\omega \in (0, \pi]$ different from the origin, such that their spectral density satisfies

$$f_X(\omega + \lambda) \sim G_X|\lambda|^{-2d} \qquad \text{as } \lambda \to 0. \tag{36}$$

A time series with such a spectral density displays cycles of period $2\pi/\omega$, which are more persistent the larger d is. The condition $d < \frac{1}{2}$ entails stationarity by the integrability of f_X. The autocovariances of such processes show an asymptotic slow decay typical of long memory but with oscillations that depend on the frequency ω where the spectral pole or zero occurs, so that

$$\gamma_j \sim c_X \cos(j\omega) j^{2d-1} \qquad \text{as } j \to \infty$$

(see, e.g., Chung (1996), Andel (1986), who introduced the Gegenbauer ARMA (GARMA) processes, or Gray, Zhang and Woodward (1989)). Oppenheim, Ould Haye and Viano (2000) and Lindholdt (2002) show that the seasonal long memory that has been found in many macroeconomic time series can be explained by cross-sectional aggregation and structural changes, providing ways of generating parametric seasonal long memory models. Arteche and Robinson (1999) called this property Seasonal or Cyclical Long Memory (SCLM) and investigated semiparametric inference for SCLM processes based on versions of the LP and LW estimates. When two-sided estimates are used, the asymptotic variance should be adapted since, in fact, we are using $2m$ different periodograms instead of the usual m when considering the zero frequency long memory. Arteche (2002) addresses the issue of testing for equal memory parameters when more than one seasonal frequency is considered.

Arteche and Robinson (2000) have further introduced Seasonal or Cyclical Asymmetric Long Memory (SCALM), for which

$$\begin{aligned} f_X(\omega + \lambda) &\sim G_{X1}\lambda^{-2d_1} \qquad \text{as } \lambda \to 0^+ \\ f_X(\omega - \lambda) &\sim G_{X2}\lambda^{-2d_2} \qquad \text{as } \lambda \to 0^+, \end{aligned}$$

where $\omega \in (0, \pi)$, $0 < G_{Xi} < \infty, |d_i| < \frac{1}{2}, i = 1, 2$, and it is permitted that $d_1 \neq d_2$ and/or $G_{X1} \neq G_{X2}$. This (semi)parameterization shows that the extension of the concept of long memory from $\omega = 0$ to any ω between 0 and π broadens the scope for modelling, since the spectrum is symmetric about zero and π. The spectral asymmetry involves a different persistence for cycles of period just shorter and just larger than $2\pi/\omega$. Arteche and Robinson (2000) have discussed semiparametric inference based on one-sided LP and LW estimates for both memory parameters d_1 and d_2. When d_1 and d_2 have opposite signs there is very strong leakage from the peak, which indicates strong persistence, to the zero at the other side of the singularity, affecting noticeably semiparametric inference in finite samples. To alleviate this problem, Arteche and Velasco (2004) find similar benefits of tapering as those for treating symmetric nonstationary singularities in f_X.

A related problem in some applications is the estimation of the location ω of the pole when $d > 0$. Hidalgo and Soulier (2004) employed a semiparametric model for f_X around ω,

$$f_X(\lambda) = |1 - e^{i(\lambda-\omega)}|^{-d}|1 - e^{i(\lambda+\omega)}|^{-d} f^*(\lambda),$$

to generate behavior such as (36). If f^* is smooth this model allows for poles where the exponent of the singularity is defined as $\alpha = d$ if $\omega \in (0, \pi)$ and as $\alpha = 2d$ if $\omega \in \{0, \pi\}$. The estimate of ω they propose is the maximum of the periodogram,

$$\hat{\omega}_T = \frac{2\pi}{T} \arg \max_{1 \leq j \leq \tilde{T}} I_X(\lambda_j),$$

where $\tilde{T} = [(T-1)/2]$. This estimate is consistent for Gaussian time series and its convergence rate is close to the parametric rate T obtained by Giraitis, Hidalgo and Robinson (2001), but its asymptotic distribution is unknown. Hidalgo (2001) investigates the asymptotic distribution of an alternative estimate of ω which also has a rate of convergence close to the parametric rate T, provided the process X_t has enough finite moments. Furthermore, Hidalgo and Soulier (2004) show that the LP estimate of d when we plug in the estimate $\hat{\omega}_T$ is robust to estimation of the location of the pole, with the usual asymptotic properties. This result relies on the symmetry of the peak $f_X(\lambda)$ around ω and on the use of both sides of the periodogram around ω when $\omega \in (0, \pi)$.

4 Developments

In this section we consider two fields where the semiparametric methodology of memory estimation has been widely developed and applied to solve inference problems on economic time series where parametric models are often difficult to justify. These are fractionally cointegrated systems and nonlinear models of conditional heteroskedasticity for time series with persistent volatility. In the first case, the new challenges are related to the treatment of nonstationary series of unknown degree of integration, together with the analysis of vector time series with degenerate long run dynamics in the case of cointegration. In the second problem, the nonlinearity produces difficulties in applying usual semiparametric methods, so ad-hoc modifications have been developed.

4.1 Fractional cointegration

We consider a $P \times 1$ fractionally integrated vector

$$\mathbf{Z}_t = \mu + diag\left\{(1-L)^{-d_1}, \ldots, (1-L)^{-d_P}\right\} \mathbf{u}_t 1_{t>0}(t), \quad t = 0, 1, 2, \ldots,$$

with memory parameters $\mathbf{d} = (d_1, \ldots, d_P)'$, where $\mathbf{u}_t$ is a zero mean weakly dependent vector process. The concept of (fractional) cointegration establishes that a certain (non-null) linear combination $\mathbf{b}'\mathbf{Z}_t$ has less memory than the vector $\mathbf{Z}_t$ in some sense. When we allow for different memory parameters there are many ways to make precise such a definition (see, e.g., the review in Robinson and Yajima (2002)). If we partition the original vector as $\mathbf{Z}_t' = (\mathbf{X}_t', Y_t)$, then one of the simplest possibilities is to state that $\mathbf{Z}_t$ is fractionally cointegrated if there exists an $M \times 1$ vector $\boldsymbol{\beta}$, $M = P - 1$, such that $e_t = Y_t - \boldsymbol{\beta}'\mathbf{X}_t$ is $I(\delta)$ with $\delta < d_Y$. This definition implies that $d_i = d_Y$ for at least one $i = 1, \ldots, M$, since we impose the restriction that the coefficient of Y_t in $\mathbf{b}$ is not null, and leads to the linear regression representation

$$Y_t = \boldsymbol{\beta}'\mathbf{X}_t + e_t, \tag{37}$$

where $\boldsymbol{\beta}$ could be estimated by standard methods, such as OLS.

To develop this line of argument, define a version of the AP statistic introduced in Section 2.3,

$$\mathbf{F}_{ab}(n) = 2\frac{2\pi}{T}\sum_{j=1}^{n} \mathrm{Re}\left\{\mathbf{I}_{ab}(\lambda_j)\right\} - \frac{2\pi}{T}\mathbf{I}_{ab}(\pi)1\left\{n = T/2\right\}, \quad 1 \leq n \leq T/2.$$

Note that $\mathbf{F}_{ab}([T/2])$ is equal to the usual covariance matrix between a_t and b_t, $t = 1, \ldots, T$, so $\mathbf{F}_{ab}(n)$ reflects the contribution to that covariance from frequencies up to λ_n. Omitting the zero frequency implies mean correction as usual. Robinson (1994a), for stationary series, and Robinson and Marinucci (2001), for nonstationary processes, proposed the narrow band or frequency domain least squares (FDLS) coefficients

$$\boldsymbol{\beta}_n = \mathbf{F}_{\mathbf{XX}}(n)^{-1}\mathbf{F}_{\mathbf{X}Y}(n)$$

to estimate the cointegrating vector $\boldsymbol{\beta}$ in the representation (37) under the assumption of rank one cointegration (so $\boldsymbol{\beta}$ is the only direction which reduces the memory of $\mathbf{Z}_t$). See also Robinson and Marinucci (2000), Chen and Hurvich (2003a) and Robinson and Iacone (2005) for related results in the presence of deterministic trends and Marinucci (2000) for alternative estimates using continuous periodogram averages.

When $n = [T/2]$, $\boldsymbol{\beta}_n$ is the OLS estimate with intercept, but $n < [T/2]$ may be desirable. When a nondegenerating band of frequencies is considered, $n \sim CT$, $C \in (0, 0.5)$, this corresponds to the band-spectrum regression introduced by Hannan (1963). However, when the convergence condition (6) holds, $\boldsymbol{\beta}_n$ still uses an increasing number of frequencies, but in a degenerating band around the origin. This option solves the consistency problem of OLS estimates in stationary frameworks due toe simultaneity bias, and in the nonstationary case also avoids some asymptotic bias terms and focuses on the relevant frequencies for the analysis of long run relationships. The improvements depend basically on the degree of nonstationarity of the observed series. The more interesting cases analyzed in Robinson and Marinucci (2001) are the so called 'less than unit root nonstationarity', with $d_i > 0.5$, $\delta \geq 0$ and $d_i + \delta < 1$, for which

$$T^{d_i + d_{\min} - 1}\left(\beta_{i,[T/2]} - \beta_i\right) \quad \text{and} \quad T^{d_i - \delta} n^{\delta + d_{\min} - 1}\left(\beta_{i,n} - \beta_i\right) \tag{38}$$

converge to well defined nondegenerate random variables under (6), and the 'greater than unit root nonstationarity', with $d_1 = \cdots = d_M > 0.5$ and $\delta > 0$, $d_i + \delta > 1$, when

$$T^{d_i - \delta}\left(\beta_{i,[T/2]} - \beta_i\right) \quad \text{and} \quad T^{d_i - \delta}\left(\beta_{i,n} - \beta_i\right) \tag{39}$$

both converge weakly. In the well studied case of unit root cointegration, $d_1 = \cdots = d_M = 1$ and $\delta = 0$, the rate of convergence of both estimates is also $T^{d_i - \delta} = T$. The limits are functionals of fractional Brownian motions. The standardizations in (38) show that FDLS may achieve a great superiority over OLS given (6) holds, although some benefits can also be found in the unit root case.

For these results to be useful requires, on the one hand, that there exists cointegration of rank one (up to scale, only one cointegration vector exists) and, on the other hand, that the orders of integration are known or can be estimated consistently. The existence of cointegration can be deduced from the values of the memory parameters d_i, d_y and δ, so we first concentrate on this problem.

The memory of observables $\mathbf{X}_t$ can be estimated semiparametrically by any of the methods discussed in Section 2. For the cointegrating errors e_t we could use similar ideas, but two further problems arise, namely the use of the residuals $\hat{e}_t = Y_t - \hat{\boldsymbol{\beta}}_n' \mathbf{X}_t$, and the ignorance of whether these are stationary or not. Dittmann (2000) studies the finite sample performance of several residual-based tests for fractional cointegration. The effects of the use of residuals depend fundamentally on the rates of convergence (38)-(39) of the estimates of the cointegrating vectors, which can be very fast, but also arbitrarily slow when $d_{\min}$ is close to δ, even if it is assumed from the outset that cointegration exists with stationary errors. Hassler, Marmol and Velasco (2003) and Velasco (2003a) have studied the estimation of the memory parameters of the vector $(\mathbf{X}_t', e_t)'$ with the LP and LW estimates respectively, both using (FD)LS residuals $\hat{e}_t$ or their increments, $\Delta\hat{e}_t$. LW memory estimation with (nonparametric) residuals was first studied by Robinson (1997). The main conclusion is that asymptotic semiparametric inference for δ based on cointegrating residuals is not affected by $\boldsymbol{\beta}$ estimation as far as the $\hat{\boldsymbol{\beta}}_n$ are superconsistent, i.e. $d_i - \delta > 0.5$, all i. If $d_i - \delta \leq 0.5$ for some i, the semiparametric estimates of δ may remain consistent (LP estimates seem to require further pooling or tapering), but with a slower rate. In the 'greater than unit root nonstationarity' case, original residuals or increments of the residuals have to be used depending on whether $\delta < 0.5$ or $\delta > 0.5$. However, tapering renders semiparametric inference robust to the decision of which input is used. In the 'less than unit root nonstationarity' case, only original residuals should be used since $\delta < 0.5$ necessarily and there can be additional restrictions on the range of allowed bandwidths m depending on the values of d and δ.

Using the theory reviewed in Section 2 we could test hypotheses on the values of the parameters d_i and δ, but the previous restrictions under the assumption of cointegration lead to some caution in constructing a direct test of the null of no cointegration, $\delta = d$, against $\delta < d$, assuming $d_1 = \cdots = d_M$. Alternatively, Marinucci and Robinson (2001) propose a Hausman (1978) type test based on alternative LW estimates of d when $M = 1$,

$$H_m = 8m \left(\hat{d}_m^{LW} - \tilde{d}_m^{LW} \right)^2,$$

where the univariate $\hat{d}_m^{LW}$ can be based on either ΔX_t or ΔY_t, and $\tilde{d}_m^{LW}$ is the efficient restricted estimate of the memory $d_Y = d_1$ with input $(\Delta X_t, \Delta Y_t)'$, cf. (19). Then $\tilde{d}_m^{LW}$ and $\hat{d}_m^{LW}$ have asymptotic variances of 1/8 and 1/4 respectively, so their difference is expected to have asymptotic variance $1/4 - 1/8 = 1/8$ under the null of no cointegration. Note that, under this null hypothesis, the long run variance matrix $\mathbf{G}_{\mathbf{Z}}$ of $\mathbf{Z}_t$, is non-singular, and that $d = \delta$, so the distribution of H_m can be approximated by that of a χ_1^2 variable, but under the alternative $\mathbf{G}_{\mathbf{Z}}$ is singular and $\tilde{d}_m^{LW}$ will not be consistent for d.

Velasco (2003b) considers an alternative semiparametric method of estimating the degree of cointegration $\alpha = d - \delta \geq 0$ of a vector $\mathbf{Z}_t$ that avoids the use of residuals that depend on initial slope estimates. For this it is assumed that the (pseudo) spectral density matrix $\mathbf{f_Z}$ of the bivariate vector $\mathbf{Z}_t = (X_t, e_t)'$ satisfies

$$\mathbf{f_Z}(\lambda) = \lambda^{-2d} \begin{pmatrix} \Xi_{XX} & \Xi_{Xe}\lambda^{\alpha} \\ \Xi_{eX}\lambda^{\alpha} & \Xi_{ee}\lambda^{2\alpha} \end{pmatrix} (1 + o(1)) \quad \text{as } \lambda \to 0^+, \tag{40}$$

where the matrix $\boldsymbol{\Xi} = \{\Xi_{ab}\}$, $a, b \in \{X, e\}$, is hermitian and nonsingular (see also Levy (2003)). Then, using (37) and (40), it is possible to show that the squared coherence between Y_t and X_t satisfies

$$|R_{XY}(\lambda)|^2 \sim 1 - \Xi_H \lambda^{2\alpha} \quad \text{as } \lambda \to 0^+, \tag{41}$$

for a real constant $0 < \Xi_H < \infty$,

$$\Xi_H = \frac{\Xi_e}{\Xi_X}\left[1 - \frac{|\Xi_{eX}|^2}{\Xi_e \Xi_X}\right] = \frac{G_e}{G_X}\left[1 - \frac{{G_{eX}}^2}{G_e G_X}\right],$$

that depends on the (normalized) noise to signal ratio and on the coherence at zero between X_t and e_t using the long run variance $\mathbf{G_Z}$. Rearranging and taking logs in (41) we have that

$$\log(1 - |R_{XY}(\lambda)|^2) \sim \log \Xi_H + 2\alpha \log \lambda \quad \text{as } \lambda \to 0^+, \tag{42}$$

which suggests the log-coherence regression estimate of α, analogous to GPH LP regression,

$$\widehat{\alpha}_m = -\left(\sum_{j=\ell}^{m} \Lambda_j^2\right)^{-1} \sum_{j=\ell}^{m} \Lambda_j \log\left(1 - |\widehat{R}_{XY,n}(\lambda_j)|^2\right).$$

$\widehat{\alpha}_m$ uses consistent estimates of $|R_{XY}(\lambda)|^2$ at frequencies λ_j in a degenerating band around the origin,

$$|\widehat{R}_{XY,n}(\lambda_j)|^2 = \frac{|\widehat{f}_{XY,n}(\lambda_j)|^2}{\widehat{f}_{X,n}(\lambda_j)\widehat{f}_{Y,n}(\lambda_j)},$$

where $\widehat{f}_{XY,n}$, $\widehat{f}_{X,n}$, $\widehat{f}_{Y,n}$ are nonparametric estimates of the corresponding (pseudo) spectral densities with bandwidth n (see also Hidalgo (1996)). As in Robinson (1995a), a trimming of the very first $\ell - 1$ coherence estimates is allowed. This approach is valid for both stationary and nonstationary series (tapering might be used to eliminate an intercept or polynomial trend in (37) or to cover very nonstationary situations, $d \geq 1$) and it is not affected asymptotically by the endogeneity of the residuals ($\Xi_{eX} \neq 0$). However, if X_t and e_t are incoherent at zero frequency, the semiparametric model (42) provides a better approximation.

The analysis of $\widehat{\alpha}_m$ is complicated with respect to the LP memory estimate due to the nonlinear and nonparametric nature of the sample coherences $|\widehat{R}_{XY,n}(\lambda_j)|^2$. Velasco (2003b) showed the consistency of $\widehat{\alpha}_m$ and suggested to approximate its sample variability by

$$\mathrm{Var}[\widehat{\alpha}_m] \approx \left(\sum_j \Lambda_j^2\right)^{-2} 4 \sum_j \sum_k \Lambda_j \Lambda_k \,\mathrm{Cov}\left[\tanh^{-1}(|\widehat{R}_{XY,n}(\lambda_j)|), \tanh^{-1}(|\widehat{R}_{XY,n}(\lambda_k)|)\right]. \tag{43}$$

Here the transformation $\tanh^{-1}$ is variance-stabilizing because $\widehat{R}_{XY,n}$ is a sort of correlation coefficient in the frequency domain, and, when $\widehat{R}_{XY,n}$ uses spectral estimates with uniform weights over $2q + 1$ Fourier frequencies, we can approximate the covariance in (43) by

$$\mathrm{Cov}\left[\tanh^{-1}(|\widehat{R}_{XY,n}(\lambda_j)|), \tanh^{-1}(|\widehat{R}_{XY,n}(\lambda_{j+p})|)\right] \approx \frac{2q + 1 - |p|}{2(2q + 1)^2}, \quad p = 0, \pm 1, \ldots, \pm 2q,$$

and assume that estimates of $\widehat{R}_{XY,n}$ evaluated at frequencies sufficiently far apart are asymptotically uncorrelated. For tapered series this approximation has to be adjusted by $\Phi^{(v,p)}$ as for the LW memory estimates in (33).

Robinson and Yajima (2002) have investigated semiparametric methods of inference on the cointegration rank of a stationary vector. The methods proposed depend, first, on obtaining subsets of $\mathbf{Z}_t$ with the same memory by sequential testing, using modified Wald tests based on (univariate) LW semiparametric estimates to account for the degeneracy of the asymptotic distribution in case of cointegration (because $\mathbf{G_Z}$ is singular). The cointegration rank is then determined by analyzing the eigenvalues of the estimate of $\mathbf{G_Z}$, given by $\widehat{\mathbf{G}}_{\mathbf{Z},m} = \widehat{\mathbf{G}}_{\mathbf{Z},m}(\tilde{\mathbf{d}}_m)$ defined in (17), where $\tilde{\mathbf{d}}_m$ is the vector containing the univariate LW estimates of the memory of each of the components of $\mathbf{Z}_t$. A similar procedure using ELW estimation is pursued by Nielsen and Shimotsu (2004).

Following a parallel route, Chen and Hurvich (2003b) study the properties of eigenvectors of an AP matrix of differenced, tapered observations, where the bandwidth m is fixed in asymptotics. They show that the eigenvectors corresponding to the smallest eigenvalues (as many as the cointegrating rank) lie close to the space of true cointegrating vectors with high probability. An implicit assumption is that all cointegration relationships have the same memory, so Chen and Hurvich (2004) propose to separate the space of cointegrating vectors into subspaces that might yield different memory parameters. The rate of convergence for the estimated cointegrating vectors depends only on the difference between the memory parameters in the given and adjacent subspaces, and residual-based LW estimation of the memory parameters is proposed to consistently identify the cointegrating subspaces and to test for fractional cointegration.

In a related, but nonstationary, framework, Marmol and Velasco (2004) propose a test for fractional cointegration in a $P \times 1$ nonstationary fractionally integrated (NFI) vector

$$\mathbf{Z}_t = (1-L)^{-d}\left\{\mathbf{u}_t 1_{t>0}(t)\right\}, \quad t = 0, 1, 2, \ldots,$$

where $\mathbf{u}_t = \sum_{j=-\infty}^{\infty} \mathbf{A}_{t-j}\boldsymbol{\varepsilon}_j$ is a linear process with iid innovations $\boldsymbol{\varepsilon}_t$ and long-run covariance matrix $\boldsymbol{\Omega} = \mathbf{A}(1)\mathbf{A}(1)'$, $\mathbf{A}(1) = \sum_{j=-\infty}^{\infty} \mathbf{A}_j$. With the partition $\mathbf{Z}_t' = (Y_t, \mathbf{X}_t')$, the matrix $\mathbf{A}(1)$ is parameterized as

$$\mathbf{A}(1) = \begin{pmatrix} \omega_{YY}^{1/2}\left(1-\rho^2\right)^{1/2} & \rho\overline{\boldsymbol{\omega}}_{\mathbf{X}Y}'\boldsymbol{\Omega}_{\mathbf{XX}}^{-1/2} \\ 0 & \boldsymbol{\Omega}_{\mathbf{XX}}^{1/2} \end{pmatrix}, \quad \boldsymbol{\Omega}_{\mathbf{ZZ}} = \begin{pmatrix} \omega_{YY} & \boldsymbol{\omega}_{\mathbf{X}Y}' \\ \boldsymbol{\omega}_{\mathbf{X}Y} & \boldsymbol{\Omega}_{\mathbf{XX}} \end{pmatrix}$$

where $\boldsymbol{\Omega}_{\mathbf{XX}}$ is positive definite with $\omega_{YY} > 0$, $\overline{\boldsymbol{\omega}}_{\mathbf{X}Y}'$ is an $M \times 1$ vector satisfying $\overline{\boldsymbol{\omega}}_{\mathbf{X}Y}'\boldsymbol{\Omega}_{\mathbf{XX}}^{-1}\overline{\boldsymbol{\omega}}_{\mathbf{X}Y} = \omega_{YY}$, and $\rho^2 = \boldsymbol{\omega}_{\mathbf{X}Y}'\boldsymbol{\Omega}_{\mathbf{XX}}^{-1}\boldsymbol{\omega}_{\mathbf{X}Y}/\omega_{YY}$ is the squared coefficient of multiple correlation computed from $\boldsymbol{\Omega}_{\mathbf{ZZ}}$, so that $0 \leq \rho^2 \leq 1$. The long-run covariance $\boldsymbol{\omega}_{\mathbf{X}Y}$ is given by $\rho\overline{\boldsymbol{\omega}}_{\mathbf{X}Y}$, where $\overline{\boldsymbol{\omega}}_{\mathbf{X}Y}$ expresses the direction of the covariance, while $\boldsymbol{\beta}_0 = \boldsymbol{\Omega}_{\mathbf{XX}}^{-1}\boldsymbol{\omega}_{\mathbf{X}Y}$ is the projection vector of Y_t on $\mathbf{X}_t$. The parameter ρ measures the strength of the covariance and the type of long run relationship among the elements of the nonstationary $\mathbf{Z}_t$. When $\rho^2 < 1$, $\boldsymbol{\Omega}_{\mathbf{ZZ}}$ is nonsingular and we say that $\mathbf{Z}_t$ is spuriously related. This model is completed when $\rho^2 = 1$, so that $\boldsymbol{\Omega}_{\mathbf{ZZ}}$ is singular and the model is disturbed to produce a (fractionally) cointegrated vector $\mathbf{Z}_t$ with $\boldsymbol{\beta}_0'\mathbf{Z}_t$ of memory $\delta \in [d-1, d)$.

As is well known, in the spurious case usual OLS statistics of a regression of Y_t on $\mathbf{X}_t$ may lead to the conclusion that there is a meaningless linear relationship between the elements of $\mathbf{Z}_t$. This result is in part a consequence of standardization by the residual sample variance, which ignores any serial correlation (or nonstationarity) in the residual series. A first step towards a feasible cointegration test is an alternative studentization of the OLS coefficients that uses all frequencies by means of the matrix

$$\widehat{\mathbf{V}}_T = \left(\sum_{j=-\tilde{T}}^{\tilde{T}} \mathbf{I}_{\mathbf{X}}(\lambda_j) \right)^{-1} \sum_{j=-\tilde{T}}^{\tilde{T}} \mathbf{I}_{\mathbf{X}}(\lambda_j) I_{\widehat{e}}(\lambda_j) \left(\sum_{j=-\tilde{T}}^{\tilde{T}} \mathbf{I}_{\mathbf{X}}(\lambda_j) \right)^{-1},$$

where $I_{\widehat{e}}(\lambda_j)$ stands for the residual periodogram computed with the observed residuals $\widehat{e}_t = Y_t - \widehat{\boldsymbol{\beta}}_T' \mathbf{X}_t$, $\widehat{\boldsymbol{\beta}}_T$ is the OLS coefficient in (37) and $\tilde{T} = [T/2]$. The test statistic proposed by Marmol and Velasco (2004) is given by the following Wald or adjusted F statistic

$$\mathcal{W}_T = \mathcal{W}_T(\widehat{\boldsymbol{\beta}}_T, \widehat{\boldsymbol{\beta}}_{0,n}) = \frac{1}{M} \left(\widehat{\boldsymbol{\beta}}_T - \widehat{\boldsymbol{\beta}}_{0,n} \right)' \widehat{\mathbf{V}}_T^{-1} \left(\widehat{\boldsymbol{\beta}}_T - \widehat{\boldsymbol{\beta}}_{0,n} \right),$$

where the OLS estimate $\widehat{\boldsymbol{\beta}}_T$ is inconsistent under no cointegration and $\widehat{\beta}_{0,n}$ is an alternative semiparametric GLS-type estimate, which is consistent under this hypothesis,

$$\widehat{\boldsymbol{\beta}}_{0,n} = \widehat{\boldsymbol{\beta}}_{0,n}(\widehat{d}_m, \widehat{\delta}_m) = \widehat{\boldsymbol{\Omega}}_{\mathbf{XX},n}^{-1}(\widehat{d}_m) \widehat{\boldsymbol{\omega}}_{\mathbf{X}Y,n}(\widehat{\delta}_m). \tag{44}$$

Here $\widehat{\boldsymbol{\Omega}}_{\mathbf{XX},n}$ is similar to $\widehat{\mathbf{G}}_{\mathbf{X},m}$ in (17) up to a constant, but using a common d and the periodogram of the increments of $\mathbf{X}_t$,

$$\widehat{\boldsymbol{\Omega}}_{\mathbf{XX},n}(d) = \frac{2\pi}{n} \sum_{j=1}^{n} \lambda_j^{2(d-1)} \operatorname{Re}\{\mathbf{I}_{\Delta\mathbf{X}}(\lambda_j)\},$$

in the same way that

$$\widehat{\boldsymbol{\omega}}_{\mathbf{X}Y,n}(\delta) = \frac{2\pi}{n} \sum_{j=1}^{n} \lambda_j^{2(\delta-1)} \operatorname{Re}\{\mathbf{I}_{\Delta\mathbf{X}\Delta Y}(\lambda_j)\}$$

uses the cross periodogram of the increments $\Delta\mathbf{X}_t$ and ΔY_t. In (44), $\widehat{d}_m$ is a $\log T$-consistent semiparametric estimate of d, as given in Sections 2.1-2.2, based on any subset of $\Delta\mathbf{X}_t$, but $\widehat{\delta}_m$ is a consistent estimate of δ based on OLS residuals. By contrast with the customary F-statistic, constructed using the usual (time-domain) residual sum of squares, the Wald statistic $\mathcal{W}_T$ has a well defined limiting distribution under the null of a spurious relationship.

Under the null of no cointegration $\delta = d$, both semiparametric memory estimates in $\widehat{\boldsymbol{\beta}}_{0,n}$ have the same probabilistic limit and the periodograms in $\widehat{\boldsymbol{\Omega}}_{\mathbf{XX},n}(\widehat{d}_m)$ and $\widehat{\boldsymbol{\omega}}_{\mathbf{X}Y,n}(\widehat{\delta}_m)$ are (asymptotically) properly normalized, so $\widehat{\boldsymbol{\beta}}_{0,n}$ is consistent for $\boldsymbol{\beta}_0$ if $\left\{q^{d-2} + q^{\epsilon-1}\log T\right\}\log^2 T + qT^{-1} \to 0$, for $q = n, m$ and some $\epsilon > 0$, together with the usual regularity conditions on the spectral density of $\mathbf{u}_t$. However, under the alternative of fractional cointegration, $\delta < d$, $\widehat{\boldsymbol{\omega}}_{\mathbf{X}Y,n}(\widehat{\delta}_m)$ does not have the adequate normalization, and it can be shown to diverge as $T, n \to \infty$, whereas $\widehat{\boldsymbol{\Omega}}_{\mathbf{XX},n}(\widehat{d}_m)$ remains consistent for $\boldsymbol{\Omega}_{\mathbf{XX}}$. Therefore, the Wald statistic diverges with T when $0 < d - \delta < 0.5$, leading to the consistency of the test that rejects the null of no cointegration for large values of $\mathcal{W}_T$.

4.2 Nonlinear models

Many economic time series display conditional heteroskedasticity, this being the main feature of the dynamics of many asset prices, whose levels are assumed generally to form a martingale sequence. Robinson and Henry (1999) and Henry (2001) illustrate the robustness of LW and AP estimation of the memory of the levels in the presence of conditional heteroskedasticity. Recent interest has been focused on the estimation of the degree of persistency of volatility itself through a long memory parameter that describes the slowly decaying autocorrelation of nonlinear transformations of the returns of the corresponding asset. The availability of long records of high-frequency returns of many financial assets calls for the intensive use of the semiparametric methodology in the investigation of the long range properties of these time series.

Robinson (1991) proposed that the conditional volatility $\sigma_t^2 = \mathrm{Var}[X_t|I_{t-1}]$ series, where I_s is the σ-field of events generated by X_k, $k \leq s$, may display long range dependence in an ARCH(∞) specification,

$$\sigma_t^2 = \sigma^2 + \sum_{j=1}^{\infty} \theta_j X_{t-j}^2,$$

where θ_j decay slowly as the weights $\psi_j(d)$ in (4) for $d > 0$, and propose LM testing of this possibility. This has also been an issue in applied work, see e.g. Ding, Granger and Engle (1993).

Considerable effort has been put into studying parametric generalized autoregressive conditional heteroskedasticity (GARCH) specifications which actually produce long range dependence in σ_t^2 and valid inference procedures (see e.g. the fractionally integrated GARCH (FIGARCH) model of Baillie, Bollerslev and Mikkelsen (1996), the fractionally integrated exponential GARCH (FIEGARCH) of Bollerslev and Mikkelsen (1996), or Giraitis, Robinson and Surgailis (2000)), including also semiparametric proposals (Giraitis, Kokoszka, Leipus and Teyssière (2000)). However, stochastic volatility (SV) specifications have been more amenable for semiparametric analysis. Harvey (1998) and Breidt, Crato and de Lima (1998) studied a Long Memory SV (LMSV) model for asset returns defined by

$$X_t = \sigma_t \xi_t, \quad \sigma_t = \sigma \exp\left(v_t/2\right),$$

where v_t is a stationary long memory process independent of ξ_t, which is itself iid with zero mean and unit variance. The persistence in the volatility of X_t depends on the persistence of v_t. Breidt et al. (1998) proposed its estimation by a global Whittle estimate, using the linearization

$$\begin{aligned} \log X_t^2 &= \log \sigma_t^2 + \log \xi_t^2 \\ &= \log \sigma^2 + E\left[\log \xi_t^2\right] + v_t + \left\{\log \xi_t^2 - E\left[\log \xi_t^2\right]\right\} \\ &= \mu + v_t + u_t, \end{aligned} \tag{45}$$

say, where u_t is a zero mean iid random sequence and independent of v_t, whose spectral density depends on some parameters. Note that the autocovariances of $\log X_t^2$ are the same as those of v_t except at lag zero, for which it is $\sigma_v^2 + \sigma_u^2$. A justification of such procedures can be found in Hosoya (1997).

However, semiparametric methods are also natural in this context if we assume that f_v satisfies (3), especially given the difficulty of properly specifying all short run dynamics and the availability of long data sets at different sampling frequencies. Breidt et al. (1998) and Andersen and Bollerslev (1997) propose LP estimation on some nonlinear transformation of X_t, such as $\log X_t^2$ or $|X_t|$, but this violates the usual Gaussianity assumption. In the case of a LMSV, note that if v_t follows a fractional model with spectral density $|2\sin\lambda/2|^{-2d} g_v^*(\lambda)$, then $f_{\log X^2}(\lambda) = |2\sin\lambda/2|^{-2d} f^*(\lambda)$, where now

$$f^*(\lambda) = g_v^*(\lambda) + |2\sin\lambda/2|^{2d} \frac{\sigma_u^2}{2\pi} = g_v^*(0)\left\{1 + O(\lambda^{2d})\right\} \quad \text{as } \lambda \to 0+, \tag{46}$$

for smooth g^*. This justifies the use of customary semiparametric models since f^* is bounded above and away from zero (if $g_v^*(\lambda)$ is bounded for all λ and positive at $\lambda = 0$) and $f_{\log X^2}(\lambda)/f_v(\lambda) \to 1$ as $\lambda \to 0$. Deo and Hurvich (2001) show that the central limit theorem (8) for the LP estimate holds for Gaussian v_t when we replace I_X by $I_{\log X^2}$, and m is chosen to satisfy

$$\frac{\log^2 T}{m} + \frac{m^{4d+1}\log^2 m}{T^{4d}} \to 0 \ \text{ as } \quad T \to \infty, \tag{47}$$

with f^* twice differentiable. This condition corresponds to that of Robinson (1995a, Assumption 6) when $\gamma = 2d$ in (25), cf. (46). Note that this result implies that $d > 0$ (and $\gamma > 0$), so long memory in v_t is assumed. Hurvich and Soulier (2002) have extended the previous result to the case $d = 0$ for volatility persistence testing, whereas Arteche (2004) gives a similar analysis for the LW estimate leading to (11) under the usual conditions and (46)-(47).

The additive structure of f^* in (46) suggests a bias problem in the selection of the bandwidth m, much restricted when d is small. To control this problem, Sun and Phillips (2003), in the spirit of the bias reduction techniques of Section 2.4, propose enlarging the LP regression with a term in λ^{2d}, cf. (21), thus leading to the so called nonlinear LP (NLP) regression estimate, which now has no explicit expression. It is shown that the NLP estimate is consistent under (6), allowing for $\sigma_u^2 = 0$, but $d > 0$. If further

$$\frac{T^{4d(1+\epsilon)}}{m^{4d(1+\epsilon)+1}} + \frac{m^{8d+1}}{T^{8d}} \to 0 \quad \text{as} \quad T \to \infty,$$

for some $\epsilon > 0$, which allows for much larger choices of m than (47), and so faster converging estimates, then

$$2m^{1/2}\left(\hat{d}_m^{NLP} - d\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\pi^2}{6}\frac{(2d+1)^2}{4d^2}\right).$$

This limit, by contrast, reflects the increase in asymptotic variance due to the use of additional (nonlinear) regressors.

Hurvich and Ray (2003) exploit the same idea for the PLW estimate, introducing the term in λ^{2d} in (27), with $\exp\left(-p_r(\lambda_j;\boldsymbol{\theta})\right)$ replaced by $1 + \theta\lambda_j^{2d}$, and consider possibly nonstationary time series. Denoting this estimate as $\hat{d}_m^{NLW}$, Hurvich and Ray show that

$$2m^{1/2}\left(\hat{d}_m^{NLW} - d\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{(2d+1)^2}{4d^2}\right),$$

for $d \in (0, 0.75)$ if

$$\frac{T^{4d}}{m^{4d+1}} + \frac{m^{2\gamma+1}\log^2 m}{T^{2\gamma}} \to 0 \quad \text{as} \quad T \to \infty, \tag{48}$$

under (25), for linear v_t and $\gamma > 2d$. Note that typically $\gamma = 2$ for regular cases, cf. (25).

Building on this series of works, Hurvich, Moulines and Soulier (2005) consider a semiparametric specification for the spectral density of $\log X_t^2$ that nests both the LMSV and the FIEGARCH models, allowing for possible correlation between the signal and noise processes in (45) by means of the augmented correction factor

$$1 + \theta_1\lambda_j^{2d}\,\mathrm{Re}\left(\left(1 - e^{i\lambda_j}\right)^{-d}\right) + \theta_2\lambda_j^{2d}, \tag{49}$$

which replaces $\exp\left(-p_r(\lambda_j;\boldsymbol{\theta})\right)$ in the nonlinear PLW criterion (27). In this way they nest the usual LW and the NLW estimate of Hurvich and Ray (2003) setting $\theta_1 = \theta_2 = 0$ or $\theta_1 = 0$, respectively. The NLW estimate defined using the correcting factor (49), $\hat{d}_m^{N2LW}$ say, recovers basically the optimal semiparametric rate of convergence implied by (48), and its additional bias control properties have the counterpart of an increased asymptotic variance, since

$$2m^{1/2}\left(\hat{d}_m^{N2LW} - d\right) \xrightarrow{d} \mathcal{N}\left(0, (d+1)^2\frac{(2d+1)^2}{4d^2}\right),$$

for $d \in (0, 0.75)$ if, additionally to (48), $T^{4d}m^{\epsilon-4d-1} \to 0$ for some $\epsilon > 0$.

Apart from the problems of bias and bandwidth choice, other difficulties arise in semiparametric estimation of the persistence of financial time series. These include the choice of volatility measures and the role of aggregation (Bollerslev and Wright (2000)), the treatment of smooth trends and cointegration (Lobato and Velasco (2000), Christensen and Nielsen (2002)), or seasonality and efficient estimation, see e.g. Deo, Hurvich and Lu (2005). In particular Deo et al. (2005) investigate the choice of power transformations to make the distribution of $\log X_t^2$ closer to Gaussian to enhance the properties of a Whittle estimate of a LMSV model, noting that this procedure might affect the persistence of the volatility series (Dittmann and Granger (2002)).

4.3 Other areas of application

Semiparametric inference on persistence properties of time series is applied in many other fields of economic empirical analysis. Apart from descriptive and exploratory analysis, semiparametric estimation and testing of the degree of integration are key in the modelling of many macroeconomic series, specially in the presence of complex cyclical, seasonal or short run dynamics. These applies to series of output (Diebold and Rudebush (1989), Michelacci and Zaffaroni (2000)), consumption (Diebold and Rudebush (1991)), exchange rates (Cheung (1993)) and inflation (Hassler and Wolters (1995)). Following the application of a modified R/S analysis by Lo (1991), frequency and time domain semiparametric methods have been also used to document long memory in stock prices (Lee and Robinson (1996), Lobato and Savin (1997)) and the relationship of volatility with other time series, such as traded volume (Bollerslev and Jubinski (1999)).

Semiparametric estimates, despite their inefficiency, can also be used in optimization routines or in plug-in methods which do not require a fast converging, but a robust, initial estimate of the long run memory parameter. This is important in (fractional) cointegration analysis (see e.g. Robinson and Hualde (2003) or Marmol and Velasco (2004)). In this line, a main field of application of semiparametric methods is serving in the studentization of other parameter estimates, possibly of parametric nature, or in testing problems, as pursued in a general setting by Robinson (2005). A related problem is the design of efficient semiparametric estimates of regression coefficients in the presence of long memory time series as in Hidalgo and Robinson (2002) or Hualde and Robinson (2004).

5 Conclusion

There is a growing menu of semiparametric methods offered to the practitioner to analyze long memory properties of economic time series. Despite initial analyses have focused on LP estimation, mainly because of its computational appeal and the availability of approximate inference rules, LW methodology has arisen as more efficient, flexible and robust to the presence of non-Gaussian characteristics or changing conditional higher moments. However, the final performance of the semiparametric methodology depends dramatically on the bandwidth choice, specially when nonstationarity, trending of cyclical behaviours may affect the dynamics of the series under investigation. In these cases, it is a recommendable policy the use of an appropriate modification of those suggested to robustify semiparametric memory estimation. Tapering provides a simple solution, but due to the loss of efficiency implied, it might be only appropriate if long enough records are available. In the presence of substantial ignorance on the degree of integration, ELW methods can provide more efficient solutions, but these might be more sensitive to the presence of unknown mean or trends (Shimotsu, 2004). Volatility analysis based on nonlinear transformations of returns should account for the bias problem that otherwise may affect severely semiparametric inference for a wide range of bandwidths. In all cases, automatic bandwidth choices must be confronted with knowledge about cyclical and seasonal patterns which restrict in applications the validity of the basic long memory semiparametric model.

As in many other inference problems, semiparametric methods in time series analysis are of general application and apparently require limited degree of previous knowledge or study. However some care must be taken when employing these methods blindly. Following some justifications for the presence of long memory on observed time series by aggregation mechanisms of different nature, possibly involving heavy tails innovations (see the review in Diebold and Inoue (2001)), several simple models which are able to reproduce some long range dependence properties have been investigated. Many of the models developed are not properly long memory, as defined in the Introduction, but with an appropriate choice of key parameters can generate long memory features in finite samples, as described for example by the convergence rate of partial sums or correlograms (see e.g. Granger and Terasvirta (1999)). GPH's LP regression estimate is one of the benchmarks used by Gourieroux and Jasiak (2001), Diebold and Inoue (2001) and Granger and Hyung (2004) to evaluate different models, including stochastic permanent break, regime switching and occasional structural break models. It turned out that this semiparametric estimate is highly biased for the estimation and testing of the true degree of integration of the process, issuing a serious warning on the possibility that routine application of these methods lead to the finding of spurious long memory if data present some of these features. Some remedies can consist on previous application of structural break tests robust to long memory (see the revision in Banerjee and Urga (2005)) or allowance for possible breaks in memory estimation (e.g., Bos, Franses and Ooms (1999) and Choi and Zivot (2005)).

Despite these potential drawbacks, which may affect even more seriously the specification and estimation of parametric models, semiparametric inference for long memory processes has an increasing scope for the analysis of economic time series. Future developments can be expected in the derivation of (semi)automatic methods of inference, procedures for the study of multivariate possibly nonstationary and cointegrated time series, and specific techniques for the analysis of nonlinear and financial time series.

6 References

Adentstedt, R. (1974). On large sample estimation for the mean of a stationary random sequence. *Annals of Statistics* 2, 1095-1107.

Andel, J. (1986). Long-memory time series models. *Kybernetika* 22, 105-123.

Andersen, T. and T. Bollerslev (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance* 4, 115-158.

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

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

Arteche, J. (2002). Semiparametric robust tests on seasonal or cyclical long memory time series. *Journal of Time Series Analysis* 23, 251-286.

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. and P.M. Robinson (1999). Seasonal and cyclic long memory, *Asymptotics, Nonparametrics and Time Series: A Tribute to Madan Lal Puri* (S. Ghosh, ed.), Marcel Dekker, 115-145.

Arteche, J. and P.M. Robinson (2000). Semiparametric inference in seasonal and cyclical long memory processes. *Journal of Time Series Analysis* 21, 1-25.

Arteche, J. and C. Velasco (2004). Trimming and tapering semiparametric estimates in asymmetric long memory time series. *Journal of Time Series Analysis*, forthcoming.

Baillie, R.T., T. Bollerslev and H.O. Mikkelsen (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 74, 3-30.

Banerjee, A. and G. Urga (2005). Modelling structural breaks, long memory and stock market volatility: an overview. *Journal of Econometrics*, forthcoming.

Bardet, J.-M., G. Lang, E. Moulines and P. Soulier (2000). Wavelet estimator of long-range dependent processes. *Statistical Inference for Stochastic Processes* 3, 85-99.

Bhansali, R.J. and P.S. Kokoszka (2001). Estimation of the long-memory parameter: a review of recent developments and an extension. In (I.V. Basawa, C.C. Heyde and R.L. Taylor, eds.) *Proceedings of the Symposium on Inference for Stochastic Processes*, IMS Lecture Notes, 125-150.

Bollerslev, T. and P.D. Jubinski (1999). Equity trading volume and volatility: latent information arrivals and common long-run dependencies. *Journal of Business and Economic Statistics* 17, 9-21.

Bollerslev, T. and H.O. Mikkelsen (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics* 73, 151-184.

Bollerslev, T. and J.W. Wright (2000). Semiparametric estimation of long-memory volatility dependencies: the role of high-frequency data. *Journal of Econometrics* 98, 81-106.

Bos, C., S.P.H. Franses and M. Ooms (1999). Long-memory and level shifts: re-analyzing in inflation rates. *Empirical Economics* 24, 427-449.

Breidt, F.J., N. Crato and P. de Lima (1998). The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics* 83, 32-348.

Chen, W.W. and C.M. Hurvich (2003a). Estimating fractional cointegration in the presence of polynomial trends. *Journal of Econometrics* 117, 95-121.

Chen, W.W. and C.M. Hurvich (2003b). Semiparametric estimation of multivariate fractional cointegration. *Journal of the American Statistical Association* 463, 629-642.

Chen, W.W. and C.M. Hurvich (2004). Semiparametric estimation of fractional cointegrating subspaces. Preprint, Texas A&M University.

Cheung, Y.-W. (1993). Long memory in foreign exchange rates. *Journal of Business and Economic Statistics* 11, 93-101.

Choi, K. and E. Zivot (2005). Long Memory and structural breaks in the forward discount: An empirical investigation. Preprint Ohio University.

Chung, C.F. (1996). Estimating a generalized long memory process. *Journal of Econometrics* 73, 237-259.

Christensen, B.J. and M.O. Nielsen (2002). Semiparametric analysis of stationary fractional cointegration and the implied-realized volatility relation. University of Aarhus Working Paper No. 2001-4.

Delgado, M.A. and P.M. Robinson (1996a). Optimal spectral bandwidth for long memory. *Statistica Sinica* 6, 97-112.

Delgado, M.A. and P.M. Robinson (1996b). Optimal spectral kernel for long-range dependent time series. *Statistics and Probability Letters* 30, 37-43.

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

Deo, R.S., C.M. Hurvich and Y. Lu (2005). Forecasting realized volatility using a long-memory stochastic volatility model: estimation, prediction and seasonal adjustment. *Journal of Econometrics,* forthcoming.

Diebold, F.X. and A. Inoue (2001). Long memory and regime switching. *Journal of Econometrics* 105, 131-159.

Diebold, F.X. and G. Rudebush (1989). Long memory and persistence in aggregate output. *Journal of Monetary Economics* 24, 189-209.

Diebold, F.X. and G. Rudebush (1991). Is consumption too smooth: long memory and the Deaton paradox. *Review of Economics and Statistics* 73, 1-9.

Ding, Z., C.W.J. Granger and R. Engle (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83-106.

Dittmann, I. (2000). Residual-based tests for fractional cointegration:a Monte Carlo study. *Journal of Time Series Analysis* 21, 614-647.

Dittmann, I. and C.W.J. Granger (2002). Properties of nonlinear transformations of fractionally integrated processes. *Journal of Econometrics* 110, 113-133

Engle, R. and C.W.J. Granger (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica* 55, 251-277.

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

Giraitis, L., J. Hidalgo and P.M. Robinson (2001). Gaussian estimation of parametric spectral density with unknown pole. *Annals of Statistics* 29, 98-1023.

Giraitis, L., P. Kokoszka, R. Leipus and G. Teyssière (2000). Semiparametric estimation of the intensity of long memory in conditional heteroskedasticity. *Statistical Inference for Stochastic Processes* 3, 113-128.

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

Giraitis, L., P.M. Robinson and A. Samarov (2000). Adaptive semiparametric estimation of the memory parameter. *Journal of Multivariate Analysis* 72, 183-207.

Giraitis, L., P.M. Robinson and D. Surgailis (2000). A model for long memory conditional heteroscedasticity. *The Annals of Applied Probability* 10, 1002-1024.

Gourieroux, C. and J. Jasiak (2001). Memory and infrequent breaks. *Economics Letters* 70, 29-41.

Granger, C.W.J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* 16, 121-130.

Granger, C.W.J. and N. Hyung (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance* 11, 399-421.

Granger, C.W.J. and R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15-29.

Granger, C.W.J. and T. Terasvirta (1999). A simple nonlinear time series model with misleading linear properties. *Economics Letters* 62, 161-165.

Gray, H.L., N.F. Zhang and W.A. Woodward (1989). On generalized fractional processes. *Journal of Time Series Analysis* 10, 233-257.

Hannan, E.J. (1963). Regression for time series with errors of measurement. *Biometrika* 50, 293-302.

Hall, P., H.L. Koul and B.A. Turlach (1997). Note on convergence rates of semiparametric estimators of dependence index. *Annals of Statistics* 25, 1725-1739.

Harvey, A.C. (1998). Long memory in stochastic volatility. In (J. Knight and S. Satchell, eds.) *Forecasting Volatility in the Financial Markets*, pp. 307-320, Butterworth-Heinemann, London.

Hassler, U., F. Marmol and C. Velasco (2003). Residual log-periodogram inference for long run relationships. *Journal of Econometrics*, forthcoming.

Hassler, U. and J. Wolters (1995). Long memory in inflation rates: international evidence. *Journal of Business and Economic Statistics* 13, 37-45.

Hausman, J. (1978). Misspecification tests in econometrics. *Econometrica* 46, 1251-1271.

Henry, M. (2001). Robust automatic bandwidth for long memory. *Journal of Time Series Analysis* 22, 293-316.

Henry, M. and P.M. Robinson, (1996). Bandwidth choice in Gaussian semiparametric estimation of long range dependence. In (P.M. Robinson and M. Rosenblatt, eds.), *Athens Conference on*

Applied Probability and Time Series in memory of E. J. Hannan: Volume II. New York: Springer-Verlag.

Henry, M. and P. Zaffaroni (2003). The long range dependence paradigm for macroeconomics and finance. In (P. Doukhan, G. Oppenheim and M.S. Taqqu, eds.), *Theory and Applications of Long-Range Dependence*, Birkhäuser. Boston.

Heyde, C.C. and Y. Yang. (1997). On defining long-range dependence. *Journal of Applied Probability* 34, 939-944.

Hidalgo, F.J. (1996). Spectral analysis for bivariate long-memory time series. *Econometric Theory* 12, 773-792.

Hidalgo, F.J. (2001). Semiparametric estimation when the location of the pole is unknown. Preprint.

Hidalgo, F.J. and P.M. Robinson (2002). Adapting to unknown disturbance autocorrelation with long memory. *Econometrica* 70, 1545-1581.

Hidalgo, J. and P. Soulier (2004). Estimation of the location and exponent of the spectral singularity of a long memory process. *Journal of Time Series Analysis* 25, 55-81.

Hosking, J.R.M. (1981). Fractional differencing. *Biometrika* 68, 165-176.

Hosking, J.R.M. (1996). Asymptotic distributions of the sample mean, autocovariances and autocorrelations of long-memory time series. *Journal of Econometrics* 73, 261-284.

Hosoya, Y. (1997). A limit theory for long-range dependence and statistical inference on related models. *Annals of Statistics* 25, 105-137.

Hualde, J. and P.M. Robinson (2004). Semiparametric estimation of fractional cointegration. Preprint, Universidad de Navarra.

Hurst, H. (1951). Long term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* 116, 770-799.

Hurvich, C.M. (2001). Model selection for broadband semiparametric estimation of long memory in time Series. *Journal of Time Series Analysis* 22, 679-709.

Hurvich, C.M. and K.I. Beltrao (1994). Automatic semiparametric estimation of the memory parameter of a long memory time series. *Journal of Time Series Analysis* 15, 285-302.

Hurvich, C.M. and 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. and W. Chen (2000). An efficient taper for potentially overdifferenced long-memory time series. *Journal of Time Series Analysis* 21, 155-180.

Hurvich, C.M. and R.S. Deo (1999). Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long memory time series. *Journal of Time Series Analysis* 20, 331-341.

Hurvich, C.M. and B.K. Ray (1995). Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *Journal of Time Series Analysis* 16, 17-41.

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

Hurvich, C.M. and P. Soulier (2002). Testing for long memory in volatility. *Econometric Theory* 18, 1291-1308.

Hurvich, C.M., R.S. Deo and J. Brodsky (1998). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. *Journal of Time Series Analysis* 19, 19-46.

Hurvich, C.M., E. Moulines and P. Soulier (2002). The FEXP estimator for potentially non-stationary linear time series. *Stochastic Processes and Their Applications* 97, 307-340.

Hurvich, C.M., E. Moulines and P. Soulier (2005). Estimating long memory in volatility. *Econometrica*, forthcoming.

Iouditsky, A., E. Moulines and P. Soulier (2002). Adaptive estimation of the fractional differencing coefficient. *Bernoulli* 7, 699-731.

Jensen, M.J. (1999). Using wavelets to obtain a consisten ordinary least squares estimator of the long-memory parameter. *Journal of Forecasting* 18, 17-32.

Kim, C. and P.C.B. Phillips (2000). Modified log periodogram regression. Preprint Department of Economics, Yale University.

Künsch, H.R. (1987). Statistical Aspects of Self-similar Processes, in *Proceedings of the First World Congress of the Bernoulli Society* 1, 67-74, ed. by Yu. Prohorov and V.V. Sazanov. Utrecht: VNU Science Press.

Lee, D. and P.M. Robinson (1996). Semiparametric exploration of long memory in stock prices. *Journal of Statistical Planning and Inference*, 50, 155-174.

Levy, D. (2003). Cointegration in the frequency domain. *Journal of Time Series Analysis* 23, 333-339.

Lindholdt, P.M. (2002). Sources of seasonal fractional integration in macroeconomic time series. CAF Working Paper 125.

Lo, A. (1991). Long term memory in stock market prices. *Econometrica*, 59, 1279-1313.

Lobato, I.N. (1997). Consistency of the averaged cross-periodogram in long memory series. *Journal of Time Series Analysis* 18, 137-155.

Lobato, I.N. (1999). A semiparametric two-step estimator for a multivariate long memory model. *Journal of Econometrics* 90, 129-153.

Lobato, I.N. and P.M. Robinson (1996). Averaged periodogram estimation of long memory. *Journal of Econometrics* 73, 303-324.

Lobato, I.N. and P.M. Robinson (1998). A nonparametric test for I(0). *Review of Economic Studies* 65, 475-495.

Lobato, I.N. and N.E. Savin (1997). Real and spurious long memory in stock market data. *Journal of Business and Economic Statistics*, 16, 261-268.

Lobato, I.N. and C. Velasco (2000). Long memory in stock market trading volume. *Journal of Business and Economic Statistics* 18, 410-427.

Mandelbrot, B.B. and J.W. Van Ness (1968). Fractional Brownian motion, fractional noises and applications. *SIAM Review* 10, 422-437.

Mandelbrot, B.B. and J.R. Wallis (1968). Noah, Joseph and operational hydrology, *Water*

Resources Research 4, 909-918.

Marinucci, D. (2000). Spectral regression for cointegrated time series with long memory innovations. *Journal of Time Series Analysis* 21, 685-705.

Marinucci, D. and P.M. Robinson (2000). Weak convergence of multivariate fractional processes. *Stochastic Processes and their Applications* 86, 103-120.

Marinucci, D. and P.M. Robinson (2001). Semiparametric fractional cointegration analysis. *Journal of Econometrics* 105, 225-247.

Marmol, F. and C. Velasco (2004). Consistent Testing of Cointegration Relationships. *Econometrica* 72, 1809-1844.

Michelacci, C. and P. Zaffaroni (2000). (Fractional) Beta Convergence. *Journal of Monetary Economics* 45, 129-153.

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

Nielsen, M.Ø. and K. Shimotsu (2004). Determining the cointegrating rank in nonstationary fractional systems by the exact local Whittle approach. Preprint, Cornell University.

Oppenheim, G., M. Ould Haye and M.C. Viano (2000). Long memory with seasonal effects. *Statistical Inference for Stochastic Processes* 3, 53-68.

Phillips, P.C.B. (1999). Discrete Fourier transforms of fractional processes. Cowles Foundation Discussion Paper # 1243, Yale University.

Robinson, P.M. (1991). Testing for strong serial-correlation and dynamic conditional heteroskedasticity in multiple-regression. *Journal of Econometrics* 47, 67-84.

Robinson, P.M. (1994a). Semiparametric analysis of long-memory time series. *Annals of Statistics* 22, 515-539.

Robinson, P.M. (1994b). Rates of convergence and optimal spectral bandwidth for long range dependence. *Probability Theory and Related Fields* 99, 443-473.

Robinson, P.M. (1994c). Time series with strong dependence. In *Advances in Econometrics. Sixth World Congress* (C.A. Sims, ed.) 1, 47-96, Cambridge University Press.

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. *Annal Statistics* 23, 1630-1661.

Robinson, P.M. (1997). Large sample inference for nonparametric regression with dependent errors. *Annals of Statistics* 25, 2054-2083.

Robinson, P.M. (2004). The distance between rival nonstationary fractional processes. *Journal of Econometrics*, forthcoming.

Robinson, P.M. (2005). Robust covariance matrix estimation: 'HAC' estimates with long memory/antipersistence correction. *Econometric Theory*, forthcoming.

Robinson, P.M. and M. Henry (1999). Long and short memory conditional heteroscedasticity in estimating the memory parameter of levels. *Econometric Theory* 15, 299-336.

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

Robinson, P.M. and J. Hualde (2003). Cointegration in fractional systems with unknown integration orders. *Econometrica* 71, 1727-1766.

Robinson, P.M. and F. Iacone (2005). Cointegration in fractional systems with deterministic trends. *Journal of Econometrics*, forthcoming.

Robinson, P.M. and D. Marinucci (2000). The averaged periodogram for nonstationary vector time series. *Statistical Inference for Stochastic Processes* 3, 149-160.

Robinson, P.M. and D. Marinucci (2001). Narrow-band analysis of nonstationary processes. *Annals of Statistics* 29, 947-986.

Robinson, P.M. and Y. Yajima (2002). Determination of cointegrating rank in fractional systems. *Journal of Econometrics* 106, 217-241.

Shimotsu, K. (2003). Gaussian semiparametric estimation of multivariate fractionally integrated processes. No 571, Economics Discussion Papers from University of Essex, Department of Economics.

Shimotsu, K. (2004). Exact local Whittle estimation of fractional integration with unknown mean and time trend. Preprint, Queen's University.

Shimotsu, K. and P.C.B. Phillips (2000). Modified local Whittle estimation of the memory parameter in the nonstationary case. Cowles Foundation Discussion Paper #1265, Yale University.

Shimotsu, K. and P.C.B. Phillips (2002). Pooled log-periodogram regression. *Journal of Time Series Analysis* 23, 57-93.

Shimotsu, K. and P.C.B. Phillips (2004). Local Whittle estimation in nonstationary and unit root cases. *Annals of Statistics* 32, 656-692.

Shimotsu, K. and P.C.B. Phillips (2005a). Exact local Whittle estimation of fractional integration. *Annals of Statistics*, forthcoming.

Shimotsu, K. and P.C.B. Phillips (2005b). Local Whittle estimation of fractional integration and some of its variants. *Journal of Econometrics,* forthcoming

Sinai, Ya G. (1976). Self-similar probability distributions. *Theory of Probability and its Applications* 21, 64-80.

Sun Y.X. and P.C.B. Phillips (2003). Nonlinear log-periodogram regression for perturbed fractional processes. *Journal of Econometrics* 115, 355-389.

Taqqu, M.S. (2003). Fractional Brownian motion and long-range dependence. In (P. Doukhan, G. Oppenheim and M.S. Taqqu, eds.), *Theory and Applications of Long-Range Dependence*, Birkhäuser. Boston.

Twefik, A.H. and M. Kim (1992). Correlation structure of the discrete wavelet coefficients of fractional Brownian motions. *IEEE Transmission on Information Theory* IT-38, 904-909.

Velasco, C. (1999a). Semiparametric Gaussian 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. (2000). Non-Gaussian log-periodogram regression. *Econometric Theory* 16, 44-79.

Velasco, C. (2003a). Gaussian semiparametric estimation of fractional cointegration. *Journal of Time Series Analysis* 24, 345-378.

Velasco, C. (2003b). Nonparametric frequency domain analysis of non-stationary multivariate time series. *Journal of Statistical Planning and Inference* 116, 209-247.

Velasco, C. (2004). The periodogram of fractional processes. Preprint, Universidad Carlos III de Madrid.

Yajima, Y. (1988). On estimation of a regression model with long-memory stationary errors. *Annals of Statistics* 16, 791-807.

Yajima, Y. (1991). Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *Annals of Statistics* 19, 158-177.

Zhurbenko, I.G. (1979). On the efficiency of estimates of a spectral density. *Scandinavian Journal of Statistics* 6, 49 -56.