Peter M. Robinson

The estimation of misspecified long memory models

Article (Accepted version) (Refereed)

Original citation:
Robinson, Peter M. (2014) *The estimation of misspecified long memory models.* Journal of Econometrics, 178 (2). pp. 225-230. ISSN 0304-4076

DOI: 10.1016/j.jeconom.2013.08.023

The Estimation of Misspecified Long Memory Models

Peter M Robinson*

Department of Economics, London School of Economics,

Houghton Street, London WC2A 2AE, UK

March 7 2013

Abstract

We consider time series that, possibly after integer differencing or integrating or other detrending, are covariance stationary with spectral density that is regularly varying near zero frequency, and unspecified elsewhere. This semiparametric framework includes series with short, long and negative memory. We consider the consistency of the popular log-periodogram memory estimate that, conventionally but wrongly, assumes the spectral density obeys a pure power law. The local-to zero misspecification leads to increased bias, which is liable to prevent the usual central limit theorem from holding. The order of the bias is calculated for several slowly-varying factors, and some discussion of mean squared error and bandwidth choice is included.

JEL classifications: C14; C22

Keywords: Long memory; slowly-varying function; log-periodogram estimate.

*Tel. +44-20-7955-7516; fax: +44-20-7955-6592. *E-mail address:* p.m.robinson@lse.ac.uk

1. INTRODUCTION

The spectral density at low frequencies determines the long-run behaviour of stationary time series. Let the covariance stationary and invertible process z_t, $t = 0, \pm 1, ...,$ have a spectral density function $f(\lambda)$, $\lambda \in (-\pi, \pi]$, defined by

$$cov(z_t, z_{t+j}) = \int_{-\pi}^{\pi} f(\lambda) \cos(j\lambda)\, d\lambda,\ \ j = 0, \pm 1, ... \tag{1}$$

In practice, a finite realization, $z_1, ..., z_n$, may be the outcome of integer differencing or integrating or deterministic detrending of a nonstationary or non-invertible series. With $a \sim b$ meaning that $a/b \to 1$, we assume that $f(\lambda)$ is *regularly-varying* at zero frequency, that is

$$f(\lambda) \sim L\left(\frac{1}{\lambda}\right)\lambda^{-2d}\ ,\ \text{as } \lambda \to 0^+, \tag{2}$$

where $0 \leq |d| < 1/2$ and, for positive argument x, the function $L(x)$ is *slowly-varying* (in Karamata's sense), being positive and measurable on some neighbourhood $[X, \infty)$, with

$$L(cx)/L(x) \to 1 \text{ as } x \to \infty,\ \text{all } c > 0. \tag{3}$$

Detailed discussions of slowly-varying functions, and their applications in probability theory, are contained in Seneta (1974) and Bingham, Goldie and Teugels (1987). A basic property is that as $x \to \infty$ $L(x)$ can diverge, or converge to zero, or converge to a positive constant, or oscillate, and for any $a > 0$,

$$x^a L(x) \to \infty,\ x^{-a} L(x) \to 0,\ \text{as } x \to \infty. \tag{4}$$

Therefore in (2) the power law λ^{-2d} dominates the slowly varying factor $L(1/\lambda)$ so that, for any L, as $\lambda \to 0^+$ $f(\lambda)$ still diverges for $0 < d < 1/2$, and still $f(0) = 0$ for $-1/2 < d < 0$, while when $d = 0$ $f(\lambda)$ diverges when $L(x) \to \infty$ as $x \to \infty$ and $f(0) = 0$ when $L(x) \to 0$ as $x \to \infty$.

The simplest example of such L is

$$L(x) \equiv C > 0. \tag{5}$$

Others include (see Bingham, Goldie and Teugels (1987, p. 16))

$$L(x) = C \log_k x \ , \ k \geq 1, \tag{6}$$

where $\log_1 x = \log x$ and $\log_k x = \log_{k-1} \log x$, $k \geq 2$, as well as powers and rational functions of the $\log_k x$, $k \geq 1$ (e.g. $L(x) = 1/\log x$), and

$$L(x) = C \exp\left\{ \prod_{j=1}^{k} (\log_j x)^{a_j} \right\}, 0 < a_j < 1, \ j = 1, ..., k \geq 1, \tag{7}$$

$$L(x) = C \exp\{\log x / \log_2 x\}. \tag{8}$$

Let $\mathcal{A}_{j,k}$ denote the σ-field of events generated by z_t, $j \leq t \leq k$, and define $\alpha_j = \sup_{A \in \mathcal{A}_{-\infty,t}, B \in \mathcal{A}_{t+j,\infty}} |P(AB) - P(A)P(B)|$ for $j > 0$. Then if $\alpha_j \to 0$ as $j \to \infty$, z_t is said to be $\alpha-$mixing. Suppose for the purposes of this paragraph that z_t is Gaussian, in which case the coefficient of complete regularity decays at the same rate as α_j, see Ibragimov and Rozanov (1978, pp. 111, 113). Thus from Ibragimov and Rozanov (1978, pp. 178) z_t satisfying (2) cannot be $\alpha-$mixing when $d > 0$ (because not every positive power of $f(\lambda)$ is integrable). The usual examples of Gaussian $\alpha-$mixing processes have bounded spectral density, e.g. a stationary and invertible autoregressive moving average (ARMA), and thus satisfy (2) with $d = 0$ and constant L, (5). However $\alpha-$mixing does not rule out all unbounded $f(\lambda)$. From Ibragimov and Rozanov (1978, pp. 179, 180),

$$f(\lambda) = C^* \exp\left\{ \sum_{j=1}^{\infty} \frac{\cos j\lambda}{j(\log j + 1)} \right\} \tag{9}$$

for some $C^* > 0$ implies z_t is $\alpha-$mixing. The spectral density in (9) satisfies

$$f(\lambda) \sim C \log(1/\lambda) \text{ as } \lambda \to 0^+, \tag{10}$$

which corresponds to combining (6) for $k=1$ with (2) for $d=0$. Incidentally under (9) α_j decays very slowly, like $1/\log j$ (and thus does not satisfy conditions for central limit theory for statistics such as the sample mean of z_t, $1 \leq t \leq n$). From Ibragimov and Rozanov (1978, p. 180) a process with spectral density the reciprocal of the right side of (9) (which converges like $(\log(1/\lambda))^{-1}$ as $\lambda \to 0^+$) is also $\alpha-$mixing.

Under additional conditions to (2) (see Yong, 1974) the autovariance sequence satisfies

$$cov(z_t, z_{t+j}) \sim \frac{L\,(j)\,\pi}{\cos(d\pi)\Gamma(2d)} j^{2d-1}, \text{ as } j \to \infty. \tag{11}$$

The probability literature covers the asymptotic behaviour of various simple statistics under (11), in particular linear and quadratic forms (see e.g. Taqqu (1975), Dobrushin and Major (1979), Fox and Taqqu (1985, 1987)). However, the frequency domain form (2) perhaps provides greater intuitive appeal. Early empirical support for the notion of a divergent spectral density at zero frequency was noted by Granger (1966). He reported nonparametric spectral density estimates for a number of economic time series, and while these are inevitably finite at zero frequency, they are strongly peaked there, and his Figure 1 is suggestive of a spectral singularity at zero frequency. Of course such an outcome could also be consistent with nonstationarity (such as a unit root), and he did not present formulae such as (2), but clearly (2) with $d > 0$ and any L, or even with $d = 0$ and diverging L, is consistent with his "typical spectral shape". The leading methods of semiparametric estimation of the memory parameter d have also been frequency-domain. However, they have mainly focussed on the simple power law form, with (5) assumed in (2), that is

$$f(\lambda) \sim C\lambda^{-2d}, \text{ as } \lambda \to 0^+. \tag{12}$$

The leading fractional parametric models (which specify $f(\lambda)$ parametrically for all λ), namely $f(\lambda) \propto \left|1 - e^{i\lambda}\right|^{-2d}$ (Adenstedt (1974)) and its extension to fractionally-integrated ARMA (FARIMA) spectra are covered by (12). In (12) the knife-edge case $d = 0$ describes *short memory*, when a FARIMA reduces to an ARMA, while the cases $0 < d < 1/2$ and $-1/2 < d < 0$ repectively describe *long memory* and *antipersistence.* However, methods of estimating such parametric models are inconsistent when $f(\lambda)$ is misspecified, in particular high-frequency misspecification produces asymptotic bias even in estimates of the low-frequency parameter d. This drawback is overcome (at cost of slower convergence, and of requiring choice of a smoothing number) by semiparametric methods, based on (12), in particular log-periodogram and local Whittle estimates of d and C, see e.g. Geweke and Porter-Hudak (1983), Kuensch (1987), Robinson (1995a,b), where the latter two references established that both estimates are asymptotically normal for all $d \in (-1/2, 1/2)$, and with an asymptotic variance that is constant with respect to d. Thus, standard large-sample inference using these estimates is very simple to implement. Extensions to estimates based on nonstationary processes have been developed by Velasco (1999a,b) and subsequent authors.

In principle, one could specify a particular L in (2) up to an unknown scale factor as in (6)-(8), for example, and accordingly modify the estimates, and we would expect to achieve good statistical properties if L is correctly chosen. One could also imagine specifiying L up to finitely many unknown parameters, e.g. $L(x) = C(\log x)^{\theta}$ for unknown θ, and extend the semiparametric methods to estimate d, C and the additional parameter vector. However in either case the prospect of correct specification of L seems far-fetched, and of greater practical interest is the robustness of existing estimates to unknown, nonparametric, L.

Robinson (1994a) investigated asymptotic properties of the averaged periodogram statistic, and functionals of it of interest, including a semiparametric estimate of d,under (2) with unknown L. Define the discrete Fourier transform

$$w(\lambda) = (2\pi n)^{-1/2} \sum_{t=1}^{n} z_t e^{it\lambda}, \ \lambda \in (-\pi, \pi], \tag{13}$$

and the periodogram

$$I(\lambda) = |w(\lambda)|^2, \quad \lambda \in (-\pi, \pi]. \tag{14}$$

The averaged periodogram is defined as

$$\widehat{F}(\lambda) = \sum_{j=1}^{[n\lambda/2\pi]} I(\lambda_j), \; 0 < \lambda \leq \pi, \tag{15}$$

where [.] here denotes integer part and $\lambda_j = 2\pi j/n$, For a user-chosen integer $m \in [1, n/2)$ satisfying

$$1/m + m/n \to 0 \;\; \text{as } n \to \infty, \tag{16}$$

Robinson (1994a) showed that

$$\widehat{F}(\lambda_m)/F(\lambda_m) \to_p 1, \; \text{as } n \to \infty, \tag{17}$$

where

$$F(\lambda) = \int_0^\lambda f(h)dh. \tag{18}$$

For this purpose (2) was assumed but (like a good deal of the long memory literature) under the restriction $0 < d < 1/2$ (though there seems no reason why a similar result should not hold also for $-1/2 < d \leq 0$), as well as regularity conditions. Further, Robinson (1994a) proposed the following *averaged periodogram* estimate of d :

$$\widetilde{d}_q = \frac{1}{2} - \frac{\log\left\{\widehat{F}(q\lambda_m)/\widehat{F}(\lambda_m)\right\}}{2\log q}, \tag{19}$$

where q is chosen in the interval $(0, 1)$. He showed that under the same conditions as imposed for (17),

$$\widetilde{d}_q \to_p d, \; \text{as } n \to \infty. \tag{20}$$

Under somewhat stronger conditions he obtained a rate of convergence in (20), $O_p(n^{-\eta})$, for some $\eta > 0$. The property (20), like (17), holds for any slowly varying L, which is unknown to the practitioner. Intuitively both properties might be anticipated due to (3) and the ratio forms in the left hand side of (17) and in $\widetilde{d}_q$. Robinson

(1994b) discussed mean squared error and optimal choice of m in this setting. The present paper addresses the above issues with respect to the log-periodogram estimate, which, like $\widetilde{d}_q$ but unlike the local Whittle estimate, is defined in closed form, so relatively easily yields information on rates of convergence. Soulier (2010) established a lower bound for the rate of converrgence of estimates of d in (2), and proved it to be optimal, illustrating his results with the log periodogram estimate. Giraitis, Robinson and Samarov (1997) had considered similar issues with respect to (12), but Soulier (2010) found that the presence of an unanticipated L can produce much slower rates, and that unlike under (12), the log periodogram estimate is no less efficient than the local Whittle estimate, cf. Robinson (1995a, 1995b), where the asymptotic distributional results derived in the latter references may only hold alongside bandwidth choices that yield unacceptable imprecision.

The following section considers the consistency of the log-periodogram estimate. Section 3 evaluates the order of magnitude of the bias in several slowly varying examples, with some discussion of mean sqared error and bandwidth choice. Section 4 provides some concluding remarks.

2. CONSISTENCY OF LOG PERIODOGRAM ESTIMATE

We employ the version of the log-periodogram estimate proposed by Robinson (1995a) (which is slightly simpler than Geweke and Porter-Hudak's (1983)). For m as described in the previous section, define

$$\nu_j = \log j - \frac{1}{m}\sum_{k=1}^{m} \log k, \quad 1 \leq j \leq m, \tag{21}$$

and introduce the additional notation

$$\upsilon_j^{(\ell)} = \sum_{k=1}^{j} \nu_k^{\ell}, \ 1 \leq j \leq m, \tag{22}$$

for integer ℓ. The log-periodogram estimate we consider is

$$\widehat{d} = -\frac{1}{2}\sum_{j=1}^{m}\nu_j \log I(\lambda_j)/v_m^{(2)}. \tag{23}$$

Define also

$$U_j = \log\left\{I(\lambda_j)/f(\lambda_j)\right\}. \tag{24}$$

We introduce two assumptions.

Assumption 1 *As* $n \to \infty$.

$$\frac{1}{m}\sum_{j=1}^{m}\nu_j U_j = o_p(1). \tag{25}$$

The unprimitive Assumption 1 can hold under a variety of conditions, including when z_t is a Gaussian process, a linear process, or a fractional process driven by a mixing input, indeed Robinson (1995a), Hurvich, Deo and Brodsky (1998) and Velasco (1999a) establish central limit theorems for $m^{1/2}(\widehat{d} - d)$ which entail an $O_p(m^{-1/2})$ bound in (25). Strictly, these and other references assume (12) rather than the more general (2) but essentially the same arguments apply.

Assumption 2 *Uniformly in* $\delta \in (0, 1]$,

$$\left\{L\left(\frac{x}{1+\delta}\right)/L(x)\right\}^{1/\delta} \to 1 \textit{ as } x \to \infty. \tag{26}$$

For any $\delta > 0$ the slow variation property (3) implies (26), but Assumption 2 imposes uniformity. Note also from Bingham, Goldie and Teugels (1987, Theorem 1.2.1, p.6) that slow variation implies that (3) holds uniformly on each compact $c-$set in $(0, \infty)$. The parameter δ measures the discrepancy of arguments of $L(x/(1+\delta))$ and $L(x)$ so the power $1/\delta$ in (26) does not look unnatural. Bingham, Goldie and Teugels (1987, p.16) mention an "infinitely oscillating" example of L,

$$L(x) = \exp\left\{(\log x)^{1/3}\cos(\log x)^{1/3}\right\}. \tag{27}$$

This satisfies (3) but not (26), indeed one would not expect a statistical procedure to work in such circumstances. In Section 3 we show that Assumption 2 holds for several examples of L.

In the following theorem the intermediate terms on the right hand sides of (28) and (29) are (identical) expressions for bias, whose rates are obtained in the examples of Section 3.

Theorem *Let (2), (16) and Assumptions 1 and 2 hold. Then as* $n \to \infty$,

$$\begin{aligned} \widehat{d} &= d - \frac{1}{2m}\sum_{j=1}^{m} \nu_j \log\left\{L\left(\frac{1}{\lambda_j}\right)\right\} + o_p(1) && (28) \\ &= d - \frac{1}{2m}\sum_{j=1}^{m-1} \log\left\{L\left(\frac{1}{\lambda_j}\right)/L\left(\frac{1}{\lambda_{j+1}}\right)\right\} v_j^{(1)} + o_p(1) && (29) \\ &\to_p d. && (30) \end{aligned}$$

Proof For some $\varepsilon > 0$ (2) implies that we can write $f(\lambda) = L(1/\lambda)\,\lambda^{-2d}$ for $|\lambda| < \varepsilon$. Thus for sufficiently large n, (16) implies

$$\widehat{d} - d = -\frac{1}{2}\sum_{j=1}^{m} \nu_j \log\left\{L\left(\frac{1}{\lambda_j}\right)\right\}/v_m^{(2)} - \frac{1}{2}\sum_{j=1}^{m}\nu_j U_j / v_m^{(2)}. \tag{31}$$

Noting that

$$v_m^{(2)} \sim m \text{ as } m \to \infty \tag{32}$$

(see Robinson, 1995a), we have

$$\widehat{d} - d = -\frac{1}{2m + o(1)}\sum_{j=1}^{m} \nu_j \log\left\{L\left(\frac{1}{\lambda_j}\right)\right\} + o_p(1) \tag{33}$$

by Assumption 2. By Abel summation by parts, definition (22) and the identity $v_m^{(1)} = 0$,

$$\sum_{j=1}^{m} \nu_j \log\left\{L\left(\frac{1}{\lambda_j}\right)\right\} = \sum_{j=1}^{m-1} \log\left\{L\left(\frac{1}{\lambda_j}\right)/L\left(\frac{1}{\lambda_{j+1}}\right)\right\} v_j^{(1)}. \tag{34}$$

A bound for the absolute value of (34) is

$$\sum_{j=1}^{m-1} \left|\log\left\{L\left(\frac{1}{\lambda_{j+1}}\right)/L\left(\frac{1}{\lambda_j}\right)\right\}\right| (-v_j^{(1)}) \tag{35}$$

because $v_j^{(1)} < 0$ for $1 \le j < m$. For $1 \le j < m$,

$$-v_j^{(1)} = \frac{j}{m}\sum_{k=1}^{m} \log k \; - \sum_{k=1}^{j} \log k. \tag{36}$$

For $r \ge 1$,

$$\sum_{k=1}^{r} \log k \le \int_{1}^{r+1} \log x dx = (r+1)\log(r+1) - r \tag{37}$$

and

$$\sum_{k=1}^{r} \log k \ge \int_{0}^{r} \log x dx = r \log r - r. \tag{38}$$

Thus

$$-v_j^{(1)} \le \frac{j}{m}\left\{(m+1)\log(m+1) - m\right\} - j\log j + j. \tag{39}$$

By Assumption 2 there exists $\varepsilon > 0$ independent of δ such that for large enough x

$$\frac{1}{\delta}\left|\log\left\{L\left(\frac{x}{1+\delta}\right)/L\left(x\right)\right\}\right| < \varepsilon. \tag{40}$$

Thus, taking $x = n/(2\pi j)$, $\delta = 1/j$ there exists $\varepsilon > 0$ independent of j such that for large enough n/m

$$j\left|\log\left\{L\left(\frac{1}{\lambda_{j+1}}\right)/L\left(\frac{1}{\lambda_j}\right)\right\}\right| < \varepsilon, \quad 1 \le j \; \le m. \tag{41}$$

Thus for large enough n, in view of (16) and using (38) again and (as frequently in the following section) the inequality $|\log(1+y)| \le |y|$, (35) is bounded by

$$\begin{aligned}
& \varepsilon\sum_{j=1}^{m-1}\left\{\frac{1}{m}\left\{(m+1)\log(m+1) - m\right\} - (\log j - 1)\right\} \\
\le \;& \varepsilon\left(\frac{m-1}{m}\left\{(m+1)\log(m+1) - m\right\} - (m-1)\log(m-1) + 2(m-1)\right) \\
\le \;& \varepsilon\left[(m-1)\left\{(1+\frac{1}{m})\log(m+1) - \log(m-1)\right\} + m\right] \\
\le \;& \varepsilon\left[m\log((m+1)\,/\,(m-1)) + 2m\right] \\
\le \;& \varepsilon\left[m\log(1+\frac{2}{m-1}) + 2m\right] \le \varepsilon\left[\frac{2m}{m-1} + 2m\right] \le 3\varepsilon m.
\end{aligned} \tag{42}$$

From (31), (34), (35) and arbitrariness of ε the proof is completed.

3. EXAMPLES AND RATES

The paragraph following Assumption 2 argues that the assumption does not much strengthen the slow variation property of L, but it is nevertheless desirable to check it in several cases, and this will desirably indicate rates of convergence. Throughout the derivations it is understood that x is chosen arbitrarily large and $\delta \in (0,1]$.

1. $L(x) = C(1 + Dx^{-\beta})$, $0 < \beta \leq 2$, $D \neq 0$.

This is actually a case of (12), and was assumed in the central limit theorem for $\widehat{d}$ of Robinson (1995a), because some refinement of (12) is necessary in order to get a rate of convergence and thence limit distribution theory. The order of the bias is thus already known in this case and we consider it here only to verify that estimating the bias by approximating (35) produces a sharp outcome. We have

$$\log\left\{L\left(\frac{x}{1+\delta}\right)/L\left(x\right)\right\} = \log\left\{(1 + D(x/(1+\delta))^{-\beta})/(1 + Dx^{-\beta})\right\}. \qquad (43)$$

This has absolute value

$$\begin{aligned} & \left|\log\left\{1 + Dx^{-\beta}((1+\delta)^{\beta} - 1))/(1 + Dx^{-\beta})\right\}\right| \leq \left|Dx^{-\beta}((1+\delta)^{\beta} - 1))/(1 + Dx^{-\beta})\right| \\ \leq \; & 8\left|D\right|\delta x^{-\beta} \leq K\delta x^{-\beta}, \end{aligned} \qquad (44)$$

where K denotes a generic positive constant, and, here and subsequently, we use the inequalities $|(1+y)^a - 1| \leq 4y$ for $y \in (0,1]$, $a \in (0,1/2]$, and $1 + Dy \geq 1/2$ for small enough positive y. Thus Assumption 2 is checked. Further, (44) implies that the modulus in (35) is bounded by $K(n/j)^{\beta}/j \leq K(n/m)^{\beta}/j$, rather than (as in (41)) ε/j, so the calculation in (42) implies that

$$\frac{1}{2m}\sum_{j=1}^{m}\nu_j \log\left\{L\left(\frac{1}{\lambda_j}\right)\right\} = O((\frac{m}{n})^{\beta}). \qquad (45)$$

This accords with the bias calculation implicit in Robinson (1995a) so the bound (45) is in fact sharp. For the central limit theorem for $m^{1/2}(\widehat{d} - d)$ one needs at least that

$m^{2\beta+1}/n^{2\beta} \to 0$ as $n \to \infty$, while on the other hand the asymptotic mean squared error (MSE) of $\widehat{d}$ is of form $a/m + b(m/n)^{2\beta}$, for $a, b > 0$, producing the optimal rate for m, $n^{2\beta/(2\beta+1)}$, for example $n^{4/5}$ in the case $\beta = 2$ mostly considered in the bandwidth choice literature.

2. $L(x) = C(1 + D(\log x)^{-1})$, $D \neq 0$.

Again (12) is satisfied, but there is less local smoothness than in the preceding example 1. We have

$$\begin{aligned}\log\left\{L\left(\frac{x}{1+\delta}\right)/L(x)\right\} &= \log\left\{(1 + D(\log(x/(1+\delta)))^{-1})/(1 + D(\log x)^{-1})\right\} \\ &= \log\left\{1 + D(\log x)^{-1}\left((1 - \log(1+\delta)/\log x)^{-1} - 1\right)\right. \\ &\qquad \left./(1 + D(\log x)^{-1})\right\}. \end{aligned} \tag{46}$$

This is bounded in absolute value by

$$\begin{aligned} & 2|D|(\log x)^{-1}\left((1 - \log(1+\delta)/\log x)^{-1} - 1\right) \\ \leq\ & 4|D|\log(1+\delta)(\log x)^{-2} \leq K\delta(\log x)^{-2}, \end{aligned} \tag{47}$$

using the inequality $(1-y)^{-1} - 1 \leq 2y$ for small enough positive y. Thus we have checked Assumption 2. Also, arguing as in example 1, (47) gives

$$\frac{1}{2m}\sum_{j=1}^{m}\nu_j \log\left\{L\left(\frac{1}{\lambda_j}\right)\right\} = O((\log(n/m))^{-2}). \tag{48}$$

No central limit theorem for $m^{1/2}(\widehat{d} - d)$ is thus possible unless $(\log(n/m))^{-2} = o(m^{-1/2})$, for which a necessary condition is $m = o((\log n)^4)$. The MSE of $\widehat{d}$ is $a/m + b(\log(n/m))^{-4}$, which is of order $(\log n)^{-4}$ when m is chosen to increase like $(\log n)^{\beta}$ for any $\beta \geq 4$.

3. $L(x) = C(\log x)^{\beta}$, $\beta \neq 0$.

This generalizes (6) with $k = 1$. We have

$$\begin{aligned}\log\left\{L\left(\frac{x}{1+\delta}\right)/L\left(x\right)\right\} &= \beta\log\left\{\log\left(\frac{x}{1+\delta}\right)/\log x\right\}\\ &= \beta\log\left\{1-\log\left(1+\delta\right)/\log x\right\}, \end{aligned}\tag{49}$$

which is bounded in absolute value by

$$\left|\beta\log\left(1+\delta\right)/\log x\right| \leq K\delta/\log x. \tag{50}$$

Thus Assumption 2 holds, and arguing as before

$$\frac{1}{2m}\sum_{j=1}^{m}\nu_j\log\left\{L\left(\frac{1}{\lambda_j}\right)\right\} = O((\log\left(n/m\right))^{-1}). \tag{51}$$

Direct integral approximation of the left side leads to the same result, so the bound in (51) appears to be sharp. It is interesting to note that the rate is independent of the power β, and is half as good as in example 2, where (12) held. In the present case the central limit theorem for $m^{1/2}(\widehat{d}-d)$ would require $(\log\left(n/m\right))^{-1} = o(m^{-1/2})$, for which a necessary condition is $m = o((\log n)^2)$. The MSE of $\widehat{d}$ is of order $(\log n)^{-2}$ when m is chosen to increase like $(\log n)^{\beta}$ for any $\beta \geq 2$.

4. $L(x) = C\log_k x,\ k \geq 2$.

This possibility was mentioned in (6), and discussed by Soulier (2010) in case $k = 2$. We have

$$\begin{aligned}\log\left\{L\left(\frac{x}{1+\delta}\right)/L\left(x\right)\right\} &= \log\left\{\log_k\left(\frac{x}{1+\delta}\right)/\log_k x\right\}\\ &= \log_k\left\{\log(\frac{x}{1+\delta})\right\} - \log_{k+1}x\\ &= \log_k\left\{\log x(1-\log(1+\delta)/\log x)\right\} - \log_{k+1}x\\ &= \log_{k-1}\left\{\log_2 x + \log(1-\log(1+\delta)/\log x)\right\} - \log_{k+1}x\\ &= \log_{k-1}\left\{\log_2 x(1+\log(1-\log(1+\delta)/\log x)/\log_2 x)\right\}\\ &\quad - \log_{k+1}x. \end{aligned}\tag{52}$$

For $k = 2$ this has absolute value

$$\begin{aligned}
|\log \{1 + \log(1 - \log(1+\delta)/\log x)/\log_2 x\}| &\leq |\log \{1 - \log(1+\delta)/\log x\} / \log_2 x| \\
&\leq \log(1+\delta)/(\log x \log_2 x) \\
&\leq \delta/(\log x \log_2 x). \qquad (53)
\end{aligned}$$

For $k \geq 3$ (52) is

$$\begin{aligned}
& \log_{k-2} \{\log_3 x + \log(1 + \log(1 - \log(1+\delta)/\log x)/\log_2 x)\} - \log_{k+1} x \\
= \; & \log_{k-2} \{\log_3 x(1 + \log(1 + \log(1 - \log(1+\delta)/\log x)/\log_2 x)/\log_3 x)\} \\
& - \log_{k+1} x, \qquad (54)
\end{aligned}$$

and by continuing the arguments in (52) and (53) it is eventually seen that (54) is bounded in absolute value by

$$\delta/\prod_{j=1}^{k} \log_j x. \qquad (55)$$

Thus Assumption 2 holds, and

$$\frac{1}{2m}\sum_{j=1}^{m}\nu_j \log\left\{L\left(\frac{1}{\lambda_j}\right)\right\} = O(\left|\prod_{j=1}^{k}\log_j(n/m)\right|^{-1}). \qquad (56)$$

The rate improves with increasing k as expected, albeit slowly.

5. $L(x) = C \exp\left\{(\log x)^{\beta}\right\}$, $0 < \beta < 1$.

This is a special case of (7). We have

$$\begin{aligned}
\log\left\{L\left(\frac{x}{1+\delta}\right)/L(x)\right\} &= (\log\left(\frac{x}{1+\delta}\right))^{\beta} - (\log x)^{\beta} \\
&= (\log x)^{\beta}\left\{(1 - \log(1+\delta)/\log x)^{\beta} - 1\right\}. \qquad (57)
\end{aligned}$$

This is bounded in absolute value by

$$K(\log x)^{\beta}\log(1+\delta)/\log x \leq K\delta(\log x)^{\beta-1}, \qquad (58)$$

to check Assumption 2, and arguing as before

$$\frac{1}{2m}\sum_{j=1}^{m}\nu_j \log\left\{L\left(\frac{1}{\lambda_j}\right)\right\} = O(|\log(n/m)|^{\beta-1}). \tag{59}$$

6. $L(x) = C\exp\{\log x/\log_2 x\}$.

This is (8). We have

$$\begin{aligned}\log\left\{L\left(\frac{x}{1+\delta}\right)/L(x)\right\} &= \log\left(\frac{x}{1+\delta}\right)/\log_2\left(\frac{x}{1+\delta}\right) - \log x/\log_2 x \\ &= \left\{\log\left(\frac{x}{1+\delta}\right)\log_2 x - \log_2\left(\frac{x}{1+\delta}\right)\log x\right\} \\ &\quad /\log_2\left(\frac{x}{1+\delta}\right)\log_2 x. \end{aligned} \tag{60}$$

The numerator is

$$\begin{aligned} &\{\log x - \log(1+\delta)\}\log_2 x - \{\log(\log x - \log(1+\delta))\}\log x \\ =\ &\{\log x - \log(1+\delta)\}\log_2 x - \{\log(\log x(1-\log(1+\delta)/\log x)\}\log x \\ =\ &-\log(1+\delta)\log_2 x - \{\log(1-\log(1+\delta)/\log x\}\log x, \end{aligned} \tag{61}$$

which is bounded in absolute vale by $K\delta\log_2 x$. The denominator of (60) is

$$\begin{aligned}\log(\log x - \log(1+\delta))\log_2 x &= \log(\log x(1-\log(1+\delta))/\log x)\log_2 x \\ &= \{\log_2 x + \log(1-\log(1+\delta))/\log x\}\log_2 x \\ &\sim (\log_2 x)^2. \end{aligned} \tag{62}$$

Thus Assumpton 2 is checked, and arguing as before

$$\frac{1}{2m}\sum_{j=1}^{m}\nu_j \log\left\{L\left(\frac{1}{\lambda_j}\right)\right\} = O(|\log_2(n/m)|^{-1}), \tag{63}$$

the slowest rate of any of our examples.

4. FINAL COMMENTS

We have considered the consistency of the semiparametric log-periodogram regression memory estimate in the presence of an unanticipated slowly-varying factor in the spectral density, under a general condition on the function, and verified this condition and calculated convergence rates in several examples. As implied by the results of Soulier (2010), these convergence rates are mostly slow, to the extent that unless the bandwidth m grows extremely slowly the bias will be too large to allow the central limit theorem to hold. Practically this might suggest picking m very small, unless n is extremely large, but the effect would likely be unacceptable imprecision in the estimate. Soulier (2010) discussed the bandwidth choice issue, with numerical illustrations.

Similar results hold for the original log-periodogram estimate of Geweke and Porter-Hudak (1983). Note that ν_j in (21) is identical to $\log \lambda_j - m^{-1}\Sigma_{k=1}^{m} \log \lambda_k$, and Geweke and Porter-Hudak's (1983) version replaces ν_j by $2\log(\sin \lambda_j/2) - 2m^{-1}\Sigma_{k=1}^{m} \log(\sin \lambda_k/2)$, where $2\log(\sin \lambda/2) = \log \lambda + O(\lambda^2)$ as $\lambda \to 0^+$. Similar results also hold for improved modifications of log-periodogram regression (Moulines and Soulier, 1999) and for the local Whittle estimate and its modified versions. We also anticipate similar outcomes for extensionss of these estimates that allow for possible nonstationarity or non-invertibility (see e.g. Velasco, 1999a, 1999b).

If we are to be concerned about the effect of a possible slowly-varying factor on inference on long memory, we might also worry about its effect on nonparametric spectral estimation and conventional autocorrelation-consistent variance estimation. These are both based on the assumption of a finite, positive spectral density. When there is actually a divergent slowly-varying factor a nonparametric spectral estimate at zero frequency will lose consistency, while a divergent or convergent-to-zero slowly-varying factor would appear to invalidate the usual autocorrelation-robust rules of inference, though in view of (17) appropriate ones can be constructed (see Robinson, 1994a).

The possibility of investigating the presence of a slowly-varying factor and its form might be pursued. Since $\widehat{d}$ is at least consistent for d, the normalized periodograms

$I(\lambda_j)\lambda_j^{2\widehat{d}}$ might be employed in nonparametric estimation of L, or in a hypothesis test. However the slow convergence of $\widehat{d}$ could prove an obstacle, and even if asymptotically valid procedures can be developed, they would surely require an extremely long time series, while it may be recalled that under (12), Robinson (1995a) found that estimates of d and C are asymptotically perfectly correlated.

Acknowledgment

I thank two referees for comments which have led to presentational improvements, and Phillipe Soulier for alerting me to the reference Soulier (2010).

References

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

Bingham, N.H., Goldie, C.M., Teugels, J.L. 1987. Regular Variation. Cambridge University Press, Cambridge.

Dobrushin, R.L., Major, P., 1979. Non-central limit theorems for nonlinear functionals of Gaussian fields. Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete 50, 27-52.

Fox, R., Taqqu, M.S., 1985. Non-central limit theorems for quadratic forms in random variables having long-range dependence. Annals of Probability 13, 482-486.

Fox, R., Taqqu, M.S., 1987. Central limit theorems for quadratic forms in random variables having long-range dependence. Probability Theory and Related Fields 74, 213-240.

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

Giraitis, L., Robinson, P.M., Samarov, A., 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.

Granger, C.W.J., 1966. The typical spectral shape of an economic variable. Econometrica 34, 150-161.

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

Ibragimov, I.A., Rozanov, Y., 1978. Gaussian Random Processes. Springer-Verlag, New York.

Kuensch, H.R., 1987. Statistical aspects of self-similar processes. Proceedings First World Congress of the Bernoulli Society. VNU Science Press, Utrecht.

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

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., 1995a. Log-periodogram regression of time series with long range dependence. Annals of Statistics 23, 1048-1072.

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

Seneta, E., 1976. Functions of Regular Variation. Springer, New York.

Soulier, P., 2010. Best attainable rates of convergence for the estimation of the memory parameter. In Dependence in Probability and Statistics, Lecture Notes in Statistics 200, Springer, Berlin, pp. 43–57.

Taqqu, M.S., 1975. Weak convergence to fractional Brownian motion and to the Rosenblatt process. Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete 31, 287-302.

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

Velasco, C., 1999b. Semiparametric Gaussian estimation of non-stationary time series. Journal of Time Series Analysis 20, 87-127.

Yong, C.H., 1974. Asymptotic Behaviour of Trigonometric Series. Chinese University of Hong Kong, Hong Kong.