A Wavelet Based Joint Estimator of the Parameters of Long-Range Dependence.

Darryl VEITCH [(1)] Patrice ABRY [(2)]

(1) - *Software Engineering Research Centre, Level 2, 723 Swanston Street, Carlton, Victoria 3053, - Australia*
tel: (+613) 9282 2440 - Fax: (+1) (+613) 9282 2444
Email : darryl@serc.rmit.edu.au - *Web :* http://www.serc.rmit.edu.au/~darryl

(2) - *CNRS URA 1325 - Laboratoire de Physique - Ecole Normale Supérieure de Lyon - 46, allée d'Italie 69 364 LYON Cedex 07 - France*
tel: (+33) 4 72 72 84 93 - Fax: (+33) 4 72 72 80 80
Email : pabry@physique.ens-lyon.fr - *Web :* http://www.physique.ens-lyon.fr/ts

Abstract

A joint estimator is presented for the two parameters that define the long-range dependence phenomenon in the simplest case. The estimator is based on the coefficients of a discrete wavelet decomposition, improving a recently proposed wavelet-based estimator of the scaling parameter [4], as well as extending it to include the associated power parameter. An important feature is its conceptual and practical simplicity, consisting essentially in measuring the slope and the intercept of a linear fit after a discrete wavelet transform is performed, a very fast ($O(n)$) operation. Under well justified technical idealisations the estimator is shown to be unbiased and of minimum or close to minimum variance for the scale parameter, and asymptotically unbiased and efficient for the second parameter. Through theoretical arguments and numerical simulations it is shown that in practice, even for small data sets, the bias is very small and the variance close to optimal for both parameters. Closed form expressions are given for the covariance matrix of the estimator as a function of data length, and are shown by simulation to be very accurate even when the technical idealisations are not satisfied. Comparisons are made against two maximum likelihood estimators. In terms of robustness and computational cost the wavelet estimator is found to be clearly superior and statistically its performance is comparable. We apply the tool to the analysis of Ethernet teletraffic data, completing an earlier study on the scaling parameter alone.

Keywords: Packet Traffic, Wavelet Decomposition, Time-Scale Analysis, Long Range Dependence, Hurst Parameter, Parameter Estimation, Telecommunications Networks.

I. Introduction

Motivation. The phenomenon of *Long Range Dependence* (LRD) has recently attracted strong interest in telecommunications, with the discovery of self-similar and long range dependent properties in data and communications traffic of diverse types [19] [23] [20]. The investigation of the impact of this on telecommunication network performance (for example [21] [9] [13]) has highlighted the need for accurate and computationally effective estimation methods for LRD parameters. Both real-time estimation from limited measurements, and off-line analysis of enormous data sets, are aspects which call in particular for computationally efficient techniques.

The long range dependence phenomenon. A common definition of long range dependence is the slow, power-law like decrease at large lag of the autocovariance function of a stationary stochastic process $\{x_t\}$ given by $\gamma_x(k) \sim c_\gamma |k|^{-(1-\alpha)}$, $\alpha \in (0,1)$. Equivalently, it can be defined as the power-law divergence at the origin of its spectrum:

$$f_x(\nu) \sim c_f |\nu|^{-\alpha}, \quad |\nu| \to 0, \tag{1}$$

where $f_x(\nu)$ satisfies, in the case of discrete time processes, $\gamma_x(0) = \sigma_x^2 = \int_{-1/2}^{1/2} f_x(\nu)\, d\nu$, σ_x^2 being the variance (or power) of x_t.

Each of these definitions include two parameters: (α, c_γ) or (α, c_f) respectively, which are equivalent as $c_f = 2(2\pi)^{-\alpha} c_\gamma \Gamma(\alpha) \sin((1-\alpha)\pi/2)$, where Γ is the Euler function. In each pair α is the most important, as it defines the existence of the phenomenon itself, and governs the characteristic scaling behaviour of a LRD process as well as statistics derived from it. Consequently, the emphasis has traditionally been on the estimation of α or equivalently of the Hurst parameter $H = (1+\alpha)/2$, together with the range of lags or frequencies where the power-law relation holds. The second parameter, c_γ or c_f, or other equivalent choices, is an independent quantitative parameter with the dimensions of variance which has received far less attention, and its estimation and importance have been largely neglected. This is unfortunate because in applications the second parameter plays a major role in fixing the absolute *size* of LRD generated effects, the general *character* of which is determined by H. Estimating c_γ is therefore an issue of importance for quantitative analysis, and yet is non-trivial, being fraught with the same statistical difficulties intrinsic to H estimation. We emphasize that the second parameter is a independent parameter that must be estimated. It is not related to the variance of $x(t)$ except in a rigid parametric context. This paper reports on a wavelet based approach to the joint estimation of LRD parameters, specifically (α, c_f).

Importance of the second parameter. A powerful example of the importance of the second parameter, valid in any LRD context, is the role it plays in the elementary problem of mean estimation. For LRD processes the classical asymptotic expression σ_x^2/n for the variance of the sample mean with sample size n, is replaced by $\frac{2c_\gamma n^\alpha}{(1+\alpha)\alpha} \cdot \frac{1}{n}$, ([8], p.160). Since the variance is proportional to c_γ, it follows immediately that the confidence intervals about sample mean estimates are essentially proportional to $\sqrt{c_\gamma}$. The value of the second parameter is therefore critical even for the simplest practical estimation issues for LRD processes.

In application areas such as telecommunications the second parameter also plays an important role, over and above the obvious importance of reliable mean estimates. It appears directly in expressions for metrics measuring performance in queues with LRD input, which have recently attracted considerable attention. For example in both [21] and [9] storage models are considered with stationary LRD input, resulting in Weibullian-like tails for the stationary queue content distribution V. More precisely $P(V > v) = O(\exp(-\kappa^2\lambda^{2H}v^{2(1-H)}/2\Sigma^2))$ where λ is related to the utilisation of the system, and $\kappa = H^{-H}(1-H)^{H-1}$ is an almost constant function of $H = (\alpha+1)/2$. The variance of the work $W(t)$ arriving in the interval $[0,t]$ behaves as $\mathrm{Var}\,(W(t)) \sim \Sigma^2 t^{2H}$, and it is not difficult to show that the constant Σ^2 is generically related to the c_γ of the LRD input as $\Sigma^2 = c_\gamma/(H(2H-1))$. Thus, the term in the exponential is inversely proportional to c_γ (or c_f), so an increase in c_f at constant H increased the mass in the tail. Thus increases in c_f increase queueing delays, consistent with the interpretation of the former as a measure of the size of LRD effects.

The above examples illustrate that the availability of a reliable estimate of c_f is crucial in the theoretical and practical study of LRD phenomena.

Choice of the second parameter. To the choices of the second parameter defined above it is natural to add the normalised forms $\underline{c}_\gamma = c_\gamma/\sigma_x^2$ and $\underline{c}_f = c_f/\sigma_x^2$ which characterise the correlation function $\gamma_x(k)/\sigma_x^2$ and its spectral equivalent respectively. At a theoretical level all four of these are equivalent choices for the second long range-dependent parameter. In practice however, where H and σ_x^2 are not known but must be estimated also, each of these is a separate estimation problem. In this paper we choose to concentrate on frequency domain estimation and we consider σ_x^2 to be fixed, as we are concerned with the parameters of long range dependence only. Together these imply c_f as the choice of second parameter. For theoretical reasons we also consider a related quantity, $c_f C$, to be defined in the next section. This quantity also happens to be that considered in the parametric joint estimator described by Wornell [29], enabling a very direct comparison.

There are of course other quantities of interest which are related to c_f and which may be more important to estimate directly in other contexts. One example is the variance σ_ϵ^2 of the innovations of a linear process, which also corresponds to the minimal mean squared prediction error amongst all linear predictors (see Beran [8]). Since $\sigma_\epsilon^2 = \exp\Big(\int_{-1/2}^{1/2} \log f(\nu)\, d\nu\Big)$, this is in general quite a different quantity to c_f, however in the case of the fractional ARIMA process farima0d0 [17] $c_f = \sigma_\epsilon^2 (2\pi)^{-\alpha}$, a relatively simple function. The discrete Whittle estimator offers a joint estimation of $(H, \sigma_\epsilon^2/2\pi)$, so we take the opportunity to compare the wavelet estimator against one based on the discrete Whittle in this context. Another reason to compare against the discrete Whittle estimator is that it has been widely used in the analysis of network traffic.

The wavelet estimator. A semi-parametric wavelet based estimator for the Hurst parameter with excellent statistical, computational, and robustness properties has already been reported in [3], [4] (see also [6]). The joint estimator described here is based on the same approach, where several properties of the wavelet decomposition combine to reduce LRD in the time domain to short range dependence (SRD) in the wavelet representation. Key properties are the band pass nature of the analysing wavelets, the fact that the analysing family of wavelets (and scaling functions) are generated from the change of scale operator which matches the power-law form of LRD spectra, and the fact that the number of vanishing 'moments' of the wavelets can be controlled.

In this paper we improve upon the estimator for the scaling exponent α reported in [3], [4], [6], and extend it to the joint case (α, c_f). Under reasonable additional technical idealisations, we first show that the related joint estimator $(\hat{\alpha}, \widehat{c_f C})$ is unbiased and asymptotically efficient, and give explicit formulae for the covariance matrix and Cramer Rao bound. It is shown that the bound is very close to being attained in all known cases. Based on the results for $(\hat{\alpha}, \widehat{c_f C})$, approximate formulae for the expectation and covariance matrix of $(\hat{\alpha}, \hat{c}_f)$ are presented as well as theoretical arguments indicating that $\hat{c}_f$ is asymptotically unbiased and efficient. Simulation results are presented confirming the accuracy of the approximations, and the virtual lack of bias even for data of moderate length. The covariance of the estimators however is negative and quite large. A preliminary version of this work was presented at Gretsi '97 [27].

Because the discrete wavelet transform can be performed using the Multi-Resolution Analysis algorithm, data can be split into blocks, analysed and recombined, so that memory problems are not encountered in treating data of arbitrary length n. The run time complexity of the estimator is very low, only $O(n)$, making it very suitable for the analysis of very large data sets.

The estimator inherits the robustness advantages of the earlier H estimator, namely insensitivity to the form of the finite dimensional distributions, as well as to the presence of arbitrary smooth deterministic trends [4]. This latter advantage is particularly important in the context of LRD.

Outline. In section 2 a brief background in relevant wavelet theory is given, followed by the definition of the estimators and the derivation of their properties. In section 3 simulation results are presented whilst in section 4 comparisons are made against the performance of the discrete Whittle estimator [8] and the maximum likelihood wavelet based estimator of Wornell [29]. In section 5 the estimator is applied to large Ethernet data sets, complementing the results reported in [4]. In section 6 we conclude the paper with a discussion on outstanding issues. Finally in section 7 we provide a compact summary of the steps involved in the estimation to assist the reader in implementing it.

II. The Wavelet Estimator

A. Foundation of the Method

Discrete Wavelet Decomposition and LRD processes. We denote by $d_x(j,k) = \langle x, \psi_{j,k} \rangle$ the coefficients of the discrete or non-redundant wavelet transform of a process or 'signal' x. The family of wavelet basis functions $\left\{\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k)\right\}$, $j = 1, \ldots, J$, $k \in \mathcal{Z}$ is generated from the *mother wavelet* ψ_0, itself defined via a multiresolution framework [11].

In the analysis of the LRD phenomenon, the following two features, (F1,F2), of such a family play key roles.

F1: The basis is constructed from the dilation (change of scale) operator: $\psi_{j,0}(t) = 2^{-j/2}\psi_0(2^{-j}t)$. This means that the analyzing family exhibits, by contruction, a scale invariance feature. The LRD phenomenon can be understood as the absence of any characteristic frequency (and therefore scale) in the range of frequencies close to the origin. The LRD property can thus be interpreted as a scale invariance characteristic which is efficiently analysed by wavelets.

F2: ψ_0 has a number N of *zero* or *vanishing moments* which can be freely chosen provided $N \geq 1$. By definition this means that $\int t^k \psi_0(t) dt \equiv 0$, $k = 0, \ldots, N-1$ (but not for $k \geq N$), or equivalently that the Fourier Transform of ψ_0 satisfies $|\Psi_0(\nu)| = O(|\nu|^N)$ at the origin. This property can be used to control divergences arising with processes having power-law spectra at the origin.

For a process with a power-law spectrum such as a LRD process, these features engender the following key properties of the wavelet coefficients $d_x(j,k)$ over a range of scales 2^j, $j = j_1 \ldots j_2$ where the power-law scaling holds (see [3], [4], [6] for more details).

P1: Due to F1, the scale invariance (the power-law behaviour) is captured exactly:

$$\mathbb{E} d_x(j,\cdot)^2 = 2^{j\alpha} c_f C \tag{2}$$

where

$$C = \int |\nu|^{-\alpha} |\Psi_0(\nu)|^2 d\nu. \tag{3}$$

This exact recovery of a power-law is not a trivial effect and results directly from the dilation operator underlying the design of the wavelet basis. (Time-frequency or periodogram based estimates would not exhibit such a feature [1], [3].)

P2: Due to F1 and F2, the $d_x(j,k)$ are a collection of random variables which are *quasi-decorrelated* [15]. In particular, the long-range dependence present in the time domain representation is completely absent in the wavelet coefficient plane $\{j,k\}$.

Property P2 deserves elaboration. It has been shown that correlations in the time-scale plane decay at least hyperbolically in all directions [15], [26] with exponents controlled by the number of vanishing moments and corresponding to short range dependence. Since by definition the *octave* $j = \log_2(\text{scale})$, this implies *exponential* decay in octave j.

In this paper $\log_2$ will denote the logarithm base 2, whereas ln will denote natural logarithms.

Intuitive basis of the estimator. The starting point for the analysis is equation (2). Rewritting it as $\log_2(\mathbb{E} d_x(j,\cdot)^2) = j\alpha + \log_2(c_f C)$ strongly suggests a linear regression approach for estimating (α, c_f), where clearly the slope of the regression would estimate α and the intercept would be related to c_f. This idea of using a log-log plot is an obvious one, and common to many contexts when an exponent is the object of interest. The real issue is to what extent the promise of this simple linear form is realised in the resulting estimator, once the inevitable complications are taken into account.

• The first, essential, complication is of course that $\mathbb{E}d_x(j,\cdot)^2$, a second order quantity that can be related to the spectrum of x, is not known but must be estimated. In the present context this is the principal difficulty as it is well known [8] that the estimation of second order (and other) quantities in a long range dependent context is a delicate task. Here however property P2, the quasi-decorrelation of the $d_x(j,k)$, allows us to effectively use the simple "time average"

$$\mu_j = \frac{1}{n_j}\sum_{k=1}^{n_j} d_x^2(j,k), \tag{4}$$

where n_j is the number of coefficients at octave j available to be analysed. This quantity is an unbiased and efficient estimator of $\mathbb{E}d_x(j,\cdot)^2$ [3], [4]. (Note that μ_j is the sample variance of $d_x(j,\cdot)$, since from F2 the expectation of $d_x(j,\cdot)$ is identically zero for each j. This interesting fact plays no direct role in our analysis however.)

• The second complication is the non-linearity introduced by the $\log_2$, which biases the estimation. We will see below how this problem also can be circumvented under reasonable hypotheses. Simplying things slightly, we confirm that the fundamental approach underlying our estimator is indeed a linear regression of $\log_2(\mu_j)$ on $\log_2(2^j) = j$. A weighted linear regression will be used as the variances of the $\log_2(\mu_j)$ vary with j.

Improvement on the previous wavelet-based estimator of α. As stated above, a semi-parametric wavelet based estimator for α has recently been proposed in [4]. Here a more detailed analysis of the statistics of $\log_2(\mu_j)$ allows us to improve the estimator in two ways. First the bias due to the logarithm discussed in the previous paragraph is corrected for explicitly. Second, explicit closed form solutions for the variances of the $\log_2(\mu_j)$ have been obtained, and are used in the weighted regression, rather than asymptotic formulae. These refinements result in an estimator which is strictly unbiased, even for data of small length, rather than only asymptotically unbiased, and which is of lower variance. Although theoretically important, in practice these improvements are small to negligible in most cases. They are appreciable when the number of available wavelet coefficients n_j for each octave in the scaling range, is small. In contrast however, the above refinements have important implications for the estimation of the second parameter c_f.

B. Analytic Elements

Linear regression. We recall some standard results on one dimensional weighted linear regression of the random variables y_j on the deterministic independent variables x_j, $j = j_1 \ldots j_2$. Sums will always be taken over this range.

The fundamental hypothesis of linear regression is $\mathbb{E}y_j = bx_j + a$. Define the quantities $S = \sum 1/\sigma_j^2$, $S_x = \sum x_j/\sigma_j^2$ and $S_{xx} = \sum x_j^2/\sigma_j^2$, where σ_j^2 is an arbitrary weight associated with y_j. The usual unbiased estimator $(\hat{b},\hat{a})$ of (b,a) is

$$\hat{b} = \frac{\sum y_j(Sx_j - S_x)/\sigma_j^2}{SS_{xx} - S_x^2} \equiv \sum w_j y_j, \tag{5}$$

$$\hat{a} = \frac{\sum y_j(S_{xx} - S_x x_j)/\sigma_j^2}{SS_{xx} - S_x^2} \equiv \sum v_j y_j, \tag{6}$$

where the weights w_j et v_j satisfy $\sum w_j = \sum j v_j = 0$, $\sum j w_j = \sum v_j = 1$. Note that these conditions imply that there are always both positive and negative v_j and w_j.

If in addition the y_j are mutually independent then the covariance matrix is given by

$$\begin{aligned} \mathrm{Var}(\hat{b}) &= \sum \sigma_j^2 w_j^2 = \frac{S}{SS_{xx} - S_x^2}, &(7)\\ \mathrm{Var}(\hat{a}) &= \sum \sigma_j^2 v_j{}^2 = \frac{S_{xx}}{SS_{xx} - S_x^2}, &(8)\\ \mathrm{Cov}(\hat{a}, \hat{b}) &= \sum \sigma_j^2 w_j v_j = \frac{-S_x}{SS_{xx} - S_x^2}, &(9)\\ r &= -S_x / \sqrt{SS_{xx}}, &(10) \end{aligned}$$

where r is the correlation coefficient. If $x_j \geq 0$ for each j it is easy to see that r will be negative, and large in magnitude if x_1 is large, as a small change in the slope "to the right" will result in an amplified change of opposite sign in the intercept.

Finally if we set $\sigma_j^2 = \mathrm{Var}(y_j)$, then $(\hat{b}, \hat{a})$ is the minimum variance unbiased estimator (MVUE) [18] with covariance matrix as above.

Note that in the event of small errors in the values of the σ_j^2 and small correlations between the y_j, the estimator remains unbiased and its covariance matrix can be accurately estimated by the expressions just given.

The statistic $\log_2(\mu_j)$. Thus far we have indicated that $\log_2(\mu_j)$ is the variable y_j of the desired linear regression satisfying $\mathbb{E}y_j = bj + a$. Since $\mathbb{E}\log_2(\mu_j) \neq \log_2(\mathbb{E}\mu_j) = j\alpha + \log_2(c_f C)$ in general, this cannot be exactly true, although under the conditions ID1–ID3 below, and also assuming n_j large, it can be established that $\log_2(\mu_j) \overset{\mathrm{d}}{\sim} N\Big(j\alpha + \log_2 c_f C, \frac{2^{j+1}}{n \ln^2 2}\Big)$ [3], [4], where $\overset{\mathrm{d}}{\sim}$ signifies equality in distribution and $N(\mu, \sigma^2)$ is a Gaussian random variable. In a LRD context however, the large scales are usually the most important to consider, and it is precisely there that the n_j are *not* large. Here we remove the condition on n_j by examining the distribution of $\log_2(\mu_j)$ in more detail.

Throughout the analysis it is instructive to bear in mind that the number of available detail coefficients n_j essentially decreases by half as the scale is doubled, that is $n_{j+1} \approx n_j/2$, and therefore that $n_j \approx n2^{-j}$ where n is the length of the initial data.

We assume that the following supplementary idealisations hold true.

ID1: The process x, and hence the processes $d_x(j, \cdot)$, are Gaussian.

ID2: For fixed j the process $d_x(j, \cdot)$ is iid.

ID3: The processes $d_x(j, \cdot)$ and $d_x(j', \cdot)$, $j \neq j'$, are independent.

Hypothesis ID1 is justified by numerical evidence which shows that the method is very insensitive to the form of the marginal distributions of x [4]. Hypotheses ID2 and ID3 are both well justified by property P2 (they are separated to make it clearer which properties are needed where).

These extra conditions, whilst appearing very restrictive at first glance, are in fact very reasonable in practical terms, as borne out in simulations. The reason for this is that the underlying effectiveness of the method is based on P1 and P2, ID1–ID3 being added only to extend the quantitative analysis.

Let the density of a Chi-squared variate $X_\nu \overset{\mathrm{d}}{\sim} \chi^2_\nu$ be denoted by $f_\nu(x) = \Big(\frac{1}{2^{\nu/2}\Gamma(\nu/2)}\Big) x^{\nu/2-1} e^{-x/2}$. The mean and variance of such a variate are ν and 2ν respectively. Also set $z_j = 2^{j\alpha} c_f C$.

From ID1 and ID2 and equations 2 and 4 we have

$$\mu_j \overset{\mathrm{d}}{\sim} \frac{z_j}{n_j} X_{n_j}, \tag{11}$$

where $\mathbb{E}\mu_j = z_j$ as μ_j is unbiased, and therefore

$$\begin{aligned}\log_2(\mu_j) &\overset{\mathrm{d}}{\sim} \log_2 z_j - \log_2 n_j + \log_2 X_{n_j} \\ &\overset{\mathrm{d}}{\sim} j\alpha + \log_2 c_f C - \log_2 n_j + \ln X_{n_j} \,/\, \ln 2. \qquad (12)\end{aligned}$$

Thus the study of $\log_2(\mu_j)$ reduces to that of the logarithm of a Chi-squared variable. Using the relations $\int_0^\infty x^{\nu-1} e^{-\mu x} \ln x \, dx = \frac{1}{\mu^\nu}\Gamma(\nu)[\psi(\nu) - \ln\mu]$, $\mathrm{Re}\,\mu > 0$, $\mathrm{Re}\,\nu > 0$ ([16], equation 1,§4.352), and $\int_0^\infty x^{\nu-1} e^{-\mu x} (\ln x)^2 \, dx = \frac{1}{\mu^\nu}\Gamma(\nu)[(\psi(\nu) - \ln\mu)^2 + \zeta(2,\nu)]$, $\mathrm{Re}\,\mu > 0$, $\mathrm{Re}\,\nu > 0$ ([16] equation 2,§4.358), where $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the Psi function and $\zeta(z,\nu)$ a generalised Riemann Zeta function, it is straightforward to show from the definition of $f_\nu(x)$ above that

$$\mathbb{E}\ln X_\nu = \psi(\nu/2) + \ln 2, \qquad (13)$$
$$\mathrm{Var}\,(\ln X_\nu) = \zeta(2, \nu/2). \qquad (14)$$

It follows that

$$\mathbb{E}\log_2(\mu_j) = j\alpha + \log_2 c_f C + g_j, \qquad (15)$$
$$\mathrm{Var}\,(\log_2(\mu_j)) = \zeta(2, n_j/2)/\ln^2 2, \qquad (16)$$

where the term

$$g_j = \psi(n_j/2)/\ln 2 - \log_2(n_j/2), \qquad (17)$$

a negative function of n_j only, can be easily calculated for all values of n_j.

For future reference we record here the asymptotic form for n_j large of the quantities above:

$$g_j \sim \frac{-1}{n_j \ln 2} \qquad (18)$$
$$\mathrm{Var}\,(\log_2(\mu_j)) \sim \frac{2}{n_j \ln^2 2} \qquad (19)$$

C. Definition of the Estimator(s)

In this section we define the LRD estimator $(\hat{\alpha}, \hat{c}_f)$, and another, related estimator, $(\hat{\alpha}, \widehat{c_f C})$. We are interested in the quantity $\widehat{c_f C}$ because, like $\hat{\alpha}$, its statistical properties are entirely independent of the specific form of the mother wavelet, depending only on the general properties P1 and P2. It is nonetheless very closely related to $\hat{c}_f$, yet has the advantage of being amenable to a detailed analysis. The study of $\hat{c}_f$ is rendered far more complex because of the wavelet dependence, which enters explicitly via the integral C of equation 3.

• Define the variables y_j as

$$y_j \equiv \log_2(\mu_j) - g_j. \qquad (20)$$

From the above discussion it is clear that under ID1 and ID2 they obey

$$\mathbb{E}y_j = j\alpha + \log_2 c_f C, \qquad (21)$$
$$\mathrm{Var}\,(y_j) = \zeta(2, n_j/2)/\ln^2 2, \qquad (22)$$

and thus satisfy the requirements for a weighted linear regression.

Perform then a weighted regression estimation $(\hat{b}, \hat{a})$ of y_j on $j = x_j$ according to equations 5–6, with $\sigma_j^2 = \mathrm{Var}\,(y_j)$. This notation will be fixed in the sequel.

The joint **unbiased** estimator $(\hat{\alpha}, \widehat{c_f C})$ is then given by

$$\begin{aligned} \hat{\alpha} &= \hat{b}, && (23) \\ \widehat{c_f C} &= p \cdot 2^{\hat{a}} && (24) \end{aligned}$$

where

$$p = \prod \frac{\Gamma(n_j/2) n_j^{v_j}}{\Gamma(v_j + n_j/2)\, 2^{v_j(1-g_j)}} = \prod \frac{\Gamma(n_j/2)\exp(\psi(n_j/2)v_j)}{\Gamma(v_j + n_j/2)}$$

is a bias correcting factor.

• Define the estimator $\hat{C}$ of the integral $C(\alpha, \Psi_0) = \int |\nu|^{-\alpha} |\Psi_0(\nu)|^2 d\nu$ as

$$\hat{C} = \begin{cases} C(0, \Psi_0) & \hat{\alpha} \le 0 \\ C(\hat{\alpha}, \Psi_0) & 0 < \hat{\alpha} < 1 \\ C(1, \Psi_0) & \hat{\alpha} \ge 1 \end{cases} \tag{25}$$

• We can now define $(\hat{\alpha}, \hat{c}_f)$ as

$$\begin{aligned} \hat{\alpha} &= \hat{b}, && (26) \\ \hat{c}_f &= \widehat{c_f C} / \hat{C}. && (27) \end{aligned}$$

An example of the regression fit is given in the log-log plot in figure 1, which we term the *Logscale Diagram*. The 95% confidence intervals for each y_j, shown as vertical lines at each octave j, are seen to increase with j. This can be understood from equation 19, remembering that $n_j \approx n 2^{-j}$. The intervals are derived from the variances σ_j^2 under gaussian assumptions.

D. Summary of the principal statistical properties

The main properties of the estimators defined above, under ID1-ID3, are summarized here. Proofs and further details, including explicit expressions for the covariance matrices, are given in the subsections to follow.

• $(\hat{\alpha}, \widehat{c_f C})$ is **unbiased** even for data of finite size. (ID3 is not needed in the case of $\hat{\alpha}$). For $\widehat{c_f C}$ the result holds only for sets of octaves $\{j\}$ such that

$$2v_j / n_j > -1$$

is satisfied for each j within it, where v_j are the regression coefficients given by equation 8 (typically this set will be of the form (j_1, j_2) for some j_1, j_2, as generally either all octaves satisfy the condition, or all but a small contiguous collection of the largest octaves). The estimation of $\hat{\alpha}$ has no such restriction. The estimator is efficient, non-diagonal with negative correlation, and attains the Cramer-Rao lower bound in the limit $v_j / n_j \to 0$ for each j selected.

• $\hat{C}$ is asymptotically unbiased, and efficient.

• $\hat{c}_f$ is asymptotically unbiased, and efficient. The same conditions on j hold as for $\widehat{c_f C}$. The properties of $\hat{c}_f$ are closely related to those of $\widehat{c_f C}$, for example the correlation coefficient of $(\hat{\alpha}, \hat{c}_f)$ is also negative and large in magnitude.

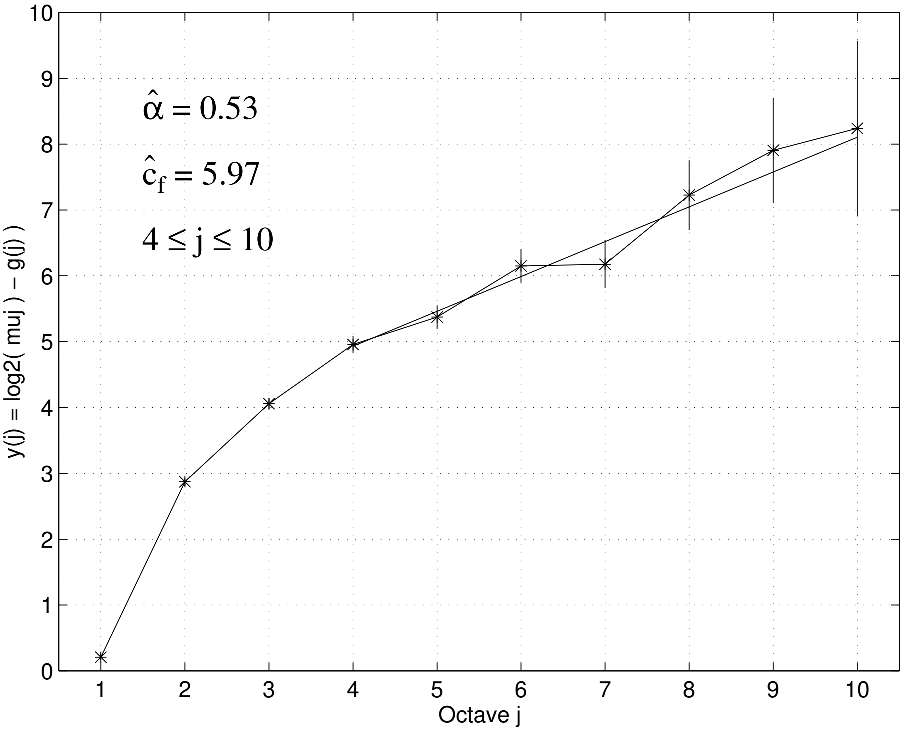

Fig. 1. **Linear regression.** An example of the $y_j = \log_2(\mu_j) - g_j$ against j plot and regression line for a LRD process with strong SRD. The vertical bars at each octave give 95% confidence intervals for the y_j. The series is simulated farima(0,d,2) with $d = 0.25$ ($\alpha = 0.50$) and $\Psi = [-2, -1]$ implying $c_f = 6.38$. Selecting $(j_1, j_2) = (4, 10)$ identifies the relevant scaling range allowing an accurate estimation despite the strong SRD: $\hat{\alpha} = 0.53 \pm 0.07$, $\hat{c}_f = 6.0$ with $4.5 < \hat{c}_f < 7.8$.

E. Statistical performance of $(\hat{\alpha}, \widehat{\mathbf{c_f C}})$

Throughout the remainder of this section we assume the supplementary idealisations ID1–ID3. As usual, all sums and products over j are over the range (j_1, j_2). The basic result underlying the close to optimal performance of the estimator is the optimality of the linear regression $(\hat{b}, \hat{a})$, which can be exploited thanks to the choice of y_j given by equation 20, and the fact that the variances σ_j^2 of y_j are known.

Bias. Since for both $(\hat{\alpha}, \widehat{c_f C})$ and $(\hat{\alpha}, \hat{c}_f)$ the estimator for α is just $\hat{b}$, by construction we have:

$$\mathbb{E}\hat{\alpha} = \alpha\,.$$

Since $\hat{a}$ estimates $\log_2(\widehat{c_f C})$ without bias, it is natural to suggest $2^{\hat{a}}$ as an estimator for $\widehat{c_f C}$, although bias will be introduced by the non-linearity. We claim that the form of the factor p in the definition of $\widehat{c_f C}$ corrects for this bias whilst simultaneously improving the variance.

To show this, first rewrite $\hat{a}$ as

$$\hat{a} \;=\; \sum v_j y_j = \sum v_j (\log_2(\mu_j) - g_j) = \log_2 \prod \mu_j^{v_j} - \sum v_j g_j .$$

Thus, recalling that X_{n_j} is a Chi-squared variable with n_j degrees of freedom,

$$\begin{aligned} \widehat{c_f C} = p\,2^{\hat{a}} \;&=\; p \cdot 2^{-\sum v_j g_j} \prod \mu_j^{v_j} \\ &\stackrel{\mathrm{d}}{\sim}\; p \cdot \prod 2^{-v_j g_j} \prod \left(\frac{z_j}{n_j}\right)^{v_j} \prod X_{n_j}^{v_j} \end{aligned}$$

$$
\begin{aligned}
&= \frac{p \cdot \prod n_j^{v_j}}{\prod 2^{v_j} \exp(\psi(n_j/2)v_j)} \cdot \frac{c_f C}{\prod n_j^{v_j}} \prod X_{n_j}^{v_j} \\
&= c_f C \cdot \prod \frac{\Gamma(n_j/2)}{2^{v_j}\Gamma(v_j + n_j/2)} \prod X_{n_j}^{v_j}, \qquad (28)
\end{aligned}
$$

where the definition of p has been used in the last line, and the relations $\sum v_j = 1$, $\sum j v_j = 0$ have been used to show that $\prod z_j^{v_j} = c_f C$.

Since under ID3 the μ_j are mutually independent, the expectation of $\widehat{c_f C}$ reduces essentially to the product of the expectations of real valued powers of Chi-squared variables. The key result is

$$
\begin{aligned}
\mathbb{E} X_\nu^d &= \int_0^\infty x^d f_\nu(x) dx \\
&= \frac{2^d \Gamma(d + \nu/2))}{\Gamma(\nu/2)} \int_0^\infty f_{2d+\nu}(x) dx \\
&= \frac{2^d \Gamma(d + \nu/2))}{\Gamma(\nu/2)} \qquad (29)
\end{aligned}
$$

provided that $(2d + \nu) > 0$. It is now immediately clear from equation 28 that $\widehat{c_f C}$ is unbiased, provided that $2v_j/n_j > -1$ for each j. If this condition fails then the expectation of $\widehat{c_f C}$ is infinite! This reveals an unexpected but very important aspect of $\widehat{c_f C}$, namely that those scales where $2v_j/n_j \leq -1$ must not be included in the analysis. The intuitive reason behind this is that the estimator involves a raising to a power, which can blow up if there is too much variance in the intercept $\hat{a}$ (equation 8). The scales at risk are those with the largest j value, where σ_j^2 is the largest. It is difficult however to derive an exact formula for those octaves which satisfy the condition, so in practice each octave needs to be tested for separately. Numerical evidence strongly suggests however that the 'safe' set of octaves take the form (j_1, j_2) for some j_1, j_2, and that problems occur only when a small number of octaves are included, each with a small n_j. If the test fails, a new set of octaves must be chosen and the regression recalculated.

Variance. The variance of $\hat{\alpha}$ is that of $\hat{b}$ and is therefore, using ID3, given by using equation 7 with equation 22.

Now consider the variance of $\widehat{c_f C}$. From equation 28 and noting that letting $d \mapsto 2d$ in equation 29 yields the second moment of X_ν^d, again using ID3 we have

$$
\begin{aligned}
\mathrm{Var}\,(\widehat{c_f C}) &= (c_f C)^2 \left(\prod \frac{\Gamma(n_j/2)}{2^{v_j}\Gamma(v_j + n_j/2)} \right)^2 \left[\mathbb{E}(\prod X_{n_j}^{v_j})^2 - (\mathbb{E} \prod X_{n_j}^{v_j})^2 \right] \\
&= (c_f C)^2 \left(\prod \frac{\Gamma(n_j/2)}{2^{v_j}\Gamma(v_j + n_j/2)} \right)^2 \left[\prod \mathbb{E}(X_{n_j}^{2v_j}) - (\prod \mathbb{E} X_{n_j}^{v_j})^2 \right] \\
&= (c_f C)^2 \left[\left(\prod \frac{\Gamma(n_j/2)}{2^{v_j}\Gamma(v_j + n_j/2)} \right)^2 \prod \frac{2^{2v_j}\Gamma(2v_j + n_j/2)}{\Gamma(n_j/2)} - 1 \right] \\
&= (c_f C)^2 \left[\prod \frac{\Gamma(2v_j + n_j/2)\Gamma(n_j/2)}{\Gamma(v_j + n_j/2)^2} - 1 \right] \qquad (30)
\end{aligned}
$$

provided that $4v_j/n_j > -1$ which must be satisfied for each j to ensure the existence of the variance (implying that of the expectation).

To see that the variance is reduced by the introduction of the bias correction factor p, recall the expression $p = \prod \frac{\Gamma(n_j/2) n_j^{v_j}}{\Gamma(v_j+n_j/2)\, 2^{v_j(1-g_j)}}$ and consider the two regimes of v_j/n_j small and v_j/n_j 'large', that is $4v_j/n_j \approx -1$. In the former case it is not difficult to show from the asymptotic relations given in appendix 2 that p tends to one, which has neutral implications with respect to the variance. In the other case we now show that p can be small, leading to a reduction of variance. Combine the power terms in the product as $n_j^{v_j}/2^{v_j(1-g_j)} = 2^{v_j[\log_2(n_j)-(1-g_j)]}$, and assume that $n_j > 2$, as $n_j = 1, 2$ are special cases that pose no difficulties. If $4v_j/n_j \approx -1$ then $v_j < 0$ and since it can be shown that $\log_2(n_j) - (1 - g_j)$ is positive for $n_j > 2$, it follows that p is bounded from above by $\prod \frac{\Gamma(n_j/2)}{\Gamma(v_j+n_j/2)}$. Clearly if $4v_j/n_j \approx -1$ then $v_j + n_j/2 \approx -v_j$ which can be smaller than 1 (and often is in practice), in which case the corresponding factor in the bound and thus p will also be small as $\Gamma(\cdot)$ diverges at the origin.

Covariance of $(\hat{\alpha}, \widehat{c_f C})$**.** The calculation of the covariance is slightly more involved. First rewrite $\hat{b}$ as

$$\begin{aligned}
\hat{b} = \sum w_j y_j &= \sum w_j(\log_2(\mu_j) - g_j) \\
&\overset{\mathrm{d}}{\sim} \sum w_j \Big(\log_2 \frac{z_j}{n_j} X_{n_j} - g_j\Big) \\
&= \alpha + \sum w_j \log_2 X_{n_j} - \sum w_j (g_j + \log_2 n_j),
\end{aligned} \tag{31}$$

where we have used $\sum w_j = 0$, $\sum j w_j = 1$ in showing that $\sum w_j \log_2 z_j = \alpha$. The other main result needed,

$$\mathbb{E} X_\nu^d \ln X_\nu = \mathbb{E} X_\nu^d \,\mathbb{E} \ln X_{2d+\nu} = \mathbb{E} X_\nu^d \,[\psi(d + \nu/2) + \ln 2],$$

is easily established using equations 29 and 13 and the discussion above the latter. Putting these results together and recalling that $\log_2 x = \ln x / \ln 2$ we have

$$\begin{aligned}
&\mathrm{Cov}\,(\hat{\alpha}, \widehat{c_f C}) \\
= \;& \mathbb{E}\Big((\widehat{c_f C} - c_f C)(\hat{\alpha} - \alpha)\Big) = \mathbb{E}\Big(\widehat{c_f C}\,(\hat{\alpha} - \alpha)\Big) \\
= \;& \mathbb{E}\Big(c_f C \,\big(\mathbb{E} \prod X_{n_j}^{v_j}\big)^{-1} \prod X_{n_j}^{v_j} \cdot \big(\sum w_k \log_2 X_{n_k} - \sum w_k (g_k + \log_2 n_k)\big)\Big) \\
= \;& c_f C \left[\big(\mathbb{E} \prod X_{n_j}^{v_j}\big)^{-1} \mathbb{E}\left(\prod_j X^{v_j} \sum_k w_k \log_2 X_{n_k}\right) - \sum w_k (g_k + \log_2 n_k)\right] \\
= \;& c_f C \left[\big(\mathbb{E} \prod X_{n_j}^{v_j}\big)^{-1} \mathbb{E}\Big(\sum_k w_k \prod_{j \neq k} X_{n_j}^{v_j} (X_{n_k}^{v_k} \log_2 X_{n_k})\Big) - \sum w_k (g_k + \log_2 n_k)\right] \\
= \;& c_f C \left[\big(\mathbb{E} \prod X_{n_j}^{v_j}\big)^{-1} \sum_k w_k \prod_{j \neq k} \mathbb{E}(X_{n_j}^{v_j}) \mathbb{E}(X_{n_k}^{v_k} \log_2 X_{n_k}) - \sum w_k (\psi(n_k/2) + 1)/\ln 2\right] \\
= \;& c_f C \cdot \frac{1}{\ln 2} \sum_j w_j \Big(\psi(v_j + n_j/2) - \psi(n_j/2)\Big).
\end{aligned} \tag{32}$$

F. The Cramer-Rao Bound and asymptotics of $(\hat{\alpha}, \widehat{c_f C})$

The Cramer-Rao Lower Bound. The bound on the covariance matrix of an unbiased estimator of $(\alpha, c_f C)$ under ID1–ID3 is given by equation 46 of appendix 1:

$$\mathrm{Var}(\hat{\alpha}, \widehat{c_f C}) \geq \frac{2}{(I_2 I_0 - I_1^2)} \begin{bmatrix} I_0 & -(c_f C)\, I_1 \\ -(c_f C)\, I_1 & (c_f C)^2\, I_2 \end{bmatrix}, \tag{33}$$

where $I_k = \ln^k 2 \sum_j j^k n_j, \quad k = 0, 1, 2.$

The limit $\mathbf{v_j/n_j \to 0}$. In order to compare this lower bound with the covariance matrix for $(\hat{\alpha}, \widehat{c_f C})$ whose elements are given above, it is useful to consider the limit of n_j large in comparison to v_j. This limit, studied in [3], [4], [6] in the context of α only, corresponds to the limiting case of Gaussian behaviour for y_j. The limit is interesting because in practice the assumption that n_j is large is a very good one except, possibly, for the n_j corresponding to the largest j. In appendix 2 we show that in this limit the covariance matrix is given by

$$\mathrm{Var}(\hat{\alpha}, \widehat{c_f C}) = \begin{bmatrix} \mathrm{Var}(\hat{b}) & (c_f C)\ln 2 \cdot \mathrm{Cov}(\hat{b}, \hat{a}) \\ (c_f C)\ln 2 \cdot \mathrm{Cov}(\hat{b}, \hat{a}) & (c_f C)^2 \ln^2 2 \cdot \mathrm{Var}(\hat{a}) \end{bmatrix}.$$

Note that the correlation coefficient corresponding to the limiting covariance is equal to that of $(\hat{b}, \hat{a})$.

Observe that, in view of equation 19, for large n_j $I_0 = 2S/\ln^2 2$, $I_1 = 2S_x/\ln 2$ and $I_2 = 2S_{xx}$, where the S, S_x and S_{xx} are defined in section 2.2. Using equations 7–9, the limiting covariance matrix can therefore be rewritten in terms of the I_k and turns out to be just the right hand side of equation 33. Thus, in the limit where $v_j/n_j \to 0$ and large n_j, $(\hat{\alpha}, \widehat{c_f C})$ attains the bound and is therefore the **MVUE** under ID1–ID3.

Moreover, in the fully asymptotic case of $n \to \infty$, where we let $j_2 \to \infty$ at a rate such that $v_j/n_j \to 0$ for each j in $j_1 \ldots j_2$, j_1 fixed, an explicit form for the bound can be found, given in equation 47 from appendix 1, and is seen to be non diagonal (equation 48). This asymptotic form with $j_1 = 1$ is identical to that obtained in [29], indicating that asymptotically, if all the octaves in the data can be (and are) used, our estimator matches the performance of a maximum likelihood estimator. More details are given in section 4.

Finally, away from this limit the Cramer Rao bound is clearly not attained, and the question of whether $(\hat{\alpha}, \widehat{c_f C})$ remains the MVUE or not is therefore an open one. We can nonetheless compare the covariance matrix and the CRLB. From figure 2 it can be seen that the covariance of $(\hat{\alpha}, \widehat{c_f C})$ is very close to the bound even when the conditions of the limit are not satisfied. Hence even if $(\hat{\alpha}, \widehat{c_f C})$ is not the MVUE in general, its performance is very close to it.

G. Statistical Performance of $(\hat{\alpha}, \hat{\mathbf{c}}_\mathbf{f})$

Despite the complex, highly non-linear nature of $(\hat{\alpha}, \widehat{c_f C})$, due to the special structure of the problem we were able to determine the expectations and variances exactly in terms of known special functions. For $(\hat{\alpha}, \hat{c}_f) = (\hat{\alpha}, \widehat{c_f C}/\hat{C})$ this cannot be done even when using, as we do, the approximation $C1(\alpha)$ for $C(\alpha, \Psi_0)$ (equation 34 below) to bypass the issue of wavelet-basis dependence. Because of the small variation of $C1(\alpha)$ over the range $\alpha \in [0, 1]$ however, and the lack of bias and low variance of $\hat{\alpha}$, we expect $\hat{c}_f$ to have properties which essentially mirror those of $\widehat{c_f C}$. We now derive expressions which support this expectation and which are borne out in simulations.

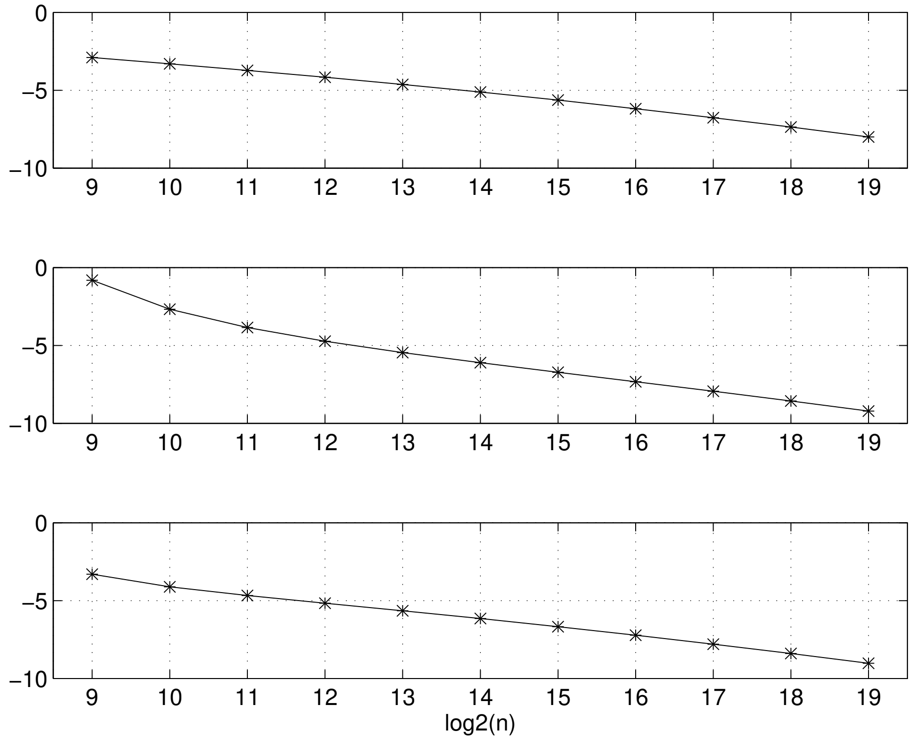

Fig. 2. **Variance against CRLB.** $\log_2(|\mathrm{Cov}_{ij} - \mathrm{CRLB}_{ij}|/\mathrm{CRLB}_{ij})$ as a function of $\log_2(n)$ with $j_1 = 3$ for, from top to bottom, $\mathrm{Var}(\hat{\alpha})$, $\mathrm{Var}(\widehat{c_f C})$ and $\mathrm{Cov}(\hat{\alpha}, \widehat{c_f C})$. It is clearly seen that the elements of the covariance matrix rapidly converge (as $\sim n^{-1/2}$ for $\mathrm{Var}(\hat{\alpha})$ and $\mathrm{Cov}(\hat{\alpha}, \widehat{c_f C})$ and as $\sim n^{-2/3}$ for $\mathrm{Var}(\widehat{c_f C})$) to those of the Cramer Rao Lower Bound. The actual variance must therefore be close to that of the MVUE, even for small size data.

Approximating the integral C. The integral $C(\alpha, \Psi_0)$ is a function of the (Fourier transform of the) mother wavelet and α only, and can be accurately and rapidly calculated via the multiresolution algorithm, provided α is known. In order to calculate the statistical properties of the estimate $\hat{C} = C(\hat{\alpha}, \Psi_0)$ however, we need a simplifying approximation for $C(\alpha, \Psi_0)$. It can be proven from a theorem in [7] that, for interpolating or orthonormal families of wavelet (like the Daubechies's that we generally use here), when the number of vanishing moments N is large one has $|\Psi_0(\nu)|^2 \to \mathbf{1}_{[\mathbf{1/2},\mathbf{1}]}$, where $\mathbf{1}$ is the indicator function. This convergence implies a convergence of the integral since the behaviour at the origin is well behaved for $N \geq 1$. Setting $|\Psi_0(\nu)|^2 = \mathbf{1}_{[\mathbf{1/2},\mathbf{1}]}$ and integrating, we obtain the following approximation $C1(\alpha)$ for $C(\alpha, \Psi_0)$, N large

$$C1(\alpha) = \frac{2 - 2^{\alpha}}{1 - \alpha}, \tag{34}$$

a positive, monotone increasing function of $\alpha \in I\!\!R$ (the singularity at $\alpha = 1$ is removable). It varies so little over the domain of interest, mapping $[0, 1]$ to $[1, 2\ln 2]$, that it could be well approximated in its turn by the affine function $1 + (2\ln 2 - 1)\alpha$. This observation gives insight into the distribution of $\hat{C}$, although it is not useful for the study of $\hat{c}_f$. As can be seen in figure 3 the approximation $C1(\alpha)$ is reasonable even for $N = 1$ (note the scale) and rapidly becomes very accurate.

Approximating the statistical performance. By using $C1(\alpha)$ to replace $C(\alpha, \Psi_0)$ we bypass several difficulties in the approximation of the properties of $(\hat{\alpha}, \hat{c}_f)$, however there are others. Although we expect α to lie in the range $[0, 1]$, the estimator $\hat{\alpha}$ takes real values. Hence the integral $\hat{C} = C(\hat{\alpha}, \Psi_0)$ is not necessarily defined for a given fixed

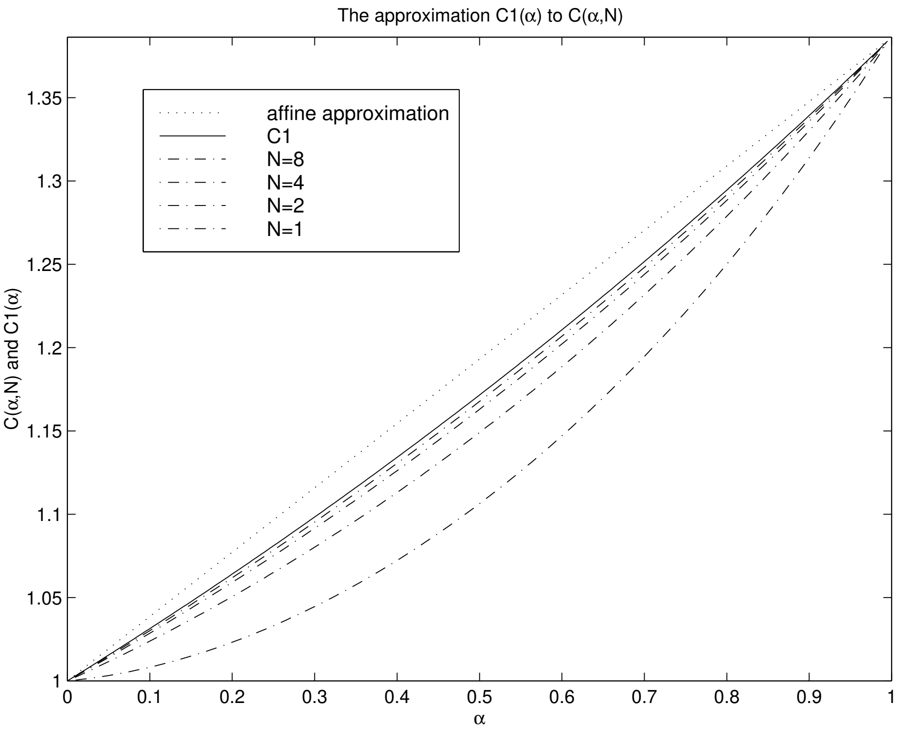

Fig. 3. **Accuracy of** $C1(\alpha)$**.** It can be seen that the approximation $C1(\alpha)$ (solid line) is rapidly approached by the exact value $C(\alpha, \Psi_0) = C(\alpha, N)$ with increasing N. The lowest curve is for $N = 1$.

N, as $\hat{\alpha}$ can be too large. Conversely, $C(\hat{\alpha}, \Psi_0)$ approaches zero for negative values of $\hat{\alpha}$, generating very large sample values, and hence variance, for $\hat{c}_f$. Indeed, if the affine function $1 + (2\ln 2 - 1)\hat{\alpha}$ plotted in figure 3, a quite acceptable approximation for $\hat{C}$, were used, the expectation of $\hat{c}_f$ would be infinite! For these reasons the definition of $\hat{C}$ bounds its values between $[C(0, \Psi_0), C(1, \Psi_0)] \approx [1, 2\ln 2]$, via a check on $\hat{\alpha}$. Note that each of c_f, $\hat{c}_f$ and $\widehat{c_f C}/C1(\hat{\alpha})$ are positive quantities.

Our approximation of the properties of $(\hat{\alpha}, \hat{c}_f)$ is based on the following result [22]. If a function $g(x, y)$ of the jointly distributed random variable (x, y) is sufficiently smooth near the joint mean (μ_x, μ_y), then its mean and variance can be approximated in terms of the mean, variance and covariance of (x, y) as

$$\mathbb{E}g(x,y) \approx g + \frac{1}{2}\left(\frac{\partial^2 g}{\partial x^2}\sigma_x^2 + 2\frac{\partial^2 g}{\partial x \partial y}\sigma_{xy} + \frac{\partial^2 g}{\partial y^2}\sigma_y^2\right) \tag{35}$$

$$\mathrm{Var}\,(g(x,y)) \approx \left(\frac{\partial g}{\partial x}\right)^2 \sigma_x^2 + 2\left(\frac{\partial g}{\partial x}\right)\left(\frac{\partial g}{\partial y}\right)\sigma_{xy} + \left(\frac{\partial g}{\partial y}\right)^2 \sigma_y^2 \tag{36}$$

where g and its derivatives are evaluated at $(x, y) = (\mu_x, \mu_y)$.

Let $h(y) = 1/C1(y)$ and set $g(x, y) = xh(y)$, where for stylistic clarity we put $x = \widehat{c_f C}$, $y = \hat{\alpha}$, and so $\hat{c}_f = g(\widehat{c_f C}, \hat{\alpha})$ and $(\mu_x, \mu_y) = (c_f C, \alpha)$. The mean and variance of the function $g(x, y)$ will be calculated with respect to the joint distribution of $(\hat{\alpha}, \widehat{c_f C})$ via the results above. It is immediately seen that $\partial g/\partial x = h(y)$ and $\partial^2 g/\partial x^2 = 0$. The other derivatives can be easily calculated, and evaluated at $(x,y)=(c_f C,\alpha)$ to obtain

$$\begin{aligned} \mathbb{E}\,\hat{c}_f &\approx \frac{c_f C}{C1(\alpha)}+\frac{1}{2}\left(2h'(\alpha)\mathrm{Cov}\,(\hat{\alpha},\widehat{c_f C})+c_f C h''(\alpha)\mathrm{Var}\,(\hat{\alpha})\right), \\ \mathrm{Var}\,(\hat{c}_f) &\approx \frac{\mathrm{Var}\,(\widehat{c_f C})}{C1(\alpha)^2}+2c_f h'(\alpha)\mathrm{Cov}\,(\hat{\alpha},\widehat{c_f C})+(c_f C)^2 h'(\alpha)^2\mathrm{Var}\,\hat{\alpha}, \end{aligned} \tag{37}$$

where $h'(y)$ and $h''(y)$ are given by

$$\begin{aligned} h'(y)\equiv\frac{dh(y)}{dy} &= \frac{2^y h(y)\ln 2-1}{2-2^y}, \\ h''(y)\equiv\frac{d^2h(y)}{dy^2} &= 2^y\ln 2\frac{h(y)\ln 2+2h'(y)}{2-2^y}. \end{aligned}$$

To estimate the covariance we use equation 35 with $g(x,y)=xh(y)(y-\mu_y)$, $h(y)$ as above, yielding

$$\mathrm{Cov}\,(\hat{\alpha},\hat{c}_f)\approx\frac{\mathrm{Cov}\,(\hat{\alpha},\widehat{c_f C})}{C1(\alpha)}+c_f C h'(\alpha)\mathrm{Var}\,(\hat{\alpha}). \tag{38}$$

The expressions above for the mean and variance of $\hat{c}_f$ and the covariance can be read as consisting of first order terms corresponding to constant C, plus correction terms. Since $-h'(y)$ and $h''(y)$ are both positive, monotone and almost constant over the range $y\in[0,1)$ with values ≈ 0.27 and ≈ 0.056 respectively, these correction terms are generally quite small. (Of course in practice the correction terms cannot be calculated exactly but must be estimated because they are themselves functions of the unknown parameters.)

These approximations suggest that the statistical properties of $\hat{c}_f$ are very close to those of $\widehat{c_f C}$. The numerical simulations in the next section confirm this expectation. Furthermore, as the elements of the covariance matrix of $(\hat{\alpha},\widehat{c_f C})$ decrease as $1/n$ for large n, it follows immediately from equations 37–38 that the same is true of $(\hat{\alpha},\hat{c}_f)$, and also that in the limit the expectation of $\hat{c}_f$ is just c_f. That is, the estimator $(\hat{\alpha},\hat{c}_f)$ is **asymptotically unbiased and efficient**.

We complete this subsection with an approximation of the fully asymptotic form of the covariance matrix, that is where $n,\ j_2\to\infty$, for use in section 4. Using the asymptotic form for the covariance of $(\hat{\alpha},\widehat{c_f C})$ of equation 47 from appendix 1, together with the elements of the approximate covariance matrix of $(\hat{\alpha},\hat{c}_f)$ given above, an asymptotic approximation is given by

$$\frac{1}{n\ln^2 2\,2^{1-j_1}}\begin{bmatrix} 1 & -c_f((1+j_1)\ln 2-\frac{h'(\alpha)}{h(\alpha)}) \\ -c_f((1+j_1)\ln 2-\frac{h'(\alpha)}{h(\alpha)}) & c_f^2\Big((3+2j_1+j_1^2)\ln^2 2-2(1+j_1)\frac{h'(\alpha)}{h(\alpha)}\ln 2+\Big(\frac{h'(\alpha)}{h(\alpha)}\Big)^2\Big) \end{bmatrix},$$

where $C(\alpha)$ has been identified with $C1(\alpha)$. This matrix is a function of α, however using the properties of h and h' above it is easily seen that the range of the α dependence is very small. In fact $h'(\alpha)/h(\alpha)\approx -1/3$ and a good approximation in the $j_1=3$ case (for example) is given by

$$\mathrm{Var}_{\mathrm{as},\,j_1=3}(\hat{\alpha},\hat{c}_f)\approx\frac{4}{n\ln^2 2}\begin{bmatrix} 1 & -3.1c_f \\ -3.1c_f & 10.6c_f^2 \end{bmatrix}. \tag{39}$$

with a corresponding correlation coefficient of $r\approx -0.95$.

H. Asymptotic distributions and confidence intervals

Distributions of the estimators. We recall the following results valid in the limit v_j/n_j small:

$$\begin{aligned} g_j &\to 0 \\ y_j &\to \log_2(\mu_j) \overset{\mathrm{d}}{\sim} N\Big(j\alpha + \log_2 c_f C,\ \frac{2^{j+1}}{n \ln^2 2}\Big) \\ \mathrm{Var}\,(y_j) &\to \frac{2}{n_j \ln^2 2} \\ p &\to 1 \end{aligned}$$

We have restated above that for n_j large, or more precisely v_j/n_j large, that the distribution of y_j is approximately normal. It is worth re-emphasizing that since $n_{j+1} \approx n_j/2$, only those octaves corresponding to the largest scales in the system will have a small n_j, and therefore a distribution which is markedly non-Gaussian with a corresponding variance which is abnormally large. Since $\hat{\alpha}$ consists of a sum of the y_j (equation 5), most of which are approximately Gaussian and weighted according to the (known) variances, for the purposes of confidence interval estimation $\hat{\alpha}$ can be considered as Gaussian. The same is true of $\hat{C}$ as $C(\alpha, \Psi_0)$ is well approximated by $1 + (2\ln 2 - 1)\alpha$ (for appropriate wavelet families).

We have shown that is reasonable to assume that $\hat{c}_f$ and $\widehat{c_f C}$ have very similar properties. For v_j/n_j large we have $p \to 1$ and so $\widehat{c_f C} \approx 2^{\hat{a}}$. Now the arguments concerning $\hat{b}$ from the previous paragraph apply equally to $\hat{a}$ (equation 6), so we can conclude that $\hat{a}$ is approximately Gaussian and therefore that $\widehat{c_f C}$ and $\hat{c}_f$ are approximately lognormally distributed.

Confidence intervals for the y_j, $\hat{\alpha}$, $\widehat{c_f C}$ and $\hat{c}_f$ have been calculated using these arguments.

III. Numerical Simulations

Simulation protocol. The aim of this section is to illustrate the accuracy of the theoretical statistical performance of $(\hat{\alpha}, \widehat{c_f C})$ and $(\hat{\alpha}, \hat{c}_f)$. The bias and covariance matrices are examined as functions of the size n of the data, and compared against the theoretical results. Recall that the theoretical properties were derived under ID1, ID2 and ID3. In the simulations below we retain ID1, that is we simulate Gaussian time series, but test robustness with respect to mild departures from ID2 and ID3. An example of the quality of the estimation under more severe departures from ID2 and ID3 is provided in figure 1. Because of the choice of scale issue discussed below, we leave a systematic study of such cases to a future paper [5].

Issues of non-stationarity are not considered here. Robustness with respect to smooth deterministic trends has already been addressed in [4], and a statistical test for stationarity [28] is under study.

Choice of scales. Since the definition of LRD in equation 1 is an asymptotic one, the wavelet based estimators are by nature semi-parametric, implying the need to choose the scales $j = j_1 \dots j_2$ over which the power-law behaviour in equation 1, and more particularly in equation 2, holds. In particular the choice of j_1: the decision of where short range dependence 'ends' and long range dependence 'begins', is a difficult one and not without a subjective element, as usual in a semi-parametric problem. The question of how to make a statistically justifiable, automated choice of j_1, j_2 is the subject of ongoing research [5] and is beyond the scope of this paper.

Nonetheless we must address the issue of choosing a 'safe' value of j_1 in our simulations or no meaningful statistical conclusions will be able to be drawn. In order to do this we simulate fractional Gaussian noise (fGn), which, as is well known, has a spectrum which is a uniform power-law over almost the full range of frequencies. (Indeed, the spectral synthesis method we use actually attempts to make the spectrum a uniform power-law.) Thus, for the simulated series we know in advance that the LRD 'begins' almost immediately. Based on prior studies of fGn we know that small but significant departures from the power-law scaling can occur for $j_1 = 1, 2$. We therefore choose $j_1 = 3$ in all our simulations.

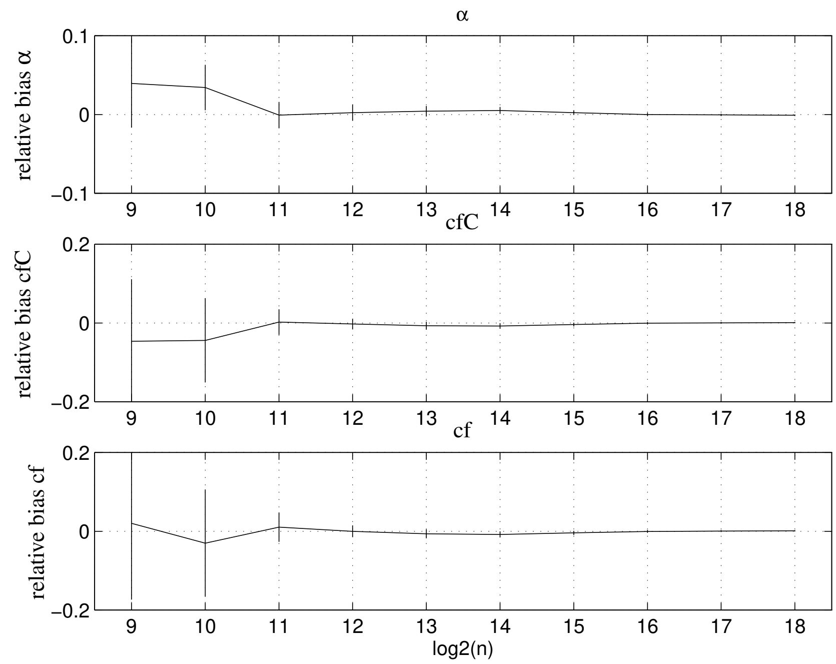

Fig. 4. **Relative biases of $\hat{\alpha}$, $\widehat{c_f C}$ and $\hat{c}_f$.** Relative biases of α, $c_f C$ and c_f are given as functions of $\log_2(n)$. Bias is seen to be small in each case even for small n (for instance, less than 5% for $n = 2^{10}$) and decreases rapidly for large n. The bias of $\hat{c}_f$ is no worse than that of $\widehat{c_f C}$ despite its additional nonlinearity. Empirical 95% confidence intervals have been added.

Initialization of the pyramidal algorithm. It is well known (see for example [2]) that the computation of the coefficients $d_x(j,k)$ of the wavelet transform from the fast recursive pyramidal algorithm requires the computation of an initial approximation sequence $a_x(0,k)$. This initial series is to be derived from the full process x_t itself. In actual data analysis we typically have access only to a digitally sampled version $x(k)$ of x_t, and the problem arises of how to estimate $a_x(0,k)$ from it. Very often, one is reduced to the crude simplification $a_x(0,k) = x(k)$, resulting in errors in the $d_x(j,k)$.

It is known [1], [15] that initialization errors are significant on the first octaves but quickly decrease with increasing j. In the LRD context, by its very nature, we are typically not concerned with small j, so the initialization issue can be overlooked. For instance, in the analysis of the Ethernet data presented in the next section, j_1 is always larger than 6. However, the initialization problem does exist and may require care if ever the smallest j's are needed (as can be the case of the fGn, for instance).

Evaluating the integral C. It is well known [11], [1] that the filter-bank based pyramidal algorithm, which underlies the reconstruction phase of the wavelet transform algorithm, allows an extremely precise approximation of the time-shape of the mother wavelet ψ_0 to be obtained in order n time. From this we simply compute its fast Fourier transform, and finally the desired integral defining C (equation 3) is estimated by a trapezoidal method. The divergence of the spectrum at the origin, $|\nu|^{-\alpha}$, is balanced by the smoothness of the Fourier transform $|\Psi_0(\nu)| = O(|\nu|^N)$ due to the N vanishing moments, so the integral is well behaved and poses no numerical difficulties.

Number of coefficients at each scale. It is clear from the nature of the multiresolution algorithm that in theory the number of available detail coefficients decreases by half as the scale is doubled, that is $n_{j+1} = n_j/2$. In practice, for most implementations of the discrete wavelet transform (see however [10]) there are border effects which render this relationship slightly optimistic at each octave. We therefore did not assume the above relation in the analysis. In the simulations also the actual number of wavelet coefficients at each scale are taken into account, enabling a correct comparison with the theoretical predictions.

Results. The results presented in figures 4, 5 and 6 are derived, for each n in the range $n = 2^9, 2^{10} \ldots, 2^{19}$, from 500 independent trials of simulated fGn. The parameters values chosen were $(\alpha, c_f) = (0.5, 8.5)$ and we selected *Daubechies 3* wavelets, that is with $N = 3$ vanishing moments. From these we compute that $C = 1.1590$ and $c_f C = 9.8517$.

Figure 4 shows relative biases as functions of n. As predicted, those for $\hat{\alpha}$ and $\widehat{c_f C}$ are very close to 0, illustrating that the estimator provides us with close to unbiased estimates even for very small ($n = 2^9$) sized data. For $\hat{c}_f$ also, which we could only show to be unbiased asymptotically, we see (lowest subplot) that in practice the relative bias is numerically very close to that of $\widehat{c_f C}$ for finite data and, moreover, exhibits the same behaviour with respect to n. This clearly shows that the biases for $\hat{c}_f$ and $\widehat{c_f C}$ are comparable and that $\hat{c}_f$ can be regarded as unbiased in practice.

Figure 5 compares the variances of the estimates against the actual variances from simulation as functions of $\log_2 n$. They are seen to be very close even in the case of the variance of $\hat{c}_f$ where the theoretical value is an approximation. They illustrate particularly well the $1/n$ decrease in the variances with n, which is a key, non-trivial feature in the context of LRD. A careful examination of the variances shows that the departure from theoretical predictions is larger at small n. This can be understood in terms of the residual correlation of the wavelet coefficients, i.e., the fact that ID1 and ID2 are not rigourously satisfied.

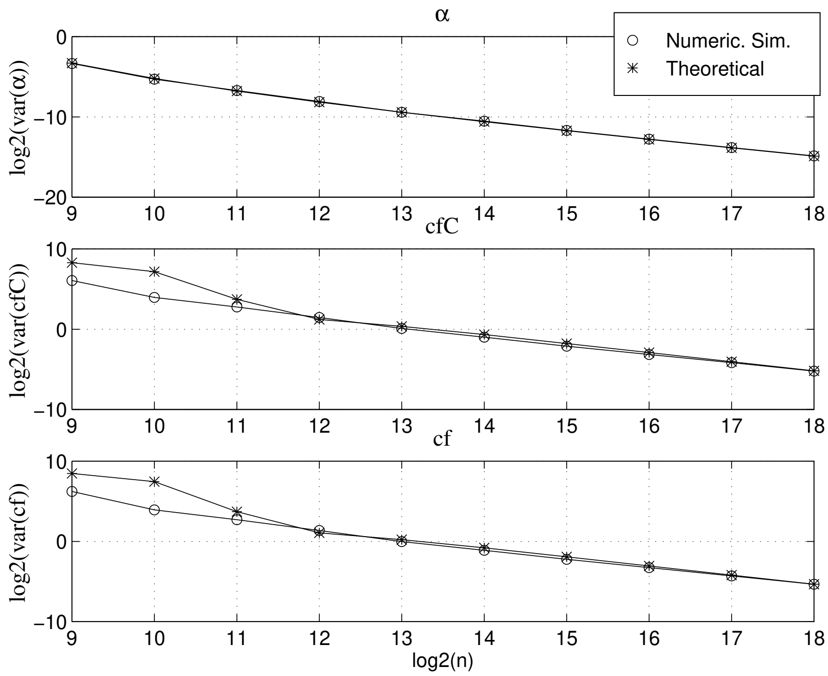

Fig. 5. **Variance of $\hat{\alpha}$, $\widehat{c_f C}$ and $\hat{c}_f$.** The $\log_2$ of the variances of $\hat{\alpha}$, $\widehat{c_f C}$ and $\hat{c}_f$ are given as functions of $\log_2(n)$. Agreement between numerical a nd the oretical performance is satisfying even for small n. For c_f, the approximate expression is close to the empirical variance. Moreover, the variances of $\hat{c}_f$ and $\widehat{c_f C}$ are very much alike, supporting the idea that the statistical properties of $\hat{c}_f$ and $\widehat{c_f C}$ can be regarded as equivalent. Finally, all three variances decrease as $1/n$.

Figure 6 shows the covariances of the estimates as functions of $\log_2 n$. For $(\hat{\alpha}, \widehat{c_f C})$ the agreement with the theoretical value is excellent. The second subplot illustrates again both the validity of the approximate calculations presented above for $(\hat{\alpha}, \hat{c}_f)$, and the fact that the statistical performance of $\hat{c}_f$ is very close to that of $\widehat{c_f C}$.

As shown in appendix 2, in the limit of large n the covariance of $(\hat{\alpha}, \widehat{c_f C})$ is controlled by that of $(\hat{b}, \hat{a})$, which as shown in appendix 1 (equation 48) tells us that the limiting form of the covariance matrix of our estimate is not diagonal. The correlation coefficient is typically negative and large in magnitude as seen in the figure.

To give an idea of how the results presented above vary with (α, c_f), in figure 7 surface plots are given showing the variation of key quantities over a sizeable portion of the (α, c_f) plane. Each calculated point on the surface is an average over 500 realisations, with the length of each series fixed at $n = 2^{15} = 32,768$. The relative biases of each of $\hat{\alpha}$ and $\hat{c}_f$ are shown in the top row. It is seen that the dependence on (α, c_f) is weak, an advantageous property. Regarding variances, equation 39 predicts that under ID1–ID3 the variance of $\hat{\alpha}$ is independent of (α, c_f) whereas that of $\hat{c}_f$ is proportional to the square of c_f. The plots in the bottom row of the standard deviation of $\hat{\alpha}$ and the *relative* standard deviation of $\hat{c}_f$ show that these idealisations are justified in practice.

Summarising, we have: i) excellent agreement between numerical and theoretical performance for $(\hat{\alpha}, \widehat{c_f C})$, ii) excellent agreement between numerical and approximate theoretical performances for $(\hat{\alpha}, \hat{c}_f)$, iii) statistical performance of $(\hat{\alpha}, \hat{c}_f)$ which is very close to that of $(\hat{\alpha}, \widehat{c_f C})$, iv) estimators which exhibit very low bias even for very small n, v) variances which decrease as $1/n$, vi) limiting forms for the covariance matrices which are not diagonal and basically controlled by those of $(\hat{b}, \hat{a})$, and finally, vii) residual biases which are small and almost independent of (α, c_f).

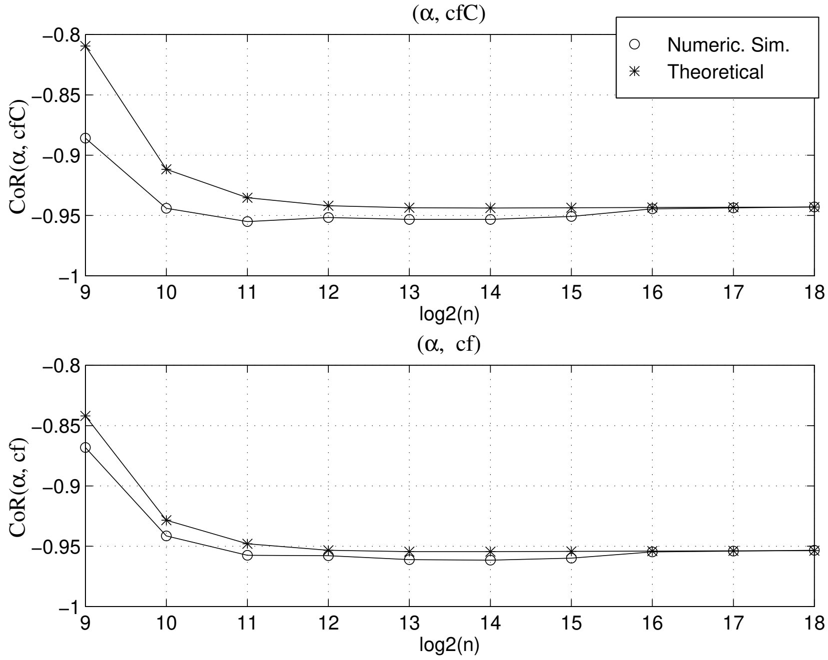

Fig. 6. **Correlation coefficients for** $(\hat{\alpha}, \widehat{c_f C})$ **and** $(\hat{\alpha}, \hat{c}_f)$**.** The correlation coefficients for $(\hat{\alpha}, \widehat{c_f C})$ and $(\hat{\alpha}, \hat{c}_f)$ are given as functions of $\log_2(n)$. Agreement between numerical and theoretical performance is satisfying even for small n. The correlation in $(\hat{\alpha}, \hat{c}_f)$ is hardly stronger than that in $(\hat{\alpha}, \widehat{c_f C})$ despite the appearance of $\hat{\alpha}$ in $\hat{c}_f$.

IV. Comparison against Other Estimators

A. *Generalities*

When evaluating the quality of an estimator both statistical and computational issues need to be addressed. It is often the case that an estimator has desirable properties in one, but not both of these categories. For example semi-parametric or non-parametric estimators of α (or equivalently, H) such as the variogram, the R/S method and the periodogram are simple conceptually and have desirable computational properties, but suffer from bias and/or high variance. On the other hand parametric methods, whilst offering excellent statistical performance, generally suffer from severe computational disadvantages, as well as being restricted to particular model classes. The wavelet based estimator of α however presented in [3], [4] performs well in both the statistical and computational senses. These issues are treated in more detail in [4], and also in [8]. The reader can also find in [25] an informative simulation based study of the finite data statistical properties of a wide variety of estimators of H.

In this paper we wish to compare against other *joint* estimators, of which few are available. The two selected happen to both be based on Maximum likelihood Estimation (MLE), the best known fully parametric method. The second of these, the discrete Whittle method, is a pertinent choice as it has been widely considered in the telecommunications literature as the best available. We shall not enter here into a thorough discussion of parametric versus semi-parametric estimation, but make the following points before entering into the detailed comparisons.

• **Statistical comparison:** A semi-parametric estimator can not in general out-perform a parametric one statistically when the data fits the assumed model class. The most it can do is equal its performance asymptotically. Approximations to parametric estimators however, such as those necessary in MLE estimation with large data sets, are typically only

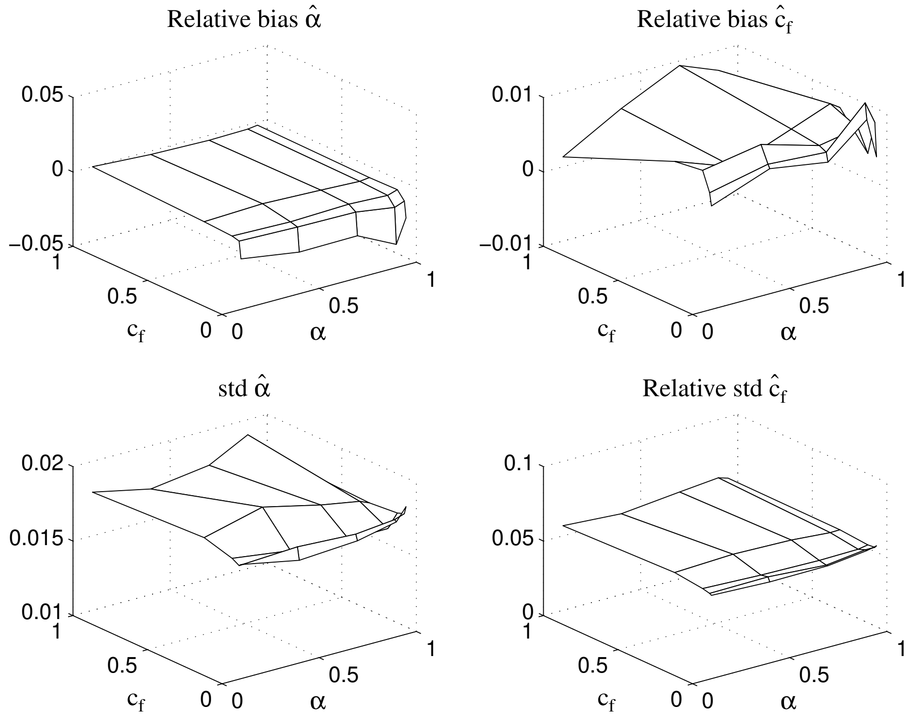

Fig. 7. **Variation with** (α, c_f). The relative biases of $\hat{\alpha}$ and $\hat{c}_f$ in the top row are seen to be quasi-independant of (α, c_f). In line with theoretical predictions, the standard deviation for $\hat{\alpha}$ and the relative standard deviation for $\hat{c}_f$ have the same property. The surfaces are sampled over $\{(\alpha, c_f)\}$ with $\alpha \in \{0.10, 0.40, 0.70, 0.90, 0.95\}$, $c_f \in \{2^{-8}, 2^{-6}, 2^{-4}, 2^{-2}, 2^{-0}\}$, each point being obtained from 500 realizations with $n = 2^{15}$. *Daubechies3* wavelets were used.

optimal *asymptotically*. For finite data sets the semi-parametric estimator may therefore perform as well or better, even if the data fits the underlying parametric model.

• **Robustness comparison:** A parametric estimator is very tightly tied to the assumed model class, and in general will perform poorly if those assumptions are invalid. Semi-parametric estimators are by definition much more robust in this respect in general. Inseparable from this robustness is the need for an analysis phase before the estimation phase: a semi-parametric estimator cannot, fortunately, be applied blindly, nor automatically, until after the first phase. In the LRD context the preliminary phase involves the examination of the Logscale Diagram (log-log plot) in order to choose the scale range where scaling is observed to occur. The wavelet based joint estimator, like that for α alone, enjoys significant additional robustness advantages, notably with respect to the elimination of smooth deterministic trends including non-discontinuous mean level changes [4], and variance changes [24].

• **Computational comparison:** Maximum likelihood estimation involves huge computational complexity and simplifications are always needed to make it practical. These simplifications remain computationally intensive and involve minimisation procedures with attendant convergence problems. In contrast, for the wavelet based joint estimator, after a discrete wavelet transform is performed the estimation involves only simple calculations, no minimization is needed. The transform can be calculated with the multiresolution (pyramidal) algorithm which has the very low complexity of $O(n)$.

As the wavelet estimator is clearly superior in robustness and computational terms, it remains to see how it compares statistically. In the comparisons below we compare the performance of two MLE estimators under optimal conditions against the performance of the wavelet estimator under optimal conditions, assuming therefore that the hypotheses of both are satisfied. We also make comments for the finite data case but there we do not have analytic results for the MLE estimators.

B. Wavelet-based MLE

The parametric family. In [29], [30] Wornell *et al.* consider processes which satisfy idealisations ID1, ID2 and ID3 and whose wavelet detail coefficients have variance taking the two parameter form

$$\mathbb{E}d_x(j,\cdot)^2 = 2^{j\alpha}V \tag{40}$$

for *all* octaves j. A MLE approach is applied to the joint estimation of (α, V). Such a family is very close to the fractional Gaussian noise for which the above relation holds except at the very smallest scales.

It is important to understand that although the estimation of (α, V) is defined and performed in the wavelet domain, and despite the formal similarity between the two problems in other respects, that the approach of [29], [30] is very different to that of the present paper theoretically, statistically and computationally. For example although a wavelet transform is performed to estimate (α, V), the computational properties of the wavelet-based MLE are determined by the much slower minimisation procedure.

Statistical performance The wavelet-based MLE is shown to be asymptotically unbiased and attains its asymptotic Cramer Rao bound. The Cramer Rao lower bound of $(\hat{\alpha}, \hat{V})$ is clearly the same as that of $(\hat{\alpha}, \widehat{c_f C})$ with $j_1 = 1$ and j_2 the largest possible. Its asymptotic form is therefore given by setting $j_1 = 1$ in equation 47 in appendix 1, namely

$$\mathrm{Var}_{\mathrm{as}}(\hat{\alpha}, \widehat{c_f C}) = \frac{1}{n \ln^2 2} \begin{bmatrix} 1 & -2c_f C \ln 2 \\ -2c_f C \ln 2 & 6(c_f C)^2 \ln^2 2 \end{bmatrix} \tag{41}$$

as derived in [29] by a different method.

Comparison We assume ID1–ID3 so that the exact results for both estimators hold. Since the Cramer Rao bounds are the same a comparison is particularly meaningful. If in $(\hat{\alpha}, \widehat{c_f C})$ we use $j_1 = 1$ and j_2 the largest possible for the given data then asymptotically the performance of the two estimators is the same. For finite data we have shown that $(\hat{\alpha}, \widehat{c_f C})$ is unbiased and is close to the bound, which is not the case for the wavelet-based MLE.

If we cannot choose the full range of scales then the MLE estimator will have a smaller variance asymptotically and for finite data there will be competition between this fact and the good finite data performance of $(\hat{\alpha}, \widehat{c_f C})$.

In equation 40 the quantity V has no physical interpretation but is simply a model parameter. Our approach identifies V as $c_f C$, gives insight into its meaning, and through it allows an estimation of c_f, a parameter of physical interest.

The wavelet-based MLE has the advantage of being able to include the estimation of additive noise provided however that one has a precise stochastic model for it. There is no corresponding ability in our estimator however as already mentioned it is robust to the presence of noise via the choice of scales j_1, j_2 over which the analysis is performed. Outside of this range $\mathbb{E}d_x(j,\cdot)^2$ may depart very significantly from the power-law form (see figure 1).

C. Frequency-based MLE (discrete Whittle)

The parametric family. In [8] Beran applies a discrete form of the Whittle estimator to Gaussian fractional ARIMA processes, which have known parametrised spectral densities. In general none of the parameters correspond to c_f, precluding a direct comparison. In the case of the two parameter farima0d0 process however, as discussed in the introduction, a connection to c_f can be made which is not too complex. Note that the definition of the spectral density of Beran is not that used here, the relation being $f_x(\nu) = 2\pi f_x^B(2\pi\nu)$. This seemingly innocent scale change and renormalisation makes an important difference to the estimation problem, as it implies a corresponding change in the definition of c_f, a point ripe for confusion. Beran estimates (α, θ_1) based on $f_x^B(\omega) = \theta_1|1-e^{\imath\omega}|^{-\alpha} \sim \theta_1|\omega|^{-\alpha}$, where $\theta_1 = \sigma_\epsilon^2/2\pi$, a notation we retain to facilitate comparison. In this scenario the parameter θ_1 is none other than the 'c_f' corresponding to f_x^B (for farima0d0), and the asymptotic covariance matrix, equation 43, is diagonal. Based on definition 1 however we have $c_f = \theta_1(2\pi)^{1-\alpha}$, and because of this α dependence the covariance matrix of $(\hat{\alpha}, \hat{c}_f)$, as we saw in section 2, is far from diagonal. In our notation the two parameter family in question is that with a spectrum of the form

$$\begin{aligned} f_x(\nu) &= 2\pi\theta_1|1 - e^{2\pi\imath\nu}|^{-\alpha} \\ &\overset{\nu\to 0}{\sim} \theta_1(2\pi)^{1-\alpha}|\nu|^{-\alpha} \\ &= c_f|\nu|^{-\alpha} \end{aligned} \tag{42}$$

In [8] an iterative minimization scheme based on a discretised form of Whittle's original approximate MLE approach is applied to the joint estimation of (α, θ_1).

Statistical performance The discrete Whittle MLE estimator of (α, θ_1) is shown to be asymptotically unbiased and efficient. The asymptotic covariance matrix is given by ([8] page 107)

$$\mathrm{Var}_{\mathrm{as}}(\hat{\alpha}, \hat{\theta}_1) = \frac{1}{n}\begin{bmatrix} 24/\pi^2 & 0 \\ 0 & 2\theta_1^2 \end{bmatrix}. \tag{43}$$

An estimate of $\hat{c}_f$ can be based on an initial estimate of (α, θ_1) followed by setting $\hat{c}_f = \hat{\theta}_1(2\pi)^{1-\hat{\alpha}}$. Using the approximation technique of section 2.6, we can derive from the matrix above an approximation to the covariance matrix of this Whittle based estimator of (α, c_f). The result is

$$\mathrm{Var}_{\mathrm{as}}(\hat{\alpha}, \hat{c}_f) \approx \frac{1}{n}\begin{bmatrix} 24/\pi^2 & -24c_f\ln(2\pi)/\pi^2 \\ -24c_f\ln(2\pi)/\pi^2 & 2c_f^2\left(1 + 12\ln^2(2\pi)/\pi^2\right) \end{bmatrix}. \tag{44}$$

The correlation coefficient is given by $r \approx -0.897$.

Comparison Although asymptotically both estimators are unbiased and attain their respective Cramer Rao bounds, the problems are not equivalent statistically so the bounds are not equal, resulting in a difference in performance. In addition, to apply the semi-parametric wavelet estimator here, a choice of j_1 must be made. Although for farima0d0 the asymptotic power-law for the spectrum is attained quickly, j_1 cannot be chosen equal to one. From prior studies we know that $j_1 = 3$ is an acceptable value.

Again for the wavelet based estimation we assume ID1–ID3 and use $j_1 = 3$ with j_2 the largest possible for the given data. The asymptotic performance in this case is approximately given by equation 39. Comparing with equation 44 one sees that the variances decrease asymptotically as $1/n$ in both cases, however the constants differ. The discrete Whittle based estimator has a lower variance for α and a much lower variance for

c_f. Specifically, they differ by constant factors given by Var($\hat{\alpha}$,MLE)/Var($\hat{\alpha}$,Wavelet)$=6\ln^2 2/\pi^2 \approx 0.29$ and Var($\hat{c}_f$,MLE)/Var($\hat{c}_f$,Wavelet)≈ 0.12. Note that if $j_1 = 2$ had been used instead the variance of the wavelet estimator would have decreased markedly, at the price of introducing only a small bias. The omission of the first two octaves makes such a significant difference to the variance of the wavelet estimation because y_1 and y_2 have the smallest variances of all the y_j, and hence the largest weighting factors in the estimation if included. By the same token however, it is precisely at these small scales (high frequencies) where in practice the parametric modelling assumptions of the MLE estimator are likely to fail. In such a case parametric estimators can produce completely erroneous results unless the form of the departure is both known, and taken into account explicitly. In contrast, the wavelet estimator will provide reliable estimates provided only that j_1 is chosen sufficiently large. A knowledge of the form of the departure from the asymptotic behaviour is not required.

Comparing the covariances of equations 39 and 44 we note that in each case the correlation coefficients are negative and large in magnitude (equation 48). This suggests that the high correlation arising from the wavelet based estimation is not an undesirable consequence of the linear regression underlying it, but an inherent feature of the parameters being measured.

For finite data we have shown that $\hat{\alpha}$ is unbiased with optimal or close to optimal variance and that $\hat{c}_f$ has a very small bias and variance. It is known however that for finite data the discrete Whittle estimator is biased ([4]). For finite data there will be competition between the lower asymptotic variance of the discrete Whittle estimator and the very good finite data performance of $(\hat{\alpha}, \hat{c}_f)$.

V. Application to Ethernet data

Background. As discussed in the introduction, the presence of LRD in traffic data of diverse types is now well accepted. Some of the most detailed evidence comes from the Ethernet traces of Leyland *et al.* described in [19]. It is not the purpose of this paper to present a detailed analysis of real traffic, however we give a brief analysis of two of these Ethernet traces, for two reasons. First, the traces in question have become in some sense de facto standards in the field of traffic analysis. It is therefore of interest to publish estimates of LRD parameters for them using the wavelet estimator, to serve as a basis for comparison with other estimators, and as a reference. The second reason is to complement previous work performed on them. In [4] we presented a thorough analysis of the structure of selected Ethernet traffic traces with respect to the measurement of the Hurst parameter, using an earlier form of the wavelet-based estimator. A main motivation was the design of compact, highly accurate models for Ethernet traffic. The central, highly unobvious, finding, was that quite different aspects of the traces (to be described below) had the same value of the Hurst parameter, whereas a priori most of them need not have even been long range dependent! It was in attempting to use this finding to design traffic models that it was realised that knowledge of c_f was an essential, yet neglected, factor, thus motivating the present paper. We therefore take the opportunity to return to these traces to complement results on H from [4] with measurements of c_f. (Note that the values of H reported below differ slightly from those in [4] as here the new, more accurate estimator is used, and the scale ranges (j_1, j_2) have been more carefully chosen.) This is the first time we are aware of that the value of c_f has been reported in traffic data, and we know of no a priori argument indicating whether it should be constant across different traces or different aspects of the same trace, nor even what its order of magnitude could be. As a simple example of the utility of such information, without values of c_f confidence intervals for mean estimates cannot be calculated, and therefore questions of stationarity of the mean across the traces cannot even begin to be addressed. Such issues impact on stationarity questions in general and therefore on model choice and validation.

In this section we use H rather than α (recall $H = (1+\alpha)/2$) as it is more customary in tele-traffic studies. In addition, as different time series are to be compared, we present the normalised form $\underline{c}_f = c_f/\sigma_x^2$ which characterises the correlation function, rather than c_f, where σ_x^2 denotes the variance (power) of x_t.

Description and modelling of the data. The Bellcore Ethernet traces have been described in detail elsewhere [19], [4]. Each file consists of 1 million rows in two columns. The first column gives the timestamp (measured from the beginning of the trace) for the *end* of the frame in seconds. The second column gives the integer size in bytes of the frame. The actual traffic consists therefore of an alternating sequence of disjoint frames and silent periods. Frames have a maximum and minimum size, whereas silences have a minimum but no maximum duration. Since silences are not restricted to be multiples of bytes, a process x_t general enough to fully capture the Ethernet arrival process must be defined in continuous time. We do not do this but consider discretised versions where the process is averaged within a window of size δ. This simplifying procedure is common practice and was justified in [4].

pAug	W_δ	C_δ	A	F	S
(j_1, j_2)	(6,14)	(7,14)	(9,16)	(9,16)	(9,16)
$\hat{H}$	0.813	0.870	0.839	0.839	0.837
$[H_{min}, H_{max}]$	[0.796,0.830]	[0.845, 0.895]	[0.814, 0.864]	[0.814, 0.864]	[0.812, 0.862]
$\hat{\underline{c}}_f$	0.185	0.104	0.051	0.129	0.053
$[\hat{\underline{c}}_{f,min}, \hat{\underline{c}}_{f,max}]$	[0.155, 0.219]	[0.077, 0.138]	[0.035, 0.071]	[0.088, 0.183]	[0.039, 0.080]

TABLE I

JOINT ESTIMATION OF $(\hat{\alpha}, \hat{\underline{c}}_f)$ FOR THE PAUG TRACE. $\delta = 0.0120$ SECONDS. THE H VALUES ARE ESSENTIALLY THE SAME AND THOSE FOR c_f ALSO, EXCEPT FOR A AND S WHICH ARE HALF AS LARGE.

pOct	W_δ	C_δ	A	F	S
(j_1, j_2)	(8,14)	(7,14)	(6,12)	(8,14)	(7,14)
$\hat{H}$	0.768	0.789	0.774	0.808	0.778
$[H_{min}, H_{max}]$	[0.731, 0.806]	[0.764, 0.814]	[0.758, 0.790]	[0.775, 0.842]	[0.756, 0.800]
$\hat{\underline{c}}_f$	0.198	0.216	0.148	0.119	0.133
$[\hat{\underline{c}}_{f,min}, \hat{\underline{c}}_{f,max}]$	[0.119, 0.310]	[0.160, 0.285]	[0.125, 0.173]	[0.076, 0.179]	[0.102, 0.171]

TABLE II

JOINT ESTIMATION $(\hat{\alpha}, \hat{\underline{c}}_f)$ FOR THE POCT TRACE. $\delta = 0.0067$ SECONDS. THE H VALUES ARE ESSENTIALLY UNIFORM ACROSS THE DIFFERENT SERIES AND THOSE FOR c_f ALSO.

We analyse five differents aspects of the traces, that is time series derived from the full arrival process x_t: W_δ, C_δ, A, F, and S. Each of these are assumed stationary. Although these series are not unrelated, they are by no means equivalent and give information regarding different aspects of the data of interest from a modelling perspective. The first two are directly related to the full process x_t and are discretised versions of underlying continuous time processes. The number of bytes (Work) arriving in subsequent time intervals of width δ defines W_δ, whereas C_δ Counts the number of frames in the same interval. The other three are intrinsically discrete time series indirectly related to x_t. They are the lists of the successive inter-Arrival times (the time gap between the beginnings of successive frames), Frame sizes, and Silence durations respectively. In [4] x_t and a seventh series, a point process, were also considered. We do not consider these here as W_δ represents x_t adequately, and for point processes the interpretation of c_f is problematic.

Discussion of results. For each of the two traces 'pAug' and 'pOct' the joint estimation was performed on the five different kinds of series, and the results recorded in the tables 1 and 2. The *daubechies 3* wavelet with $N = 3$ was used, ensuring the elimination of any linear or quadratic trends, if present (see [4]). It can be seen that for each trace the estimates for H are highly coherent whereas those for c_f differ slightly. It seems that the remarkable equality of behaviour across the five different series holds true for LRD properties as a whole, and not just for H. The only significant exception is that in the pAug trace where A and S have clearly lower values for $\underline{c}_f$ with confidence intervals separated from those of the other series. This is significant since it was F and A which were identified as the main series of use in model design, and their values differ by a factor of two.

VI. Conclusion

Because the long range dependence phenomenon as it is currently defined consists basically in the power-law behaviour of certain second-order statistics [6], there are a group of estimation techniques which consist essentially in the measurement of the slope in a log-log plot to extract the scale parameter α. Such methods have the advantage of being conceptually simple and practical to implement but generally exhibit extremely poor statistical properties [25]. These statistical disadvantages would be inherited in any obvious extension to the joint case based on the intercept of the said log-log plots, though such extensions have not been attempted. On the other hand, MLE techniques are characterised by excellent statistical performance but involve minimization procedures which are computationally complex and slow. The wavelet-based joint estimator for the parameters of LRD proposed here exhibits the advantages of both categories, but without their drawbacks. It moreover displays greater robustness [4] than either of the two. This results from the matching of the analysed phenomenon - LRD is essentially a scale invariance feature - to the analysing tool - the wavelet transform is essentially a scale invariant analysis method. More precisely two key features of the wavelet analysing family, namely its generation from a dilation operator and the existence of vanishing moments, result in two advantageous properties enjoyed by the wavelet coefficients of a LRD process: their variance exactly reproduces the power-law defining LRD, and they are stationary and almost independent. The statistical performance of the estimator follows from these two key properties and is, as we have shown, comparable to that of maximum likelihood estimators, even when their parametric assumptions are fully satisfied. Under departures from the parametric model the wavelet based estimator often performs better, without the need for knowledge of the form of the departures. The wavelet estimator performs well provided only that the data is large enough for the asymptotic behaviour to be present across at least three octaves, and that a cuttoff scale j_1 is correctly identified. The asymptotic nature of approximations to MLE methods such as the discrete Whittle however, mean that the parametric based estimation can at times be poor for smaller data sets.

There are two important areas where more work needs to be done. A vital practical aspect of the semi-parametric wavelet estimator is the choice of the scales over which the analysis is performed. This range of scales, unknown a priori, should correspond to where the signal exhibits its asymptotic power-law behaviour, that is the long range dependent or scaling regime. As mentioned earlier, the development of a well founded statistical basis for the automatic selection of the scaling range is under study [5]. An effective solution to this problem is necessary to benefit fully from the inherent robustness of the method to departures from the asymptotic power-law model which can be caused by additive random noise and strong and/or 'long' short-range correlations.

The second area is that of various aspects of non-stationarity. Specifically, a test for non-stationarity with respect to the LRD parameters is lacking and is an important first step. This problem is currently under study [28], as is robustness with respect to level shifts (sudden changes in the mean) [24], [5]. The questions of defining, and then analysing, detecting and estimating possible time variation of the LRD parameters are also of importance. Finally, in some situations long range dependencies may appear in processes whose second-order statistics do not exist. Preliminary studies [12] suggest that variations of this estimation tool could be useful in such circumstances.

VII. Summary

The objective:

Let x_t be a second order stationary *long range dependent* stochastic process, that is its spectrum is well defined and reads:

$$f_x(\nu) \sim c_f |\nu|^{-\alpha}, \;\; |\nu| \to 0.$$

We wish to jointly estimate α and c_f.

Definitions:

- *Wavelet coefficients*

$$\begin{array}{rcl} d_x(j,k) & = & \langle x, \psi_{j,k} \rangle \quad \text{(there are } n_j \text{ coefficients at octave } j) \\ \mu_j & = & 1/n_j \sum_k d_x^2(j,k) \\ \sigma_j^2 & = & \mathrm{Var}\,(\log_2(\mu_j)) = \zeta(2, n_j/2)/\ln^2 2 \; \simeq \; 2/(n_j \ln^2 2) \end{array}$$

- *Weighted Linear Regression*

$$\begin{array}{rclcrclcrcl} S & = & \sum 1/\sigma_j^2 & , & S_x & = & \sum j/\sigma_j^2 & , & S_{xx} & = & \sum j^2/\sigma_j^2, \\ w_j & = & \frac{(Sj - S_x)/\sigma_j^2}{S S_{xx} - S_x^2} & , & v_j & = & \frac{(S_{xx} - S_x j)/\sigma_j^2}{S S_{xx} - S_x^2} & , & g_j & = & \psi(n_j/2)/\ln 2 - \log_2(n_j/2), \\ y_j & = & \log_2(\mu_j) - g_j & , & \hat{b} & = & \sum w_j y_j & , & \hat{a} & = & \sum v_j y_j. \end{array}$$

- *Estimators*

$$\begin{array}{rcl} \hat{\alpha} & = & \hat{b} \\ \widehat{c_f C} & = & p \cdot 2^{\hat{a}}, \quad \text{where } \; p = \prod \frac{\Gamma(n_j/2)\exp(\psi(n_j/2) v_j)}{\Gamma(v_j + n_j/2)} \\ \hat{C} & = & \left\{ \begin{array}{lr} \int |\Psi_0(\nu)|^2 d\nu & \hat{\alpha} \le 0 \\ \int |\nu|^{-\hat{\alpha}} |\Psi_0(\nu)|^2 d\nu & 0 < \hat{\alpha} < 1 \\ \int |\nu|^{-1} |\Psi_0(\nu)|^2 d\nu & \hat{\alpha} \ge 1 \end{array} \right. \\ \hat{c}_f & = & \widehat{c_f C}/\hat{C} \end{array}$$

Statistical Performance

Assume that the scaling range has been correctly chosen, that is that a set of octaves, say $j \in [j_1, j_2]$, have been found such that for each j the spectrum has the power-law form above, and therefore that $\mathbb{E} d_x(j,\cdot)^2 = 2^{j\alpha} c_f C$ in this range. Under the additional technical idealisations,

- **ID1:** The process x, and hence the processes $d_x(j,\cdot)$, are Gaussian.
- **ID2:** For fixed j the process $d_x(j,\cdot)$ is iid.
- **ID3:** The processes $d_x(j,\cdot)$ and $d_x(j',\cdot)$, $j \neq j'$, are independent.

we have the following statistical properties of the estimators:

- $(\hat{\alpha}, \widehat{c_f C})$ is **unbiased** even for data of finite size (ID3 is not needed in the case of $\hat{\alpha}$). For $\widehat{c_f C}$ the result holds only if $2v_j/n_j > -1$ for each j. If this condition is not satisfied a new set of octaves must be chosen and the regression recalculated. The estimation of $\hat{\alpha}$ has no such restriction. The estimator is efficient, non-diagonal with negative correlation, and attains the Cramer-Rao lower bound in the limit $v_j/n_j \to 0$ for each j selected.
- $\hat{C}$ is asymptotically unbiased, and efficient.
- $\hat{c}_f$ is asymptotically unbiased, and efficient. The same conditions on j hold as for $\widehat{c_f C}$. The properties of $\hat{c}_f$ are closely related to those of $\widehat{c_f C}$, for example the correlation coefficient of $(\hat{\alpha}, \hat{c}_f)$ is also negative and large in magnitude.

Acknowledgements

The first author gratefully acknowledges the support of Ericsson Australia. The second author has been partially supported by the CNRS grant TL97035, Programme Télécommunications.

Appendix

A. The Cramer-Rao bound of $(\hat{\alpha}, \widehat{c_f C})$

Recall that by $f_\nu(x) = \left(\frac{1}{2^{\nu/2}\Gamma(\nu/2)}\right) x^{\nu/2-1}e^{-x/2}$ we denote the density of a Chi-squared variate, with mean and variance ν and 2ν respectively.

The Cramer-Rao bound is with respect to the joint density of the μ_j, which is just the product of the (rescaled) individual Chi-squared densities under ID1–ID3. Put $C_f = c_f C$ and define the constants $a_j \equiv \frac{n_j}{z_j} = \frac{n_j 2^{-j\alpha}}{C_f}$. From equation 11 the joint density can be written as

$$f_{\vec{\mu}}(\vec{x}) = \prod_j a_j f_{n_j}(x_j a_j) \tag{45}$$

where x_j is a dummy variable associated with μ_j.

The Fisher information matrix is $[\vec{I}(\vec{\theta})]_{ik} = -\mathbb{E}\left[\frac{\partial^2 \ln f_{\vec{\mu}}(\vec{x};\vec{\theta})}{\partial\theta_i \partial\theta_k}\right]$ and the Cramer-Rao bound is just its inverse [18]. Here $\vec{\theta} = (\alpha, C_f)$.

Let $L = \ln f_{\vec{\mu}}(\vec{x})$. We have $\frac{\partial L}{\partial a_j} = \frac{1}{a_j}[1 + G_j(x_j a_j)]$, where $G_j(x) = x f'_{n_j}(x)/f_{n_j}(x)$. Using the relations $\frac{\partial a_j}{\partial \alpha} = -j \ln 2 \cdot a_j$ and $\frac{\partial a_j}{\partial C_f} = -a_j/C_f$, we readily obtain

$$\begin{aligned}
\frac{\partial^2 L}{\partial^2 \alpha} &= \ln^2 2 \sum_j j^2 x_j a_j G'_j(x_j a_j), \\
\frac{\partial^2 L}{\partial \alpha \partial C_f} &= \frac{\ln 2}{C_f} \sum_j j x_j a_j G'_j(x_j a_j), \\
\frac{\partial^2 L}{\partial^2 C_f} &= \frac{1}{C_f^2}\left[\sum_j x_j a_j G'_j(x_j a_j) + 1 + G_j(x_j a_j)\right].
\end{aligned}$$

It is easily verified that $G_j(x) = n_j/2 - 1 - x/2$, and therefore that $G'_j(x) = -1/2$, and $\mathbb{E}G_j(x_j a_j) = -1$ since $\mathbb{E}x = z_j$. The information matrix is therefore given by

$$\vec{I}(\alpha, C_f) = \frac{1}{2}\begin{bmatrix} I_2 & \frac{1}{C_f} I_1 \\ \frac{1}{C_f} I_1 & \frac{1}{C_f^2} I_0 \end{bmatrix}$$

where $I_k = \ln^k 2 \sum_j j^k n_j, \quad k = 0, 1, 2$. By inverting we obtain finally

$$\mathrm{Var}(\hat{\alpha}, \widehat{c_f C}) \geq \frac{2}{(I_2 I_0 - I_1^2)} \begin{bmatrix} I_0 & -(c_f C)\, I_1 \\ -(c_f C)\, I_1 & (c_f C)^2\, I_2 \end{bmatrix}. \tag{46}$$

It is apparent that the bound is not diagonal, but exhibits negative correlation. To show this explicitly, assume that $j = j_1 \dots j_2$ and that $n_j \approx n 2^{-j}$, where n is the size of the data, holds exactly. Set $J = j_2 - j_1 + 1$, the number of octaves used in the analysis. Using the identities $\sum_{k=1}^K 2^{-k} = 1 - 2^{-K}$, $\sum_{k=1}^K k 2^{-k} = 2 - 2^{-K}(K+2)$ and $\sum_{k=1}^K k^2 2^{-k} = 6 - 2^{-K}(K^2 + 4K + 6)$, ([16], equations 0.113 and 0.114^6 [1]) , after a certain amount of algebra the elements of the bound can be written explicitly as

$$\begin{aligned}
\mathrm{Var}(\hat{\alpha}) &\geq (1 - 2^{-J})/F \\
\mathrm{Var}(\widehat{c_f C}) &\geq C_f^2 \ln^2 2\, [(3 + 2j_1 + j_1^2) - 2^{-J}(J^2 + 2J(1 + j_1) + (3 + 2j_1 + j_1^2))]/F \\
\mathrm{Cov} &= -C_f \ln 2\, [(1 + j_1) - 2^{-J}(J + 1 + j_1)]/F
\end{aligned}$$

[1] note the error in equation 0.114^6, the n^2 in the numerator should be x^2

where $F = F(n, j_1, J) = n \ln^2 2 \; 2^{1-j_1} \cdot (1 - (J^2/2 + 2)2^{-J} + 2^{-2J})$. We consider two limit cases to understand the behaviour of r, the correlation coefficient corresponding to the bound.

First consider the asymptotic limit of $n \to \infty$, allowing the limit $J \to \infty$ to be taken simultaneously with j_1 fixed. The asymptotic Cramer Rao bound is simply

$$\mathrm{Var}_{\mathrm{as}}(\hat{\alpha}, \widehat{c_f C}) \geq \frac{1}{n(\ln^2 2)\, 2^{1-j_1}} \begin{bmatrix} 1 & -(1+j_1)c_f C \ln 2 \\ -(1+j_1)c_f C \ln 2 & (3 + 2j_1 + j_1^2)(c_f C)^2 \ln^2 2 \end{bmatrix} \tag{47}$$

for which

$$r^2 = \frac{1 + 2j_1 + j_1^2}{3 + 2j_1 + j_1^2}. \tag{48}$$

It is clear that r is typically large in magnitude, independent of n. For example $j_1 = 1$ gives $r \approx -0.81650$, and for j_1 large $|r| \sim 1 - 1/j_1^2$.

B. **Limiting form of Cov**$(\hat{\alpha}, \hat{c}_f)$ **when** $v_j/n_j \to 0$.

For $\hat{\alpha}$ there is nothing to show, so consider first the variance of $\widehat{c_f C}$. Using the expansion $\Gamma(z) \sim \sqrt{2\pi} e^{-z} z^{z-1/2}(1 + 1/12z)$ for $z \to \infty$ ($|\arg z| < \pi$), and $(1 + d/z)^z \sim e^d(1 - d^2/2z)$, one can show that $\Gamma(z + d) \sim \sqrt{2\pi} e^{-z} z^{z+d-1/2}(1 + 1/12z(1 + 6d(d-1)))$ and so $\Gamma(2d + z)\Gamma(z)/\Gamma(d+z)^2 = 1 + d^2/z + o(1/z)$.

Continuing from equation 30, we can therefore write

$$\begin{aligned}
\mathrm{Var}\,(\widehat{c_f C}) \quad &= \quad (c_f C)^2 \left[\prod \frac{\Gamma(2v_j + n_j/2)\Gamma(n_j/2)}{\Gamma(v_j + n_j/2)^2} - 1 \right] \\
&\overset{n_{j_2} \to \infty}{\sim} \quad (c_f C)^2 \left[\prod (1 + 2{v_j}^2/n_j) - 1 \right] \\
&\overset{n_{j_2} \to \infty}{\sim} \quad (c_f C)^2 \left[1 + 2 \sum {v_j}^2/n_j - 1 \right] \\
&= \quad (c_f C)^2 \ln^2 2 \cdot \sum {v_j}^2 \sigma_j^2 \\
&= \quad (c_f C)^2 \ln^2 2 \cdot \mathrm{Var}(\hat{a})
\end{aligned}$$

where we have used equation 19 in the second last step and equation 8 to identify the expression as the variance.

Now consider the covariance. Using the expansion $\psi(z) \sim \ln z - 1/2z$ for $z \to \infty$ ($|\arg z| < \pi$), and $\ln(x + \epsilon) \sim \ln x + \epsilon/x$ for $\epsilon/x \to 0$, $x, \epsilon \in I\!R^+$, it follows that $\psi(x+\epsilon) - \psi(x) \sim \epsilon/x$ when x, $x/\epsilon \to \infty$.

Continuing from equation 32, we can therefore write

$$\begin{aligned}
\mathrm{Cov}\,(\hat{\alpha}, \widehat{c_f C}) \quad &= \quad c_f C \cdot \frac{1}{\ln 2} \sum w_j \Big(\psi(v_j + n_j/2) - \psi(n_j/2) \Big) \\
&\overset{n_{j_2} \to \infty}{\sim} \quad c_f C \cdot \frac{1}{\ln 2} \sum w_j \frac{2v_j}{n_j} \\
&= \quad c_f C \frac{1}{\ln 2} \ln^2 2 \cdot \mathrm{Cov}(\hat{b}, \hat{a}) \\
&= \quad c_f C \ln 2 \cdot \mathrm{Cov}(\hat{b}, \hat{a}),
\end{aligned}$$

using equations 19 and 9. Collecting these last two results it is seen that the limiting covariance matrix is given by:

$$\lim_{v_j/n_j \to 0} \mathrm{Var}(\hat{\alpha}, \widehat{c_f C}) = \begin{bmatrix} \mathrm{Var}(\hat{b}) & (c_f C) \ln 2 \cdot \mathrm{Cov}(\hat{b}, \hat{a}) \\ (c_f C) \ln 2 \cdot \mathrm{Cov}(\hat{b}, \hat{a}) & (c_f C)^2 \ln^2 2 \cdot \mathrm{Var}(\hat{a}) \end{bmatrix} \tag{49}$$

and we observe that the correlation coefficient corresponding to 49 is equal to that of $(\hat{b}, \hat{a})$, and therefore negative.

References

[1] P.Abry, Ondelettes et Turbulences - Multirésolutions, algorithmes de décompositions, invariance d'échelle et signaux de pressions. Diderot, Editeur des Sciences et des Arts, Paris, 1997, ISBN : 2-84134-064-3.

[2] P.Abry, P.Flandrin, On the initialization of the discrete wavelet transform. *IEEE Signal Processing Letters*, Vol. 1, n. 2, pp. 32–34, 1994.

[3] P.Abry, P.Gonçalvès and P.Flandrin, Wavelets, Spectrum estimation, $1/f$ processes. Wavelets and Statistics, *Lectures Note in Statistics*, Vol. 105, pp. 15–30, 1995.

[4] P.Abry, D.Veitch, Wavelet Analysis of Long Range Dependent Traffic. *IEEE Trans. on Info. Theory*, (Dec. 1997).

[5] P.Abry, D.Veitch, M.Taqqu, On the automatic selection of scaling range in semi-parametric LRD parameter estimation and the detection and compensation of level shift non-stationarities. In preparation.

[6] P. Abry, D. Veitch, and P. Flandrin, Long-range dependence: revisiting aggregation with wavelets, *Journal of Time Series Analysis*, 19(3):253–266, 1998.

[7] A.Aldroubi and M.Unser, Families of multiresolution and wavelet spaces with optimal properties, *Num. Func. Anal. Opt.*, Vol. **14**, (1993), pp.417-446.

[8] J.Beran, Statistics for Long-Memory Processes, Chapman & Hall (1994)

[9] F.Brichet, J.Roberts, A.Simonian, D.Veitch, Heavy Traffic Analysis of a Storage Model with Long Range Dependent On/Off Sources, *Queueing Systems and their Applications*, **23** 197–215, (1996).

[10] A.Cohen, I.Daubechies and P.Vial, Wavelets on the Interval and Fast Wavelet Transforms, *Applied and Computational Harmonic Analysis*, Vol. 1, No.1, pp.54–81, (1993).

[11] I.Daubechies, Ten Lectures on Wavelets. SIAM, (1992).

[12] L. Delbeke, P.Abry, Stochastic integral representation and properties of the wavelet coefficients of linear fractional stable motion, submitted to *Stochastic Processes and Applications*, (1997)

[13] A.Erramilli, O.Narayan, W.Willinger, Experimental Queueing Analysis with Long-Range Dependent Packet Traffic, IEEE/ACM Transactions on Networking, Vol. 2, April (1996).

[14] P.Flandrin, On the spectrum of fractional Brownian motions, *IEEE Trans. on Info. Theory*, **IT-35**, (1989) pp.197-199.

[15] P.Flandrin, Wavelet analysis and synthesis of fractional Brownian motion, *IEEE Trans. on Info. Theory* **IT-38**, (1992) pp.910-917.

[16] I.S.Gradshteyn and I.M.Ryzhik, Table of integrals, series and products, Academic Press, corrected and enlarged edition, (1980).

[17] J.R.M.Hosking, "Fractional differencing," *Biometrika*, Vol. 68, N 1, pp. 165–176, (1981).

[18] S.M.Kay, Fundamentals of statistical signal processing, Prentice-Hall (1993).

[19] W.Leland, M.Taqqu, W.Willinger, D.Wilson, On the self similar nature of Ethernet traffic (extended version), *IEEE/ACM Transactions on Networking*, February (1994), pp.1-15

[20] K. S.Meier–Hellstern, P.E.Wirth, Y-L.Yan, and D.A.Hoeflin. "Traffic Models for ISDN Data Users: Office Automation Application," Proceedings, *ITC13*, pp 167, Copenhagen 1991.

[21] I.Norros, A storage model with self-similar input, *Queueing Systems* 16 (1994), pp.387–396.

[22] A.Papoulis, Probability, Random Variables, and Stochastic Processes McGraw-Hill, Second Edition, (1984).

[23] V.Paxson, S.Floyd, wide-area traffic: the failure of Poisson modelling, *Proceedings of SIGCOMM '94* (1994).

[24] M.Roughan, D.Veitch, Measuring Long-Range Dependence under Changing Traffic Conditions, *Proceedings of INFOCOM '99* (1999).

[25] M.S.Taqqu, V.Teverosky, W.Willinger, Estimators for long-range dependence: an empirical study *Fractals*, 1996.

[26] A.H.Tewfik, M.Kim Correlation Structure of the Discrete Wavelet Coefficients of fractional Brownian motion. *IEEE Trans. Info. Theory*, **IT-38**, (1992) pp.904–909.

[27] D.Veitch, P.Abry, Estimation Conjointe en Ondelette des Paramètres du Phénomène de Dépendance Longue, *16ieme Colloque GRETSI* Grenoble, France, pp.1451–1454, September (1997).

[28] D.Veitch, P.Abry, A Statistical Test for the Time Constancy of Scaling Exponents. *submitted for publication* Jan. (1999).

[29] G.W.Wornell, A.V.Oppenheim, Estimation of Fractal Signals from Noisy Measurements Using Wavelets, *IEEE Trans. on Signal Proc.*, vol. 40, No. 3, March (1992), pp.611–623.

[30] G.W.Wornell, Signal processing with Fractals - A wavelet-based approach, Prentice-hall, 1995.