

Some aspects of statistical inference for long memory processes

Alcuni aspetti dell'inferenza statistica per processi con memoria lunga

Domenico Marinucci

Dip. di Studi Geoeconomici, Linguistici, Statistici e Storici per l'Analisi Regionale
Università di Roma "La Sapienza"
marinucc@scec.eco.uniroma1.it

Riassunto: Vengono discussi alcuni problemi di inferenza statistica che sorgono nell'analisi dei processi stocastici con memoria lunga. In particolare verranno analizzati i problemi di stima sia in ambito parametrico che in ambito semiparametrico. Verranno inoltre discusse alcune questioni relative ai processi a varianza infinita e agli effetti di trasformazioni non lineari.

Keywords: Parametric and semiparametric methods; frequency domain analysis.

1. Introduction

Let $\{x_t\}$, $t = 0, \pm 1, \pm 2, \dots$ denote a zero mean, covariance stationary process with autocovariance function

$$\gamma(\tau) = Ex_0x_\tau, \tau = 0, \pm 1, \pm 2, \dots$$

Introduce the spectral density

$$f(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma(\tau) \exp(-i\lambda\tau), 0 \leq \lambda \leq \pi.$$

We shall say x_t is long memory if

$$f(\lambda) \sim G\lambda^{-2d}, G > 0, 0 < d < \frac{1}{2}, \text{ as } \lambda \downarrow 0, \quad (1)$$

$\sim$ denoting that the ratio of the left- and right-hand sides tends to unity. Alternatively, long memory can be defined by the related condition

$$\gamma(\tau) \sim G'\tau^{2d-1}, \text{ as } \tau \rightarrow \infty. \quad (2)$$

(1) and (2) can be shown to be equivalent under mild regularity conditions, entailing basically that $f(\lambda)$ has no singularities at seasonal frequencies; in general, (1) is slightly more general than (2). Both assumptions are in marked contrast with the standard properties of many traditional time series models; for instance, for the broad class of stationary autoregressive moving averages the spectral density is bounded and bounded away from zero at all frequencies and the decay of autocovariances is exponential. (1) and (2) can be readily interpreted as imposing a slow decay of innovations and a dominance of very long run components in the observed behaviour of x_t . Such characteristics have made long memory model a very popular tool for a wide range of applications, including in

particular telecommunications data, hydrology, and economics. At the same time, long memory models have drawn an enormous amount of attention in the theoretical literature. It is impossible to provide a comprehensive survey of recent developments in a few pages; a very good and updated reference is the survey paper in the forthcoming book by Robinson (forthcoming). Our purpose here is more limited, namely we focus only on the estimation of the structural parameters and on the effects of nonlinear transformations on the estimates. This paper shall hence be divided into four sections: §2 reviews parametric estimates, §3 discusses semiparametric methods; §4 relaxes (1) and (2) to consider models with long memory and infinite variance whereas §5 focusses on the effects of nonlinear transformations on the memory properties of the processes involved and on statistical inference. Of course, due to space limitations we do not even attempt to completeness, and we refer to the original papers for proofs and details.

2. Parametric estimates

Most parametric estimates for long memory series have been developed in the frequency domain. For a parametric estimate, we must assume first that the functional form of the spectral density is known, up to a vector of finitely many unknown parameters θ . For instance, consider the highly popular autoregressive fractionally integrated moving average (*ARFIMA*(p, d, q)) class, defined by

$$(1 - L)^d x_t = \frac{\vartheta(L)}{\varphi(L)} u_t, \quad Eu_t = 0, \quad Eu_t^2 = \sigma_u^2, \quad d \in (0, \frac{1}{2}),$$

where L is the lag operator, u_t is an independent and identically distributed (*i.i.d.*) sequence,

$$\vartheta(L) = 1 + \vartheta_1 L + \dots + \vartheta_q L^q, \quad \varphi(L) = 1 - \varphi_1 L - \dots - \varphi_p L^p,$$

are polynomials whose roots lie outside the unit circle, and formally

$$(1 - L)^d = \sum_{k=1}^{\infty} \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)} L^k, \quad \Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx.$$

In this case the spectral density of x_t is known up to a vector θ of $p + q + 2$ parameters, namely

$$f(\lambda) = f(\lambda; \theta) = |1 - e^{i\lambda}|^{-2d} \frac{|\vartheta(e^{i\lambda})|^2 \sigma_u^2}{|\varphi(e^{i\lambda})|^2 2\pi},$$

$$\theta = (\varphi_1, \dots, \varphi_p; \vartheta_1, \dots, \vartheta_q; d, \sigma_u^2)'.$$

Let us now introduce the well-known periodogram, which we define as

$$I_n(\lambda) = \frac{1}{2\pi} \sum_{\tau=-n+1}^{n-1} c(\tau) \exp(-i\lambda\tau),$$

where the sample autocovariance $c(\tau)$ is defined as

$$c(\tau) = \begin{cases} n^{-1} \sum_{t=1}^{n-\tau} x_t x_{t+\tau}, & \text{for } \tau \geq 0 \\ c(-\tau), & \text{for } \tau < 0 \end{cases}.$$

Now consider the Whittle approximation to (minus) the Gaussian log-likelihood, i.e.

$$\mathcal{L}(\theta; x_1, x_2, \dots, x_n) = \int_{-\pi}^{\pi} \left\{ \frac{I_n(\lambda)}{f(\lambda; \theta)} + \ln f(\lambda; \theta) \right\} d\lambda. \quad (3)$$

Fox and Taqqu (1986) study the asymptotic behaviour (under Gaussianity) of the approximate maximum likelihood estimator of θ defined implicitly by

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} \mathcal{L}(\theta; x_1, x_2, \dots, x_n), \quad (4)$$

where Θ is some compact set of admissible values. Write θ_0 for the “true” value of the vector of parameters in the data generating process for the observed data; under mild regularity conditions, Fox and Taqqu (1986) show that, as $n \rightarrow \infty$,

$$\begin{aligned} \lim_{n \rightarrow \infty} \hat{\theta}_n &= \theta_0, \text{ with probability one,} \\ n^{1/2}(\hat{\theta}_n - \theta_0) &\xrightarrow{d} N(0, 4\pi W^{-1}(\theta_0)), \end{aligned} \quad (5)$$

where $W(\theta_0)$ is a $(p + q + 2) \times (p + q + 2)$ matrix with elements

$$W(\theta_0) = \int_{-\pi}^{\pi} \left[f(\lambda; \theta) \frac{\partial^2}{\partial \theta \partial \theta'} f^{-1}(\lambda; \theta) \right]_{\theta=\theta_0} d\lambda. \quad (6)$$

(6) can be clearly interpreted as a frequency-domain analogue of the Fisher information matrix, and in this sense it is natural to conjecture that $\hat{\theta}_n$ is also asymptotically efficient. This conjecture is proved correct by Dahlhaus (1989), where it is shown that $\hat{\theta}_n$ is asymptotically equivalent to maximum likelihood in the time domain and it achieves asymptotically the Cramer-Rao lower bound for variance. The proof of (5) is hard; it exploits standard arguments for implicitly defined extremum estimates, together with a central limit theorem for the quadratic form

$$\int_{-\pi}^{\pi} \{I_n(\lambda) - EI_n(\lambda)\} g(\lambda) d\lambda, \quad (7)$$

for the special case where $g(\lambda) = \partial f^{-1}(\lambda; \theta) / \partial \theta$. The asymptotic behaviour of (7) is discussed in Fox and Taqqu (1985), Fox and Taqqu (1987), and it is found not to be Gaussian, in general, for arbitrary $g(\lambda)$. Gaussianity is achieved due to the special nature of $\partial f^{-1}(\lambda; \theta) / \partial \theta$, which has a zero at the origin matching exactly the singularity in the spectral density of x_t . The work of Fox and Taqqu (1986) was extended to linear processes with finite fourth moments by Giraitis and Surgailis (1990). Further extensions are provided by Hosoya (1997), who considers multivariate processes and innovations that satisfy a mild mixing conditions. Kokoszka and Taqqu (1996) focus instead on the case where the innovations are *i.i.d.* but do not have finite variance, i.e. they belong to the domain of attraction of so-called α -stable laws; this assumption is useful for applications in finance and telecommunications. The limiting distribution, however, is extremely complicated, and depends on several nuisance parameters. Further work in the area includes the recent papers by Giraitis and Taqqu (1998), Giraitis and Taqqu (1999); extensions to nonstationary circumstances are discussed by Velasco and Robinson (2000).

3. Semiparametric estimates

In view of the results surveyed in the previous Section, we can say in short that the problem of statistical inference for long memory series has been successfully addressed in the presence of a parametric model, at least under linearity assumptions. It should be argued, however, that knowledge of the full dynamics of x_t is a rather restrictive assumption for many, if not most, practical applications. For instance, if an $ARFIMA(p, d, q)$ model is believed to represent appropriately the behaviour of a given series (an assumption by itself restrictive), then the results of Section 2 require that the exact order of lags is known for both the AR and MA part, otherwise inconsistent estimates will arise. Little is known on the behaviour of model diagnostic criteria (such as the AIC or BIC) in the presence of long memory, and this is particularly disturbing for applications, as misspecifications of the short-run dynamics (the $ARMA$ part) can result in inconsistencies in the estimate of the parameter d , which only relates to the long-run behaviour of the series and can be the only parameter of interest for many applications. These considerations motivate the introduction of a number of semiparametric procedures based upon assumption (1) alone, i.e. where no condition is imposed on the behaviour of the spectral density outside from an arbitrarily small band around the origin. The first of such procedures was the so-called log periodogram regression estimate, discussed by Geweke and Porter-Hudak (1983) and analyzed rigorously by Robinson (1995b). Consider the identity

$$\log I_n(\lambda_j) = \log G - 2d \log \lambda_j + \log \frac{f(\lambda_j)}{G\lambda_j^{-2d}} + \log \frac{I_n(\lambda_j)}{f(\lambda_j)}, \quad j = 1, 2, \dots, m < n, \quad (8)$$

where $\lambda_j = 2\pi j/n$ represent the Fourier frequencies. It is convenient to focus on these frequencies alone, as in this case no mean correction is necessary; this can be particularly advantageous in the presence of long memory, as it is known that the sample mean is less than $\sqrt{n}$ -consistent under these circumstances. Now introduce the bandwidth condition

$$\frac{1}{m} + \frac{m}{n} \rightarrow 0, \text{ as } n \rightarrow \infty. \quad (9)$$

In words, this means we are focussing only on a degenerating band of frequencies around the origin. Under (1) and (9), it is natural to expect $\log f(\lambda_j)/G\lambda_j^{-2d}$ to go to zero; on the other hand, Geweke and Porter-Hudak (1983) conjectured $\log I_n(\lambda_j)/f(\lambda_j)$ to be approximately *i.i.d.* as $n \rightarrow \infty$, as it happens with short memory, so that (8) might be dealt with as a standard regression model with approximately white noise residuals. The conjecture on $\log I_n(\lambda_j)/f(\lambda_j)$ was proved false by Robinson (1995b); the regression idea is however retained by this author, who investigates the asymptotic behavior of the OLS-like estimate

$$\begin{aligned} \hat{d}_n &= \left\{ \sum_{j=1}^m \Lambda_j^2 \right\}^{-1} \sum_{j=1}^m \Lambda_j \log I_n(\lambda_j), \\ \Lambda_j &= \log \lambda_j - \frac{1}{m} \sum_{j=1}^m \log \lambda_j. \end{aligned} \quad (10)$$

Under Gaussianity, and no condition on $f(\lambda)$ away from the origin other than integrability (a consequence of stationarity), Robinson (1995b) proves that

$$m^{1/2} \left\{ \hat{d}_n - d \right\} \xrightarrow{d} N\left(0, \frac{\pi^2}{24}\right), \text{ as } n \rightarrow \infty.$$

Some remarks are needed. The asymptotic variance of the estimate can be lowered, i.e. the asymptotic efficiency improved, by pooling together neighbouring frequencies of the periodogram (the lower bound for the variance of the pooled estimate is $1/4$, but this value cannot be achieved exactly in this class). Gaussianity is a restrictive assumption, but it is very hard to relax for this kind of estimates, which require a complicated non-linear transformation of the periodogram and a demanding analysis of its higher order cumulants; Velasco (2000) managed to relax Gaussianity to linear processes with finite moments of all orders and a smooth distribution function. The estimate is computationally very convenient, and it was extended by Robinson (1995b) to the vector case as well; under these circumstances, further efficiency gains can be achieved by a GLS-type procedure, where a priori constraints on the parameters for the various components are included in the specification. The analysis of asymptotic variance should not obscure the fact that the relative efficiency is zero with respect to correctly specified parametric estimates, due to the bandwidth condition $m/n \rightarrow 0$. This is natural and expected for a semiparametric method, where large part of the information is discarded (the periodogram values for $j = m + 1, \dots, n/2$); it can be viewed as a necessary price to pay for the lack of any assumption other than (1). Further efficiency improvement can be obtained in an alternative semiparametric procedure, which was proposed by Robinson (1995a) and labelled semiparametric Whittle estimate. The idea is to focus on the discretized Whittle likelihood

$$\mathcal{L}_{n,m}(G, d; x_1, x_2, \dots, x_n) = \sum_{j=1}^m \left\{ \frac{I_n(\lambda_j)}{G\lambda_j^{-2d}} + \log G\lambda_j^{-2d} \right\}. \quad (11)$$

For $m = n/2$, (11) would represent nothing else than a discretized version of the likelihood function (3), in the presence of the parametric model (fractional Gaussian noise) $f(\lambda; G, d) = G\lambda^{-2d}$, $\lambda \in (0, 2\pi)$. The motivation is here different, however; we assume that only (1) hold, and maximize the likelihood only on the smallest $m < n/2$ frequencies, under the bandwidth condition (9), i.e. we focus on the extremum estimate

$$\tilde{d}_n = \arg \min_{d \in (0, 1/2), G > 0} \mathcal{L}_{n,m}(G, d; x_1, x_2, \dots, x_n).$$

In words, we are maximizing the likelihood only on the frequencies corresponding to long run behaviour, because only on those frequencies we are able to propose a reliable functional (parametric) form for the spectral density. A rigorous asymptotic theory can be provided under the only assumptions that x_t satisfies a linear in martingales representations, with finite fourth moments (ironically, no Gaussianity is needed for these estimates based upon the Gaussian likelihood). Robinson (1995a) is then able to show that

$$\sqrt{m} \left\{ \tilde{d}_n - d \right\} \xrightarrow{d} N\left(0, \frac{1}{4}\right). \quad (12)$$

(12) allows for standard inference procedures under very general conditions; as for (10), there is clearly an issue for the optimal choice of the bandwidth parameter m , and this is addressed by Giraitis *et al.* (1997). More recently, some work has also been devoted to the possibility to improve the approximation of the asymptotic distribution in finite samples, by means of Edgeworth expansion techniques (Giraitis and Robinson, 2000); again, the mathematical details are quite demanding. The extension of the log-periodogram and Whittle procedure to (some form of) nonstationary processes is provided in two papers by Velasco (1999b), Velasco (1999a).

An approach which we can view as somewhat intermediate between fully parametric and semiparametric procedures has been recently advocated by Moulines and Soulier (1999). The idea is to entertain a log-periodogram regression on the lowest m frequencies, as in Robinson (1995b); the remaining periodogram ordinates are not discarded, but the spectral density in the corresponding region is expanded into orthogonal components and estimated nonparametrically. More precisely, Moulines and Soulier (1999) focus on the regression model

$$\log I_n(\lambda_j) = \log G - 2d \log \lambda_j + \sum_{k=1}^p \vartheta_k \cos \{k\lambda_j\}, \quad j = 1, 2, \dots, n/2,$$

where $1/p + p/n \rightarrow 0$ as $n \rightarrow \infty$; in words, and as expected for a nonparametric estimate, for consistency the number of regressors p must increase with n , but at a smaller rate. The asymptotic distribution of the estimate of d from this regression model is again Gaussian, with a consistency rate smaller than $\sqrt{n}$, but it can achieve $\sqrt{n/\log n}$ in the presence of a sufficiently smooth spectral density $f(\cdot)$.

Other semiparametric methods, such as those by Hall *et al.* (1997) or Giraitis *et al.* (1999) have focused instead on a time domain analysis, for instance by regressing the (log-) sample autocovariances over the (log-) lag τ , see (2). The analysis of the asymptotics for such procedures is rather complicated, however, due to the lack of any orthogonality structure among autocovariances at different lags, and the resulting asymptotic distributions are almost intractable for empirical applications.

4. Parametric estimates for infinite variance models

In many empirical applications, especially for telecommunications and finance, the assumption that the variance of the innovations is finite is not innocuous. As in (Kokoszka and Taqqu, 1995, 1996), it is possible to consider the infinite variance *ARFIMA* model

$$(1 - L)^d x_t = \frac{\vartheta(L)}{\varphi(L)} z_t, \quad E z_t = 0, \quad d \in (0, \frac{1}{2}) \quad (13)$$

where the z_t are *i.i.d.* with mean zero and in the domain of attraction of an α -stable law with $1 < \alpha < 2$, that is

$$P(|z_t| > z) = z^{-\alpha} \ell(z) \text{ as } z \rightarrow \infty,$$

where $\ell(\cdot)$ is a slowly varying function, and

$$\frac{P(z_t > z)}{P(|z_t| > z)} = a, \quad \frac{P(z_t < -z)}{P(|z_t| > z)} = b, \text{ as } z \rightarrow \infty, \quad a + b = 1;$$

recall a symmetric random variable Y is α -stable if $E \exp(iuY) = \exp(-\sigma^\alpha |u|^\alpha)$, for some scale parameter σ . Of course, conditions (1) and (2) no longer provide useful characterizations of long memory behaviour here, because the spectral density and the autocovariance sequence are no longer well-defined if the process is not square integrable. However Kokoszka and Taqqu (1995) show that if $\vartheta(z)$ and $\varphi(z)$ have no root in common

and satisfy the usual stationarity and invertibility conditions, there exist a unique moving average process

$$x_t = \sum_{j=0}^{\infty} c_j z_{t-j} ,$$

which satisfies (13). The coefficients c_j are defined by the equality

$$\sum_{j=0}^{\infty} c_j z^j = \frac{\vartheta(z)}{(1-z)^d \varphi(z)} , |z| < 1 ;$$

note that $c_j \sim c j^{d-1}$, some $c > 0$ as $j \rightarrow \infty$. We can then define the *power transfer function* of x_t as

$$g(\lambda; \theta) = \left| \frac{\vartheta(e^{-i\lambda}; \theta)}{(1 - e^{-i\lambda})^d \varphi(e^{-i\lambda}; \theta)} \right|^2 , \theta = (\varphi_1, \dots, \varphi_p; \vartheta_1, \dots, \vartheta_q; d)' ,$$

which would be proportional to the spectral density of x_t , if the variance of z_t were finite. It is thus natural to consider the parametric pseudo-maximum likelihood estimator defined implicitly by

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} \mathcal{L}'(\theta; x_1, x_2, \dots, x_n) , \quad (14)$$

where now

$$\mathcal{L}'(\theta; x_1, x_2, \dots, x_n) = \int_{-\pi}^{\pi} \frac{I_n(\lambda)}{g(\lambda; \theta)} d\lambda . \quad (15)$$

Kokoszka and Taqqu (1996) study the asymptotic behaviour of (14)/(15), and prove under regularity conditions that the estimator is consistent with rate $(n/\log n)^{1/\alpha}$; the limiting distribution, however, is a complicated linear combination of infinite variance symmetric α -stable variables. Semiparametric estimates of the memory parameter in the presence of infinite variance innovations are discussed also by Peng (2001). A general approach to modelling long-memory time series with finite or infinite variance is discussed by Leipus and Viano (2001), whereas computer generation of infinite variance *ARFIMA* series is considered by Kokoszka and Taqqu (2001).

5. The analysis of nonlinear transformations

For many practical applications, it is of great relevance to investigate the property of a transformed process, and the nature of statistical inference under such circumstances. The results we present below are essentially due to Taqqu (1975), Taqqu (1979) and to Dobrushin and Major (1979), under the assumption of Gaussianity; the non-Gaussian case is more complicated, some results being provided for instance by Surgailis (2000). Assume $\{x_t\}$ is a Gaussian long memory process with zero mean and unit variance, as defined by (1) or (2); we analyze here the behaviour of

$$y_t := g(x_t) - E g(x_t) ,$$

where $g(\cdot)$ is any (measurable) function which satisfies $E g^2(x_t) < \infty$. Of course, y_t is a zero mean, strictly stationary process, which is non-Gaussian unless $g(\cdot)$ is linear: we want to verify what its memory properties are, and how to perform statistical inference.

The crucial tool for this aim are the family of so-called Hermite polynomials, which are defined as

$$H_m(u) = (-1)^m \exp\left(\frac{1}{2}u^2\right) \frac{d^m}{du^m} \exp\left(-\frac{1}{2}u^2\right), m = 1, 2, \dots$$

Straightforward computations show that the first few are given by

$$H_0(u) = 1, H_1(u) = u, H_2(u) = u^2 - 1, H_3(u) = u^3 - 3u, H_4(u) = u^4 - 6u^2 + 3, \dots$$

It can be verified that these polynomials satisfy the recursion

$$H'_m(u) = mH_{m-1}(u), EH_m(X) \equiv 0, \text{ for all } m, X \sim N(0, 1).$$

Hermite polynomials have several properties that make them an extremely valuable tool for inference. In particular, for any zero mean Gaussian variables ε, η , we have

$$EH_p(\varepsilon)H_q(\eta) = p! \{E\varepsilon\eta\}^p \delta_p^q, \delta_p^q = \begin{cases} 1, & \text{for } p = q \\ 0, & \text{for } p \neq q \end{cases}, \quad (16)$$

i.e. they are uncorrelated whenever p is different from q , and otherwise the covariance is (apart from a constant factor) the p -th power of the covariance of arguments. Based upon this, it can be shown that y_t is equal in mean square to

$$y_t = \sum_{k=k_0}^{\infty} \frac{J_k}{k!} H_k(x_t), \quad (17)$$

where the J_k are given by the standard least squares projection coefficients

$$J_k := Ey_t H_k(x_t) = E \{g(x_t) - Eg(x_t)\} H_k(x_t),$$

and k_0 (the *Hermite rank* of $g(\cdot)$) is the order of the first non-zero J_k ; for instance, $k_0 = 1$ for a linear $g(\cdot)$. Heuristically, k_0 is unity for most (but by no means all) odd functions ($g(x) = x, \sin x, x^3, \dots$), and it is equal to two for most (but by no means not all) even functions ($g(x) = \cos x, x^2, \log|x|, \dots$). By using (16), it is then easy to see that

$$\begin{aligned} Ey_0 y_\tau &= E \left[\left\{ \sum_{k=k_0}^{\infty} \frac{J_k}{k!} H_k(x_t) \right\} \left\{ \sum_{k=k_0}^{\infty} \frac{J_k}{k!} H_k(x_0) \right\} \right] = \sum_{k=k_0}^{\infty} \frac{J_k^2}{(k!)^2} EH_k(x_0) H_k(x_t) \\ &= \frac{J_{k_0}^2}{k_0!} \gamma^{k_0}(\tau) + \sum_{k=k_0+1}^{\infty} \frac{J_k^2}{k!} \gamma^k(\tau). \end{aligned} \quad (18)$$

Now in view of (2), we have $\gamma^k(\tau) \sim C\tau^{k(2d-1)}$, for some $C > 0$, as $\tau \rightarrow \infty$; hence the leading term in (18) is clearly $\gamma^{k_0}(\tau) \sim C\tau^{k_0(2d-1)}$, whereas the remainder series can be shown to be of a smaller order. In this sense, y_t can be viewed as a long memory series with parameter $2d' - 1 = k_0(2d - 1)$; more precisely, y_t can be interpreted as of order

$$d' = \begin{cases} k_0 d + (1 - k_0)/2, & \text{if } k_0 d + (1 - k_0)/2 > 0 \\ 0, & \text{if } k_0 d + (1 - k_0)/2 < 0 \end{cases}.$$

It should be noted that

$$d' = \max \left\{ d + (k_0 - 1)(d - \frac{1}{2}), 0 \right\} \leq d$$

always, i.e. transformations of Gaussian processes cannot increase their memory.

Let us now discuss statistical inference. To fix the ideas, we start from the simplest possible problem, i.e. the asymptotic behaviour of the sample mean $\bar{y}_n = n^{-1} \sum_{t=1}^n y_t$. From (17), (18) and algebraic manipulations it is possible to show that the only relevant term for the asymptotic behaviour of $\bar{y}_n$ is

$$\frac{J_{k_0}}{k_0!} \left\{ \frac{1}{n} \sum_{t=1}^n H_{k_0}(x_t) \right\},$$

the remaining terms in the expansion (17) giving no contribution asymptotically due to the smaller order of their variance: this is the *reduction principle* of Taqqu (1975), Taqqu (1979) and Dobrushin and Major (1979). Apart from the deterministic factor $(J_{k_0}/k_0!)$, the limit will therefore be the same as for the sample mean of $H_{k_0}(x_t)$; this is known to converge at rate $n^{d^1-1/2}$, the limiting distribution being Gaussian for $k_0 = 1$, and non-Gaussian otherwise. More precisely, the limiting distribution is equal to the value at 1 of the so-called *Hermite process of order k_0* , whose special case for $k_0 = 1$ is the (Gaussian) *fractional Brownian motion*. It is natural to question how the same sort of ideas can be exploited to investigate the behaviour of likelihood estimates with transformed data. This is the task entertained for instance by Giraitis and Taqqu (1999). These authors investigate the behaviour of the Whittle estimates (3)-(4), in the case where one focusses on $y_t = G(x_t)$, for $G(\cdot)$ a finite polynomial. Their basic argument is again based upon Hermite expansions and a reduction principle argument, as above; the proofs are made much more complicated, however, by the nature of the quadratic forms involved. The results are notable for their diversity across the value of the memory parameter, and the nonlinear functions involved; it emerges that the very neat behaviour described in Section 2 for the Gaussian/linear case is rather special, and indeed non-Gaussian asymptotic distributions can arise, whereas the normalizing factor need not always be $\sqrt{n}$. Extensions to more general (non-polynomials) filters are being developed by the same authors, but they are not available yet in the published literature.

References

- Dahlhaus R. (1989) Efficient parameter estimation for self-similar processes, *Annals of Statistics*, 17, 1749–1766.
- Dobrushin R.L. and Major P.. (1979) Non-central limit theorems for non-linear functionals of gaussian fields, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 50, 27–52.
- Fox R. and Taqqu M. (1985) Non-central limit theorems for quadratic forms in random variables having long-range dependence, *Annals of Probability*, 13, 428–436.
- Fox R. and Taqqu M. (1986) Large sample properties of parameter estimates for strongly dependent stationary gaussian time series, *Annals of Statistics*, 14, 517–532.
- Fox R. and Taqqu M. (1987) Central limit theorems for quadratic forms in random variables having long-range dependence, *Probability Theory and Related Fields*, 74, 213–240.

- Geweke J. and Porter-Hudak S. (1983) The estimation and application of long memory time series models, *Journal of Time Series Analysis*, 4, 221–238.
- Giraitis L. and Robinson P. (2000) Edgeworth expansions for semiparametric whittle estimation of long memory, *preprint*.
- Giraitis L., Robinson P. and 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.
- Giraitis L., Robinson P. and Surgailis D. (1999) Variance-type estimation of long memory, *Stochastic Processes and their Applications*, 80, 1–24.
- Giraitis L. and Surgailis D. (1990) A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of whittle's estimate, *Probability theory and Related Fields*, 86, 87–104.
- Giraitis L. and Taqqu M. (1998) Central limit theorems for quadratic forms with time-domain conditions, *Annals of Probability*, 26, 377–398.
- Giraitis L. and Taqqu M. (1999) Whittle estimator for finite variance non-gaussian time series with long memory, *Annals of Statistics*, 27, 178–203.
- Hall P., Koul H. and Turlach B. (1997) Note on convergence rates of semiparametric estimators of dependence index, *Annals of Statistics*, 25, 1725–1739.
- Hosoya Y. (1997) A limit theory of long range dependence and statistical inference in related models, *Annals of Statistics*, 25, 105–137.
- Kokoszka P. and Taqqu M. (1995) Fractional arima with stable innovations, *Stochastic Processes and their Applications*, 45, 169–182.
- Kokoszka P. and Taqqu M. (1996) Parameter estimation for infinite variance fractional arima, *Annals of Statistics*, 24, 1880–1913.
- Moulines E. and Soulier P. (1999) Broadband log-periodogram regression for time series with long range dependence, *Annals of Statistics*, 27, 1415–1439.
- Peng L. (2001) Semi-parametric estimation of long range dependence index in infinite variance time series, *Statistics and Probability Letters*, 51, 101–109.
- Robinson P. (1995a) Gaussian semiparametric estimation of long range dependence, *Annals of Statistics*, 23, 1630–1661.
- Robinson P. (1995b) Log-periodogram regression of time series with long range dependence, *Annals of Statistics*, 23, 1048–1072.
- Robinson P. (forthcoming) *Time Series with Long Memory*, Oxford University Press.
- Surgailis D. (2000) Long range dependence and appell rank, *Annals of Probability*, 28, 478–497.
- Taqqu M. (1975) Weak convergence to fractional brownian motion and to the rosenblatt process, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte gebiete*, 31, 287–302.
- Taqqu M. (1979) Convergence of integrated processes of arbitrary hermite rank, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 50, 53–83.
- Velasco C. (1999a) Gaussian semiparametric estimation of non-stationary time series, *Journal of Time Series Analysis*, 20, 87–127.
- Velasco C. (1999b) Non-stationary log-periodogram regression, *Journal of Econometrics*, 91, 325–371.
- Velasco C. (2000) Non-gaussian log-periodogram regression, *Econometric Theory*, 16, 44–79.
- Velasco C. and Robinson P. (2000) Whittle pseudo-maximum likelihood estimation for nonstationary time series, *J. Amer. Stat. Assoc.*, 95, 1229–1243.